Journal Pre-proofs

Genome-Wide Evaluation of Transcriptomic Responses of Human Tissues to Smoke: A Systems Biology Study

Seyed Sajad Sohrabi, Marzieh Rashidipour, Seyyed Mohsen Sohrabi, Morteza Hadizadeh, Mohammad-Hossein Shams, Mohsen Mohammadi

PII: S0378-1119(21)00709-5
DOI: https://doi.org/10.1016/j.gene.2021.146114
Reference: GENE 146114

To appear in: *Gene Gene*

Received Date: 30 July 2021
Revised Date: 29 November 2021
Accepted Date: 6 December 2021

Please cite this article as: S. Sajad Sohrabi, M. Rashidipour, S. Mohsen Sohrabi, M. Hadizadeh, M-H. Shams, M. Mohammadi, Genome-Wide Evaluation of Transcriptomic Responses of Human Tissues to Smoke: A Systems Biology Study, *Gene Gene* (2022), doi: https://doi.org/10.1016/j.gene.2021.146114

Genome-Wide Evaluation of Transcriptomic Responses of Human Tissues to Smoke: A Systems Biology Study

Seyed Sajad Sohrabi[a], Marzieh Rashidipour[b, c*], Seyyed Mohsen Sohrabi[d], Morteza Hadizadeh[e], Mohammad-Hossein Shams[f], Mohsen Mohammadi[g]

[a] Department of Production Engineering and Plant Genetics, Faculty of Agriculture, Lorestan University, Khorramabad, Iran.

[b] Nutritional Health Research Center, Lorestan University of Medical Sciences, Khorramabad, Iran.

[c] Environmental Health Research Center, Lorestan University of Medical Sciences, Khorramabad, Iran.

[d] Department of Production Engineering and Plant Genetics, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran.

[e] Physiology Research Center, Institute of Basic and Clinical Physiology Sciences, Kerman University of Medical Sciences, Kerman, Iran.

[f] Department of Immunology, Faculty of Medicine, Lorestan University of Medical Sciences, Khorramabad, Iran.

[g] Razi Herbal Medicines Research Center and Department of Pharmacognosy and Pharmaceutical Biotechnology, Faculty of Pharmacy, Lorestan University of Medical Sciences, Khorramabad, Iran.

Corresponding author:

Marzieh Rashidipour, Nutritional Health Research Center, Lorestan University of Medical Sciences, Khorramabad, Iran

E-mail: M_rashidi80@yahoo.com; TelFax: +98 66-33204007

Abstract

The harmful compounds in various sources of smoke threaten human health. So far, many studies have investigated the effects of compounds of smoke on transcriptome changes in different human tissues. However, no study has been conducted on the effects of these compounds on transcriptome changes in different human tissues simultaneously. Hence, the present study was conducted to identify smoke-related genes (SRGs) and their response mechanisms to smoke in various human cells and tissues using systems biology based methods. A total of 6,484 SRGs were identified in the studied tissues, among which 4,095 SRGs were up-regulated and 2,389 SRGs were down-regulated. Totally, 459 SRGs were smoke-related transcription factors (SRTFs). Gene regulatory network analysis showed that the studied cells and tissues have different gene regulation and responses to compounds of smoke. The comparison of different tissues revealed no common SRG among the all studied tissues. However, the *CYP1B1* gene was common among seven cells and tissues, and had the same expression trend. Network analysis showed that the *CYP1B1* is a hub gene among SRGs in various cells and tissues. To the best of our knowledge, for the first time, our results showed that compounds of smoke induce and increase the expression of *CYP1B1* key gene in all target and non-target tissues of human. Moreover, despite the specific characteristics of *CYP1B1* gene and its identical expression trend in target and non-target tissues, it can be used as a biomarker for diagnosis and prognosis.

Keywords: *CYP1B1* Gene, Smoke, Smoke-related Genes, Systems Biology, Transcriptome, Network Analysis

1. Introduction

Human is constantly exposed to various sources of smoke-related pollution in nature and in their living environment, including smoke emitted from burning of fossil fuels, soot, vehicle smoke, smoke produced by various industries, and tobacco smoke (Mulholland, 2002; Klote, 2016). According to the definition proposed by national fire prevention association (NFPA), smoke is the result of pyrolysis and combustion of particles that is spread in solid, liquid, and gas form into the air (Mulholland, 2002; Klote, 2016). Different sources of smoke have different compounds, some of which are extremely harmful to the human health. Tobacco smoke contains more than 4,000 chemicals and more than 50 carcinogens. Combustion of wood produces large amounts of carbon monoxide and carbon dioxide. Burning plastic at high temperatures leads to toxic vapor. Soot from burning various materials may be harmless on its own, but toxic gases can be absorbed on its surface and enter the human lung (Lee and Bye, 2019). In the process of cooking and storing food, the materials are exposed to smoke and toxic compounds of smoke can enter food depending on fuel source, furnace temperature, and other factors (Essumang et al., 2013). Various chemical compounds, such as carbon monoxide, benzene, formaldehyde, hydrogen cyanide, polyaromatic hydrocarbons, toluene, ethylbenzene, xylene, nicotine, phenanthrene, anthracene, chlorine, hydrochloric acid, isocyanate, and nitrosamines are among the most harmful compounds of smoke (Bullen, 2008; Moir et al., 2008; Essumang et al., 2013; Chambers et al., 2018; Lee and Bye, 2019).

Chemical compounds of various types of smoke including fossil fuel smoke, soot, vehicle smoke, industrial smoke, and tobacco smoke enter the human body through breath, skin, and digestive tract (Chambers et al., 2018). Toxicity and pathogenicity of smoke depends on its chemical compounds, such as irritating gases and thermal agents. Previous studies have reported a relationship between chemical compounds in tobacco smoke and diseases like cancer, cardiovascular diseases, rheumatoid arthritis, lung diseases, diabetes, infertility, and other diseases (Mostafa, 2010; Ligthart et al., 2016; Sudano et al., 2018; Ishikawa and Terao, 2020; Nemmar et al., 2020). However, despite significant technological advances, the relationship between diseases, xenobiotics, and synthetic compounds found in other sources of smoke has still remained unclear (Cho and Yoon, 2017)

Identification of the mechanism of action of harmful agents is the first step of preventing disease, and identification of disease causing power of these agents in human is a major challenge

(Afshari et al., 1999; Van Hummelen and Sasaki, 2010). Primarily, animal models, such as rodents have been used to evaluate the risk of chemical compounds for human, but today *in vitro* and *in silico* methods have been developed for these cases and the scientific field of toxicogenomics has been established (Liu et al., 2019). Toxicogenomics is an interdisciplinary scientific field consisting of toxicology and genomics describing changes in genome, transcriptome and proteome of different organisms in response to chemical compounds (Hook, 2010; Liu et al., 2019). In this scientific field, techniques, such as microarray and next-generation sequencing (NGS) are used to understand cellular and molecular mechanisms involved in responding to chemicals (Liu et al., 2019; Rao et al., 2019). So far, microarray and NGS techniques have been used in many experiments to investigate molecular responses of various organisms to different types of smoke. Various studies have investigated the effect of cigarette smoke on transcriptome changes in cells, tissues, and organs of human, mouse ,and rat including leukocytes, alveolar macrophages, lung epitheliuml cells, gingival epitheliuml cells, nasal epitheliuml cells, keratinocytes, monocytes, T cells, monocytes, placenta, whole lung, heart, liver, muscle, testis and ovary (Gebel et al., 2006; Büttner et al., 2007; Huuskonen et al., 2008; Talhout et al., 2011; Yauk et al., 2012; Sobinoff et al., 2013; Xiao et al., 2013; Kogel et al., 2014; Sobinoff et al., 2014; Wang et al., 2017; Woo et al., 2017; Percoco et al., 2020). Such studies have also investigated the transcriptome responses of various cells, tissues, and organs of human, mouse, and rat in exposure to smoke of coal, gasoline and diesel vehicles (Gong et al., 2007; Peretz et al., 2007; Gottipolu et al., 2009; Stevens et al., 2010; Pettit et al., 2012; Yokota et al., 2013; Schisler et al., 2015; Wang et al., 2015; Rynning et al., 2018; Drizik et al., 2020). The large amount of information generated by these studies is stored in databases, such as gene expression omnibus (GEO) and sequence read archive (SRA) and are available to the public (Leinonen et al., 2010; Barrett et al., 2012). Despite many studies conducted on the effects of smoke on cells, tissues, and organs of different organisms, to the best of our knowledge, to date, no comprehensive research has investigated the effects of different sources of smoke simultaneously. Furthermore, no previous study has identified genes and gene regulatory networks in various organisms in response to different sources of smoke. Identification of the genes involved in responding to different sources of smoke, although can be useful for understanding the mechanisms, but is not enough on its own and in some cases, may generate unreliable results. Analyzing genes and investigating their position and role in gene regulatory

networks facilitate the understanding of various mechanisms and presents reliable results (Hardt et al., 2016; Kim, 2017). In fact, only the investigation of genes in gene regulatory networks can determine their true biological functions in response to various pollutants (Iida and Takemoto, 2018). Previous studies have shown high power of system biology methods in identifying the relationship between different pollutants and biological functions, diseases, specific biological pathways, and gene regulatory factors (Perkins et al., 2011; Chen et al., 2014; Villeneuve et al., 2014; Zhernovkov et al., 2019).

Accordingly, this study was conducted to investigate the effects of different sources of smoke on transcriptome changes in human cells, tissues, and organs and to find specific genes related to smoke. In the present study, genes that regulate response mechanisms to different sources of smoke will also be identified and investigated.

2. Material and Methods

2.1. Data Collection and Expression Analysis

In this study, all microarray datasets related to the effect of different smoke sources on cells, tissues, or organs of human were obtained from the GEO database of the NCBI site. A total of 61 datasets were obtained that included the effects of cigarette, tobacco, coal, and diesel smokes on the cells, tissues and organs of human (Supplementary file). We used the GEO2R tool to analyze the normalized samples in the dataset and identify differentially expressed genes (DEGs) (Barrett et al., 2012). In these analyses, smoke-exposed samples were compared to control samples to detect DEGs. The DEGs were filtered with FDR < 0.05 and $|\text{Log 2 (fold change)}| \geq 1$. The missing gene symbols, ambiguous names, and duplicate values were manually eliminated from DEGs. The filtered DEGs were considered as smoke-related genes (SRGs) and transferred to the SRGs pool. The SRGs pool was further filtered and the duplicate SRGs were removed. The common SRGs among cells, tissues and organs were identified using a Venn diagram.

2.2. Gene Ontology Analysis, Pathway and Disease Enrichment of SRGs

The SRGs were classified among different biological processes, molecular functions, cellular components, pathways and diseases using the Comparative Toxicogenomics Database (CTD)

(Davis et al., 2018) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2016). Classification results with p-value < 0.01 considered significant.

2.3. Identification of Smoke-Related Transcription Factors (SRTFs)

Transcription factors (TFs) regulate chromatin and transcription by forming a complex system that controls the expression of genes in the cell. The human transcription factors database and Tfcheckpoint server were searched to identify SRTFs in the SRGs list (Chawla et al., 2013; Lambert et al., 2018). The TFs in the human transcription factors database and TFcheckpoint database are manually assessed for DNA binding activity and their role in regulating RNA polymerase II (Chawla et al., 2013). In this study, only those SRTFs that met mentioned criteria were selected for subsequent evaluations.

2.4. Finding Hub SRGs in the Gene Regulatory Network

Finding hub SRGs was performed using gene regulatory network analysis. For this purpose, the SRGs list, including SRTFs, was imported to the Enrichr and their target genes were identified. The Enrichr results were filtered with a p-value < 0.01 (Lachmann et al., 2010; Kuleshov et al., 2016). The filtered results were used to construct gene regulatory network based on the interaction between SRGs and their target genes. The gene regulatory network was constructed using Cytoscape 3.8.2 software (Franz et al., 2015). The gene regulatory network analysis and detection of hub genes was performed by Cytohubba plug-in of Cytoscape software (Chin et al., 2014). This plug-in of Cytoscape software uses 12 specific methods, including betweenness, bottleneck, closeness, clustering coefficient, degree, dmnc, eccentricity, epc, mcc, mnc, radiality and stress to find important hub nodes in the networks (Chin et al., 2014). The hub SRGs were identified based on 12 different methods in the Cytohubba plug-in of Cytoscape software.

2.5. Expression Analysis of Identified Hub SRGs in Different Tissues

The HPA tissue gene expression profiles data-set was obtained from the human transcriptome database of NCBI. This data-set contains transcriptome of 27 different tissues of human that are generated by high-throughput sequencing (Fagerberg et al., 2014). The RPKM values of SRGs and hub SRGs were extracted from the transcriptome of the human adrenal, appendix, bone marrow, brain, colon, duodenum, endometrium, esophagus, gall bladder, heart, kidney, liver,

lung, lymph node, ovary, pancreas, placenta, prostate, salivary gland, skin, small intestine, spleen, stomach, testis, thyroid, and urinary bladder tissues. The mean of RPKM values of SRGs and hub SRGs in each tissue was considered as the tissue-specific expression.

2.6. Validation of the Results

The batch query software in CTD database was used to validate the results (Davis et al., 2018). This software applies a text-mining algorithm to generate the results. The hub SRGs were imported to the batch query software and their expression trend (up- and down-regulation) was determined under smoke exposure.

3. Results

3.1. Identification of Smoke-Related Genes (SRGs)

Among 61 studied microarray data-sets, the analysis results of the data-sets related to common tissues were merged, and after excluding the repeated genes, they were used at the next stages. After investigating all microarray data-sets, a total of 6,484 smoke-related genes (SRGs) were identified in nine tissues and cells, including alveolar epithelium, alveolar macrophages, buccal epithelium, keratinocytes, lymphocytes, monocytes, nasal epithelium, placenta, and small and long airways epithelium (Fig. 1A).

In alveolar epithelium, 2,997 SRGs with different expression were identified. In alveolar macrophages, 246 SRGs with different expression were observed. Fifty-three SRGs with different expression were found in buccal epithelium. A total of 1,071 SRGs with different expression were observed in keratinocytes. In lymphocytes, 12 SRGs with different expression were observed. A total of 178 SRGs with different expression were identified in monocytes. In nasal epithelium, three SRGs with different expression were induced. Eleven SRGs with different expression were identified in placenta. In small and long airways epithelium, 1,913 SRGs were induced with different expression (Fig. 1B). The highest and lowest number of SRGs was observed in alveolar epithelium and nasal epithelium with different expression, respectively (Fig. 1B). The number of up-regulated SRGs was higher than down-regulated ones in alveolar epithelium, alveolar macrophages, buccal epithelium, lymphocytes, and small and long airways epithelium (Fig. 1B). In keratinocytes, monocytes, nasal epithelium, and placenta the number of

down-regulated SRGs was higher than up-regulated SRGs (Fig. 1B). In lymphocytes, no down-regulated SRG was observed, and all SRGs were up-regulated (Fig. 1B).

The common SRGs among the tissues were identified using Venn diagram. Venn diagram showed no common gene among nine studied tissues and cells, but a common gene, namely, cytochrome P450 1B1 (*CYP1B1*) was identified among seven tissues and cells, including alveolar epithelium, alveolar macrophages, buccal epithelium, keratinocytes, lymphocytes, nasal epithelium, and small and long airways epithelium. Investigation of *CYP1B1* common SRG expression among seven tissues and cells showed that this gene was up-regulated in alveolar epithelium, buccal epithelium, keratinocytes, lymphocytes, and small and long airways epithelium exposed to smoke. It was down-regulated in alveolar macrophages and nasal epithelium exposed to smoke (Fig. 2A). The *CYP1B1* gene showed almost the same expression trend among seven tissues and cells (Fig. 2A). Investigation of *CYP1B1* gene expression among different human tissues showed that the highest and lowest expression of this gene was in endometrium and salivary gland tissues, respectively (Fig. 2B).

3.2. Identification of Smoke-related Transcription Factors (SRTFs)

Investigation of smoke-related transcription factors (SRTFs) in SRGs led to the identification of 459 SRTFs (Fig. 3A). In alveolar epithelium, 216 SRTFs with different expression were identified. In alveolar macrophages, 15 SRTFs with different expression were observed. Three SRTFs with different expression were found in buccal epithelium. A total of 57 SRTFs with different expression were induced in keratinocytes. No SRTF with different expression was observed in lymphocytes. Thirteen SRTFs with different expression were identified in monocytes. No SRTFs with different expression were induced in the nasal epithelium. One SRTF with different expression was identified in placenta. In small and long airways epithelium, 154 SRTFs were induced with different expression (Fig. 3B). The highest and lowest number of SRTFs were observed in alveolar epithelium and placenta with different expression, respectively (Fig. 3B). The number of up-regulated SRTFs was higher than down-regulated ones in alveolar epithelium, alveolar macrophages, buccal epithelium, and small and long airways epithelium (Fig. 3B). In keratinocytes, monocytes, and placenta down-regulated SRTFs were higher than up-regulated ones (Fig. 3B). No up-regulated SRTF was observed in placenta (Fig. 3B).

The common SRTFs among cells and tissues were identified using Venn diagram. Venn diagram showed no common gene among nine studied cells and tissues, but identified a common gene (*NR4A2*) among four cells and tissues, including alveolar epithelium, keratinocytes, monocytes, and small and long airways epithelium. The *NR4A2* common SRTFs showed up-regulation in keratinocytes and monocytes and down-regulation in alveolar epithelium and small and long airways exposed to smoke (Fig. 4A). Our results showed a different trend for *NR4A2* expression among four studied cells and tissues (Fig. 4A). Investigation of *NR4A2* gene expression among 27 different human tissues showed that the highest and lowest expression of this gene was in adrenal gland and pancreas tissues, respectively (Fig. 4B).

3.3. Gene Ontology, Enrichment of Diseases and Metabolic Pathways of SRGs

Gene ontology analysis classified the identified SRGs into three different groups: biological processes, cellular components, and molecular functions (Table 1). In the biological processes group, the highest numbers of SRGs were classified into cellular process, biological regulation, and metabolic process. The highest numbers of SRGs in the cellular components group were enriched in cell, intracellular, and intracellular parts (Table 1). The dominant terms in the molecular function group included binding and protein binding (Table 1). Metabolic pathway analysis revealed that the identified SRGs were significantly enriched in signal transduction, immune system, and metabolism (Table 2). Disease enrichment analysis showed that cancers, digestive system diseases, and nervous system diseases are the most common diseases associated with SRGs (Table 3).

3.4. Gene Regulatory Network Analysis

Gene regulatory network analysis was performed for all SRGs and for SRGs of five cells and tissues which had the highest number of SRGs including alveolar epithelium, alveolar macrophages, keratinocytes, monocytes, and small and long airways epithelium. Moreover, the gene regulatory network of common *CYP1B1* gene was constructed to further investigate the role of this gene. Investigation of constructed gene regulatory networks for SRGs using CytoHubba plug-in revealed different hub genes for studied cells and tissues (Fig. 5). The identified hub genes had the highest interaction with other SRGs in gene regulatory networks. Investigation of

the regulatory network of common gene of *CYP1B1* showed that this gene is a hub gene and has the highest interaction with other genes in the network (Fig. 6).

3.5. Validation of Results

The results of analyses were confirmed using text-mining and RNA-seq data. For this purpose, the batch query available on the comparative toxicogenomic database (CTD) site and the search engine available on Expression Atlas site were used to confirm the results. The results of text-mining and RNA-seq data analysis confirmed the SRGs identified in this study. These results also showed that the *CYP1B1* is one of SRGs that has the same expression trend in human, mouse, and rat in different cells and tissues exposed to smoke (Fig. 7). These results can confirm the role of identified SRGs in the response of different cells and tissues to smoke.

4. Discussion

Chemical compounds of fossil fuels smoke, soot, vehicle smoke, industrial smoke, and tobacco smoke enter the human body through breath, skin, and digestive tract and influence various cells and tissues (Chambers et al., 2018). In this study, transcriptome changes in different human cells and tissues exposed to various sources of smoke were investigated and analyzed. After removing the duplicate genes from all microarray data-sets, a total of 6,484 SRGs were identified that their expression changes under exposure of the smoke. The number of up-regulated SRGs (4,095) was significantly higher than down-regulated ones (2,389). These results indicate that smoke stimulate more SRGs to up-regulation. Gene ontology analysis revealed that the identified SRGs are more involved in regulatory and metabolic processes in the cell. Metabolic pathway enrichment analysis confirmed the results of gene ontology and showed that the identified SRGs are involved in the regulatory, metabolic, and immune processes. Cellular regulation involves a wide range of processes from cell division to response to various stimuli. The induction and alteration of these processes can reflect the effect of different compounds on the cell. The greater stimulation and alteration in cell regulatory processes by a compound, increases its potential risks for the cell. A large number of the identified SRGs are involved in regulatory processes that can reflect short- and long-term harmful effects of compounds in different types of smoke. Smoke contains strong oxidizing compounds, such as oxygen radicals, nitrogen, and volatile aldehydes that can change the natural metabolism of cells and cause serious damages to vital

molecules. On the one hand, smoke-exposed cells counteract damages caused by these oxidizing compounds by changing their metabolic processes (Menon et al., 2011; Sabra et al., 2017; Solanki et al., 2018; Reigle et al., 2021). Functional analysis of SRGs showed that they are mainly part of genes stimulated in cancers, digestive system diseases, and nervous system diseases. Smoke contains more than 100 carcinogens, such as polycyclic aromatic hydrocarbons (Pfeifer et al., 2002; Godschalk et al., 2003; Mayne and Lippman, 2005). These compounds are particularly associated with cancers and other diseases of respiratory and gastrointestinal tract as the organs with the highest contact with compounds of smoke (Gandini et al., 2008; Talhout et al., 2011). Furthermore, strong oxidizing compounds of smoke may cause nervous system diseases such as Alzheimer's and Parkinson's diseases, aging, and other neurological disorders due to the stimulation of reactive oxygen species (ROS) production (Uttara et al., 2009; Cho and Yoon, 2017).

The studied cells and tissues had different numbers of SRGs as well as different responses to smoke. Based on the number of SRGs in each tissue, these responses can be grouped from severe to moderate, and poor. Alveolar epithelium, small and long airways epithelium, and keratinocytes with the highest number of SRGs showed strong responses to compounds of smoke, respectively. Alveolar macrophages and monocytes showed moderate responses to these compounds. Finally, buccal epithelium, lymphocytes, nasal epithelium, and placenta showed a poor response to these compounds with the lowest number of SRGs, respectively. A similar trend was observed for SRTFs, and interestingly, compounds of smoke did not stimulate any SRTFs in lymphocytes and nasal epithelium. Differences in expression of regulatory elements, such as SRTFs, among the studied cells and tissues indicate variations in the type of gene regulation and response to smoke. Transcriptomic responses may reflect sensitivity of different cells and tissues to smoke. This may also be related to the mechanism of action of smoke compounds in different cells and tissues. Unlike cell genome, which Maintains its structure, transcriptome is highly dynamic and responds in the face of harmful compounds (Joseph, 2017). Thus, smoke-related transcriptomic changes may be a primary index of toxicity of smoke compounds (Waring et al., 2001; Hamadeh et al., 2002; Heinloth et al., 2004; Otava et al., 2015; Joseph, 2017). Gene network analysis confirmed the results of other analyses and it was found that the studied cells and tissues had different gene regulation processes and responses to compounds of smoke.

Different regulatory network and hub genes indicate that different cells and tissues induce different genes in the face of smoke compounds to counteract harmful effects of them.

Comparison of different tissues showed no common SRG among all the studied tissues. This could be another reason for different sensitivity and responses of various cells and tissues to smoke. However, *CYP1B1* gene was common among the seven studied cells and tissues and showed the same expression pattern. The results of text-mining and data of Expression Atlas site showed that *CYP1B1* was induced by smoke and was up-regulated in most tissues of human and model organisms such as mouse and rat. Network analysis showed that *CYP1B1* gene is a hub gene in the network. *CYP1B1* gene produces an enzyme that belongs to the superfamily of cytochrome P450 enzymes. This gene is a single copy gene and is located in human chromosome 2 (Tang et al., 1996). *CYP1B1* gene is 4,667 bp in length and consists of two exons and one intron. Coding sequence of this gene produces a 543-amino acid protein that has no signal peptide and has a functional domain of cytochrome P450 in its structure. *CYP1B1* gene has the highest expression in endometrium and the lowest expression in salivary gland. *CYP1B1* is a monooxygenase involved in various processes, such as lipid production, particularly cholesterol, metabolism, and breakdown a wide range of toxic and carcinogenic drugs and chemical compounds, including polycyclic aromatic hydrocarbons (Murray et al., 2001; A Faiq et al., 2014). Various studies have shown a positive correlation between up-regulation of *CYP1B1* gene and genotoxicity as well as oxidative stress caused by smoke compounds that can cause malignancies in tissues (Shimada, 2006; Shimada and Guengerich, 2006; Jeyabalan et al., 2011; Jung et al., 2012; Shang et al., 2013; Hussain et al., 2014).

Due to the different responses of rat tissues to smoke, it is possible that other factors such as epigenetic changes are also involved in the response to smoke. Previous studies have shown that chronic exposure to chemicals in environment such as smoke, in addition to genetic changes, also can lead to epigenetic changes (Soza-Ried et al., 2019). Some smoke-related epigenetic changes are tissue-specific, but others are common in different tissues (Bakulski et al., 2019). Joehanes et al. detected cigarette smoking has a large influence on genome-wide methylation. Several differentially methylated genes were unique in terms of the biological impacts of smoking, and they might be medical targets for the prevention or treatment of tobacco-related illnesses. Methylation at these locations may create complications as sensitive and persistent indicators of lifetime tobacco smoke exposure (Joehanes et al., 2016).

Hammouz et al. discovered that smoking has a significant influence on *E2F* (a transcription factor involved in the control of the G1/S transition in humans), and that sex and smoking status may have a biological effect on epithelial-mesenchymal transition-related genes such as *HEY2*, *OLFM1*, *SFRP1*, and *STRAP* (Hammouz et al., 2020). Moreover, Richter et al. revealed that in response to tobacco smoke exposure, hypo-methylation of *AHRR* and *CYP1B1* is an essential regulatory mechanism of xenobiotic metabolism of the masticatory mucosa (Richter et al., 2019). Results of a similar study showed that cigarette smoke induces hypo-methylation of *CYP1B1*, *CYP1A1*, *ALDH3A1*, and *SFRP2* genes in human small airway epithelium (Buro-Auriemma et al., 2013).

As well as in several studies, arsenic has been identified as the cause of epigenetic changes and genome instabiltiy (Mass and Wang, 1997; Sciandrello et al., 2011; Mauro et al., 2016; Bjørklund et al., 2018). Arsenic is one of the most harmful compounds in various types of smoke (Ferreccio et al., 2013). Arsenic has a carcinogenic effect in humans and is found within a range of 40–120 μg per unfiltered cigarette (Soza-Ried et al., 2019). Mass and Wang discovered that arsenic induces hyper-methylation of the *p53* gene (Mass and Wang, 1997). Mahna reported that chronic exposure of arsenic induces methylation status of the p53 gene (Mahna et al., 2021). Therefore, DNA methylation is affected by arsenic exposure, especially in the *p53* promoter region, which might be connected to the process of arsenic carcinogenesis and the reported higher cancer incidence later in life (Intarasunanont et al., 2012). Arsenic induces genome-wide hypomethylation that can lead to genomic instability. Arsenic can also alter the methylation of CpG islands in some specific gene promoter and reduce or increase their expression. As a result of these changes, cells may undergo epigenetic reprogramming at both the gene and genome level (Sciandrello et al., 2011; Mauro et al., 2016). Production of reactive oxygen species (ROS) by arsenite can lead to increase DNA damage and genomic instability. Arsenite induces genomic instability through chromosomal abnormalities. On the other hand, ROS produced by Arsenite can increase the severity of this genomic instability. Also, DNA-protein crosslinks, mutations through oxidative DNA adduct formation, single- or double-stranded DNA breaks are other side effects of Arsenite-produced ROS (Sciandrello et al., 2011; Rao et al., 2017). Arsenic exposure also decreases the expression of *CYP1B1* gene in human (Xie et al., 2020). Moreover, the *CYP1B1* locus harboured DNA methylation changes have been observed in exposure to arsenic (Rojas et al., 2015). In laboratory research and human population studies, changes in DNA

methylation, histone post-translational modifications, and microRNAs have been identified following arsenic exposure (Bjørklund et al., 2018). Although many epidemiologic evidence and several studies have been performed on the epigenetic effects of arsenic and other environmental contaminants, further studies are needed to identify their toxicity mechanisms.

In this study, for the first time, we found that compounds of smoke induce and increase expression of *CYP1B1* gene in all target and non-target tissues of smoke. This increase in *CYP1B1* gene expression in target and non-target tissues indicated that smoke compounds can causes extensive damage to human (Hussain et al., 2014). Other members of P450 cytochrome family are also induced in response to smoke compounds in various tissues, some of them are specific to particular tissues and the others are induced in few tissues (Parke, 1994; Gonzalez, 2005; Guengerich, 2005; Klaunig et al., 2010). However, our findings revealed that *CYP1B1* is the only cytochrome induced in most tissues of human, rat, and mouse in response to smoke compounds. The *CYP1B1* gene contains xenobiotic response elements in its promoter to respond to smoke compounds. These elements are induced by aryl hydrocarbon receptors in response to smoke compounds and increase the expression of *CYP1B1* gene (Sutter and Greenlee, 1992; Alexander et al., 1997; Zhang et al., 1998; Shehin et al., 2000; Chang et al., 2007). Our findings and the results of previous studies suggest that the *CYP1B1* gene expression may be regulated by epigenetic mechanisms in tissues or cells exposed to smoke, however, further studies are needed to better understand this epigenetic mechanism.

It is not clear whether a specific compound or a set of compounds in smoke stimulates aryl hydrocarbon receptors and expression of *CYP1B1* gene. On the other hand, the mechanism of uniform induction of this gene in target tissues that directly exposed to compounds of smoke and in non-target tissues that indirectly exposed to them has not been identified. Possibly, it occurs through the one or more unknown inter-tissue signaling molecules. Although compounds of different types of smoke induce epigenetic changes and expressions of very few genes in some tissues, induction of single genes, such as *CYP1B1* or other members of cytochrome P450 family with high damaging effects alone is adequate to determine the harmful effects of these compounds on human.

Conclusion

In the present study, SRGs were identified in different human cells and tissues using system biology-based methods. Our results showed that there are very little SRGs common among the different human cells and tissues, and they use different response mechanisms in face of smoke. The number of SRGs and type of response mechanism in each cell and tissue indicate its sensitivity to smoke. We also found that compounds of smoke induce and increase the expression of *CYP1B1* gene in all target and non-target tissues of human. Induction of single genes such as *CYP1B1*, which have high damaging effects on all target and non-target tissues, indicates extensive damage of smoke on human. However, the uniform mechanism of induction of this gene in target and non-target tissues in face of smoke has not been determined. Therefore, the role of genetic and epigenetic modifications associated with smoke exposure need further studies. Our findings can help discover mechanism of action of messenger molecules in response of target and non-target tissues to smoke. The key gene of *CYP1B1* identified in this study can be investigated in diagnostic or therapeutic studies. Moreover, despite the specific characteristics of this gene and its identical expression trend in target and non-target tissues, it can be used as a biomarker for diagnosis and prognosis.

5. Compliance with ethical standards

The present research does not involve human participants and/or animals.

Funding

The author(s) received no specific funding for this work.

Conflicts of interest/Competing interests

The authors have declared that no competing interests exist.

Figures legends

Fig. 1 Identification of smoke-related genes (SRGs) in various cells and tissues. A- Total number of identified SRGs. B- Number of SRGs identified in each cell or tissue. The numbers above and below of each column indicate the number of up-regulated (red) and down-regulated (blue) genes, respectively.

Fig. 2 Expression analysis of *CYP1B1* common SRG. A- Expression analysis of *CYP1B1* gene among seven cells and tissues exposed to smoke (red and blue columns indicate up-regulation and down-regulation in the studied cells and tissues, respectively. B- Expression analysis of *CYP1B1* gene among 27 different human tissues (columns show absolute value of gene expression)

Fig. 3 Identification of smoke-related transcription factors (SRTFs) in various cells and tissues. A- Total number of identified SRTFs. B- Number of SRTFs identified in each cell or tissue. The numbers above and below of each column indicate the number of up-regulated (red) and down-regulated (blue) genes, respectively.

Fig. 4 Expression analysis of *NR4A2* common SRTF. A- Expression analysis of *NR4A2* gene among four cells and tissues exposed to smoke (red and blue columns show up-regulation and down-regulation in the studied cells and tissues, respectively, B- Expression analysis of *NR4A2* gene e among 27 different human tissues (columns show absolute value of gene expression)

Fig. 5 Hub gene modules of SRGs regulatory network in different cells and tissues. Squares show hub genes. From red to yellow color, the rank of hub genes decreases in the network.

Fig. 6 The regulatory network of *CYP1B1* gene. Red and yellow rectangles represent hub genes and blue rectangles represent the other genes. From red to yellow, the rank of hub genes decreases in the network.

Fig. 7 Confirmation of analysis results for *CYP1B1* common SRG using text-mining and RNA-seq data. Red and blue show the up-regulation and down-regulation of *CYP1B1* gene in different tissues of human, mouse and rat, respectively.

References

A Faiq, M., Dada, R., Sharma, R., Saluja, D. and Dada, T., 2014. CYP1B1: a unique gene with unique characteristics. Current drug metabolism 15, 893-914.

Afshari, C.A., Nuwaysir, E.F. and Barrett, J.C., 1999. Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. Cancer Research 59, 4759-4760.

Alexander, D.L., Eltom, S.E. and Jefcoate, C.R., 1997. Ah receptor regulation of CYP1B1 expression in primary mouse embryo-derived cells. Cancer Research 57, 4498-4506.

Bakulski, K.M., Dou, J., Lin, N., London, S. and Colacino, J., 2019. DNA methylation signature of smoking in lung cancer is enriched for exposure signatures in newborn and adult blood. Scientific reports 9, 1-13.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M. and Holko, M., 2012. NCBI GEO: archive for functional genomics data sets—update. Nucleic acids research 41, D991-D995.

Bjørklund, G., Aaseth, J., Chirumbolo, S., Urbina, M.A., Uddin, R.J.E.g. and health, 2018. Effects of arsenic toxicity beyond epigenetic modifications. 40, 955-965.

Bullen, C., 2008. Impact of tobacco smoking and smoking cessation on cardiovascular risk and disease. Expert review of cardiovascular therapy 6, 883-895.

Buro-Auriemma, L.J., Salit, J., Hackett, N.R., Walters, M.S., Strulovici-Barel, Y., Staudt, M.R., Fuller, J., Mahmoud, M., Stevenson, C.S. and Hilton, H., 2013. Cigarette smoking induces small airway epithelial epigenetic changes with corresponding modulation of gene expression. Human molecular genetics 22, 4726-4738.

Büttner, P., Mosig, S. and Funke, H., 2007. Gene expression profiles of T lymphocytes are sensitive to the influence of heavy smoking: A pilot study. Immunogenetics 59, 37-43.

Chambers, D., Reese, C., Thornburg, L., Sanchez, E., Rafson, J., Blount, B., Ruhl III, J. and De Jesus, V., 2018. Distinguishing petroleum (crude oil and fuel) from smoke exposure within populations based on the relative blood levels of benzene, toluene, ethylbenzene, and xylenes (BTEX), styrene and 2, 5-dimethylfuran by pattern recognition using artificial neural networks. Environmental science & technology 52, 308-316.

Chang, J.T., Chang, H., Chen, P.-H., Lin, S.-L. and Lin, P., 2007. Requirement of aryl hydrocarbon receptor overexpression for CYP1B1 up-regulation and cell growth in human lung adenocarcinomas. Clinical Cancer Research 13, 38-45.

Chawla, K., Tripathi, S., Thommesen, L., Lægreid, A. and Kuiper, M., 2013. TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. Bioinformatics 29, 2519-2520.

Chen, J.C., Alvarez, M.J., Talos, F., Dhruv, H., Rieckhof, G.E., Iyer, A., Diefes, K.L., Aldape, K., Berens, M. and Shen, M.M., 2014. Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. Cell 159, 402-414.

Chin, C.-H., Chen, S.-H., Wu, H.-H., Ho, C.-W., Ko, M.-T. and Lin, C.-Y., 2014. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC systems biology 8, S11.

Cho, S. and Yoon, J.-Y., 2017. Organ-on-a-chip for assessing environmental toxicants. Current opinion in biotechnology 45, 34-42.

Davis, A.P., Grondin, C.J., Johnson, R.J., Sciaky, D., McMorran, R., Wiegers, J., Wiegers, T.C. and Mattingly, C.J., 2018. The comparative toxicogenomics database: Update 2019. Nucleic acids research 47, D948-D954.

Drizik, E., Corbett, S., Zheng, Y., Vermeulen, R., Dai, Y., Hu, W., Ren, D., Duan, H., Niu, Y. and Xu, J., 2020. Transcriptomic changes in the nasal epithelium associated with diesel engine exhaust exposure. Environment International 137, 105506.

Essumang, D., Dodoo, D. and Adjei, J., 2013. Effect of smoke generation sources and smoke curing duration on the levels of polycyclic aromatic hydrocarbon (PAH) in different suites of fish. Food and chemical toxicology 58, 86-94.

Fagerberg, L., Hallström, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A. and Edlund, K., 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Molecular & Cellular Proteomics 13, 397-406.

Ferreccio, C., Yuan, Y., Calle, J., Benítez, H., Parra, R.L., Acevedo, J., Smith, A.H., Liaw, J. and Steinmaus, C., 2013. Arsenic, tobacco smoke, and occupation: associations of multiple agents with lung and bladder cancer. Epidemiology (Cambridge, Mass.) 24, 898.

Franz, M., Lopes, C.T., Huck, G., Dong, Y., Sumer, O. and Bader, G.D., 2015. Cytoscape. js: a graph theory library for visualisation and analysis. Bioinformatics 32, 309-311.

Gandini, S., Botteri, E., Iodice, S., Boniol, M., Lowenfels, A.B., Maisonneuve, P. and Boyle, P., 2008. Tobacco smoking and cancer: a meta-analysis. International journal of cancer 122, 155-164.

Gebel, S., Gerstmayer, B., Kuhl, P., Borlak, J.r., Meurrens, K. and Müller, T., 2006. The kinetics of transcriptomic changes induced by cigarette smoke in rat lungs reveals a specific program of defense, inflammation, and circadian clock gene expression. Toxicological Sciences 93, 422-431.

Godschalk, R.W., Van Schooten, F.-J. and Bartsch, H., 2003. A critical evaluation of DNA adducts as biological markers for human exposure to polycyclic aromatic compounds. BMB Reports 36, 1-11.

Gong, K.W., Zhao, W., Li, N., Barajas, B., Kleinman, M., Sioutas, C., Horvath, S., Lusis, A.J., Nel, A. and Araujo, J.A., 2007. Air-pollutant chemicals and oxidized lipids exhibit genome-wide synergistic effects on endothelial cells. Genome biology 8, R149.

Gonzalez, F.J., 2005. Role of cytochromes P450 in chemical toxicity and oxidative stress: studies with CYP2E1. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 569, 101-110.

Gottipolu, R.R., Wallenborn, J.G., Karoly, E.D., Schladweiler, M.C., Ledbetter, A.D., Krantz, T., Linak, W.P., Nyska, A., Johnson, J.A. and Thomas, R., 2009. One-month diesel exhaust inhalation produces hypertensive gene expression pattern in healthy rats. Environmental health perspectives 117, 38-46.

Guengerich, F.P., 2005. Cytochromes P450. Metabolism of Drugs and Other Xenobiotics; Anzenbacher, PE, Zanger, UME, Eds.

Hamadeh, H.K., Bushel, P.R., Jayadev, S., DiSorbo, O., Bennett, L., Li, L., Tennant, R., Stoll, R., Barrett, J.C. and Paules, R.S., 2002. Prediction of compound signature using high density gene expression profiling. Toxicological Sciences 67, 232-240.

Hammouz, R.Y., Kostanek, J.K., Dudzisz, A., Witas, P., Orzechowska, M. and Bednarek, A.K.J.J.o.a.g., 2020. Differential expression of lung adenocarcinoma transcriptome with signature of tobacco exposure. 61, 421-437.

Hardt, C., Beber, M.E., Rasche, A., Kamburov, A., Hebels, D., Kleinjans, J. and Herwig, R., 2016. ToxDB: pathway-level interpretation of drug-treatment data. Database 2016.

Heinloth, A.N., Irwin, R.D., Boorman, G.A., Nettesheim, P., Fannin, R.D., Sieber, S.O., Snell, M.L., Tucker, C.J., Li, L. and Travlos, G.S., 2004. Gene expression profiling of rat livers reveals indicators of potential adverse effects. Toxicological Sciences 80, 193-202.

Hook, S., 2010. Promise and progress in environmental genomics: a status report on the applications of microarray studies in ecologically relevant fish species. J Fish Biol 77, 1999-2022.

Hussain, T., Al-Attas, O.S., Al-Daghri, N.M., Mohammed, A.A., De Rosas, E., Ibrahim, S., Vinodson, B., Ansari, M.G. and El-Din, K.I.A., 2014. Induction of CYP1A1, CYP1A2, CYP1B1, increased oxidative stress and inflammation in the lung and liver tissues of rats exposed to incense smoke. Molecular and cellular biochemistry 391, 127-136.

Huuskonen, P., Storvik, M., Reinisalo, M., Honkakoski, P., Rysä, J., Hakkola, J. and Pasanen, M., 2008. Microarray analysis of the global alterations in the gene expression in the placentas from cigarette-smoking mothers. Clinical Pharmacology & Therapeutics 83, 542-550.

Iida, M. and Takemoto, K., 2018. A network biology-based approach to evaluating the effect of environmental contaminants on human interactome and diseases. Ecotoxicology and environmental safety 160, 316-327.

Intarasunanont, P., Navasumrit, P., Waraprasit, S., Chaisatra, K., Suk, W.A., Mahidol, C. and Ruchirawat, M.J.E.H., 2012. Effects of arsenic exposure on DNA methylation in cord blood samples from newborn babies and in a human lymphoblast cell line. 11, 1-13.

Ishikawa, Y. and Terao, C., 2020. The Impact of Cigarette Smoking on Risk of Rheumatoid Arthritis: A Narrative Review. Cells 9, 475.

Jeyabalan, J., Vadhanam, M.V., Ravoori, S. and Gupta, R.C., 2011. Sustained overexpression of CYP1A1 and 1B1 and steady accumulation of DNA adducts by low-dose, continuous exposure to benzo [a] pyrene by polymeric implants. Chemical research in toxicology 24, 1937-1943.

Joehanes, R., Just, A.C., Marioni, R.E., Pilling, L.C., Reynolds, L.M., Mandaviya, P.R., Guan, W., Xu, T., Elks, C.E. and Aslibekyan, S.J.C.c.g., 2016. Epigenetic signatures of cigarette smoking. 9, 436-447.

Joseph, P., 2017. Transcriptomics in toxicology. Food and Chemical Toxicology 109, 650-662.

Jung, M.H., Kim, H.R., Park, Y.J., Park, D.S., Chung, K.H. and Oh, S.M., 2012. Genotoxic effects and oxidative stress induced by organic extracts of particulate matter (PM10) collected from a subway tunnel in Seoul, Korea. Mutation Research/Genetic Toxicology and Environmental Mutagenesis 749, 39-47.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K., 2016. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic acids research 45, D353-D361.

Kim, S., 2017. Identifying dynamic pathway interactions based on clinical information. Computational Biology and Chemistry 68, 260-265.

Klaunig, J.E., Kamendulis, L.M. and Hocevar, B.A., 2010. Oxidative stress and oxidative damage in carcinogenesis. Toxicologic pathology 38, 96-109.

Klote, J.H., 2016. Smoke control, SFPE handbook of fire protection engineering. Springer, pp. 1785-1823.

Kogel, U., Schlage, W.K., Martin, F., Xiang, Y., Ansari, S., Leroy, P., Vanscheeuwijck, P., Gebel, S., Buettner, A. and Wyss, C., 2014. A 28-day rat inhalation study with an integrated molecular toxicology endpoint demonstrates reduced exposure effects for a prototypic modified risk tobacco product compared with conventional cigarettes. Food and chemical toxicology 68, 204-217.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M. and Lachmann, A., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic acids research 44, W90-W97.

Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R. and Ma'ayan, A., 2010. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics 26, 2438-2444.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R. and Weirauch, M.T., 2018. The human transcription factors. Cell 172, 650-665.

Lee, A. and Bye, M., 2019. Lung injury from hydrocarbon aspiration and smoke inhalation, Kendig's disorders of the respiratory tract in children. Elsevier, pp. 626-633. e2.

Leinonen, R., Sugawara, H., Shumway, M. and Collaboration, I.N.S.D., 2010. The sequence read archive. Nucleic acids research 39, D19-D21.

Ligthart, S., Steenaard, R.V., Peters, M.J., van Meurs, J.B., Sijbrands, E.J., Uitterlinden, A.G., Bonder, M.J., Hofman, A., Franco, O.H. and Dehghan, A., 2016. Tobacco smoking is associated with DNA methylation of diabetes susceptibility genes. Diabetologia 59, 998-1006.

Liu, Z., Huang, R., Roberts, R. and Tong, W., 2019. Toxicogenomics: a 2020 vision. Trends in Pharmacological Sciences 40, 92-103.

Mahna, D., Puri, S. and Sharma, S., 2021. DNA methylation modifications: mediation to stipulate pesticide toxicity. International Journal of Environmental Science and Technology 18, 531-544.

Mass, M.J. and Wang, L.J.M.R.R.i.M.R., 1997. Arsenic alters cytosine methylation patterns of the promoter of the tumor suppressor gene p53 in human lung cells: a model for a mechanism of carcinogenesis. 386, 263-277.

Mauro, M., Caradonna, F. and Klein, C.B., 2016. Dysregulation of DNA methylation induced by past arsenic treatment causes persistent genomic instability in mammalian cells. Environmental and molecular mutagenesis 57, 137-150.

Mayne, S.T. and Lippman, S.M., 2005. Cigarettes: a smoking gun in cancer chemoprevention. Oxford University Press.

Menon, R., Fortunato, S., Yu, J., Milne, G., Sanchez, S., Drobek, C., Lappas, M. and Taylor, R., 2011. Cigarette smoke induces oxidative stress and apoptosis in normal term fetal membranes. Placenta 32, 317-322.

Moir, D., Rickert, W.S., Levasseur, G., Larose, Y., Maertens, R., White, P. and Desjardins, S., 2008. A comparison of mainstream and sidestream marijuana and tobacco cigarette smoke produced under two machine smoking conditions. Chemical research in toxicology 21, 494-502.

Mostafa, T., 2010. Cigarette smoking and male infertility. Journal of Advanced Research 1, 179-186.

Mulholland, G.W., 2002. Smoke production and properties. SFPE handbook of fire protection engineering 3.

Murray, G.I., Melvin, W.T., Greenlee, W.F. and Burke, M.D., 2001. Regulation, function, and tissue-specific expression of cytochrome P450 CYP1B1. Annual review of pharmacology and toxicology 41, 297-316.

Nemmar, A., Al-Salam, S., Beegam, S., Yuvaraju, P. and Ali, B.H., 2020. Comparative Study on Pulmonary Toxicity in Mice Induced by Exposure to Unflavoured and Apple-and Strawberry-Flavoured Tobacco Waterpipe Smoke. Oxidative Medicine and Cellular Longevity 2020.

Otava, M., Shkedy, Z., Talloen, W., Verheyen, G.R. and Kasim, A., 2015. Identification of in vitro and in vivo disconnects using transcriptomic data. BMC genomics 16, 1-10.

Parke, D.V., 1994. The cytochromes P450 and mechanisms of chemical carcinogenesis. Environmental Health Perspectives 102, 852-853.

Percoco, G., Patatian, A., Eudier, F., Grisel, M., Bader, T., Lati, E., Savary, G., Picard, C. and Benech, P., 2020. Impact of cigarette smoke on physical-chemical and molecular proprieties of human skin in an ex vivo model. Experimental Dermatology.

Peretz, A., Peck, E.C., Bammler, T.K., Beyer, R.P., Sullivan, J.H., Trenga, C.A., Srinouanprachnah, S., Farin, F.M. and Kaufman, J.D., 2007. Diesel exhaust inhalation and assessment of peripheral

blood mononuclear cell gene transcription effects: an exploratory study of healthy human volunteers. Inhalation toxicology 19, 1107-1119.

Perkins, E.J., Chipman, J.K., Edwards, S., Habib, T., Falciani, F., Taylor, R., Van Aggelen, G., Vulpe, C., Antczak, P. and Loguinov, A., 2011. Reverse engineering adverse outcome pathways. Environmental toxicology and chemistry 30, 22-38.

Pettit, A.P., Brooks, A., Laumbach, R., Fiedler, N., Wang, Q., Strickland, P.O., Madura, K., Zhang, J. and Kipen, H.M., 2012. Alteration of peripheral blood monocyte gene expression in humans following diesel exhaust inhalation. Inhalation toxicology 24, 172-181.

Pfeifer, G.P., Denissenko, M.F., Olivier, M., Tretyakova, N., Hecht, S.S. and Hainaut, P., 2002. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene 21, 7435-7451.

Rao, C.V., Pal, S., Mohammed, A., Farooqui, M., Doescher, M.P., Asch, A.S. and Yamada, H.Y., 2017. Biological effects and epidemiological consequences of arsenic exposure, and reagents that can ameliorate arsenic damage in vivo. Oncotarget 8, 57605.

Rao, M.S., Van Vleet, T.R., Ciurlionis, R., Buck, W.R., Mittelstadt, S.W., Blomme, E.A. and Liguori, M.J., 2019. Comparison of RNA-seq and microarray gene expression platforms for the toxicogenomic evaluation of liver from short-term rat toxicity studies. Frontiers in genetics 9, 636.

Reigle, J., Secic, D., Biesiada, J., Wetzel, C., Shamsaei, B., Chu, J., Zang, Y., Zhang, X., Talbot, N.J. and Bischoff, M.E., 2021. Tobacco smoking induces metabolic reprogramming of renal cell carcinoma. The Journal of clinical investigation 131.

Richter, G.M., Kruppa, J., Munz, M., Wiehe, R., Häsler, R., Franke, A., Martins, O., Jockel-Schneider, Y., Bruckmann, C. and Dommisch, H.J.C.e., 2019. A combined epigenome-and transcriptome-wide association study of the oral masticatory mucosa assigns CYP1B1 a central role for epithelial health in smokers. 11, 1-18.

Rojas, D., Rager, J.E., Smeester, L., Bailey, K.A., Drobná, Z., Rubio-Andrade, M., Stýblo, M., García-Vargas, G. and Fry, R.C., 2015. Prenatal arsenic exposure and the epigenome: identifying sites of 5-methylcytosine alterations that predict functional changes in gene expression in newborn cord blood and subsequent birth outcomes. Toxicological sciences 143, 97-106.

Rynning, I., Neca, J., Vrbova, K., Libalova, H., Rossner Jr, P., Holme, J.A., Gützkow, K.B., Afanou, A.K.J., Arnoldussen, Y.J. and Hruba, E., 2018. In vitro transformation of human bronchial epithelial cells by diesel exhaust particles: gene expression profiling and early toxic responses. Toxicological Sciences 166, 51-64.

Sabra, S., Gratacós, E. and Roig, M.D.G., 2017. Smoking-induced changes in the maternal immune, endocrine, and metabolic pathways and their impact on fetal growth: a topical review. Fetal diagnosis and therapy 41, 241-250.

Schisler, J.C., Ronnebaum, S.M., Madden, M., Channell, M., Campen, M. and Willis, M.S., 2015. Endothelial inflammatory transcriptional responses to an altered plasma exposome following inhalation of diesel emissions. Inhalation toxicology 27, 272-280.

Sciandrello, G., Mauro, M., Catanzaro, I., Saverini, M., Caradonna, F. and Barbata, G., 2011. Long-lasting genomic instability following arsenite exposure in mammalian cells: The role of reactive oxygen species. Environmental and molecular mutagenesis 52, 562-568.

Shang, Y., Fan, L., Feng, J., Lv, S., Wu, M., Li, B. and Zang, Y.-S., 2013. Genotoxic and inflammatory effects of organic extracts from traffic-related particulate matter in human lung epithelial A549 cells: the role of quinones. Toxicology in Vitro 27, 922-931.

Shehin, S.E., Stephenson, R.O. and Greenlee, W.F., 2000. Transcriptional regulation of the human CYP1B1 gene: evidence for involvement of an aryl hydrocarbon receptor response element in constitutive expression. Journal of Biological Chemistry 275, 6770-6776.

Shimada, T., 2006. Xenobiotic-metabolizing enzymes involved in activation and detoxification of carcinogenic polycyclic aromatic hydrocarbons. Drug metabolism and pharmacokinetics 21, 257-276.

Shimada, T. and Guengerich, F.P., 2006. Inhibition of human cytochrome P450 1A1-, 1A2-, and 1B1-mediated activation of procarcinogens to genotoxic metabolites by polycyclic aromatic hydrocarbons. Chemical research in toxicology 19, 288-294.

Sobinoff, A.P., Beckett, E.L., Jarnicki, A.G., Sutherland, J.M., McCluskey, A., Hansbro, P.M. and McLaughlin, E.A., 2013. Scrambled and fried: cigarette smoke exposure causes antral follicle destruction and oocyte dysfunction through oxidative stress. Toxicology and applied pharmacology 271, 156-167.

Sobinoff, A.P., Sutherland, J.M., Beckett, E.L., Stanger, S.J., Johnson, R., Jarnicki, A.G., McCluskey, A., St John, J.C., Hansbro, P.M. and McLaughlin, E.A., 2014. Damaging legacy: maternal cigarette smoking has long-term consequences for male offspring fertility. Human reproduction 29, 2719-2735.

Solanki, H.S., Babu, N., Jain, A.P., Bhat, M.Y., Puttamallesh, V.N., Advani, J., Raja, R., Mangalaparthi, K.K., Kumar, M.M. and Prasad, T.K., 2018. Cigarette smoke induces mitochondrial metabolic reprogramming in lung cells. Mitochondrion 40, 58-70.

Soza-Ried, C., Bustamante, E., Caglevic, C., Rolfo, C., Sirera, R. and Marsiglia, H.J.C.r.i.o.h., 2019. Oncogenic role of arsenic exposure in lung cancer: A forgotten risk factor. 139, 128-133.

Stevens, T., Hester, S. and Gilmour, M.I., 2010. Differential Transcriptional changes in Mice exposed to chemically Distinct Diesel samples. Biomedical informatics insights 3, BII. S5363.

Sudano, I., Barthelmes, J. and Kubli, B., 2018. Smoking, smoking cessation and cardiovascular risk. Cardiovascular Medicine 21, 274-277.

Sutter, T. and Greenlee, W., 1992. Classification of members of the Ah gene battery. Chemosphere 25, 223-226.

Talhout, R., Schulz, T., Florek, E., Van Benthem, J., Wester, P. and Opperhuizen, A., 2011. Hazardous compounds in tobacco smoke. International journal of environmental research and public health 8, 613-628.

Tang, Y.M., Wo, Y.-Y.P., Stewart, J., Hawkins, A.L., Griffin, C.A., Sutter, T.R. and Greenlee, W.F., 1996. Isolation and characterization of the human cytochrome P450 CYP1B1 gene. Journal of Biological Chemistry 271, 28324-28330.

Uttara, B., Singh, A.V., Zamboni, P. and Mahajan, R., 2009. Oxidative stress and neurodegenerative diseases: a review of upstream and downstream antioxidant therapeutic options. Current neuropharmacology 7, 65-74.

Van Hummelen, P. and Sasaki, J., 2010. State-of-the-art genomics approaches in toxicology. Mutation Research/Reviews in Mutation Research 705, 165-171.

Villeneuve, D.L., Crump, D., Garcia-Reyero, N., Hecker, M., Hutchinson, T.H., LaLone, C.A., Landesmann, B., Lettieri, T., Munn, S. and Nepelska, M., 2014. Adverse outcome pathway (AOP) development I: strategies and principles. Toxicological Sciences 142, 312-320.

Wang, G., Zhou, H., Strulovici-Barel, Y., Al-Hijji, M., Ou, X., Salit, J., Walters, M.S., Staudt, M.R., Kaner, R.J. and Crystal, R.G., 2017. Role of OSGIN1 in mediating smoking-induced autophagy in the human airway epithelium. Autophagy 13, 1205-1220.

Wang, T.W., Vermeulen, R.C., Hu, W., Liu, G., Xiao, X., Alekseyev, Y., Xu, J., Reiss, B., Steiling, K. and Downward, G.S., 2015. Gene-expression profiling of buccal epithelium among non-smoking women exposed to household air pollution from smoky coal. Carcinogenesis 36, 1494-1501.

Waring, J.F., Jolly, R.A., Ciurlionis, R., Lum, P.Y., Praestgaard, J.T., Morfitt, D.C., Buratto, B., Roberts, C., Schadt, E. and Ulrich, R.G., 2001. Clustering of hepatotoxins based on mechanism of toxicity using gene expression profiles. Toxicology and applied pharmacology 175, 28-42.

Woo, S., Gao, H., Henderson, D., Zacharias, W., Liu, G., Tran, Q.T. and Prasad, G., 2017. AKR1C1 as a biomarker for differentiating the biological effects of combustible from non-combustible tobacco products. Genes 8, 132.

Xiao, R., Perveen, Z., Rouse, R.L., Le Donne, V., Paulsen, D.B., Ambalavanan, N. and Penn, A.L., 2013. In utero exposure to second-hand smoke aggravates the response to ovalbumin in adult mice. American journal of respiratory cell and molecular biology 49, 1102-1109.

Xie, L., Hu, W.-Y., Hu, D.-P., Shi, G., Li, Y., Yang, J. and Prins, G.S., 2020. Effects of inorganic arsenic on human prostate stem-progenitor cell transformation, autophagic flux blockade, and NRF2 pathway activation. Environmental health perspectives 128, 067008.

Yauk, C.L., Williams, A., Buick, J.K., Chen, G., Maertens, R.M., Halappanavar, S. and White, P.A., 2012. Genetic toxicology and toxicogenomic analysis of three cigarette smoke condensates in vitro reveals few differences among full-flavor, blonde, and light products. Environmental and molecular mutagenesis 53, 281-296.

Yokota, S., Hori, H., Umezawa, M., Kubota, N., Niki, R., Yanagita, S. and Takeda, K., 2013. Gene expression changes in the olfactory bulb of mice induced by exposure to diesel exhaust are dependent on animal rearing environment. PLoS One 8, e70145.

Zhang, L., Savas, Ü., Alexander, D.L. and Jefcoate, C.R., 1998. Characterization of the mouse Cyp1B1 gene: identification of an enhancer region that directs aryl hydrocarbon receptor-mediated constitutive and induced expression. Journal of Biological Chemistry 273, 5174-5183.

Zhernovkov, V., Santra, T., Cassidy, H., Rukhlenko, O., Matallanas, D., Krstic, A., Kolch, W., Lobaskin, V. and Kholodenko, B.N., 2019. An Integrative Computational Approach for a Prioritization of Key Transcription Regulators Associated With Nanomaterial-Induced Toxicity. Toxicological Sciences 171, 303-314.

Figures legends

Fig. 1 Identification of smoke-related genes (SRGs) in various cells and tissues. A- Total number of identified SRGs. B- Number of SRGs identified in each cell or tissue. The numbers above and below of each column indicate the number of up-regulated (red) and down-regulated (blue) genes, respectively.

Fig. 2 Expression analysis of *CYP1B1* common SRG. A- Expression analysis of *CYP1B1* gene among seven cells and tissues exposed to smoke (red and blue columns indicate up-regulation and down-regulation in the studied cells and tissues, respectively. B- Expression analysis of *CYP1B1* gene among 27 different human tissues (columns show absolute value of gene expression)

Fig. 3 Identification of smoke-related transcription factors (SRTFs) in various cells and tissues. A- Total number of identified SRTFs. B- Number of SRTFs identified in each cell or tissue. The numbers above and below of each column indicate the number of up-regulated (red) and down-regulated (blue) genes, respectively.

Fig. 4 Expression analysis of *NR4A2* common SRTF. A- Expression analysis of *NR4A2* gene among four cells and tissues exposed to smoke (red and blue columns show up-regulation and down-regulation in the studied cells and tissues, respectively, B- Expression analysis of *NR4A2* gene e among 27 different human tissues (columns show absolute value of gene expression)

Fig. 5 Hub gene modules of SRGs regulatory network in different cells and tissues. Squares show hub genes. From red to yellow color, the rank of hub genes decreases in the network.

Fig. 6 The regulatory network of *CYP1B1* gene. Red and yellow rectangles represent hub genes and blue rectangles represent the other genes. From red to yellow, the rank of hub genes decreases in the network.

Fig. 7 Confirmation of analysis results for *CYP1B1* common SRG using text-mining and RNA-seq data. Red and blue show the up-regulation and down-regulation of *CYP1B1* gene in different tissues of human, mouse and rat, respectively.

Table 1. Gene Ontology of SRGs

GO Term Name	GO Term ID	FDR	Enriched SRGs	GO Category
Biological regulation	GO:0065007	0	3082	Biological Process
Cellular macromolecule metabolic process	GO:0044260	0	2011	
Cellular metabolic process	GO:0044237	0	2734	
Cellular process	GO:0009987	0	3990	
Cellular response to stimulus	GO:0051716	0	1992	
Macromolecule metabolic process	GO:0043170	0	2334	
Metabolic process	GO:0008152	0	2907	
Multicellular organismal process	GO:0032501	0	1965	
Nitrogen compound metabolic process	GO:0006807	0	2540	
Organic substance metabolic process	GO:0071704	0	2801	
Primary metabolic process	GO:0044238	0	2670	
Regulation of biological process	GO:0050789	0	2899	
Regulation of cellular process	GO:0050794	0	2704	
Response to stimulus	GO:0050896	0	2402	
Cell part	GO:0044464	0	4102	Cellular components
Cytoplasmic part	GO:0044444	0	2682	
Intracellular	GO:0005622	0	3683	
Intracellular membrane-bounded organelle	GO:0043231	0	2717	
Intracellular organelle part	GO:0044446	0	2629	
Intracellular part	GO:0044424	0	3683	
Membrane-bounded organelle	GO:0043227	0	2961	
Organelle part	GO:0044422	0	2691	
Binding	GO:0005488	0	4108	Molecular Function
Protein binding	GO:0005515	0	3607	
Ion binding	GO:0043167	1.13e-316	1713	
Catalytic activity	GO:0003824	2.2E-253	1381	
Cation binding	GO:0043169	1.8E-196	1175	
Organic cyclic compound binding	GO:0097159	7.9E-196	1274	
Metal ion binding	GO:0046872	1.7E-195	1161	
Heterocyclic compound binding	GO:1901363	2.8E-191	1253	

Table 2. Analysis of metabolic pathways of SRGs

Pathway	Pathway ID	FDR	Enriched SRGs
Immune System	REACT:R-HSA-168256	1.8E-102	618
Metabolism	REACT:R-HSA-1430728	8.6E-94	611
Signal Transduction	REACT:R-HSA-162582	8.04E-84	665
Metabolic pathways	KEGG:HSA01100	1.79E-59	371
Innate Immune System	REACT:R-HSA-168249	2.21E-54	366
Cytokine Signaling in Immune system	REACT:R-HSA-1280215	6.85E-45	242
Metabolism of lipids and lipoproteins	REACT:R-HSA-556833	8.85E-43	249
Gene Expression	REACT:R-HSA-74160	7.66E-37	418
Generic Transcription Pathway	REACT:R-HSA-212436	1.16E-33	242
Cell Cycle	REACT:R-HSA-1640170	4.24E-32	190
Metabolism of proteins	REACT:R-HSA-392499	2.12E-31	368
Hemostasis	REACT:R-HSA-109582	2.9E-30	190
Adaptive Immune System	REACT:R-HSA-1280218	6.95E-28	219
Post-translational protein modification	REACT:R-HSA-597592	2.66E-27	257

Transmembrane transport of small molecules	REACT:R-HSA-382551	3E-27	191
Developmental Biology	REACT:R-HSA-1266738	4.26E-22	250
Signaling by GPCR	REACT:R-HSA-372790	3.93E-16	275

Table 3. Disease enrichment analysis of SRGs

Disease Name	**Disease ID**	**FDR**	**Enriched SRGs**
Cancers	MESH:D009369	1.90E-180	871
Digestive system diseases	MESH:D004066	2.14E-161	704
Nervous system diseases	MESH:D009422	3.28E-99	577
Genetic diseases (inborn)	MESH:D030342	1.30E-74	475
Musculoskeletal diseases	MESH:D009140	1.51E-80	379
Urogenital diseases	MESH:D052776	2.56E-74	347
Skin diseases	MESH:D012871	1.53E-74	338
Metabolic diseases	MESH:D008659	2.29E-54	332
Cardiovascular diseases	MESH:D002318	3.40E-51	320

Smoke-related genes (SRGs)

Smoke-related transcription factors (SRTFs)

National fire prevention association (NFPA)

Gene expression omnibus (GEO)

Sequence read archive (SRA)

Differentially expressed genes (DEGs)

Comparative Toxicogenomics Database (CTD)

Kyoto Encyclopedia of Genes and Genomes (KEGG)

Transcription factors (TFs)

CRediT author statement

Seyed Sajad Sohrabi, Seyyed Mohsen Sohrabi and Marzieh Rashidipour: Conceptualization, Methodology, Data curation, Writing- Original draft preparation.

Morteza Hadizadeh: Methodology, Data curation.

Seyed Sajad Sohrabi and Seyyed Mohsen Sohrabi analyzed the data.

Mohammad-Hossein Shams and Mohsen Mohammadi: Prepared the data.

All authors read and approved the final manuscript.

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered

as potential competing interests:

Marzieh Rashidipour reports was provided by Nutritional Health Research Center, Lorestan University of Medical Sciences, Khorramabad, Iran.

Highlights

- 6484 smoke-related genes were identified in various human tissues.
- Identification of smoke-related hub genes in various human tissues.
- Molecular mechanisms of different tissues exposed to smoke are different.