An integrated genome-wide model for hookah constituents-gene-pathway-disease relationships

Sing-Han Huang

Institute of Bioinformatics and Systems Biology, National Chiao Tung University

Yi-Hsuan Chuang

Institute of Bioinformatics and Systems Biology, National Chiao Tung University

Yong-Chun Luo

Institute of Bioinformatics and Systems Biology, National Chiao Tung University

Wei-Ting Liu

Institute of Bioinformatics and Systems Biology, National Chiao Tung University

Jung-Yu Lee

Institute of Bioinformatics and Systems Biology, National Chiao Tung University

Terry Gordon

Department of Environmental Medicine, New York University School of Medicine

Hong Sun

Department of Environmental Medicine, New York University School of Medicine

Nikhil Pathak

Institute of Bioinformatics and Systems Biology, National Chiao Tung University

Wen-Sen Lai

Department of Otolaryngology–Head and Neck Surgery, Taichung Armed Forces General Hospital

Lung-Chi Chen

Department of Environmental Medicine, New York University School of Medicine

Jinn-Moon Yang (✉ moon@faculty.nctu.edu.tw)

Institute of Bioinformatics and Systems Biology, National Chiao Tung University

Research Article

Keywords: respiratory diseases, RNA, HiSBiM, hierarchical-systems

Posted Date: December 17th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-127747/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Hookah smoking is becoming increasingly popular worldwide and recent epidemiological studies indicate that it is significantly associated with respiratory diseases, and could contain constituents more harmful than cigarette smoking. However, the gene expression alterations and mechanisms of disease development induced by the hookah constituents are still unclear. Here, we are the first to propose a new experimental design and use RNA sequencing to profile bronchial epithelial cells from different hookah smoking exposure conditions (GSE162386). For our hookah-smoking dataset, we developed an integrated genome-wide model, hierarchical-systems biology model (HiSBiM), to systematically investigate the effects of smoking hookah and to compare the biological pathways with public cigarette smoking datasets. Our results show that the infectious disease-related pathways were enriched in hookah smoking than in cigarette smoking, which could be due to polluted hookah devices or pathogen contaminated water filters. Furthermore, we investigate the subsystem and pathway differences in hookah smoking between charcoal and electronic heating ways. We observed that cardiovascular and atherosclerosis-related pathways were activated after exposure to charcoal combustion, which may relate to combustion components carbon monoxide and polyaromatic hydrocarbons. Overall, our results show that hookah smoking could involve a higher risk in several diseases than cigarette smoking and link the relations between hookah constituents to diseases. We believe that our HiSBiM is a useful integrated method for providing multiple-level valued insights for genome-wide analysis on various omics data sets.

Introduction

Hookah smoking is becoming increasingly popular worldwide. Recent epidemiological studies have indicated that hookah smoking is significantly associated with cancer, cardiovascular and lung diseases, and may contain more harmful constituents than cigarette smoking¹⁻⁵. Besides, some reports have indicated that toxic substances of hookah smoking are generated from charcoal emissions by burning charcoal together with tobacco, which yielded about 15 folds carbon monoxide (CO) and polyaromatic hydrocarbons (PAHs) and 3 folds nicotine compared to cigarette smoking⁶⁻⁸. However, traditional epidemiology or biology alone are not sufficient to elucidate the underlying mechanisms of hookah smoking and its constituents. Large omics data has led to potential use of systems biology approaches for a better understanding of biological mechanisms of hookah smoking. Currently, the omics-based insights exploring the relationships between hookah smoking, hookah constituents and cigarette smoking are still lacking.

To address these issues, we enhanced the omics data analysis strategy, Hierarchical-Systems Biology Model (HiSBiM)⁹, to explore the relationships between cigarette, hookah and hookah constituents-gene/protein-pathway-disease. In this study, the bronchial epithelial cell line (BEAS-2B) was used to expose five different hookah conditions in 2 and 24 hours, such as charcoal with no shisha, charcoal with non-tobacco (herbal) shisha, and charcoal with tobacco shisha, and all of the samples were performed NGS to study the effects and mechanisms of hookah smoking and its constituents (e.g., CO, PAH). The HiSBiM showed that hookah smoking mainly induced cancer, bacterial or viral infections, and

cardiovascular disease-related subsystems, as well as significantly changed cell cycle-related genes (e.g., CCND1) in viral infection pathway. We also observed that CO and PAH constituents in hookah smoking are closely related to cardiovascular diseases, such as atherosclerosis. In summary, we believe that our HiSBiM model was useful in genome-wide data analysis of hookah and cigarette smoking for elaborating disease mechanisms.

Materials And Methods

The overview of our method for investigating hookah smoking and its constituents are described in Fig. 1, the key steps summarized in Fig. 1A. To explore the effect of hookah smoking and its constituents, we designed cell-based experiments for hookah exposure in two heating ways and tobacco or herbal shisha, and all of the 30 samples were performed RNA sequencing to generate gene expression profiles and have been deposited to Gene Expression Omnibus¹⁰ (GEO accession GSE162386) (Fig. 1B). Cigarette smoking datasets were collected from public omics databases (Fig. 1C). Then, we identified DEGs from our hookah-smoking gene expression profiles and public cigarette smoking datasets, and evaluated pathway enrichment based on the hypergeometric distribution (Figs. 1D and 1E). Next, the *p*-values were transformed into z-scores to quantify KEGG pathways (Fig. 1F). Finally, the meta-z-scores of subsystem- and system-levels were calculated from pathway-level z-scores to quantify subsystems and systems (Fig. 1G).

Hookah exposure experimental design

To study the harmful effects of hookah smoking on cells, we designed cell-based experiments for hookah smoking exposure with various constituents and treatment times (Fig. 1B; Table 1). The bronchial epithelial cells (BEAS-2B) were cultured and exposed with and without tobacco in two heating ways of hookah smoking, including internal control (IC), charcoal with no shisha (C), charcoal with non-tobacco (herbal) shisha (CNT), charcoal with tobacco shisha (CT), and electronic heater with tobacco shisha (ET) for 30 min while on a rocker. The exposure concentration was measured as the highest that caused no toxicity at 24 hours, as assayed by LDH release. For each exposure condition, three wells were treated with Trizol at 2 and 24 hours after exposure. RNA extraction was performed by the RNeasy Plus Mini Kit, Qiagen, according to the manufacturer's instructions.

Table 1
No. of samples of our in-house datasets.

In-house datasets (GSE162386)		
Groups	After exposure 2hr	After exposure 24hr
Control (IC)	3	3
Charcoal (C)	3	3
Charcoal + non-tobacco (CNT)	3	3
Charcoal + tobacco (CT)	3	3
Electronic + tobacco (ET)	3	3

Next generation sequencing

All of the 30 samples were performed RNA-seq to analyze whole-genome transcriptome using Illumina HiSeq 2500 platform with 50 bp single-end reads (Illumina, USA). For each sample, the reads were aligned with the human reference genome (GRCh38) using the HISAT2 package¹¹ and the mapped reads were assembled into transcripts using StringTie¹². The average mapping rate of reads was 90%. Next, the expression profiles of all transcripts were estimated and calculated based on fragments per kilobase of transcript per million fragments mapped (FPKM) by StringTie and Ballgown package¹³ in R. Expression profiles have been deposited to GEO (<https://www.ncbi.nlm.nih.gov/geo/>; accession no. GSE162386).

Cigarette smoking-related public dataset

For comparing the mechanisms between cigarette and hookah smoking, we collected three cigarette-related microarray datasets, GSE63127, E-MTAB-4742, E-MTAB-4740, from GEO and ArrayExpress databases^{10,14} (Fig. 1C; Table 2), all of which used a Affymetrix Human Genome U133 Plus 2.0 platform. GSE63127 includes 143 non-smoking (NS) and 87 smoking (S) small airway epithelium samples (Cig1); E-MTAB-4742 contained 47 non-smoking and 48 smoking oral tissues (Cig2); E-MTAB-4740 is composed of 52 non-smoking and 54 smoking nasal tissues (Cig3).

Table 2
No. of samples of public datasets used in this study.

Public datasets		
Dataset	Non-smoking	Smoking
GSE63127 (Cig1)	143	87
E-MTAB-4740 (Cig2)	47	48
E-MTAB-4740 (Cig3)	87	54

Hierarchical-systems biology model

For omics data analysis, we first identified the differentially expressed genes (DEGs) in each compared group with $|\log_2\text{FC}| > 0.58$ and $p\text{-value} < 0.05$ by using the limma package¹⁵ (see Supplementary Table S1 online). Based on our identified DEGs, the hypergeometric distribution was used for pathway enrichment analysis and pathways with $p\text{-value} < 0.05$ were considered significant^{16,17} (Fig. 1D). Here, the $p\text{-value}$ is calculated as (1):

$$p = \sum_{i=x}^n \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}} \quad (1)$$

where N is the number of genes recorded in KEGG, M is the number of DEGs, n is the number of genes in the specific KEGG pathway, and i is the number of genes that are DEGs in the specific KEGG pathway. The $p\text{-value}$ of each pathway was transformed into the z-score using the standard normal distribution (Fig. 2). We then computed the subsystem- and system-level meta-z-scores for 45 KEGG subsystems and 6 systems (Figs. 1E and 1F), such as *metabolism*, *genetic information processing*, and *human diseases*, based on our previous study⁹. The meta z-score is defined as (2):

$$\text{meta-z} = \frac{\sum_{i=1}^n z_i}{\sqrt{n}} \quad (2)$$

where z_i is the z-score of the pathway i , and n is the total number of pathways in a subsystem or system. The meta-z-score reflects the significance (enrichment) of a subsystem or system for a specific condition or disorder, such as hookah smoking.

Results

HiSBiM for analyzing the mechanisms of hookah smoking

We used HiSBiM to observe and compare omics of hookah and cigarette smoking via the hierarchical manner. There are two groups of hookah smoking, including 2 hours (H1) and 24 hours (H2) after exposure, and three cigarette smoking datasets (Cig1, 2, and 3). We observed that the two system-level displayed the differences in *Human Diseases* (mean meta-z-scores 12.4 and 5.3, respectively) and *Environmental Information Processing* (mean meta-z-scores 6.6 and 2.5) between hookah smoking and cigarette smoking (Fig. 3A).

In subsystem-level of *Human Diseases*, the top 5 subsystems of hookah compared to cigarette smoking were mainly related to *Cancers: Specific types* (mean meta-z-scores 9.6 and 2.5), *Infectious diseases: Viral* (mean meta-z-scores 6.0 and 1.4), and *Cardiovascular diseases* (mean meta-z-scores 6.22 and 3.16) (Fig. 3B). This is in agreement with hookah smoking being indicated as a risk factor for cancers,

respiratory infections and cardiovascular system³. Also, higher frequency of hookah smoking is associated with the risk increasing of esophageal cancer, and an average of six folds higher risk linked to lung cancer^{1,2,4,5}. Interestingly, we observed that the pathway with the major difference of hookah versus cigarette smoking was *Human papillomavirus infection* (mean z-scores 3.13 and 0.24) belonging to subsystem of *Infectious diseases: Viral* (Fig. 3C). These phenomena might be caused by polluted hookah devices or microbe contaminated water filters. The similar observation of hookah and papillomavirus infection relationship had been reported by some epidemiological studies^{18,19}, but still lack of omics-based analysis for exploring the mechanisms. Among the DEGs obtained here, we observed that CCND1 (log2FC: 1.21 and 1.00), CCND3 (log2FC: 0.76 and 0.59), and LAMB3 (log2FC: 0.64 and 0.47) were highly expressed after hookah exposure for 2 and 24 hours, respectively (Fig. 3D). The genes CCND1 and CCND3 are cyclin D genes, coding for cell-cycle regulatory proteins involved in the control of the G1/S phase transition²⁰. LAMB3 encodes the beta 3 subunit of laminin, reported to mediate the attachment, migration and organization of cells into tissues during embryonic development by interacting with other extracellular matrix components²¹. Also, in hookah exposed case of 24 hours, PIK3R1 was up-regulated (log2FC: 0.74) which encodes the PI3K regulatory subunit p85 α involving in proliferation, apoptosis, and metabolism functions²². For linking the associations of genes-pathway-disease of hookah smoking, we mapped hookah smoking DEGs to *Human papillomavirus infection* pathway. The results showed that hookah smoking may aggravate pathogen infection by increasing cell cycle and proliferation (Fig. 3D).

Hookah constituents-gene-pathway-disease relationships

For investigating the relationship between hookah smoking, its constituents, and cigarette smoking, we clustered eight groups, including hookah, CO and PAH (CP1 and CP2), nicotine (N1 and N2), and cigarette (Cig1, Cig2 and Cig3), based on their corresponding subsystem-level meta-z-scores (Fig. 4A). The results showed that hookah and cigarette groups were clearly distinguished into two clusters, which represent a significant difference in the mechanisms of hookah and cigarettes smoking. The hookah, CO and PAH, and nicotine groups were further divided into two clusters: 1) Hookah and CO and PAH grouped in the same cluster; 2) Nicotine groups are an independent cluster.

By comparing the top 10 subsystems, we found that CO and PAH (CP) dataset shared 7 subsystems with hookah smoking dataset (Jaccard index: 0.54) (Fig. 4B). This means that hookah constituents CO and PAH may have a high correlation with the effect mechanisms of hookah smoking use. The CO and PAH mainly effected certain specific subsystems, including *Cardiovascular diseases* (meta-z-score: 2.03), *Endocrine system* (meta-z-score: 1.96), and *Amino acid metabolism* (meta-z-score: 0.98). Furthermore, we found that *Fluid shear stress and atherosclerosis* with the highest z-score (12.17) in pathway-level of *Cardiovascular diseases* (Fig. 4C). We then mapped the DEGs of CO and PAH dataset into KEGG pathway and observed that hookah constituents could link to atherogenesis (Fig. 4D). The associations between PAH and atherogenesis also have been reported where the PAHs-exposed mice showed increased macrophage content in atherosclerotic plaques²³. Here in this pathway, we find that Transcription factor AP-1 (JUN) was activated (log2FC: 1.61) after 2 hours exposure to hookah smoking (i.e., CO and PAH) triggering promoter down-stream gene expression (e.g., chemokines). In long term effect of hookah

smoking exposure, CCL2 was up-regulated (log2FC: 0.65) which induced pro-atherogenesis, inflammatory response and adhesion of leukocytes to the endothelium.

Discussion And Conclusion

Some epidemiological studies have demonstrated that hookah smoking is related to cancer, cardiovascular and lung diseases^{1–5}. Besides, toxic substances of hookah smoking are yielded from charcoal emissions, which produces about 15 folds CO and PAHs compared to cigarette smoking^{6–8}. Currently, there is still a lack of systems biology approaches to underlying the relationship between hookah smoking, its constituents, and disease. In this current research study, we are the first to propose a new experimental design and use RNA sequencing to profile bronchial epithelial cell from hookah smoking generated with different heating ways and tobacco or herbal shisha, including 30 samples among five conditions (GEO accession GSE162386). For our hookah-smoking dataset, we developed an integrated genome-wide model, called hierarchical-systems biology model (HiSBiM) with four-layered analysis, to systematically investigate the effects of smoking hookah and to compare the biological pathways with several public cigarette smoking datasets. Based on HiSBiM, we identified DEGs and pathways induced by hookah smoking and demonstrated their main differences in infection-related subsystems with cigarette smoking. These most likely is caused by polluted hookah devices or pathogen contaminated water filters. We also showed that cardiovascular disease and atherosclerosis were activated after acute exposure to hookah smoking by charcoal combustion. These results may relate to combustion components carbon monoxide and polyaromatic hydrocarbons compared to hookah smoking by electronic way. Overall, our results show that hookah smoking could involve a higher risk in several diseases than cigarette smoking and link the relations between hookah constituents to diseases. We believe that our HiSBiM is a useful integrated method for providing multiple-level valued insights for genome-wide analysis on various omics data sets.

Declarations

Acknowledgment

We are thankful for the funding and computing resources supported by the Ministry of Science and Technology (MOST 109-2634-F-009-021) in Taiwan; supported in part by the Artificial Intelligence to Precision Health: Integrating Dynamic Physiological Signals and EMR to build a Medical Digital Twins Platform (MOST 109-2321-B-009-007); the National Health Research Institutes (NHRI-EX109-10504PI), and the Taichung Armed Forces General Hospital (108-A15); the Center for Bioinformatics Research (CBR) by Ministry of Education ATU program and National Chiao Tung University; and the Center for Intelligent Drug Systems and Smart Bio-devices (IDS2B) of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

Author contributions

S.-H.H. analyzed the omics data and wrote the manuscript; Y.-H.C. performed NGS normalization processes and generated gene expression profiles; Y.-C.L and W.-T.L. collected public omics data; J.-Y.L., W.-S.L. and N.P. provided discussion on biological meaning; T.G. and H.S. designed and performed cell-based experiments. J.-M.Y and L.-C.C. designed and supervised the study, provided the intellectual contribution, discussion and edited the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

Expression profiles have been deposited to GEO (<https://www.ncbi.nlm.nih.gov/geo/>; accession no. GSE162386).

References

1. Malik, M. A., Upadhyay, R., Mittal, R. D., Zargar, S. A. & Mittal, B. Association of xenobiotic metabolizing enzymes genetic polymorphisms with esophageal cancer in Kashmir Valley and influence of environmental factors. *Nutr Cancer* **62**, 734–742, doi:10.1080/01635581003605904 (2010).
2. Nasrollahzadeh, D. *et al.* Opium, tobacco, and alcohol use in relation to oesophageal squamous cell carcinoma in a high-risk area of Iran. *Br J Cancer* **98**, 1857–1863, doi:10.1038/sj.bjc.6604369 (2008).
3. Qasim, H. *et al.* The effects of hookah/waterpipe smoking on general health and the cardiovascular system. *Environ Health Prev Med* **24**, 58, doi:10.1186/s12199-019-0811-y (2019).
4. Lubin, J. H. *et al.* Risk of lung cancer among cigarette and pipe smokers in southern China. *Int J Cancer* **51**, 390–395, doi:10.1002/ijc.2910510310 (1992).
5. Aoun, J., Saleh, N., Waked, M., Salame, J. & Salameh, P. Lung cancer correlates in Lebanese adults: a pilot case–control study. *J Epidemiol Glob Health* **3**, 235–244, doi:10.1016/j.jegh.2013.06.005 (2013).
6. Monzer, B., Sepetdjian, E., Saliba, N. & Shihadeh, A. Charcoal emissions as a source of CO and carcinogenic PAH in mainstream narghile waterpipe smoke. *Food Chem Toxicol* **46**, 2991–2995, doi:10.1016/j.fct.2008.05.031 (2008).
7. Nguyen, T. *et al.* Charcoal burning as a source of polyaromatic hydrocarbons in waterpipe smoking. *J Environ Sci Health B* **48**, 1097–1102, doi:10.1080/03601234.2013.824300 (2013).
8. Shihadeh, A. *et al.* Toxicant content, physical properties and biological activity of waterpipe tobacco smoke and its tobacco-free alternatives. *Tob Control* **24 Suppl 1**, i22–i30, doi:10.1136/tobaccocontrol-2014-051907 (2015).

9. Yang, W. Y. *et al.* Omics-based Investigation of Diet-induced Obesity Synergized with HBx, Src, and p53 Mutation Accelerating Hepatocarcinogenesis in Zebrafish Model. *Cancers (Basel)* **11**, doi:10.3390/cancers11121899 (2019).
10. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res* **41**, D991–995, doi:10.1093/nar/gks1193 (2013).
11. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* **11**, 1650–1667, doi:10.1038/nprot.2016.095 (2016).
12. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295, doi:10.1038/nbt.3122 (2015).
13. Frazee, A. C. *et al.* Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat Biotechnol* **33**, 243–246, doi:10.1038/nbt.3172 (2015).
14. Athar, A. *et al.* ArrayExpress update - from bulk to single-cell expression data. *Nucleic Acids Res* **47**, D711–D715, doi:10.1093/nar/gky964 (2019).
15. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
16. Lin, C. Y. *et al.* Membrane protein-regulated networks across human cancers. *Nat Commun* **10**, 3131, doi:10.1038/s41467-019-10920-8 (2019).
17. Bandyopadhyay, S. *et al.* Rewiring of genetic networks in response to DNA damage. *Science* **330**, 1385–1389, doi:10.1126/science.1195618 (2010).
18. Jamshidi Makiani, M. *et al.* Relative frequency of human papillomavirus genotypes and related sociodemographic characteristics in women referred to a general hospital in Tehran, 2014–2015: A cross-sectional study. *Int J Reprod Biomed* **15**, 305–310 (2017).
19. Jamdar, F. *et al.* Prevalence of human papillomavirus infection among Iranian women using COBAS HPV DNA testing. *Infect Agent Cancer* **13**, 6, doi:10.1186/s13027-018-0178-5 (2018).
20. Buschges, R. *et al.* Amplification and expression of cyclin D genes (CCND1, CCND2 and CCND3) in human malignant gliomas. *Brain Pathol* **9**, 435–442; discussion 432–433, doi:10.1111/j.1750-3639.1999.tb00532.x (1999).
21. Vailly, J. *et al.* The genes for nicein/kalinin 125- and 100-kDa subunits, candidates for junctional epidermolysis bullosa, map to chromosomes 1q32 and 1q25-q31. *Genomics* **21**, 286–288, doi:10.1006/geno.1994.1263 (1994).
22. Chen, L. *et al.* Characterization of PIK3CA and PIK3R1 somatic mutations in Chinese breast cancer patients. *Nat Commun* **9**, 1357, doi:10.1038/s41467-018-03867-9 (2018).
23. Curfs, D. M. *et al.* Polycyclic aromatic hydrocarbons induce an inflammatory atherosclerotic plaque phenotype irrespective of their DNA binding properties. *FASEB J* **19**, 1290–1292, doi:10.1096/fj.04-2269fje (2005).

Figures

Figure 1

Overview of genome-wide analysis of cigarette and hookah constituents-gene/protein-pathway-disease relationships by using Hierarchical-Systems Biology Model (HiSBiM). (A) Steps in main procedure. (B) Cell-based hookah exposed design with five conditions (i.e., control, charcoal, charcoal+non-tobacco, charcoal+tobacco, and electronic+tobacco) and 2 different exposures (2 hours and 24 hours). For each condition and exposure, three cell samples are selected for performing RNA-seq using Illumina HiSeq 2500 platform (total 30 samples; GEO accession GSE162386). (C) Cigarette smoking datasets were collected from GEO and ArrayExpress. (D) Identification of differentially expressed genes (DEGs) for each omics dataset. (E) KEGG pathway enrichment analysis of DEGs, and their corresponding (F) z-scores of each pathway and (G) meta-z-scores (Stouffer's Z-transform method) of pathway sub-systems.

Figure 1

Overview of genome-wide analysis of cigarette and hookah constituents-gene/protein-pathway-disease relationships by using Hierarchical-Systems Biology Model (HiSBiM). (A) Steps in main procedure. (B) Cell-based hookah exposed design with five conditions (i.e., control, charcoal, charcoal+non-tobacco, charcoal+tobacco, and electronic+tobacco) and 2 different exposures (2 hours and 24 hours). For each condition and exposure, three cell samples are selected for performing RNA-seq using Illumina HiSeq 2500 platform (total 30 samples; GEO accession GSE162386). (C) Cigarette smoking datasets were collected from GEO and ArrayExpress. (D) Identification of differentially expressed genes (DEGs) for each omics dataset. (E) KEGG pathway enrichment analysis of DEGs, and their corresponding (F) z-scores of each pathway and (G) meta-z-scores (Stouffer's Z-transform method) of pathway sub-systems.

Figure 2

The concept of Hierarchical-Systems Biology Model (HiSBiM). (A) Our HiSBiM provides the biological functions based on system-, subsystem- and pathway-level omics data analysis. The 330 KEGG pathways were collected and defined into 45 subsystems and 6 systems. (B) The p-value of each pathway underwent a z-score transformation based on the standard normal distribution. We then computed the subsystem- and system-level meta-z-scores of 45 KEGG subsystems and 6 systems. The meta-z-score reflects the significance (enrichment) of a subsystem or system for a specific condition, such as hookah and cigarette.

Figure 2

The concept of Hierarchical-Systems Biology Model (HiSBiM). (A) Our HiSBiM provides the biological functions based on system-, subsystem- and pathway-level omics data analysis. The 330 KEGG pathways were collected and defined into 45 subsystems and 6 systems. (B) The p-value of each pathway underwent a z-score transformation based on the standard normal distribution. We then computed the subsystem- and system-level meta-z-scores of 45 KEGG subsystems and 6 systems. The meta-z-score reflects the significance (enrichment) of a subsystem or system for a specific condition, such as hookah and cigarette.

Figure 2

The concept of Hierarchical-Systems Biology Model (HiSBiM). (A) Our HiSBiM provides the biological functions based on system-, subsystem- and pathway-level omics data analysis. The 330 KEGG pathways were collected and defined into 45 subsystems and 6 systems. (B) The p-value of each pathway underwent a z-score transformation based on the standard normal distribution. We then computed the subsystem- and system-level meta-z-scores of 45 KEGG subsystems and 6 systems. The meta-z-score reflects the significance (enrichment) of a subsystem or system for a specific condition, such as hookah and cigarette.

Figure 3

The genome-wide analysis and comparison between hookah and cigarette smoking. (A) Comparison of meta-z-scores between hookah treatment for 2 hrs (blue) and 24 hrs (yellow) and cigarette datasets (grey) in system-level. (B) Meta-z-scores of 11 subsystems between hookah (blue) and cigarette datasets (grey) in Human Diseases system. (C) Z-scores of 11 pathways between hookah (blue) and cigarette (grey) datasets in Infectious disease: viral subsystem. (D) DEGs of hookah datasets regulated in Human

papillomavirus infection pathway. The genes are listed in order as the hookah treatment 2 versus 24 hrs (from left to right). Red and green were denoted up- and down-regulation of genes, respectively.

Figure 3

The genome-wide analysis and comparison between hookah and cigarette smoking. (A) Comparison of meta-z-scores between hookah treatment for 2 hrs (blue) and 24 hrs (yellow) and cigarette datasets (grey) in system-level. (B) Meta-z-scores of 11 subsystems between hookah (blue) and cigarette datasets

(grey) in Human Diseases system. (C) Z-scores of 11 pathways between hookah (blue) and cigarette (grey) datasets in Infectious disease: viral subsystem. (D) DEGs of hookah datasets regulated in Human papillomavirus infection pathway. The genes are listed in order as the hookah treatment 2 versus 24 hrs (from left to right). Red and green were denoted up- and down-regulation of genes, respectively.

Figure 3

The genome-wide analysis and comparison between hookah and cigarette smoking. (A) Comparison of meta-z-scores between hookah treatment for 2 hrs (blue) and 24 hrs (yellow) and cigarette datasets (grey) in system-level. (B) Meta-z-scores of 11 subsystems between hookah (blue) and cigarette datasets (grey) in Human Diseases system. (C) Z-scores of 11 pathways between hookah (blue) and cigarette (grey) datasets in Infectious disease: viral subsystem. (D) DEGs of hookah datasets regulated in Human papillomavirus infection pathway. The genes are listed in order as the hookah treatment 2 versus 24 hrs (from left to right). Red and green were denoted up- and down-regulation of genes, respectively.

CO+PAH datasets, such as Infectious diseases: Bacterial and Cancers: Specific types. (C) Z-scores of five pathways of CO+PAH dataset in Cardiovascular diseases subsystem. (D) DEGs of CO+PAH dataset were involved in Cardiovascular diseases subsystem. For each gene, the fold change values, displayed in the color scheme (red: up-regulated; green: down-regulated; white: no change), are listed in order as the CO+PAH treatment 2 versus 24 hours (from left to right).

Figure 4

The mechanisms of hookah constituents and cigarette in subsystem- and pathway-level and hookah constituents-gene-pathway-disease relationships. (A) The hierarchical clustering of hookah, its constituents (i.e., CO+PAH and nicotine), and cigarette datasets in subsystem-level. The matrix distance calculating by using Spearman's rank correlation coefficient and complete linkage. Red and blue were denoted high and low meta-z-scores, respectively. (B) Union subsystems of top 10 meta-z-scores in hookah (blue) and CO+PAH (green) datasets. The 7 of 10 subsystems were shared between hookah and CO+PAH datasets, such as Infectious diseases: Bacterial and Cancers: Specific types. (C) Z-scores of five pathways of CO+PAH dataset in Cardiovascular diseases subsystem. (D) DEGs of CO+PAH dataset were

involved in Cardiovascular diseases subsystem. For each gene, the fold change values, displayed in the color scheme (red: up-regulated; green: down-regulated; white: no change), are listed in order as the CO+PAH treatment 2 versus 24 hours (from left to right).

Figure 4

The mechanisms of hookah constituents and cigarette in subsystem- and pathway-level and hookah constituents-gene-pathway-disease relationships. (A) The hierarchical clustering of hookah, its constituents (i.e., CO+PAH and nicotine), and cigarette datasets in subsystem-level. The matrix distance calculating by using Spearman's rank correlation coefficient and complete linkage. Red and blue were denoted high and low meta-z-scores, respectively. (B) Union subsystems of top 10 meta-z-scores in hookah (blue) and CO+PAH (green) datasets. The 7 of 10 subsystems were shared between hookah and CO+PAH datasets, such as Infectious diseases: Bacterial and Cancers: Specific types. (C) Z-scores of five pathways of CO+PAH dataset in Cardiovascular diseases subsystem. (D) DEGs of CO+PAH dataset were involved in Cardiovascular diseases subsystem. For each gene, the fold change values, displayed in the

color scheme (red: up-regulated; green: down-regulated; white: no change), are listed in order as the CO+PAH treatment 2 versus 24 hours (from left to right).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterials.pdf](#)
- [Supplementarymaterials.pdf](#)