



Analysis on Ambient Volatile Organic Compounds and their Human Gene Targets

Xiao Hua¹⁺, Yi-Jing Wu²⁺, Xiaoqi Zhang¹, Shumin Cheng^{1,3}, Xiao Wang⁴, Jun Chu^{4**},
Qianli Huang^{1,3*}

¹ School of Food and Biological Engineering, Hefei University of Technology, Anhui, Hefei 230009, China

² School of Chemistry and Chemical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

³ School of Resources and Environmental Engineering, Hefei University of Technology, Anhui, Hefei 230009, China

⁴ Key Laboratory of Xin'an Medicine, Ministry of Education, Anhui University of Traditional Chinese Medicine, Anhui, Hefei 230031, China

ABSTRACT

Although some ambient Volatile Organic Compounds (VOCs) have been recognized to influence human health, the functional basis and underlying molecular mechanisms remain obscure. Here, the ambient VOCs and corresponding human genes response to the VOCs (HGRV) were analyzed integrately from several aspects. (1) Firstly, through identifying composition and connectivity of functional groups within VOCs, we found that there is significant bias in the co-occurrence patterns among the top five frequent groups (respectively methyl, carbon-carbon double bond, benzene ring, chlorine and ethylene). For instances, co-occurrence of chlorine and benzene are considerably frequent, whereas the connection between chlorine and methyl/ethylene occurred rarely. (2) Then, by examining the screened HGRV, it was unveiled that four genes (respectively, *IL6*, *BCL2*, *FOS* and *PTGS2*) may act as “hub/hotspots” in response to VOCs. And four types of VOCs (respectively, Acetaldehyde, Butyraldehyde, Formaldehyde and Styrene) were found to be shared stimulators of the hubs. (3) Moreover, three dominant function categories were detected, respectively, response to lipopolysaccharide, response to molecule of bacterial origin and response to oxidative stress. (4) Further, the results disclosed that three most significant pathways are IL-17 signaling pathway, TNF signaling pathway and Rheumatoid arthritis. These three most significantly enriched pathways in HGRV are primarily associated with immune diseases and cancer. (5) Subsequently, by analyzing the diseases potentially associated with HGRV, the results revealed that 50 kinds of diseases might be significantly associated with HGRV. Among them, top 3 significant diseases are Immune system diseases, Cancers and Allergies and autoimmune diseases respectively. And these top 3 diseases might be induced by 13 common VOC species, such as Benzene, Chlorobenzene and Formaldehyde. Thus, the above findings could be useful for advancing our understanding of mechanisms related to diseases by VOCs exposure and development of control strategies to limit their damage or attenuate their sequelae.

Keywords: VOCs; Functional group; Co-occurrence network; Pathway; Enrichment analysis; Disease.

INTRODUCTION

By the US Environmental Protection Agency (USEPA), Volatile Organic Compounds (VOCs) were described as organic compounds with initial boiling point less than or

equal to 250°C measured at a standard atmospheric pressure of 101.3 kPa (<https://www.epa.gov/indoor-air-quality-iaq/technical-overview-volatile-organic-compounds>). Generally, they include alkanes, alkenes, alkynes, aromatics, alcohols, aldehydes, ethers, ketones, esters, halogenated hydrocarbons and others (Zhang *et al.*, 2017b). Recently, VOCs have aroused wide concerns and triggered numerous studies on their sources, photochemical characteristics, risks and underlying mechanisms on human health (Guerra *et al.*, 2017; Ou-Yang *et al.*, 2017; Zhang *et al.*, 2017a; Alvim *et al.*, 2018).

VOCs can originate from both biogenic sources (BVOCs) and anthropogenic sources (AVOCs) (Ren *et al.*, 2017). Although biogenic emissions (such as from vegetations or microbes) occupy a dominant position on the global scale, VOCs are significant affected by other factors including

⁺ Both authors contributed equally to this manuscript

* Corresponding author.

Tel.: +86-182-5690-4658

E-mail address: cheneyhuang5128@gmail.com

** Corresponding author.

Tel.: +86-551-65169051

E-mail address: chuj@ahctm.edu.cn

transport and industrial emissions, as well as human activity (Mečiarová *et al.*, 2017; Zheng *et al.*, 2017). In recent decades, large amounts of VOCs are released into the atmosphere annually in especially developing countries, which has experienced rapid industrialization, urbanization, and transportation development (Wang *et al.*, 2014; Mannucci and Franchini, 2017). On source apportionment, some VOC species can be used as markers to trace the specific sources. For instance, propane is a known marker for natural gas production; isopentane is an indicator of gasoline vaporization; and propylene is a tracer of vehicle exhaust (Abeleira *et al.*, 2017; Tibaquirá *et al.*, 2018). Some investigations focus on the potential of VOCs serving as signs of life, such as non-invasive markers for human or plant phenotyping (Mochalski *et al.*, 2015; Niederbacher *et al.*, 2015). So, the VOC types in different emission inventories are associated with the sources. Moreover, for the effects of VOCs on environment, most VOCs have direct impacts on environment. The VOCs, such as benzene, toluene, and 1,3-butadiene have been proved to be air toxics (Zhang *et al.*, 2018). Also, VOCs cause damages to the ecosystems indirectly through forming secondary air pollutants such as ozone, peroxyacetyl nitrates and organic aerosol via photochemical processes (Li *et al.*, 2017). Different VOCs have divergent photochemical ozone formation potential (OFP) and secondary organic aerosol formation potential (SOAFP) (Jenkin *et al.*, 2017). Hence, the abilities of VOCs to cause pollution are at least determined partly by the source and photochemical reactivity. As known, functional groups are specific collections of atoms within organic molecules and confer characteristic properties to the molecules (Villanueva-Rosales and Dumontier, 2007). On one hand, functional groups may define the reactivity and toxicity of VOC species. On the other hand, VOCs sharing same functional groups may also be generated from common emission sources. Motivated by these considerations, it is vital to analyze functional groups within VOCs, which will help to explore functional basis of VOC-related mechanisms and develop strategies for controlling VOCs.

Along with the harms to the environment, VOCs can cause a variety of adverse health effects. For instance, exposure to BTEX (i.e., benzene, toluene, ethylbenzene, and xylene), which occur naturally in petroleum products, has been associated with diseases like the loss of pulmonary function, skin problems, immunological system disorders and acute myeloid leukemia in children (Wallner *et al.*, 2012; Montero-Montoya *et al.*, 2018). Epidemiologic studies have also suggested inflammatory and cardiorespiratory effects of ambient VOCs (Ye *et al.*, 2017). In addition to the aberrant phenotypes induced by VOCs, the molecular mechanism behind them become a hot research topic lately. Juarez *et al.* (2018) reported that, like other DNA damaging agents, Formaldehyde can form DNA-protein crosslinks, in which over 300 genes engaged. 96 genes targets with altered expression were involved in occupational exposed to benzene (Santos *et al.*, 2018). Through bisulfite sequencing of the gene promoter regions, methylation in gene promoters were found related to hematopoietic malignancy

in workers exposed to a VOC mixture (Jiménez-Garza *et al.*, 2018). So, whatever a single VOC or a mixture of VOCs, these studies have shown that their toxicity are related to alteration in the expression of specific target genes and achieved through coordinated interacting among multiple target genes (Kim *et al.*, 2011; Neghab *et al.*, 2018). Obviously, ambient VOCs are usually complex mixtures of species from different sources, which may jointly contribute to the toxic effects. However, most previous investigations considered only a limited number of VOCs and were not representative of the wide range of species found in urban air. Therefore, by considering the VOCs as a collection, systematic analyses on the genes perturbed by VOCs will facilitate mining of integrated mechanisms involved in exposure to the blend of VOCs.

Thus, the present study has been carried out from three main aspects. First, to explore functional basis of VOCs, the composition and connectivity of functional groups within VOCs were analyzed. Then, human genes response to VOCs (HGRV), through which the effects of VOCs on human health are likely to be achieved, were examined by screening the potential relations among VOC-gene pairs. Subsequently, the enriched function and pathways in HGRV were obtained by enrichment analysis based on hypergeometric distribution. Accordingly, the diseases associated with HGRV were mapped and the key VOCs potentially implicated in these diseases were detected. This study may provide useful background information to advance understanding of mechanisms related to health risks of VOCs exposure and develop control strategies on VOCs.

METHODS

Dataset of VOC Species

The dataset of VOCs was constructed by merging data from two sources, respectively, the ambient volatile organic compounds recently identified by Zhang *et al.* (2017b); and the list provided by USEPA (United States Environmental Protection Agency; <https://www.epa.gov/indoor-air-quality-iaq/volatile-organic-compounds-impact-indoor-air-quality>). Here, only the species are retrieved, and other information such as levels, sources and spatial distributions of VOCs are not considered.

Construction of Co-occurrence Network

Firstly, the functional groups in each VOC species are determined manually. Accordingly, the frequency of co-occurrence patterns among functional groups were calculated. And then, the network was constructed with *igraph* package in R language (<http://igraph.org/r/>). To be more intuitive, all the edges are weighted with the frequency of co-occurrence patterns through setting the parameter of “edge.width”. The colour bar was set based on frequency intervals. Conceivably, a thicker and darker edge denotes more stronger relation between two functional groups.

Screening of Human Genes Influenced by VOCs

Aims to advance understanding about mechanisms of environmental exposures on human health, Comparative

Toxicogenomics Database (CTD) is a community-supported genomic resource providing manually curated information about chemical-gene/protein interactions (Davis *et al.*, 2016). From CTD, a dataset including about 1,654,664 chemical-gene interactions was downloaded. Subsequently, our examined VOCs were mapped to the dataset and the candidate VOC-gene interactions were selected with the following criteria: (1) specific chemical-gene or chemical-protein interactions should be in *Homo sapiens*; (2) the expression of the gene/protein should be changed (with terms of “increases expression” or “decreases expression”) in VOC response; (3) To further reduce possible false positives, the expression of target genes can be perturbed by at least 3 types of VOCs. Thus, 1257 VOC-gene pairs were obtained, in which 373 unique genes were harboured. The screened results are provided in Table S1.

Enrichment Analyses on Biological Function and Pathway in HGRV

Generally, the gene functions are annotated with GO Terms which describes the functions of specific genes using concepts in Gene Ontology Consortium (<http://www.geneontology.org/>). Kyoto Encyclopedia of Genes and Genomes (KEGG) is a very useful database resource which provides plentiful higher order functional information (<https://www.genome.jp/kegg/>). As a compilation of manually verified pathway maps, KEGG PATHWAY displays both the molecular interactions and biochemical reactions, and usually utilized to analyze the pathway of target genes (Kanehisa *et al.*, 2016). To streamline workflow, the package clusterProfiler implemented in R was adopted to perform enrichment test for GO terms and KEGG pathways based on hypergeometric distribution. (<https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>). To avoid possible high false discovery rate (FDR) in multiple testing, *q*-values are also estimated with FDR control in functions *enrichGO* and *enrichKEGG* (Yu *et al.*, 2012). The parameters were set as follows: *OrgDb* = *org.Hs.eg.db*, *ont* = "BP", *keyType* = "ENTREZID", *pAdjustMethod* = "BH", *pvalueCutoff* = 0.05, *qvalueCutoff* = 0.01. Enriched GO Terms and pathways in HGRV (*p.adjust* < 0.01) are enclosed in Table S2.

Identification of Diseases Associated with HGRV

The diseases associated with HGRV were identified and corresponding enrichment analysis was carried out with the tool KOBAS 3.0 which combined three authoritative disease databases including Online Mendelian Inheritance in Man (OMIM), KEGG DISEASE and NHGRI GWAS Catalog (Xie *et al.*, 2011). In detail, with a particular focus on the gene-phenotype relationship, OMIM is a comprehensive compendium of human genes and genetic disorders (https://en.wikipedia.org/wiki/Online_Mendelian_Inheritance_in_Man). KEGG DISEASE is a collection of disease entries focusing on influence factors (<https://www.genome.jp/kegg/disease/>). The NHGRI GWAS Catalog provides a curated resource of SNP-trait associations (Welter *et al.*, 2013).

RESULTS AND DISCUSSION

Functional Group Composition of Volatile Organic Compounds (VOCs)

The VOC related characteristics, source apportionment, secondary organic aerosol formation potential and biochemical features of VOCs, are quite complicated and remain debated (Wang *et al.*, 2016; Thorsson *et al.*, 2018). Functional groups describe the semantics of chemical reactivity in terms of atoms and their connectivity, which are categorized by their specific composition and chemical behavior (Villanueva-Rosales and Dumontier, 2007). The composition and connectivity of functional groups within VOCs may be associated with their sources, photochemical properties and biochemical roles. First, the functional groups within 128 VOC species were identified (see Table S3). 36 VOCs contain single functional group, 74 VOCs have 2 groups and 18 VOCs harbor 3 groups. This suggests that the examined VOCs are mainly small molecules composed of one or two functional groups. Among the individual groups, the dominant five species are respectively, methyl (*n* = 42), carbon-carbon double bond (*n* = 32), benzene ring (*n* = 31), chlorine (*n* = 25) and ethylene (*n* = 20). In contrast, the species isobutyl and tertiary pentyl (*n* = 1) contributed the least to VOCs, followed by the isopropyl and cyano (*n* = 2). For the alkyl group, methyl is much more frequent than the group composed of more than 2 carbons. The results indicated that the hydrocarbons in VOCs are more inclined to forming benzene rings than long chains. For the halogens in VOCs, the frequency of chlorine (Cl) is much higher than fluorine (F) and bromine (Br). Probably, the reaction between chlorine and other small molecular (such as ethane, acetone and propane) can often be more vigorous than bromine (Sherwen *et al.*, 2016). This may also result from the fact that, chlorine is commonly in industry and make products like household bleach or swimming pool shock, from which halogenated VOCs emissions (Odabasi, 2008).

Next, to further investigate the connectivity of functional groups in VOCs, co-occurrence patterns of functional groups were counted and depicted in Fig. 1. In the co-occurrence network, the nodes represent functional groups, and all the edges are weighted by the frequency of co-occurrence patterns. A thick edge denotes strong relation between two functional groups. The connection between the benzene ring and methyl was the most frequent (*n* = 13), followed by carbon-carbon double bond and chlorine (*n* = 11), carbon-carbon double bond and methyl (*n* = 11), ethylene and methyl (*n* = 11), benzene ring and chlorine (*n* = 8), and benzene ring and ethylene (*n* = 6). Although the groups (respectively methyl, carbon-carbon double bond, benzene ring, chlorine and ethylene) occupied top five individually, there is significant bias in the co-occurrence patterns among them. For instances, co-occurrence of chlorine and benzene are considerably frequent, whereas the connection between chlorine and methyl/ethylene occurred rarely. This indirectly illustrates the composition features of VOCs, but it does not mean that rare pattern is not important. Methylene chloride is a good example. As an

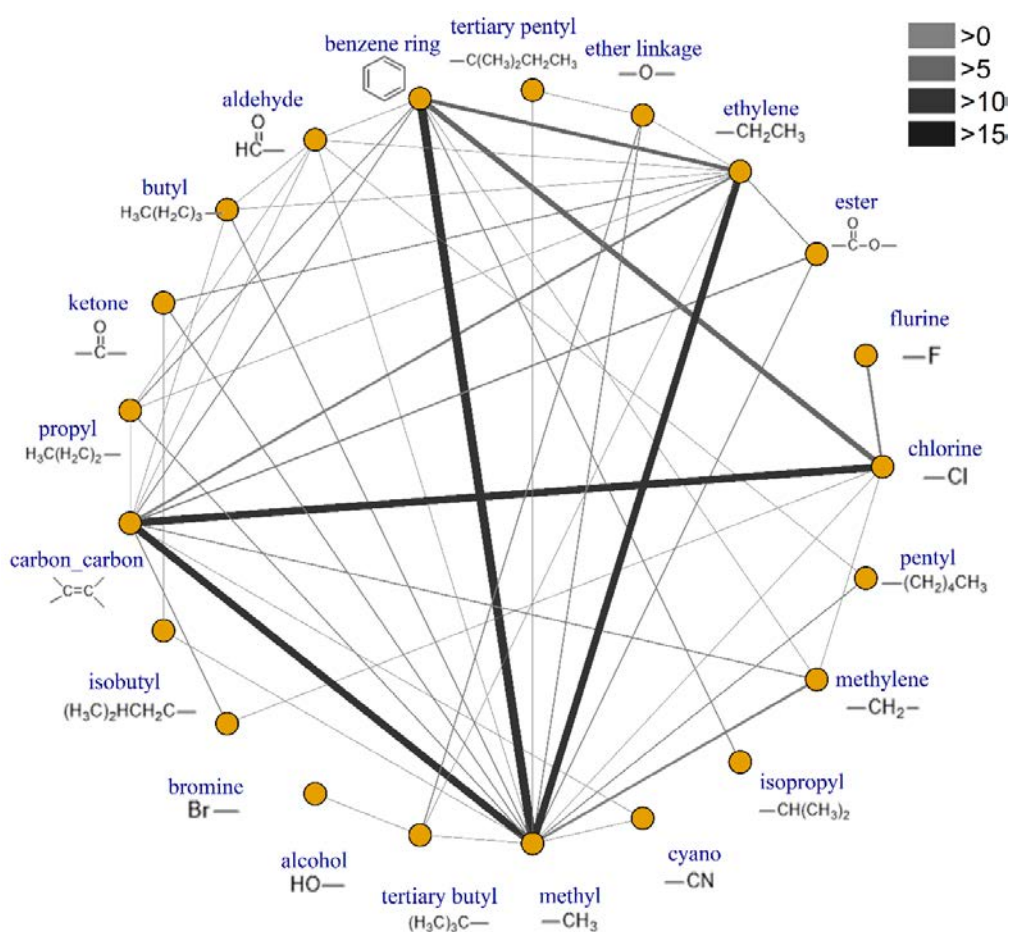


Fig. 1. Co-occurrence network of functional groups in examined VOCs. Each functional group is represented as a node and each co-occurrence of a pair of groups is demonstrated as a link. The number of times that a pair of groups co-occurs in the VOCs constitutes the weight of the link connecting the pair. The network displays a robust view of interactions among functional groups in examined VOCs. It facilitates uncovering meaningful component and connection bias based on the patterns and strength of links among functional groups.

VOC emission from industrial sources (such as Fermentation Exhaust from Penicillin Production), Methylene chloride can be used in adhesive removers and paint stripping products (Morose *et al.*, 2017; Guo *et al.*, 2018). After following the inhalation of methylene chloride, animal studies have shown increased ratio of cancers such as lung cancer and benign mammary gland tumors (<https://www.epa.gov/haps/health-effects-notebook-hazardous-air-pollutants>). In human, methylene chloride is utilized as a marker in exhaled breath VOCs to facilitate non-invasive diagnosis of chronic kidney disease, diabetes mellitus and healthy subjects (Saidi *et al.*, 2018). Hence, although it is possible to draw some conclusions about the chemical properties on molecules once their functional groups have been assessed, their roles on human health belongs to another kind of scenery. In the following sections, we will explore the potential mechanisms of VOCs on human health through analyzing their influenced genes.

Identification of Human Genes Influenced by VOCs

Although some VOCs have been recognized or suspected to influence human health, our understanding of the

underlying molecular mechanisms is still limited (Soni *et al.*, 2018). To explore possible mechanisms, the human genes response to VOCs were selected mainly based on the criteria that the expression of candidate gene was perturbed by at least three types of VOCs (see Methods Section). The frequency distribution of human genes in different groups are depicted in Fig. S1. The group “G3” denotes that the expression of individual gene in this group can influenced by three types of VOCs, and so forth. Obviously, there are only one gene (named *IL6*) in group “G9” and 3 genes (respectively, *BCL2*, *FOS* and *PTGS2*) in group “G8”. This implies that, in our datasets, these four genes are “hotspots” in response to VOCs. Interestingly, a plentiful literature has linked these genes to a variety of complex diseases. For instances, Interleukin 6 (*IL-6*) is a key cytokine which can not only stimulate immune responses and inflammation processes, but also play roles in regulation of tumor microenvironment (Choy and Rose-John, 2017; Rose-John *et al.*, 2017). *IL-6* was found associated with multiple diseases such as Alzheimer's Disease, depression, rheumatoid arthritis, diabetes and breast cancer (Hunter and Jones, 2015; Korneev *et al.*, 2017). B cell lymphoma 2 (*BCL-2*) is a cell survival

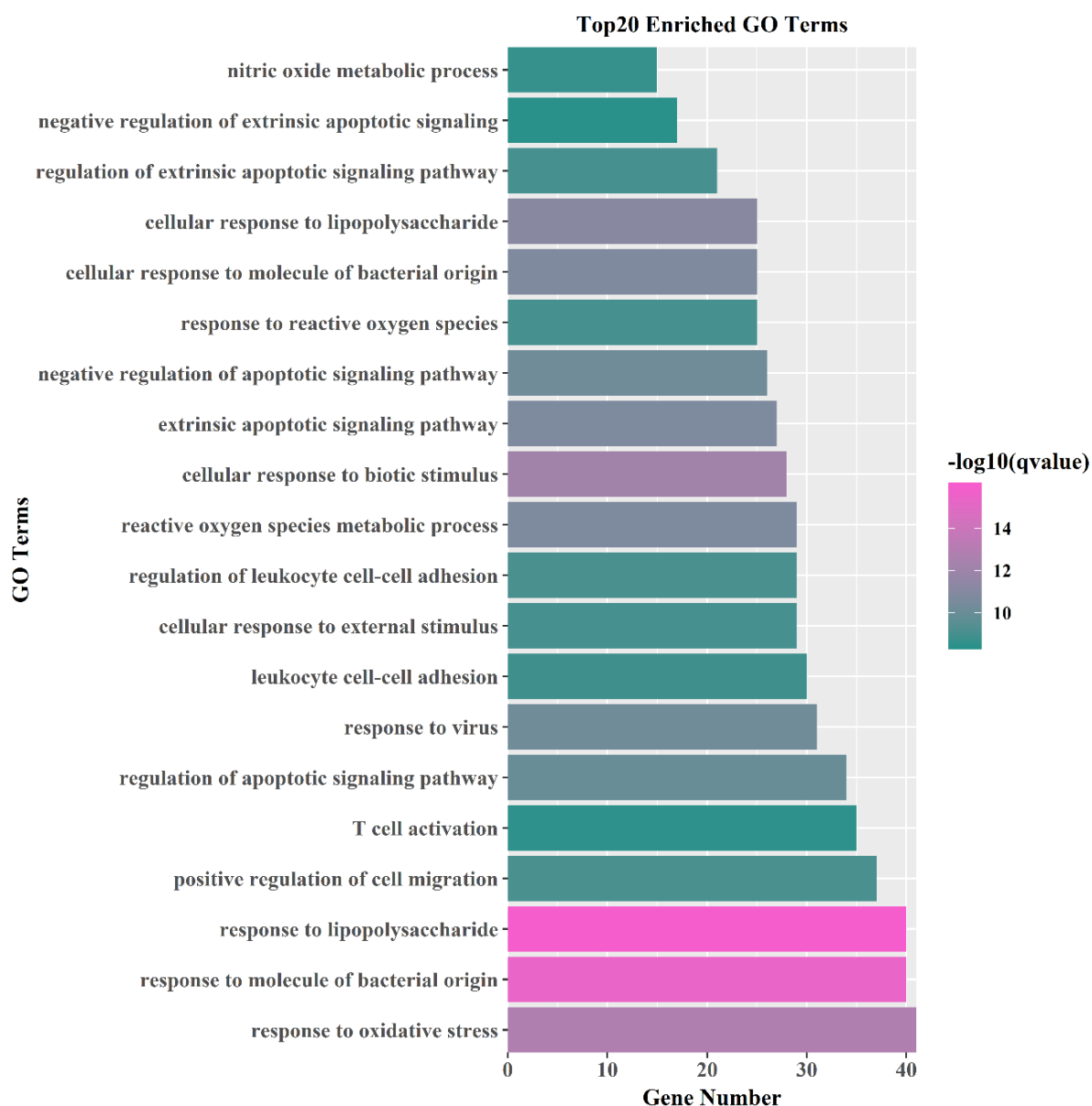


Fig. 3. Distribution of top 20 enriched Gene ontology (GO) Terms. The figure demonstrates top 20 significant GO terms (biological processes) associated with HGRV. The vertical axis represents the GO category, and the horizontal axis denotes the number of HGRV genes in the significant GO terms. The degree of color saturation of each bar is correlated with the qvalue. The lower the displayed qvalue, the more significant it is for the GO term to be.

production and gene expression) as stimulated by bacterial-derived molecules (<https://www.ebi.ac.uk/QuickGO/term/GO:0002237>). These two processes are mainly associated with immune response. When exposure to BTEX, oxidative stress produced may be one of the main mechanisms causing diseases like the loss of pulmonary function (Wallner *et al.*, 2012). However, as reviewed by Sies *et al.* (2017) oxidative stress is two sided: excessive oxidant level causing damage and normal physiological oxidant level governing vital life processes through redox signaling. Hence, VOCs may have negative or positive impacts on human health under different circumstances. Concrete mechanisms warrant future investigations. Moreover, the three canonical processes contain 40, 40, and 41 genes in HGRV, respectively.

Interestingly, both categories (response to lipopolysaccharide and response to molecule of bacterial origin) enclose 3 hub genes mentioned above, respectively, *IL6*, *FOS* and *PTGS2*. And all the four hub genes are embedded in the third process (response to oxidative stress).

Thus, on one hand, in line with our finding in the above section, the hub genes play crucial roles in enriched functional categories, such as immune and oxidative stress response. On the other hand, in addition to hub genes, these functional processes include multiple other HGRV. This implies that the implementation of the corresponding biological functions depends on the coordinated cooperation between these genes.

Analysis of Pathways Enriched in HGRV

To further unveil the coordinated cooperation between HGRV, we conducted the pathway mapping with KEGG database and recognized the enriched pathways in dataset HGRV. Top 20 enriched pathways were illustrated in bubble chart (See Fig. 4), in which size and color of the bubble represent amount of HGRV in individual enriched pathway and corresponding enrichment significance. Three most significant pathways are IL-17 signaling pathway ($p = 9.38E-12$), TNF signaling pathway ($p = 9.88E-10$) and Rheumatoid arthritis ($p = 1.92E-09$). Aberrant regulation of IL-17 (Interleukin 17) signaling pathway is associated with immunopathology, autoimmune disease and cancer

(Amatya et al., 2017). As a cell signaling protein, TNF (Tumor necrosis factor) is involved in systemic inflammatory response (such as response to lipopolysaccharide and other bacterial products), which in turn causes various types of human diseases including autoimmune disorders, Alzheimer's disease and cancer (Menon and Gaestel, 2017). Rheumatoid arthritis (RA) is a chronic autoimmune disease that primarily affects joints. It has been proven a critical role for tumor necrosis factor α and interleukin 6 in RA pathogenesis (McInnes and Schett, 2017). So, these three most significantly enriched pathways in HGRV are primarily associated with immune diseases and cancer, which is consistent with our forgoing analysis.

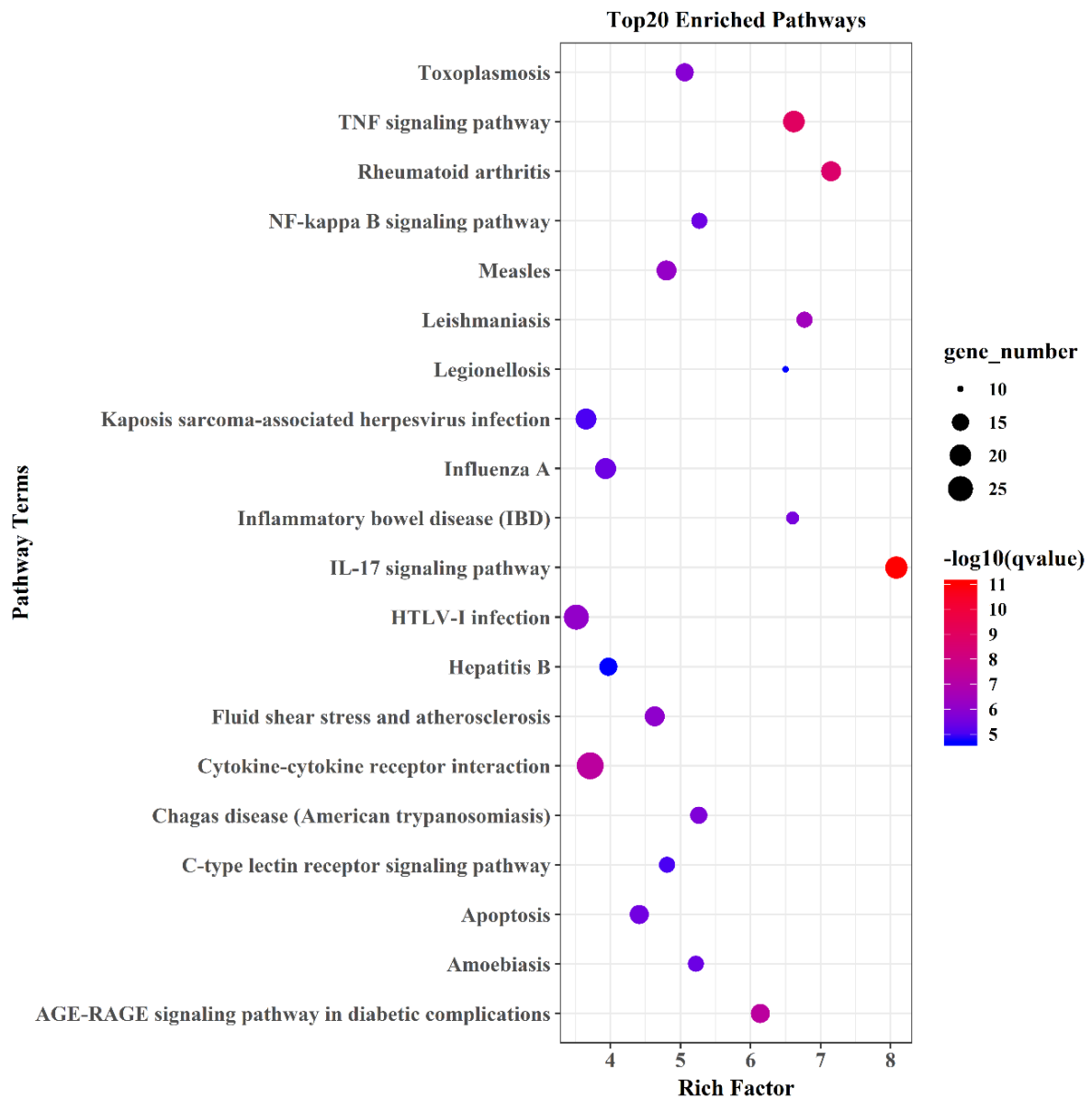


Fig. 4. Distribution of top 20 enriched KEGG pathways. Advanced bubble chart shows enrichment of HGRV in KEGG pathways. Y-axis label represents pathway category, and X-axis label denotes rich factor (rich factor = amount of HGRV in the pathway/amount of all human genes annotated in this pathway). Size and color of the bubble represent amount of HGRV enriched in pathway and enrichment significance, respectively. Greater- $\text{Log}_{10}(qvalue)$ scores correlated with increased statistical significance.

Furthermore, to reveal which HGRV are inlaid in these pathways, intuitive pathway maps were drawn and visualized with R package. Here, we use the pathway map of Rheumatoid arthritis as example. As displayed in Fig. 5, each box is given a gene symbol, and the color scale represents the normalized number of VOCs found for each HGRV. Clearly, the Rheumatoid arthritis may cause by more VOCs through “hotspot” genes marked in red. The genes marked in green, such as CCL20 and VEGF, can only response to fewer VOC types. Although other genes (marked in white) are also related to this disease, they are probably not VOC sensor. In addition, 17 sensor genes in Rheumatoid arthritis pathway can respond to multiple VOCs. On the VOC etiology (endogenous/exogenous), it is worth further research that which type of VOC or VOC combination is linked with an increased risk of Rheumatoid arthritis. Besides, many other disease pathways, such as Leishmaniasis, Measles, HTLV-I infection, Toxoplasmosis and Chagas disease (American trypanosomiasis), are enclosed in Top 20 enriched pathways. Probably, through modulating these pathways, the HGRV may participate in a variety of diseases. This inspires further analysis on HGRV related diseases.

Detection of Diseases Associated with HGRV

To analysis the diseases potentially associated with HGRV, annotation and identification of enriched diseases were performed with tool KOBAS 3.0 (Xie et al., 2011). The results revealed that 50 kinds of diseases might be significantly associated with HGRV ($p < 0.01$; see Table S2). Among them, top 3 significant diseases are Immune system diseases ($p = 9.48E-18$), Cancers ($p = 7.72E-14$) and Allergies and autoimmune diseases ($p = 8.27E-12$) respectively. It suggests that VOCs are more prone to induce these three representative diseases. This is consistent with our previous finding that significantly enriched pathways in HGRV are primarily associated with immune diseases and cancer. Moreover, through further capturing the relations between VOCs and diseases, we found these top 3 diseases might be induced by different VOC collections, respectively, Immune system diseases by 18 VOCs, Cancers by 15 VOCs and Allergies and autoimmune diseases by 17 VOCs. The common VOCs related to these three diseases are Acetaldehyde, Acetone, Acrolein, Benzene, butyraldehyde, Carbon Tetrachloride, chlorobenzene, Formaldehyde, propionaldehyde, Styrene, Tetrachloroethylene, Toluene and Trichloroethylene. On one hand, this result provides a caution that these VOCs, especially when present in our surrounding environments, should be paid more serious attention to avoid potential adverse health effects. And on other hand, it also signifies that there may exist common mechanisms among these diseases. This can be further sustained by the proposition that cancer can weaken the immune system through spreading into the bone marrow which makes blood cells help to fight infection (Thorsson et al., 2018). It is also worth mentioning that only one type of VOC (named Vinyl Chloride) is specific to Cancers. As narrated by International Agency for Research on Cancer (IARC; <https://www.cancer.gov/about-cancer/causes-preven>

tion/risk/substances/vinyl-chloride), Vinyl Chloride exposure is associated with an increased risk of hepatic angiosarcoma, as well as brain and lung cancers, lymphoma and leukemia. This alludes that dissimilar to immune-mediated mechanisms, specific pathway/mechanism may be implicated in cancers aroused from Vinyl Chloride.

CONCLUSIONS

In the present study, the ambient Volatile Organic Compounds and their human gene targets was analyzed from several perspectives, mainly including functional group composition of VOCs, human genes response to VOCs (HGRV), the biological function annotation and enriched GO Terms, enriched KEGG pathways and the diseases associated with HGRV. The results can be summarized as follows: (1) Through identifying composition and connectivity of functional groups within VOCs, it was revealed that there is significant bias in the co-occurrence patterns among the top five frequent groups (respectively methyl, carbon-carbon double bond, benzene ring, chlorine and ethylene). For instances, co-occurrence of chlorine and benzene are considerably frequent, whereas the connection between chlorine and methyl/ethylene occurred rarely. (2) In response to VOCs, four genes (respectively, *IL6*, *BCL2*, *FOS* and *PTGS2*) are “hotspots”. And by examining the VOCs which influenced the expression of these four hub genes, it was indicated that four types of VOCs (respectively, Acetaldehyde, Butyraldehyde, Formaldehyde and Styrene) may act as shared stimulators of the hotspots. (3) Then, through enrichment analysis on biological functions of HGRV, three dominant function categories were detected, respectively, response to lipopolysaccharide, response to molecule of bacterial origin and response to oxidative stress. And the hub genes play crucial roles in enriched functional categories, such as immune and oxidative stress response. (4) Further, the results revealed that three most significant pathways are IL-17 signaling pathway, TNF signaling pathway and Rheumatoid arthritis. These three most significantly enriched pathways in HGRV are primarily associated with immune diseases and cancer, which is consistent with our forgoing analysis. Probably, through modulating enriched pathways, the HGRV may participate in a variety of diseases. (5) Subsequently, by analyzing the diseases potentially associated with HGRV, the results revealed that 50 kinds of diseases might be significantly associated with HGRV ($p < 0.01$). Among them, top 3 significant diseases are Immune system diseases, Cancers and Allergies and autoimmune diseases respectively. The results further disclosed that these top 3 diseases might be induced by common VOCs, including Acetaldehyde, Acetone, Acrolein, Benzene, butyraldehyde, Carbon Tetrachloride, chlorobenzene, Formaldehyde, propionaldehyde, Styrene, Tetrachloroethylene, Toluene and Trichloroethylene. Overall, the findings in this study could be useful for abroad understanding of mechanisms related to diseases by VOCs exposure and development of control strategies on VOCs.

ACKNOWLEDGMENTS

This research was supported by the Fundamental Research Funds for the Central Universities (Grant No. JZ2017YYPY0899), Anhui Province 2017 Public Welfare Technology Application Research Program and Key Project for Excellent Youth Talent Support Program of Anhui Province Colleges and Universities.

SUPPLEMENTARY MATERIAL

Supplementary data associated with this article can be found in the online version at <http://www.aaqr.org>.

REFERENCES

- Abeira, A., Pollack, I., Sive, B., Zhou, Y., Fischer, E. and Farmer, D. (2017). Source characterization of volatile organic compounds in the Colorado Northern Front Range Metropolitan Area during spring and summer 2015. *J. Geophys. Res.* 122: 3595–3613.
- Alvim, D.S., Gatti, L.V., Corrêa, S.M., Chiquetto, J.B., Santos, G.M., de Souza Rossatti, C., Pretto, A., Rozante, J.R., Figueroa, S.N. and Pendharkar, J. (2018). Determining VOCs reactivity for ozone forming potential in the megacity of São Paulo. *Aerosol Air Qual. Res.* 18: 2460–2474.
- Amatya, N., Garg, A.V. and Gaffen, S.L. (2017). IL-17 signaling: The yin and the yang. *Trends Immunol.* 38: 310–322.
- Ashkenazi, A., Fairbrother, W.J., Levenson, J.D. and Souers, A.J. (2017). From basic apoptosis discoveries to advanced selective BCL-2 family inhibitors. *Nat. Rev. Drug Discovery* 16: 273.
- Benelli, R., Venè, R. and Ferrari, N. (2018). PTGS2 (cyclooxygenase-2), a complex target for colorectal cancer prevention and therapy. *Transl. Res.* 196: 42–61.
- Brooks, P. and Schuebel, K. (2017). Timeless insights into prevention of acetaldehyde genotoxicity? *Cell Cycle* 16: 308–309.
- Bryant, C.E., Spring, D.R., Gangloff, M. and Gay, N.J. (2010). The molecular basis of the host response to lipopolysaccharide. *Nat. Rev. Microbiol.* 8: 8–14.
- Choy, E. and Rose-John, S. (2017). Interleukin-6 as a multifunctional regulator: Inflammation, immune response, and fibrosis. *J. Scleroderma Relat. Disord.* 2: S1–S5.
- Davis, A.P., Grondin, C.J., Johnson, R.J., Sciaky, D., King, B.L., McMorran, R., Wieggers, J., Wieggers, T.C. and Mattingly, C.J. (2016). The comparative toxicogenomics database: update 2017. *Nucleic Acids Res.* 45: D972–D978.
- Guerra, F.D., Smith, G.D. Jr, Alexis, F. and Whitehead, D.C. (2017). A survey of VOC emissions from rendering plants. *Aerosol Air Qual. Res.* 17: 209–217.
- Guo, B., Yao, R., Zhang, S., Ma, L., Kang, H. and Wang, S. (2018). Pollution condition and health risk assessment of VOCs in fermentation exhaust from penicillin production. *Environ. Sci.* 39: 3102–3109. (in Chinese)
- Hunter, C.A. and Jones, S.A. (2015). IL-6 as a keystone cytokine in health and disease. *Nat. Immunol.* 16: 448.
- Jenkin, M., Derwent, R. and Wallington, T. (2017). Photochemical ozone creation potentials for volatile organic compounds: Rationalization and estimation. *Atmos. Environ.* 163: 128–137.
- Jiménez-Garza, O., Guo, L., Byun, H.-M., Carrieri, M., Bartolucci, G.B., Barrón-Vivanco, B.S. and Baccarelli, A.A. (2018). Aberrant promoter methylation in genes related to hematopoietic malignancy in workers exposed to a VOC mixture. *Toxicol. Appl. Pharmacol.* 339: 65–72.
- Juarez, E., Chambwe, N., Tang, W., Mitchell, A.D., Owen, N., Kumari, A., Monnat, R.J. and McCullough, A. (2018). An RNAi screen in human cell lines reveals conserved DNA damage repair pathways that mitigate formaldehyde sensitivity. *bioRxiv* 310730.
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2016). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45: D353–D361.
- Kim, J.H., Moon, J.Y., Park, E.Y., Lee, K.H. and Hong, Y.C. (2011). Changes in oxidative stress biomarker and gene expression levels in workers exposed to volatile organic compounds. *Ind. Health* 49: 8–14.
- Korneev, K.V., Atrekhany, K.S.N., Drutskaya, M.S., Grivennikov, S.I., Kuprash, D.V. and Nedospasov, S.A. (2017). TLR-signaling and proinflammatory cytokines as drivers of tumorigenesis. *Cytokine* 89: 127–135.
- Li, B., Ho, S.S.H., Xue, Y., Huang, Y., Wang, L., Cheng, Y., Dai, W., Zhong, H., Cao, J. and Lee, S. (2017). Characterizations of volatile organic compounds (VOCs) from vehicular emissions at roadside environment: The first comprehensive study in Northwestern China. *Atmos. Environ.* 161: 1–12.
- Liu, Z.G., Jiang, G., Tang, J., Wang, H., Feng, G., Chen, F., Tu, Z., Liu, G., Zhao, Y. and Peng, M.J. (2016). c-Fos over-expression promotes radioresistance and predicts poor prognosis in malignant glioma. *Oncotarget* 7: 65946.
- Mannucci, P.M. and Franchini, M. (2017). Health effects of ambient air pollution in developing countries. *Int. J. Environ. Res. Public Health* 14: 1048.
- McInnes, I.B. and Schett, G. (2017). Pathogenetic insights from the treatment of rheumatoid arthritis. *The Lancet* 389: 2328–2337.
- Mečiarová, E., Vilčeková, S., Křídlová Burdová, E. and Kiselák, J. (2017). Factors effecting the total volatile organic compound (TVOC) concentrations in slovak households. *Int. J. Environ. Res. Public Health* 14: 1443.
- Menon, M.B. and Gaestel, M. (2017). MK2-TNF-signaling comes full circle. *Trends Biochem. Sci.* 43: 170–179.
- Mochalski, P., Unterkofler, K., Teschl, G. and Amann, A. (2015). Potential of volatile organic compounds as markers of entrapped humans for use in urban search-and-rescue operations. *TrAC, Trends Anal. Chem.* 68: 88–106.
- Montero-Montoya, R., López-Vargas, R. and Arellano-Aguilar, O. (2018). Volatile organic compounds in air:

- Sources, distribution, exposure and associated illnesses in children. *Ann. Global Health* 84: 225–238.
- Morose, G., Marshall, J., McCarthy, A., Harripersaud, V. and Giarrosso, A. (2017). Assessment of safer and effective alternatives to methylene chloride for paint stripping products. TURI Report 2017-102, Toxics Use Reduction Institute, University of Massachusetts Lowell, USA.
- Muhammad, N., Bhattacharya, S., Steele, R., Phillips, N. and Ray, R.B. (2016). Involvement of c-Fos in the promotion of cancer stem-like cell properties in head and neck squamous cell carcinoma. *Clin. Cancer Res., in Press*.
- Neghab, M., Nourozi, M.A., Shahtaheri, S.J., Mansoori, Y., Bazzaz, J.T. and Nedjat, S. (2018). Effects of Genetic Polymorphism on Susceptibility to Nephrotoxic Properties of BTEXs Compounds. *J. Occup. Environ. Med.* 60: e377–e382.
- Nett, R.J., Edwards, N.T., Ruder, A.M., Bertke, S.J., Keumala, I., Cox-Ganser, J. and Cummings, K.J. (2017). Deaths from nonmalignant respiratory disease in styrene-exposed workers: Does obliterative bronchiolitis contribute to mortality? *Ann. Ame. Thoracic Soc.* 14: 810–811.
- Niederbacher, B., Winkler, J.B. and Schnitzler, J.P. (2015). Volatile organic compounds as non-invasive markers for plant phenotyping. *J. Exp. Bot.* 66: 5403–5416.
- Odabasi, M. (2008). Halogenated volatile organic compounds from the use of chlorine-bleach-containing household products. *Environ. Sci. Technol.* 42: 1445–1451.
- Orhan, H., van Holland, B., Krab, B., Moeken, J., Vermeulen, N.P., Hollander, P. and Meerman, J.H. (2004). Evaluation of a multi-parameter biomarker set for oxidative damage in man: increased urinary excretion of lipid, protein and DNA oxidation products after one hour of exercise. *Free Radical Res.* 38: 1269–1279.
- Ou-Yang, C.F., Chang, C.C., Wang, J.L., Shimada, K., Hatakeyama, S., Kato, S., Chiu, J.Y., Sheu, G.R. and Lin, N.H. (2017). Characteristics of summertime volatile organic compounds in the lower free troposphere: Background measurements at Mt. Fuji. *Aerosol Air Qual. Res.* 17: 3037–3051.
- Pinto-Proença, R., Santos, M., Fonseca, C., Fernandes, J., Gaspar, M.F. and Proença, R. (2018). Conjunctival melanoma: Association of cyclooxygenase-2 tumor expression to prognosis. *Graefe's Arch. Clin. Exp. Ophthalmol.* 256: 989–995.
- Pontel, L. (2018). Formaldehyde metabolism and its impact on human health. *Curr. Opin Toxicol.* 9: 28–34.
- Reed, J.C. (2018). Bcl-2 on the brink of breakthroughs in cancer treatment. *Cell Death Differ.* 25: 3.
- Ren, Y., Qu, Z., Du, Y., Xu, R., Ma, D., Yang, G., Shi, Y., Fan, X., Tani, A., Guo, P., Ge, Y. and Chang, J. (2017). Air quality and health effects of biogenic volatile organic compounds emissions from urban green spaces and the mitigation strategies. *Environ. Pollut.* 230: 849–861.
- Rose-John, S., Winthrop, K. and Calabrese, L. (2017). The role of IL-6 in host defence against infections: Immunobiology and clinical implications. *Nat. Rev. Rheumatol.* 13: 399–409.
- Saidi, T., Zaim, O., Moufid, M., El Bari, N., Ionescu, R. and Bouchikhi, B. (2018). Exhaled breath analysis using electronic nose and gas chromatography-mass spectrometry for non-invasive diagnosis of chronic kidney disease, diabetes mellitus and healthy subjects. *Sens. Actuators, B* 257: 178–188.
- Santos, M., Larentis, A. and Perini, J. (2018). 1500 Prediction and characterisation of biomarker network for benzene exposure. *Occup. Environ. Med.* 75: A411.
- Sherwen, T., Schmidt, J.A., Evans, M.J., Carpenter, L.J., Großmann, K., Eastham, S.D., Jacob, D.J., Dix, B., Koenig, T.K. and Sinreich, R. (2016). Global impacts of tropospheric halogens (Cl, Br, I) on oxidants and composition in GEOS-Chem. *Atmos. Chem. Phys.* 16: 12239–12271.
- Sies, H., Berndt, C. and Jones, D.P. (2017). Oxidative stress. *Annu. Rev. Biochem.* 86: 715–748.
- Soni, V., Singh, P., Shree, V. and Goel, V. (2018). Effects of VOCs on human health. In *Air pollution and control*, Sharma, N., Agarwal, A.K., Eastwood, P., Gupta, T. and Singh, A.P. (Eds.), Springer Singapore, Singapore, pp. 119–142.
- Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Yang, T.H.O., Porta-Pardo, E., Gao, G.F., Plaisier, C.L. and Eddy, J.A., et al. (2018). The immune landscape of cancer. *Immunity* 48: 812–830.e814.
- Tibaquirá, J.E., Huertas, J.I., Ospina, S., Quirama, L.F. and Niño, J.E. (2018). The effect of using ethanol-gasoline blends on the mechanical, energy and environmental performance of in-use vehicles. *Energies* 11: 221.
- Tong, Z. (2017). Urine formaldehyde: A non-invasive marker for Alzheimer's disease. *J. Alzheimers Dis. Parkinsonism* 7: 345.
- Villanueva-Rosales, N. and Dumontier, M. (2007). Describing chemical functional groups in OWL-DL for the classification of chemical compounds. Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions, June 6-7, 2007, Innsbruck, Austria.
- Wallner, P., Kundi, M., Moshhammer, H., Piegler, K., Hohenblum, P., Scharf, S., Fröhlich, M., Damberger, B., Tappler, P. and Hutter, H.P. (2012). Indoor air in schools and lung function of Austrian school children. *J. Environ. Monit.* 14: 1976–1982.
- Wang, G., Cheng, S., Wei, W., Zhou, Y., Yao, S. and Zhang, H. (2016). Characteristics and source apportionment of VOCs in the suburban area of Beijing, China. *Atmos. Pollut. Res.* 7: 711–724.
- Wang, H., Qiao, Y.Z., Chen, C.H., Lu, J., Dai, H.X., Qiao, L.P., Lou, S.R., Huang, C., Li, L. and Jing, S. (2014). Source profiles and chemical reactivity of volatile organic compounds from solvent use in Shanghai, China. *Aerosol Air Qual. Res.* 14: 301–310.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T. and Hindorff, L. (2013). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic*

- Acids Res.* 42: D1001–D1006.
- Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C.Y. and Wei, L. (2011). KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39: W316–322.
- Ye, D., Klein, M., Chang, H.H., Sarnat, J.A., Mulholland, J.A., Edgerton, E.S., Winkler, A., Tolbert, P.E. and Sarnat, S.E. (2017). Estimating acute cardiorespiratory effects of ambient volatile organic compounds. *Epidemiology*. *Epidemiology* 28: 197–206.
- Yu, G., Wang, L.G., Han, Y. and He, Q.Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* 16: 284–287.
- Zhang, G., Wang, N., Jiang, X. and Zhao, Y. (2017a). Characterization of ambient volatile organic compounds (VOCs) in the area adjacent to a petroleum refinery in Jinan, China. *Aerosol Air Qual. Res.* 17: 944–950.
- Zhang, X., Xue, Z., Li, H., Yan, L., Yang, Y., Wang, Y., Duan, J., Li, L., Chai, F. and Cheng, M. (2017b). Ambient volatile organic compounds pollution in China. *J. Environ. Sci.* 55: 69–75.
- Zhang, Y., Li, R., Fu, H., Zhou, D. and Chen, J. (2018). Observation and analysis of atmospheric volatile organic compounds in a typical petrochemical area in Yangtze River Delta, China. *J. Environ. Sci.* 71: 233–248.
- Zheng, C., Shen, J., Zhang, Y., Zhu, X., Wu, X., Chen, L. and Gao, X. (2017). Atmospheric emission characteristics and control policies of anthropogenic VOCs from industrial sources in Yangtze River Delta Region, China. *Aerosol Air Qual. Res.* 17: 2263–2275.

Received for review, August 29, 2018

Revised, September 23, 2018

Accepted, September 23, 2018