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

Doo Seok Kang1, Nahyun Lee3, Dong Yeop Shin1, Yu Jin Jang1, Su-Hyon Lee3, Kyung-Min Lim2, Yeon-Soon Ahn4, Cheol Min Lee5* & Young Rok Seo1*

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 reactivity1. The common toxic effects of formaldehyde include irritation and damage to contacted tissues2. Formaldehyde induces various levels of genotoxic effects, such as N2-hydroxymethyl-dG adducts, DNA–protein crosslinks, and chromosomal damage3,4. The International Agency for Research on Cancer (IARC) has classified formaldehyde as a human carcinogen that causes nasopharyngeal cancer and leukemia1. 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 purposes2. Humans are exposed to very low concentrations of environmental formaldehyde in daily life, primarily through inhalation5. Unlike the exposure in daily life, occupational exposure occurs at a wide range of concentrations, as much as 20.94 mg/m3 in an anatomical laboratory and 60.77 ppm in a factory, in extreme cases2,6,7. Therefore, people working in high-risk occupational environments are likely to be exposed to high concentrations for an extended time8. 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.

1Department of Life Science, Dongguk University Biomi Campus, Goyang, Gyeonggi 10326, Republic of Korea. 2College of Pharmacy, Ewha Womans University, Seoul 03760, Republic of Korea. 3R&D Institute, Biosolution Co., Ltd., Seoul 01811, Republic of Korea. 4Department of Preventive Medicine and Institute of Occupational and Environmental Medicine, Wonju College of Medicine, Yonsei University, Wonju, Gangwon 26426, Republic of Korea. 5Department 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
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. 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.

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, 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. 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.

**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). 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.

---

**Figure 1.** Flow chart of the steps used in this study.
**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. 

**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. 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. "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).

**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 °C and 5% CO₂ 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, 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-dimethylthiazol-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% CO₂ 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 2ΔΔC_t method. 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 24 h (upregulated: 1,432, downregulated: 876), and 3,621 genes at 13 weeks (upregulated: 1,712, downregulated: 1,909). Among all exposure groups, 258 genes were 815 genes significantly expressed at 6 h exposure (upregulated: 354, downregulated: 461), 2,308 genes at 24 h (upregulated: 1,432, downregulated: 876), 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 = 4.07×10⁻⁴), ErbB Signaling (z-score = 1.13, p = 5.37×10⁻³), and NRF2-mediated Oxidative Stress Response (z-score = 0.378, p = 4.47×10⁻²) exhibited positive activity patterns, and Ferroptosis...
sis Signaling Pathway (z-score = –1.41, p = 4.47 × 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, PLAU, PTGS2, and TIMP1) and major entities were selected.
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 study21. The MTT assay result showed that the tissue viability was > 90% at concentrations up to 5 mM and dropped significantly at concentrations above 50 μM (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).

### Table 1. List of top 10 canonical pathways for the common DEGs

| 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, SLCA15A, SLCA11 |
| ErbB Signaling | 3.27 | 0.0814 | 1.134 | ARFEG, EREG, HBECE, MAP2K3, MAP2K6, NRG1, PIK3C2G |
| Cell Cycle Regulation by BTG Family Proteins | 2.69 | 0.121 | NULL | CCNE2, NOCT, PPM1L, PPP2TB |
| Valine Degradation I | 2.39 | 0.15 | NULL | ACADSB, BGT2, 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 | FM01, FM02, GSTM2, HMOX1, MAP2K3, MAP2K6, MAP3K8, PIK3C2G, PPM1L, PPP2TB |
| NRF2-mediated Oxidative Stress Response | 2.35 | 0.0471 | 0.378 | CB1R, FM01, FOSL1, GSTM2, HMOX1, MAFF, MAP2K3, MAP2K6, PIK3C2G |
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.
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\(^2\). These concentrations are very high and unlikely to be encountered in daily life\(^2,3\). 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\(^2,5-9\). 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

---

Figure 4. (continued)
as cell proliferation and apoptosis. Elevated AREG expression is associated with airway inflammatory diseases, such as chronic obstructive pulmonary disease (COPD), asthma, and several types of cancer. 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. 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. 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. 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. PTGS2 expression and cytokines were increased in the respiratory epithelium and alveolar macrophages of asthmatics. Persistent increased PTGS2 expression is thought to be induced by the inflammatory response after formaldehyde exposure. PLAUR has been identified as an asthma, pneumonia, and COPD susceptibility gene. PLAUR expression was elevated in the bronchial biopsy of asthmatics and is required for epithelial wound repair. TIMP1 mainly functions as an important endogenous inhibitor in regulating matrix engineering technologies. Additionally, studies incorporating a sophisticated aerosol exposure system (e.g., physiological conditions have led to advanced in vitro studies, such as 3D cell culture and microfluidics/micro-conditions. Efforts to reduce the gap between animal testing and 2D cell culture models and reflect in vivo 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. 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 disease 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.

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 and presents dual pro- or antitumor activity. The Notch1 gene positively regulates Nrrarp 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. 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. This study performed a toxicogenomic in silico analysis based on in vivo data and validated the result in the 3D ALL 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. Moreover, there could be false-positive results and missing interactions due to potential bias toward well-studied interactions. Finally, further studies are needed to identify all possible
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.

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. 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.

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.

Received: 10 January 2022; Accepted: 21 March 2022
Published online: 04 April 2022

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/10.1080/1932026312466459 (2013).
5. Nandan, A., Siddiqui, N. A., Singh, C. & Aeri, A. Occupational and environmental impact of indoor air pollutant for different occupancies: A review. Toxicol. Environ. Health Sci. 13, 303–322. https://doi.org/10.1017/S113303-021-00102-9 (2021).
6. Stewart, P. A., Cubitt, D. & Blair, A. Formaldehyde levels in seven industries. Appl. Ind. Hygiene 2, 231–236. https://doi.org/10.1080/08828302.1987.10388468 (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/10.1017/S113303-021-00087-5 (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. Manangbanag, 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. Bioinformatics 24, 249–264. https://doi.org/10.1093/bioinformatics/btm252 (2008).
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/btq304 (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/gkv097 (2015).
17. Kramer, J., 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. Nikitina, 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. Genet. Environ. 43, 13. https://doi.org/10.1186/s41021-021-00183-5 (2021).
20. Assenov, Y., Ramirez, F., Schellhorn, 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 CT) Method. Methods 25, 402–408. https://doi.org/10.1016/meth.2001.1262 (2001).
23. Korns, 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., Wittek, 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/10.1097/0000039689.1986.9938338 (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/10.1097/0000039689.1990.11018748 (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, 535–39. https://doi.org/10.4103/0973-7406.155787 (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).
Acknowledgements
This study was supported Korea Environment Industry and Technology Institute (KEITI) through The Chemical Accident Prevention Technology Development Program, funded by Korea Ministry of Environment (MOE) (2017001970001) and (2018001350006).

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
This article was funded by Yeon-Soon Ahn (Grant no. 2017001970001) and Cheol Min Lee (Grant no. 2018001350006).

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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022