Transcriptomic Changes of The Buccal Epithelium Exposed To Combustion Emissions of Smoky Coal in Xuanwei and Fuyuan, A Rural Region of China

Qin Yang

Central South University

Wancheng Guo

Central South University

Hui Li (✉ lihuiscience@163.com)

Central South University <https://orcid.org/0000-0003-1086-0402>

Research Article

Keywords: buccal epithelium, differentially expressed genes (DEGs), inflammation, smoky coal, lung cancer

Posted Date: September 7th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-720980/v1>

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

[Read Full License](#)

Abstract

Indoor air pollution generated from solid fuels affects over half of the world's people. The smoky coal is one of the major sources of indoor air pollution and has been associated with the exceptionally high lung cancer incidence and mortality of non-smoking female residents in rural region Xuanwei and Fuyuan, China. Although some studies have characterized the etiology of the relationship between the smoky coal and the disease risk, the underlying molecular mechanism is still poorly understood. Thus we downloaded RNA sequence data (GSE64277) including 35 non-smoking female residents (26 smoky coal users and 9 smokeless coal users) of Xuanwei and Fuyuan, and conducted Gene Ontology Analysis, Kyoto Encyclopedia of Genes and Genomes Analysis and Weighted Gene Co-expression Network Analysis for transcriptomic changes in the buccal epithelium. The bioinformatics analysis of transcriptional data revealed that combustion emissions of smoky coal significantly altered the genes expression in the buccal epithelium. Activation of inflammation appeared to be a critical response and key risk factor in buccal epithelium associated with smoky coal-related indoor air pollution exposure. We identified 2318 differentially expressed genes, enriched-biological processes/cell components/pathways, gene co-expresssion relationships, 3 significant modules and 45 hub genes. This study contributes to understanding the uncovering molecular mechanism in the respiratory system that occurs with combustion emissions from smoky coal burning and carrying out potential effective interventions to prevent disease burdens.

1 Introduction

The combustion emissions from smoky coal burning can result in multiple respiratory system diseases and are classified as Group 1 carcinogens by the International Agency for Research on Cancer; however, the smoky coal is still an irreplaceable energy source for heating and cooking at home in many low- and middle- income countries (Gordon et al., 2014; Lan et al., 2008; Vermeulen et al., 2019). More than 90% of lung field cancer predominantly results from air pollution, smoking or asbestos. China is the world's top coal producer and consumer. Coal is widely used in China as indoor fuel for heating and cooking, especially in rural areas. Previous epidemiological and experimental studies showed that the exceptionally high lung cancer incidence and mortality of non-smoking female residents in rural countries Xuanwei and Fuyua of China has been associated with exposure to combustion emissions of local smoky coal (Hu et al., 2020; Seow et al., 2014). In 2020, lung cancer is globally the first most commonly diagnosed cancer (14.3% of total cases) and the first cause of cancer-related deaths in males (21.5% of the total cancer deaths); whereas among females, it is the second most diagnosed cancer (8.4% of total cases) and the second cause of cancer-related deaths (13.7% of the total cancer deaths) (Ferlay et al., 2021). Although a few studies have characterized the etiology of the relationship between the smoky coal and the disease risk, the underlying molecular mechanism is still largely unknown.

The concept of “field cancerization” (also named “field of injury” or “field effect”) proposes that exposure to a noxious stimulus elicits the common genetic or epigenetic changes of the entire respiratory tracts, which has been demonstrated in cigarette smokers (Saba, Halytskyy, Saleem, and Oliff, 2017; Slaughter,

Southwick, and Smejkal, 1953). Based on Gene Ontology analysis (GO), Kyoto Encyclopedia of Genes and Genomes Analysis (KEGG) and Weighted Gene Co-expression Network Analysis (WGCNA), a transcriptomic study in buccal epithelium of healthy never-smoking females who burn coal indoor from Xuanwei and Fuyuan was conducted in the current study. The transcriptome analysis revealed that the exposure to the smoky coal combustion emissions induced a strong modulation of gene transcription compared to smokeless coal in terms of differentially expressed genes (DEGs) and activated pathways. Our study identified a set of DEGs, enriched-biological processes (BP)/cell components (CC)/pathways, gene co-expression relationships, significant modules and hub genes in the buccal epithelium in response to smoky coal-related indoor air pollution exposure (SIAPE), which could provide an insight into the molecular effects associated with SIAPE in respiratory tracts. Moreover, the buccal epithelium is a promising surrogate marker for lung cancer detection (Roy et al., 2010; Saba et al., 2017). Obtaining buccal mucosa is much less invasive or uncomfortable than the bronchial mucosa (including lung). This study may be warranted to a biological basis for assessing airway health risks, particularly lung cancer, and developing prevention strategies.

2 Materials And Methods

2.1 Microarray Data Acquisition and Preprocessing

Raw RNA sequence data GSE64277 were obtained from the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>), including 35 non-smoking female residents (26 smoky coal users and 9 smokeless coal users) of Xuanwei and Fuyuan counties. These data were used to assess GO-/KEGG-/WGCNA-based analysis for transcriptomic changes in the buccal epithelium. Raw data (raw reads) of FastQC format were quality-assessed using the FastQC program. Next, Clean data (clean reads) and raw gene counts were obtained as previously reported (P. Li, Wang, Wang, Liu, and Chen, 2021; Zappaterra, Gioiosa, Chillemi, Zambonelli, and Davoli, 2021). The raw gene counts were normalized in the R language with regularized-logarithm transformation using the DESeq2 package.

2.2 Go and KEGG Analysis of the Identified DEGs

The expression difference of the normalized gene counts associated with exposure to the indoor burning of smoky versus smokeless coal were identify by volcanic plot map($p < 0.05$) and heat map $p < 0.05$, $|\text{LogFC}| > 1$ using packages in R4.0.5 version. Afterwards, DEGs $p < 0.05$ were identified to be significantly associated with SIAPE. The GO (BP and CC) and KEGG of the DEGs $p < 0.05$ were performed using the clusterProfiler v3.14.0.

2.3 WGCNA Analysis of the DEGs

WGCNA was conducted on the selected DEGs using the WGCNA package (version1.70-3) based on R language (version 4.0.5) to find modules (clusters) of highly correlated genes. WGCNA algorithm was used to build the scale-free co-expression network (Langfelder and Horvath, 2008). An adjacency matrix was build using Pearson's correlation to evaluate the correlation strength between the gene couple in the

scale-free undirected co-expression networks. A soft threshold power (β) of 9 as the scale-free topology index (R^2) reached the peak ($R^2 > 0.80$) for the first time and was chosen to raise the adjacency matrix. The adjacency matrix was transformed into a topological overlap matrix and used to quantitatively describe the similarity between the gene couple and other genes. DynamicTree Cut algorithm was used to identify the gene modules. Module eigengene (ME) is used to assess the potential correlation between modules and traits, which were named with different colors and comprised of genes having a biologically relevant role. The signed module membership, also named module-trait relationship, was evaluated by Pearson's correlation between the ME and the traits of interest. The module of the highest correlation was selected for further analyses. The reliability of individual genes within a module is established by the correlation coefficient between the gene and ME.

2.4 Construction of Protein-protein Interactions (PPI) Network and Identification of Hub Genes

The grey module with the highest module trait relationship was used to identify hub genes. To elucidate the interactions of DEGs in the most significant module, the top genes (hub genes) with the highest connectivity degree were visualized by R language (STRINGdb, igraph and ggraph packages) to develop a PPI network.

3 Results

3.1 Identify the DEGs associated with SIAPE

Transcriptome data (GSE64277) including 26 smoky coal and 9 smokeless coal users were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64277>). To investigate the underlying molecular effects of SIAPE in the buccal epithelium (GSE64277), we analyzed the DEGs according to the analysis process as shown in Fig. 1. 2318 DEGs ($p < 0.05$, Supplementary Table 1) were identified by volcanic plot map (Fig. 2a). Specifically, the top changed 8 DEGs ($p < 0.05$ and $|\text{LogFC}| > 2$) were showed as green dots in Fig. 2a and listed in Table 1. Among the 2318 genes, 116 genes ($p < 0.05$ and $|\text{LogFC}| > 1$) were selected to create a heat map (Fig. 2b). Through transcriptome cluster analysis of variation, we found two DEGs clusters in the buccal epithelium exposed to smoky versus smokeless coal (Fig. 2b).

Table 1
List of the DEGs validated the greater than 2-fold activation and $p < 0.05$

	logFC	AveExpr	t	P.Value	adj.P.Val
IL1B	2.778067341	9.044040274	3.419575863	0.001589456	0.33293579
FCGR2A	2.378067906	6.891342481	3.074389343	0.004037612	0.33974471
PLEK	2.190085508	7.593486927	3.017856339	0.004684566	0.352876645
AQP9	2.187088478	7.186263849	2.888350495	0.006553539	0.374355276
CXCR1	2.177464461	7.730265878	2.948801024	0.005607634	0.369076832
PTGS2	2.102939246	7.877227404	2.575557404	0.014317439	0.374355276
CSF3R	2.033074523	7.956106534	3.201204044	0.002880452	0.33293579
IL8	2.007668773	8.705894149	2.984117034	0.00511605	0.36207357

3.2 GO and KEGG of the DEGs

The GO analysis showed that the DEGs were enriched in many biological processes (including glucocorticoid receptor signaling pathway, regulation of translational initiation, elongation and termination, etc.) (Fig. 2C) and multiple cell components (including mitochondrial matrix, nuclear envelope, nuclear membrane, etc.) (Fig. 2D). The KEGG analysis showed that the DEGs were enriched in herpes simplex virus 1 infection, osteoclast differentiation, cell cycle, etc (Fig. 2E).

3.3 WGCNA Construction and Identification of Significant Modules

The soft-thresholding power $\beta = 9$ was calculated to establish a scale-free (scale-free $R^2 > 0.80$) co-expression network (Fig. 3a and 3b). The 2318 DEGs ($p < 0.05$) were clustered into several modules by the average linkage hierarchical clustering method (Fig. 3b and 3c). Then, the correlation analysis between the modules and SIAPE was performed. Each significant module was marked with its corresponding ME. 3 significant modules, comprised of 2318 genes (183 genes from the grey module, 482 genes from the brown module and 1653 genes from the blue module), were identified (Fig. 3d). The grey module showed the highest association with SIAPE (Fig. 3d) and thus was considered as a key module. Gene significance of the DEGs in the grey module and the module membership were calculated and plotted in a scatter diagram (Fig. 3e).

3.4 PPI network and Hub Genes

Hub genes play critical roles in pathways associated with other genes in the co-expression network and have relatively higher intramodule connectivity and gene significance than the other genes. To identify the hub genes, the 183 DEGs in the grey module were used to develop a visualized PPI network (Fig. 4). 45 hub genes with the highest intramodule connectivity were finally characterized.

4 Discussions

In this study, we analyzed expression data from buccal epithelial cells associated with SIAPE in Xuanwei and Fuyua, China and identified a set of DEGs ($p < 0.05$), enriched-BP/CC/pathways, 3 significant modules and 45 hub genes.

Activation of inflammation appeared to be a critical response and key risk factor in buccal epithelium associated with SIAPE. First, we validated 8 DEGs of the greater than 2-fold activation, and 6 (IL8, IL1B, CXCR1, FCGR2A, PTGS2, PLEK) among the 8 DEGs are inflammatory mediators. Cellular sustained inflammation leads to genomic instability and thus may induce malignant transformation of normal cells. Pro-inflammatory cytokine IL8 is also up-regulated in rural Indian women cooking with biomass fuels (Dutta, Ray, and Banerjee, 2012), and smokers with lung cancer (Beane et al., 2011). Pro-inflammatory cytokines IL1B plays a key role in mediating inflammation and further correlated with lung cancer risk (C. Li and Wang, 2013). IL1B not only plays a role in active inflammation, but also regulates the expression of the other inflammatory genes and triggers a cascade of inflammation reactions (Yang and Xiao, 2018). Then, we identified many important inflammatory mediators (for instance, CXCR5, ISG15, NR3C1, PARP12, EIF2AK2 and PARP12) from the 45 hub genes of the grey module. For instance, CXCR5 and its ligand chemokine CXCL13 play fundamental roles in inflammation (Kazanietz, Durando, and Cooke, 2019; Yoshitomi, 2020). A previous study of women in Xuanwei and Fuyua showed that smoky coal combustion-generated pollution or cigarette smoke activates CXCL13/CXCR5 signaling axis and facilitates lung cancer initiation and progression (Hussain et al., 2019; G. Z. Wang et al., 2015).

It is well-established that the causal association between inflammation and cancer. Metabolic inflammation is referred to as inflammation responding to dysregulated metabolic processes in recent years (Lonardo, Nascimbeni, Mantovani, and Targher, 2018). Meanwhile, alteration of lipid metabolism is considered as a cancer hallmark (Harjes, Kalucka, and Carmeliet, 2016; Liu et al., 2021). Our GO analysis suggested that the physiologic response to smoky coal exposure altered genes involved in fatty acid derivative biosynthetic process, glucocorticoids receptor signaling pathway and translational process (initiation, elongation and termination), etc. PTGS2 (a DEG of 2-fold activation) and hub genes ISG15, EIF2AK2, NR3C1 and PARP12, are involved in regulating inflammation and fatty acids of cancer. For instance, PTGS2 is usually negligible in most normal cells, but can be highly induced by inflammation and lung cancer (Ching, Reader, and Fulton, 2020; Hashemi Goradel, Najafi, Salehi, Farhood, and Mortezaee, 2019). The continuous overexpression of PTGS2 could significantly promote cancer by up-regulating a fatty acid derivative prostaglandin E2 (Desai, Prickril, and Rasooly, 2018). Dysregulation of EIF2AK2 is implicated in inflammation, cellular metabolism control of fatty acids and cancer by regulating translational initiation (Gal-Ben-Ari, Barrera, Ehrlich, and Rosenblum, 2018; Shen et al., 2016). PARP12 is involved in inflammation and tumor suppression by arresting mRNA translation (Grimaldi and Corda, 2019; Welsby et al., 2014). NR3C1 is a glucocorticoid receptor. Glucocorticoids/NR3C1 signaling is linked to severe lipid metabolism, restraining inflammation and transcriptional regulation (King et al., 2013; Mostafa et al., 2019; Newton, 2014; Rose and Herzig, 2013). The ubiquitin-specific peptidase USP18 removes ISG15 from substrate proteins and augments lung tumorigenicity through promoting

lipolysis and fatty acid oxidation (Liu et al., 2021). Those genes could be further used for experimental investigations of biological functions.

Exposure to the smoky coal can result in multiple respiratory system diseases such as respiratory infections, chronic lung diseases and particularly lung cancer (Lan et al., 2008; Vermeulen et al., 2019). The combustion emissions of smoky coal are complex mixtures of various components and associated with 2-fold to a more than 30-fold increase of lung cancer risk in non-smoking female residents of Xuanwei and Fuyuan in China (Vermeulen et al., 2019; Wong, Downward, et al., 2019). According to the concept of “field cancerization” (Slaughter et al., 1953) and some previous studies (Saba et al., 2017; Spira, 2010; Sridhar et al., 2008), transcriptomic changes in the bronchial airway are reflected in the buccal epithelium. Our results about the transcriptomic changes of buccal epithelium exposed to smoky coal combustion emissions will contribute to understanding the uncovering molecular mechanism in the respiratory system and carrying out potential effective interventions to prevent these disease burdens.

Abbreviations

differentially expressed genes (DEGs) · Gene Expression Omnibus (GEO) · Gene Ontology Analysis (GO) · Kyoto Encyclopedia of Genes and Genomes Analysis (KEGG) · module eigengene (ME) · protein–protein interactions (PPI) · smoky coal-related indoor air pollution exposure (SIAPE) · topological overlap matrix (TOM) · weighted gene co-expression network analysis (WGCNA)

Declarations

5.1 Ethics approval and consent to participate

Not applicable.

5.2 Consent for publication

Not applicable.

5.3 Availability of data and materials

Raw RNA sequence data GSE64277 that support the findings of this study are available from the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>).

5.4 Competing interests

The authors declare that they have no competing interests.

5.5 Acknowledgements

This work was supported by the grants from the Natural Science Foundation of Hunan Province [grant number 2020JJ4560]; Scientific Research Foundation of Hunan Provincial Education Department [grant

number 20B528]; Guidance Science and Technology Program of Shaoyang City [grant number 2019ZD07]; National Natural Science Found of China [grant number 82003286].

5.6 Authors' contributions

Qin Yang performed the literature search and wrote the manuscript. Wancheng Guo performed the analyses. Hui Li had the idea for the article and critically revised the manuscript. All authors read and approved the final manuscript.

References

1. Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J,... . Spira A (2011) Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq. *Cancer Prev Res (Phila)*, 4(6), 803–817. [https://doi.org/ 10.1158/1940-6207.CAPR-11-0212](https://doi.org/10.1158/1940-6207.CAPR-11-0212)
2. Bommel H, Haake M, Luft P, Horejs-Hoeck J, Hein H, Bartels J,... . Duschl A (2003) The diesel exhaust component pyrene induces expression of IL-8 but not of eotaxin. *Int Immunopharmacol*, 3(10–11), 1371–1379. [https://doi.org/ 10.1016/S1567-5769\(03\)00135-8](https://doi.org/10.1016/S1567-5769(03)00135-8)
3. Chen P, Li Q, Zhou Y, Lu H, Chen H, Qian M, Chen J (2021) Clinical implication of aquaporin 9 in non-small cell lung cancer patients: its expression and relationship with clinical features and prognosis. *Ir J Med Sci*. [https://doi.org/ 10.1007/s11845-021-02523-4](https://doi.org/10.1007/s11845-021-02523-4)
4. Ching MM, Reader J, Fulton AM (2020) Eicosanoids in Cancer: Prostaglandin E2 Receptor 4 in Cancer Therapeutics and Immunotherapy. *Front Pharmacol*, 11, 819. [https://doi.org/ 10.3389/fphar.2020.00819](https://doi.org/10.3389/fphar.2020.00819)
5. Chow PH, Bowen J, Yool AJ (2020) Combined Systematic Review and Transcriptomic Analyses of Mammalian Aquaporin Classes 1 to 10 as Biomarkers and Prognostic Indicators in Diverse Cancers. *Cancers (Basel)*, 12(7). [https://doi.org/ 10.3390/cancers12071911](https://doi.org/10.3390/cancers12071911)
6. Cmarik JL, Hegamyer G, Gerrard B, Dean M, Colburn NH (2000) cDNA cloning and mapping of mouse pleckstrin (Plek), a gene upregulated in transformation-resistant cells. *Genomics*, 66(2), 204–212. [https://doi.org/ 10.1006/geno.2000.6210](https://doi.org/10.1006/geno.2000.6210)
7. Cook-Mills JM, Marchese ME, Abdala-Valencia H (2011) Vascular cell adhesion molecule-1 expression and signaling during disease: regulation by reactive oxygen species and antioxidants. *Antioxid Redox Signal*, 15(6), 1607–1638. [https://doi.org/ 10.1089/ars.2010.3522](https://doi.org/10.1089/ars.2010.3522)
8. Dell KR, Walsh MP, Severson DL (1988) Characterization of bovine aortic protein kinase C with histone and platelet protein P47 as substrates. *Biochem J*, 254(2), 455–462. [https://doi.org/ 10.1042/bj2540455](https://doi.org/10.1042/bj2540455)
9. Desai SJ, Prickril B, Rasooly A (2018) Mechanisms of Phytonutrient Modulation of Cyclooxygenase-2 (COX-2) and Inflammation Related to Cancer. *Nutr Cancer*, 70(3), 350–375. [https://doi.org/ 10.1080/01635581.2018.1446091](https://doi.org/10.1080/01635581.2018.1446091)

10. Dutta A, Ray MR, Banerjee A (2012) Systemic inflammatory changes and increased oxidative stress in rural Indian women cooking with biomass fuels. *Toxicol Appl Pharmacol*, 261(3), 255–262. [https://doi.org/ 10.1016/j.taap.2012.04.004](https://doi.org/10.1016/j.taap.2012.04.004)
11. Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Pineros M, Znaor A, Bray F (2021) Cancer statistics for the year 2020: An overview. *Int J Cancer*. [https://doi.org/ 10.1002/ijc.33588](https://doi.org/10.1002/ijc.33588)
12. Gal-Ben-Ari S, Barrera I, Ehrlich M, Rosenblum K (2018) PKR: A Kinase to Remember. *Front Mol Neurosci*, 11, 480. [https://doi.org/ 10.3389/fnmol.2018.00480](https://doi.org/10.3389/fnmol.2018.00480)
13. Geng X, Pu W, Tan Y, Lu Z, Wang A, Tan L,... . Chen X (2017) Quantitative assessment of the diagnostic role of FHIT promoter methylation in non-small cell lung cancer. *Oncotarget*, 8(4), 6845–6856. [https://doi.org/ 10.18632/oncotarget.14256](https://doi.org/10.18632/oncotarget.14256)
14. Gordon SB, Bruce NG, Grigg J, Hibberd PL, Kurmi OP, Lam KB,... . Martin WJ 2nd (2014) Respiratory risks from household air pollution in low and middle income countries. *Lancet Respir Med*, 2(10), 823–860. [https://doi.org/ 10.1016/S2213-2600\(14\)70168-7](https://doi.org/10.1016/S2213-2600(14)70168-7)
15. Grimaldi G, Corda D (2019) ADP-ribosylation and intracellular traffic: an emerging role for PARP enzymes. *Biochem Soc Trans*, 47(1), 357–370. [https://doi.org/ 10.1042/BST20180416](https://doi.org/10.1042/BST20180416)
16. Grivennikov SI, Karin M (2011) Inflammatory cytokines in cancer: tumour necrosis factor and interleukin 6 take the stage. *Ann Rheum Dis*, 70 Suppl 1, i104-108. [https://doi.org/ 10.1136/ard.2010.140145](https://doi.org/10.1136/ard.2010.140145)
17. Guo C, Shao R, Correa AM, Behrens C, Johnson FM, Raso MG,... . Pataer A (2013) Prognostic significance of combinations of RNA-dependent protein kinase and EphA2 biomarkers for NSCLC. *J Thorac Oncol*, 8(3), 301–308. [https://doi.org/ 10.1097/JTO.0b013e318282def7](https://doi.org/10.1097/JTO.0b013e318282def7)
18. Harjes U, Kalucka J, Carmeliet P (2016) Targeting fatty acid metabolism in cancer and endothelial cells. *Crit Rev Oncol Hematol*, 97, 15–21. [https://doi.org/ 10.1016/j.critrevonc.2015.10.011](https://doi.org/10.1016/j.critrevonc.2015.10.011)
19. Hashemi Goradel N, Najafi M, Salehi E, Farhood B, Mortezaee K (2019) Cyclooxygenase-2 in cancer: A review. *J Cell Physiol*, 234(5), 5683–5699. [https://doi.org/ 10.1002/jcp.27411](https://doi.org/10.1002/jcp.27411)
20. Heinrich U, Pott F, Mohr U, Fuhst R, Konig J (1986) Lung tumours in rats and mice after inhalation of PAH-rich emissions. *Exp Pathol*, 29(1), 29–34. [https://doi.org/ 10.1016/s0232-1513\(86\)80003-2](https://doi.org/10.1016/s0232-1513(86)80003-2)
21. Hu W, Downward G, Wong JYY, Reiss B, Rothman N, Portengen L,... . Lan Q (2020) Characterization of outdoor air pollution from solid fuel combustion in Xuanwei and Fuyuan, a rural region of China. *Sci Rep*, 10(1), 11335. [https://doi.org/ 10.1038/s41598-020-68229-2](https://doi.org/10.1038/s41598-020-68229-2)
22. Hussain M, Adah D, Tariq M, Lu Y, Zhang J, Liu J (2019) CXCL13/CXCR5 signaling axis in cancer. *Life Sci*, 227, 175–186. [https://doi.org/ 10.1016/j.lfs.2019.04.053](https://doi.org/10.1016/j.lfs.2019.04.053)
23. Kazanietz MG, Durando M, Cooke M (2019) CXCL13 and Its Receptor CXCR5 in Cancer: Inflammation, Immune Response, and Beyond. *Front Endocrinol (Lausanne)*, 10, 471. [https://doi.org/ 10.3389/fendo.2019.00471](https://doi.org/10.3389/fendo.2019.00471)
24. Khandelwal PJ, Herman AM, Moussa CE (2011) Inflammation in the early stages of neurodegenerative pathology. *J Neuroimmunol*, 238(1–2), 1–11. [https://doi.org/ 10.1016/j.jneuroim.2011.07.002](https://doi.org/10.1016/j.jneuroim.2011.07.002)

25. Kim MR, Jang JH, Park CS, Kim TK, Kim YJ, Chung J,... . Lee S (2017) A Human Antibody That Binds to the Sixth Ig-Like Domain of VCAM-1 Blocks Lung Cancer Cell Migration In Vitro. *Int J Mol Sci*, 18(3). [https://doi.org/ 10.3390/ijms18030566](https://doi.org/10.3390/ijms18030566)
26. King EM, Chivers JE, Rider CF, Minnich A, Giembycz MA, Newton R (2013) Glucocorticoid repression of inflammatory gene expression shows differential responsiveness by transactivation- and transrepression-dependent mechanisms. *PLoS One*, 8(1), e53936. [https://doi.org/ 10.1371/journal.pone.0053936](https://doi.org/10.1371/journal.pone.0053936)
27. Kzhyshkowska J, Yin S, Liu T, Riabov V, Mitrofanova I (2016) Role of chitinase-like proteins in cancer. *Biol Chem*, 397(3), 231–247. [https://doi.org/ 10.1515/hsz-2015-0269](https://doi.org/10.1515/hsz-2015-0269)
28. Lan Q, He X, Shen M, Tian L, Liu LZ, Lai H,... . Chapman RS (2008) Variation in lung cancer risk by smoky coal subtype in Xuanwei, China. *Int J Cancer*, 123(9), 2164–2169. [https://doi.org/ 10.1002/ijc.23748](https://doi.org/10.1002/ijc.23748)
29. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559. [https://doi.org/ 10.1186/1471-2105-9-559](https://doi.org/10.1186/1471-2105-9-559)
30. Lee KM, Shen M, Chapman RS, Yeager M, Welch R, He X,... . Lan Q (2007) Polymorphisms in immunoregulatory genes, smoky coal exposure and lung cancer risk in Xuan Wei, China *Carcinogenesis*, 28(7), 1437–1441. [https://doi.org/ 10.1093/carcin/bgm030](https://doi.org/10.1093/carcin/bgm030)
31. Li C, Wang C (2013) Current evidences on IL1B polymorphisms and lung cancer susceptibility: a meta-analysis. *Tumour Biol*, 34(6), 3477–3482. [https://doi.org/ 10.1007/s13277-013-0925-6](https://doi.org/10.1007/s13277-013-0925-6)
32. Li P, Wang Y, Wang X, Liu L, Chen L (2021) Identification of Susceptible Genes for Chronic Obstructive Pulmonary Disease with Lung Adenocarcinoma by Weighted Gene Co-Expression Network Analysis. *Onco Targets Ther*, 14, 3625–3634. [https://doi.org/ 10.2147/OTT.S303544](https://doi.org/10.2147/OTT.S303544)
33. Liu X, Lu Y, Chen Z, Liu X, Hu W, Zheng L,... . Dmitrovsky E (2021) The Ubiquitin-Specific Peptidase USP18 Promotes Lipolysis, Fatty Acid Oxidation, and Lung Cancer Growth. *Mol Cancer Res*, 19(4), 667–677. [https://doi.org/ 10.1158/1541-7786.MCR-20-0579](https://doi.org/10.1158/1541-7786.MCR-20-0579)
34. Lonardo A, Nascimbeni F, Mantovani A, Targher G (2018) Hypertension, diabetes, atherosclerosis and NASH: Cause or consequence? *J Hepatol*, 68(2), 335–352. [https://doi.org/ 10.1016/j.jhep.2017.09.021](https://doi.org/10.1016/j.jhep.2017.09.021)
35. Long W, Li Q, Zhang J, Xie H (2021) Identification of key genes in the tumor microenvironment of lung adenocarcinoma. *Med Oncol*, 38(7), 83. [https://doi.org/ 10.1007/s12032-021-01529-3](https://doi.org/10.1007/s12032-021-01529-3)
36. Lv J, Xu R, Wu G, Zhang Q, Li Y, Wang P,... . Wei F (2009) Indoor and outdoor air pollution of polycyclic aromatic hydrocarbons (PAHs) in Xuanwei and Fuyuan, China. *J Environ Monit*, 11(7), 1368–1374. [https://doi.org/ 10.1039/b900382g](https://doi.org/10.1039/b900382g)
37. Merbl Y, Refour P, Patel H, Springer M, Kirschner MW (2013) Profiling of ubiquitin-like modifications reveals features of mitotic control. *Cell*, 152(5), 1160–1172. [https://doi.org/ 10.1016/j.cell.2013.02.007](https://doi.org/10.1016/j.cell.2013.02.007)
38. Mostafa MM, Rider CF, Shah S, Traves SL, Gordon PMK, Miller-Larsson A,... . Newton R (2019) Glucocorticoid-driven transcriptomes in human airway epithelial cells: commonalities, differences

- and functional insight from cell lines and primary cells. *BMC Med Genomics*, 12(1), 29. <https://doi.org/10.1186/s12920-018-0467-2>
39. Newton R (2014) Anti-inflammatory glucocorticoids: changing concepts. *Eur J Pharmacol*, 724, 231–236. <https://doi.org/10.1016/j.ejphar.2013.05.035>
 40. Pei XH, Nakanishi Y, Inoue H, Takayama K, Bai F, Hara N (2002) Polycyclic aromatic hydrocarbons induce IL-8 expression through nuclear factor kappaB activation in A549 cell line. *Cytokine* 19(5):236–241
 41. Philipp J, Sievert W, Azimzadeh O, von Toerne C, Metzger F, Posch A,.. . Tapio S (2020) Data independent acquisition mass spectrometry of irradiated mouse lung endothelial cells reveals a STAT-associated inflammatory response. *Int J Radiat Biol*, 96(5), 642–650. <https://doi.org/10.1080/09553002.2020.1712492>
 42. Podechard N, Lecureur V, Le Ferrec E, Guenon I, Sparfel L, Gilot D,.. . Fardel O (2008) Interleukin-8 induction by the environmental contaminant benzo(a)pyrene is aryl hydrocarbon receptor-dependent and leads to lung inflammation. *Toxicol Lett*, 177(2), 130–137. <https://doi.org/10.1016/j.toxlet.2008.01.006>
 43. Rose AJ, Herzig S (2013) Metabolic control through glucocorticoid hormones: an update. *Mol Cell Endocrinol*, 380(1–2), 65–78. <https://doi.org/10.1016/j.mce.2013.03.007>
 44. Roy HK, Subramanian H, Damania D, Hensing TA, Rom WN, Pass HI,.. . Backman V (2010) Optical detection of buccal epithelial nanoarchitectural alterations in patients harboring lung cancer: implications for screening. *Cancer Res*, 70(20), 7748–7754. <https://doi.org/10.1158/0008-5472.CAN-10-1686>
 45. Saba R, Halytskyy O, Saleem N, Oliff IA (2017) Buccal Epithelium, Cigarette Smoking, and Lung Cancer: Review of the Literature. *Oncology*, 93(6), 347–353. <https://doi.org/10.1159/000479796>
 46. Seow WJ, Hu W, Vermeulen R, Iii H, Downward HD, Chapman GS, Lan RS, Q (2014) Household air pollution and lung cancer in China: a review of studies in Xuanwei. *Chin J Cancer*, 33(10), 471–475. <https://doi.org/10.5732/cjc.014.10132>
 47. Shen L, Qi Z, Zhu Y, Song X, Xuan C, Ben P,.. . Yin Z (2016) Phosphorylated heat shock protein 27 promotes lipid clearance in hepatic cells through interacting with STAT3 and activating autophagy. *Cell Signal*, 28(8), 1086–1098. <https://doi.org/10.1016/j.cellsig.2016.05.008>
 48. Singh R, Mishra MK, Aggarwal H (2017) Inflammation, Immunity, and Cancer. *Mediators Inflamm*, 2017, 6027305. <https://doi.org/10.1155/2017/6027305>
 49. Slaughter, D. P., Southwick, H. W., and Smejkal, W. (1953). Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer*, 6(5), 963–968. [https://doi.org/10.1002/1097-0142\(195309\)6:5<963::aid-cnrcr2820060515>3.0.co;2-q](https://doi.org/10.1002/1097-0142(195309)6:5<963::aid-cnrcr2820060515>3.0.co;2-q)
 50. Spira A (2010) Upper airway gene expression in smokers: the mouth as a "window to the soul" of lung carcinogenesis? *Cancer Prev Res (Phila)*, 3(3), 255–258. <https://doi.org/10.1158/1940-6207.CAPR-10-0013>

51. Sridhar S, Schembri F, Zeskind J, Shah V, Gustafson AM, Steiling K,... . Spira A (2008) Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genom*, 9, 259. [https://doi.org/ 10.1186/1471-2164-9-259](https://doi.org/10.1186/1471-2164-9-259)
52. Uchiyama K, Kondo H (2005) p97/p47-Mediated biogenesis of Golgi and ER. *J Biochem*, 137(2), 115–119. [https://doi.org/ 10.1093/jb/mvi028](https://doi.org/10.1093/jb/mvi028)
53. van Rens MT, de la Riviere AB, Elbers HR, van Den Bosch JM (2000). Prognostic assessment of 2,361 patients who underwent pulmonary resection for non-small cell lung cancer, stage I, II, and IIIA. *Chest*, 117(2), 374–379. [https://doi.org/ 10.1378/chest.117.2.374](https://doi.org/10.1378/chest.117.2.374)
54. Vermeulen R, Downward GS, Zhang J, Hu W, Portengen L, Bassig BA,... . Lan Q (2019) Constituents of Household Air Pollution and Risk of Lung Cancer among Never-Smoking Women in Xuanwei and Fuyuan, China. *Environ Health Perspect*, 127(9), 97001. [https://doi.org/ 10.1289/EHP4913](https://doi.org/10.1289/EHP4913)
55. Vuong H, Cheng F, Lin CC, Zhao Z (2014) Functional consequences of somatic mutations in cancer using protein pocket-based prioritization approach. *Genome Med*, 6(10), 81. [https://doi.org/ 10.1186/s13073-014-0081-7](https://doi.org/10.1186/s13073-014-0081-7)
56. Wang GZ, Cheng X, Zhou B, Wen ZS, Huang YC, Chen HB,... . Zhou GB (2015) The chemokine CXCL13 in lung cancers associated with environmental polycyclic aromatic hydrocarbons pollution. *Elife*, 4. [https://doi.org/ 10.7554/eLife.09419](https://doi.org/10.7554/eLife.09419)
57. Wang TW, Vermeulen RC, Hu W, Liu G, Xiao X, Alekseyev Y,... . Lan Q (2015) Gene-expression profiling of buccal epithelium among non-smoking women exposed to household air pollution from smoky coal. *Carcinogenesis*, 36(12), 1494–1501. [https://doi.org/ 10.1093/carcin/bgv150](https://doi.org/10.1093/carcin/bgv150)
58. Welsby I, Hutin D, Gueydan C, Kruys V, Rongvaux A, Leo O (2014) PARP12, an interferon-stimulated gene involved in the control of protein translation and inflammation. *J Biol Chem*, 289(38), 26642–26657. [https://doi.org/ 10.1074/jbc.M114.589515](https://doi.org/10.1074/jbc.M114.589515)
59. Wong JYY, Bassig BA, Hu W, Seow WJ, Shiels MS, Ji BT,... . Rothman N (2019) Household coal combustion, indoor air pollutants, and circulating immunologic/inflammatory markers in rural China. *J Toxicol Environ Health A*, 82(6), 411–421. [https://doi.org/ 10.1080/15287394.2019.1614500](https://doi.org/10.1080/15287394.2019.1614500)
60. Wong JYY, Downward GS, Hu W, Portengen L, Seow WJ, Silverman DT,... . Lan Q (2019) Lung cancer risk by geologic coal deposits: A case-control study of female never-smokers from Xuanwei and Fuyuan, China. *Int J Cancer*, 144(12), 2918–2927. [https://doi.org/ 10.1002/ijc.32034](https://doi.org/10.1002/ijc.32034)
61. Yang B, Xiao C (2018) PM2.5 exposure significantly improves the exacerbation of A549 tumor-bearing CB17-SCID mice. *Environ Toxicol Pharmacol*, 60, 169–175. [https://doi.org/ 10.1016/j.etap.2018.04.025](https://doi.org/10.1016/j.etap.2018.04.025)
62. Yoshitomi H (2020) CXCL13-producing PD-1(hi)CXCR5(-) helper T cells in chronic inflammation. *Immunol Med*, 43(4), 156–160. [https://doi.org/ 10.1080/25785826.2020.1781998](https://doi.org/10.1080/25785826.2020.1781998)
63. Zappaterra M, Gioiosa S, Chillemi G, Zambonelli P, Davoli R (2021) Dissecting the Gene Expression Networks Associated with Variations in the Major Components of the Fatty Acid Semimembranosus Muscle Profile in Large White Heavy Pigs. *Animals (Basel)*, 11(3). [https://doi.org/ 10.3390/ani11030628](https://doi.org/10.3390/ani11030628)

Figures

Figure 1

Analysis process of this study. Gene Ontology Analysis (GO), Kyoto Encyclopedia of Genes and Genomes Analysis (KEGG) and weighted gene co-expression network analysis (WGCNA)

Figure 2

Analysis of differentially expressed genes (DEGs) between smoky and smokeless coal users from GEO (GSE64277). (a) Volcanic plot map. Red for upregulated genes (p < 0.05, LogFC > 1). Green for upregulated genes (p < 0.05, LogFC > 2). Blue for downregulated genes (p < 0.05, LogFC < -1). (b) Heat map of 116 DEGs (p < 0.05, |LogFC| > 1). The bluer the color, the lower the gene expression and the redder the color, the higher the gene expression. (c-d) GO analysis of the DEGs (p < 0.05). (c) Biological process analysis. (d) Cellular component analysis. (e) Kyoto Encyclopedia of Genes and Genomes analysis of the DEGs (p < 0.05). The x-axis shows the number of genes enriched in this term. The term color represents the degree of enrichment

Figure 3

Identification of important modules in the buccal epithelium associated with SIAPE based on WGCNA. (a) Soft-threshold power in the WGCNA by analyzing the scale-free index for various soft-threshold powers ($\beta=0.8$) and the mean connectivity. (b) Clustering dendrogram of 35 samples (26 smoky coal users and 9 smokeless coal users). (c) Dendrogram of DEGs clustered based on the dissimilarity measure. (d) The heat map displaying displaying the correlations between the gene modules and the studied trait. Three distinct modules are generated. (e) Scatter plot of gene significance in the grey module. The x-axis represents module membership, and the y-axis represents gene significance

Figure 4

Identification of hub genes from the DEGs of the grey module. 183 DEGs of the grey module were filtered into the protein–protein interactions (PPI) network that contained 45 nodes and 130 edges. The nodes represent the hub genes. The edges represent the interaction of the hub genes

Supplementary Files

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

- [SupplementaryTable1.docx](#)