Egyptian Journal of Medical Human Genetics

RESEARCH

Open Access

Pathway analysis of smoking-induced changes in buccal mucosal gene expression

Anas Khaleel[1*], Bayan Alkhawaja[2], Talal Salem Al-Qaisi[3], Lubna Alshalabi[2] and Amneh H. Tarkhan[4]

Abstract

Background: Cigarette smoking is the leading preventable cause of death worldwide, and it is the most common cause of oral cancers. This study aims to provide a deeper understanding of the molecular pathways in the oral cavity that are altered by exposure to cigarette smoke.

Methods: The gene expression dataset (accession number GSE8987, GPL96) of buccal mucosa samples from smokers ($n=5$) and never smokers ($n=5$) was downloaded from The National Center for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO) repository. Differential expression was ascertained via NCBI's GEO2R software, and Ingenuity Pathway Analysis (IPA) software was used to perform a pathway analysis.

Results: A total of 459 genes were found to be significantly differentially expressed in smoker buccal mucosa ($p<0.05$). A total of 261 genes were over-expressed while 198 genes were under-expressed. The top canonical pathways predicted by IPA were nitric oxide and reactive oxygen production at macrophages, macrophages/fibroblasts and endothelial cells in rheumatoid arthritis, and thyroid cancer pathways. The IPA upstream analysis predicted that the TP53, APP, SMAD3, and TNF proteins as well as dexamethasone drug would be top transcriptional regulators.

Conclusions: IPA highlighted critical pathways of carcinogenesis, mainly nitric oxide and reactive oxygen production at macrophages, and confirmed widespread injury in the buccal mucosa due to exposure to cigarette smoke. Our findings suggest that cigarette smoking significantly impacts gene pathways in the buccal mucosa and may highlight potential targets for treating the effects of cigarette smoking.

Keyword: Gene expression, Cigarette smoking, Ingenuity Pathway Analysis, Buccal mucosa

Background

Tobacco smoking is responsible for one in six of all deaths from non-communicable diseases, leading experts to identify tobacco control as the highest priority public health intervention [1, 2]. The prevalence of smoking has fallen around the world over the past three decades, but the absolute number of people who smoke has increased [3]. Despite a coordinated worldwide effort against smoking, there are around 1.1 billion current smokers, and it is expected that this number would reach 1.9 billion by 2025 if current smoking patterns are maintained [4].

Cigarette smoke contains over 5000 chemicals, of which 98 have been identified as carcinogenic or probably carcinogenic to humans [5]. The plethora of carcinogens in cigarette smoke perturbs biological pathways related to cellular proliferation, inflammation, and tissue injury, with strong links to various types of cancer [6, 7]. In cancer patients, cigarette smoking has been associated with an increased symptom burden as well as a reduced efficacy of chemotherapy [6, 8].

Smoking-induced differential gene expression has been well-documented in previous studies. In fact, smoking has a characteristic impact on the transcriptome, as it activates inflammatory and oxidative responses, changes airway structures, and alters gene expression across tissue types [9]. Previous studies have shown that cigarette smoking significantly alters the gene expression profiles

*Correspondence: anas.khaleel@uop.edu.jo
[1] Department of Pharmacology and Biomedical Sciences, Faculty of Pharmacy and Medical Sciences, University of Petra, Amman, Jordan
Full list of author information is available at the end of the article

Springer Open

© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

of adipose tissue, buccal cells, nasal epithelial cells, lung tissue, and whole blood [10–14].

The aim of the current study is to broaden the understanding of the molecular pathways that are altered in buccal mucosa after exposure to cigarette smoke. Gene expression data from smokers and never smokers were analyzed via Ingenuity Pathway Analysis (IPA), which is a web-based software application that identifies new targets within the context of biological systems.

Methods

Data acquisition

The microarray dataset investigated in the present study was obtained from The National Center for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO) repository (accession number GSE8987). This dataset included gene expression data of buccal mucosa samples from smokers ($n=5$) and never smokers ($n=5$) [15]. Smokers were classified as those who had smoked at least 10 cigarettes per day and who had a cumulative smoking history of at least 10 pack years [15]. Table 1 shows the gene expression data samples included in the current study.

As per the original study by Sridhar et al., buccal mucosa samples were collected from the study participants by scraping the inside of their mouths with a concave plastic tool with serrated edges. Total RNA was extracted from buccal mucosa samples using TRIzol reagent (Invitrogen, Carlsbad, CA), and RNA integrity was assessed using a denaturing agarose gel. The Affymetrix Human Genome U133A (HG-U133A) Array (Affymetrix, Santa Clara, CA) was then used to profile the gene expression of the extracted total RNA samples [15].

The demographics of the 10 subjects varied with regard to sex, age, and race. Among the 5 smokers, the mean age was 36 years old (±8 years), with 1 male and 4 females. Similarly, the mean age of the 5 never smokers was 31 years old (±9 years), with 2 males and 3 females. In terms of race, the smoker group comprised 3 Caucasians and 2 African Americans, while the never-smoker group consisted of 2 Caucasians and 3 African Americans. Demographic data for individual subjects were not provided in the dataset, but statistical comparisons of the smoker and never-smoker groups revealed not significant p values for sex ($p=0.42$), age ($p=0.36$), and race ($p=0.40$) [15].

Table 1 Gene expression data samples included in the current analysis

Sample nos.	Subject	Type of sample	Source of sample
GSM227858	Never smoker	RNA	Buccal mucosa
GSM227859	Never smoker	RNA	Buccal mucosa
GSM227860	Never smoker	RNA	Buccal mucosa
GSM227861	Smoker	RNA	Buccal mucosa
GSM227862	Smoker	RNA	Buccal mucosa
GSM227863	Smoker	RNA	Buccal mucosa
GSM227864	Smoker	RNA	Buccal mucosa
GSM227865	Smoker	RNA	Buccal mucosa
GSM227866	Never smoker	RNA	Buccal mucosa
GSM227867	Never smoker	RNA	Buccal mucosa

Identification of differentially expressed (DE) genes

The GEO2R software, which is available on the NCBI website, was used to create a list of 15,000 differentially expressed genes between smoker and never-smoker buccal samples.

The 15,000 genes were inputted into a Microsoft Excel spreadsheet and sorted by significance (Additional file 1: Table S1). After applying strict cut-off criteria ($p<0.05$ and absolute fold change between -0.5 and 1.5), the list of DE genes was narrowed down to 459 genes.

The Bioconductor package Enhanced Volcano was used to visualize the 459 DE genes in the form of a labelled volcano plot [16].

Ingenuity pathway analysis (IPA)

The list of DE genes was inputted into IPA software (QIAGEN, Hilden, Germany), where the 'core analysis' function of the software was used to interpret the data in terms of canonical pathways and upstream regulators.

Pathway and functional enrichment analysis

The Bioconductor package clusterProfiler was used to carry out an over-representation analysis of the DE genes [17, 18]. Similarly, the SIGnaling Network Open Resource 2.0 (SIGNOR 2.0) was used to explore the signaling networks that exist between the DE genes [19].

Results

Differentially expressed (DE) genes

Figure 1 displays a volcano plot of the full list of DE genes. However, only 459 genes exhibited significant differential expression, with 261 genes found to be over-expressed and 198 found to be under-expressed.

Figure 2 illustrates the chromosomal location, molecular class, and cellular location of the 459 DE genes. Chromosome 1 had the highest number of significantly DE genes ($n=63$), followed by chromosome 6 ($n=30$), chromosome 2 ($n=29$), and chromosome 19 ($n=27$). Similarly, the most represented molecular classes among the significantly DE genes were enzymes (19.6%) and transcription regulators (12%). Lastly, the majority of the significantly DE genes were located either in the cytoplasm (40.5%) or the nucleus (25.7%).

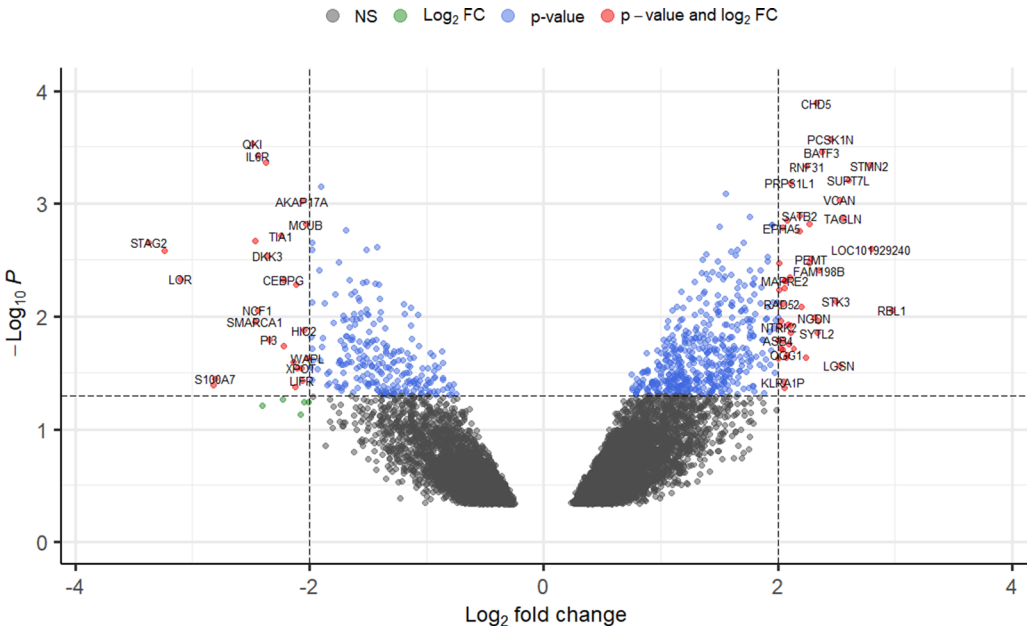

Fig. 1 Volcano plot of the most significantly differentially expressed genes in smoker buccal mucosa. Gray points are non-significant, green points have a significant log2FC, blue points have a significant *p* value, and red points have both a significant log2FC and *p* value

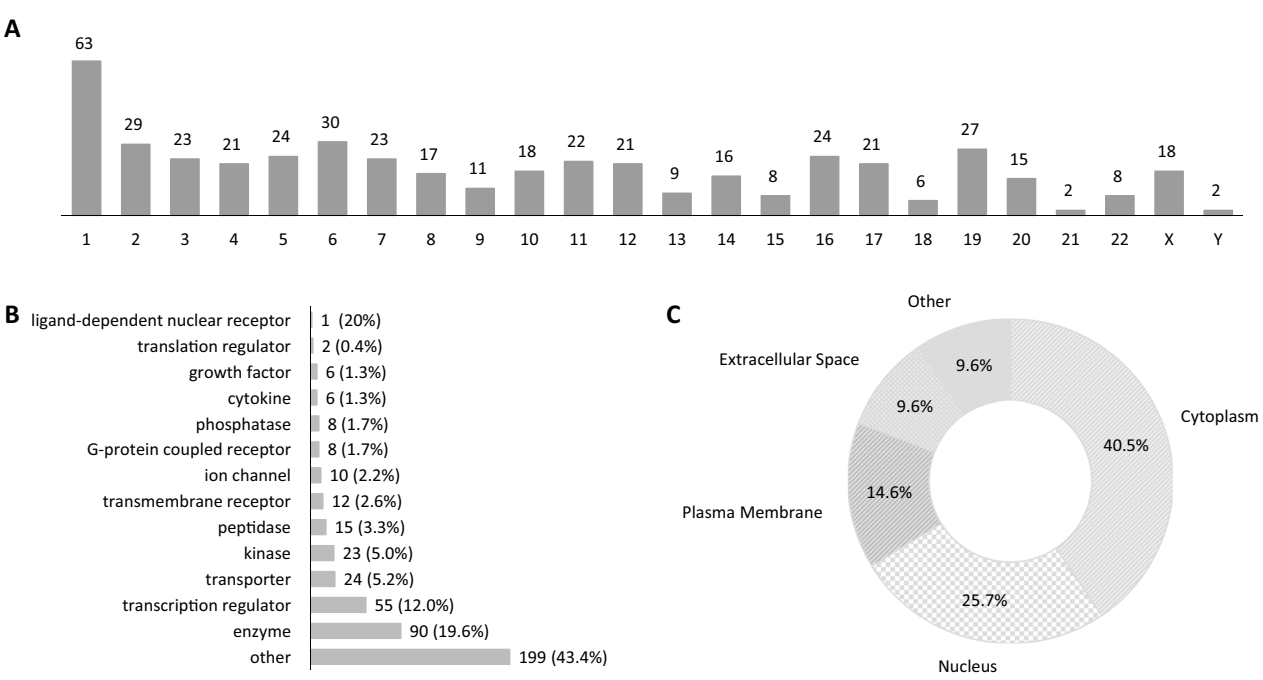

Fig. 2 Distribution of most significantly differentially expressed genes in terms of **A** chromosomal location, **B** molecular class, and **C** cellular location

Table 2 Significantly differentially expressed genes in smokers as revealed by IPA and as sorted by *p* value

NCBI ID	Gene symbol	Gene name	Chr	Molecule type	Expression	Log_2 ratio	*p* value
Top differentially expressed protein-coding genes							
26038	*CHD5*	Chromodomain helicase DNA binding protein 5	1	Enzyme	Up	2.335	1.28×10^{-04}
27344	*PCSK1N*	Proprotein convertase subtilisin/kexin type 1 inhibitor	X	Other	Up	2.45	2.72×10^{-04}
9444	*QKI*	QKI, KH domain containing RNA binding	6	Other	Down	− 2.484	2.97×10^{-04}
55509	*BATF3*	Basic leucine zipper ATF-like transcription factor 3	1	Transcription regulator	Up	2.379	3.54×10^{-04}
3570	*IL6R*	Interleukin 6 receptor	1	Transmembrane receptor	Down	− 2.444	3.80×10^{-04}
51302	*CYP39A1*	Cytochrome P450 family 39 subfamily A member 1	6	Enzyme	Down	− 2.371	4.35×10^{-04}
11075	*STMN2*	stathmin 2	8	Other	Up	2.786	4.65×10^{-04}
55072	*RNF31*	Ring finger protein 31	14	Enzyme	Up	2.249	4.74×10^{-04}
9913	*SUPT7L*	SPT7 like, STAGA complex subunit gamma	2	Transcription regulator	Up	2.605	6.26×10^{-04}
221823	*PRPS1L1*	Phosphoribosyl pyrophosphate synthetase 1 like 1	7	Kinase	Up	2.102	6.64×10^{-04}
Top differentially expressed lncRNA genes							
100505933	*ADD3-AS1*	ADD3 antisense RNA 1	10	Other	Up	2.075	1.42×10^{-03}
6315	*ATXN8OS*	ATXN8 opposite strand lncRNA	13	Other	Up	1.952	7.57×10^{-03}
100506070	*RBFADN*	RBFA downstream neighbor	18	Other	Up	1.683	9.10×10^{-03}
100131532	*LOC100131532*	Uncharacterized LOC100131532	6	Other	Up	2.136	1.93×10^{-02}
25859	*PART1*	Prostate androgen-regulated transcript 1	5	Other	Down	− 0.942	3.01×10^{-02}
100130449	*LOC100130449*	Uncharacterized LOC100130449	2	Other	Down	− 1.603	3.52×10^{-02}
Top differentially expressed pseudogenes							
388714	*FMO6P*	Flavin containing dimethylaniline monoxygenase 6, pseudogene	1	Other	Up	1.778	5.11×10^{-03}
442240	*ZNF259P1*	Zinc finger protein 259 pseudogene 1	6	Other	Up	2.054	5.65×10^{-03}
228	*ALDOAP2*	ALDOA pseudogene 2	10	Other	Down	− 1.573	1.05×10^{-02}
79986	*ZNF702P*	Zinc finger protein 702, pseudogene	19	Other	Up	1.653	1.32×10^{-02}
645,682	*POU5F1P4*	POU class 5 homeobox 1 pseudogene 4	1	Other	Up	1.521	1.46×10^{-02}
5408	*PNLIPRP2*	Pancreatic lipase related protein 2 (gene/pseudogene)	10	Enzyme	Down	− 1.437	1.61×10^{-02}
440915	*POTEKP*	POTE ankyrin domain family member K, pseudogene	2	Other	Up	1.62	1.93×10^{-02}
730092	*RRN3P1*	RRN3 pseudogene 1	16	Other	Up	1.548	1.96×10^{-02}
150244	*ZDHHC8P1*	ZDHHC8 pseudogene 1	22	Other	Up	1.559	2.27×10^{-02}
202181	*LOC202181*	SUMO interacting motifs containing 1 pseudogene	5	Other	Up	1.932	2.67×10^{-02}

Chr, chromosome; log_2 ratio, log_2 fold change ratio of the gene expression between 2 groups

Table 2 lists the most significantly DE genes between smoker and never smoker buccal mucosa samples, showing that protein-coding genes occupy the top ranks in terms of significance.

Interaction network of differentially expressed (DE) genes

Figure 3A demonstrates the interplay between the DE oncological pathways, cytokines, and genes in smoker buccal mucosa, namely the *IL2*, *EGFR*, and *ESR2* genes. Other than TIMP3, all the proteins in the pathway were predicted to be inhibited in smoker buccal mucosa.

Figure 3B illustrates the results of an interaction network analysis of the DE genes in smoker buccal mucosa. Interestingly, the *RPA1* gene was shown to have the highest number of interactions with the other DE genes in smoker buccal mucosa, but it did not have a significant level of differential expression ($p > 0.05$).

Upstream regulators

The top 20 regulators predicted by IPA included the TP53, APP, SMAD3, and TNF proteins as well

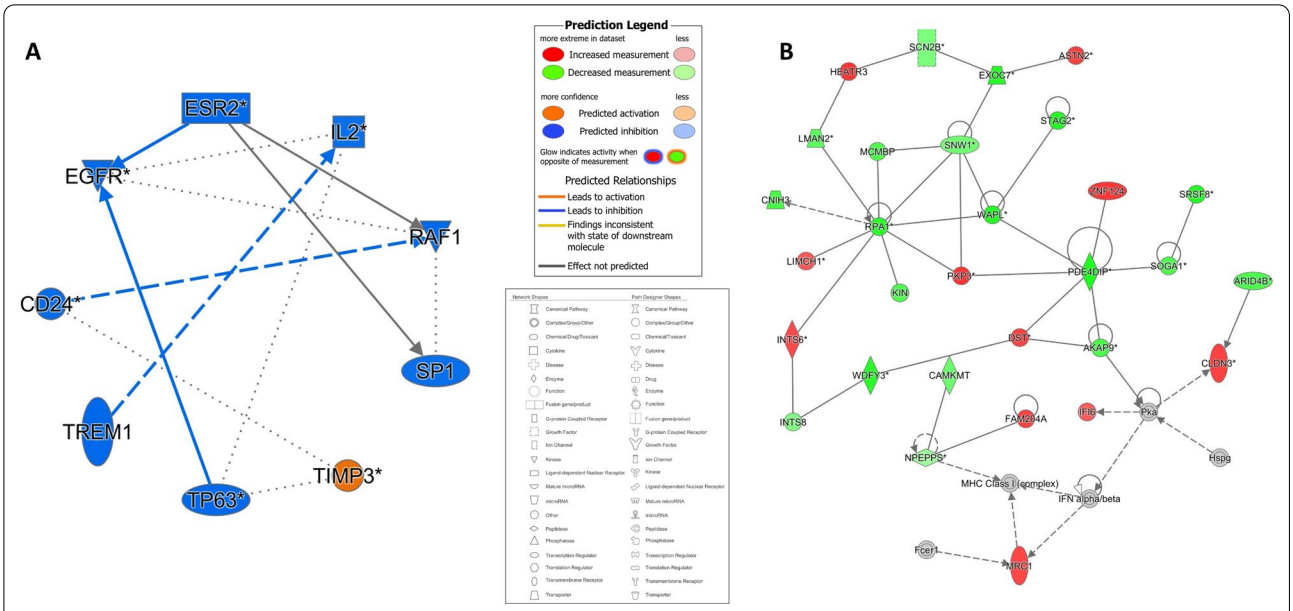

Fig. 3 **A** Graphical summary and **B** interaction network analysis of differentially expressed genes in smoker buccal mucosa. Different shapes represent the molecular class of the protein. Red and green indicate upregulation and downregulation, respectively, while blue and orange indicate inhibition and activation, respectively. A solid line indicates a direct interaction, a dashed line indicates an indirect interaction, and a dotted line indicates inferred correlation from machine-based learning. An asterisk indicates that multiple identifiers in the dataset file map to a single gene or chemical in the Global Molecular Network

Table 3 Top 20 upstream regulators revealed by Ingenuity Pathway Analysis

Upstream regulator	Molecule type	Log_2 ratio	*p* value	Z score
TP53	Transcription regulator	0.477	1.39×10^{-09}	1.028
Dexamethasone	Chemical drug	–	1.10×10^{-06}	0.438
APP	Other	1.709	1.39×10^{-06}	0.01
SMAD3	Transcription regulator	0.687	3.20×10^{-06}	− 1.508
TNF	Cytokine	0.098	5.90×10^{-06}	− 1.169
Decitabine	Chemical drug	–	6.61×10^{-06}	1.433
Tetradecanoylphorbol	Chemical drug	–	8.14×10^{-06}	− 1.916
L-type calcium channel	Complex	–	1.01×10^{-05}	− 1.98
Camptothecin	Chemical drug	–	1.53×10^{-05}	− 0.64
KDM5B	Transcription regulator	0.95	1.79×10^{-05}	1.04
OSM	Cytokine	− 0.844	2.74×10^{-05}	1.01
TGFB3	Growth factor	0.488	2.80×10^{-05}	− 0.877
Nitrofurantoin	Chemical drug	–	2.90×10^{-05}	− 1.437
Forskolin	Chemical toxicant	–	4.11×10^{-05}	− 2.25
TP63	Transcription regulator	1.346	4.15×10^{-05}	− 2.611
CSF3	Cytokine	0.301	6.00×10^{-05}	− 0.494
PPARA	Ligand-dependent nuclear receptor	0.147	6.01×10^{-05}	− 0.789
Haloperidol	Chemical drug	–	6.35×10^{-05}	− 1.591
LY294002	Chemical-kinase inhibitor	–	7.53×10^{-05}	0.822
RASSF1	Other	0.33	8.26×10^{-05}	0.333

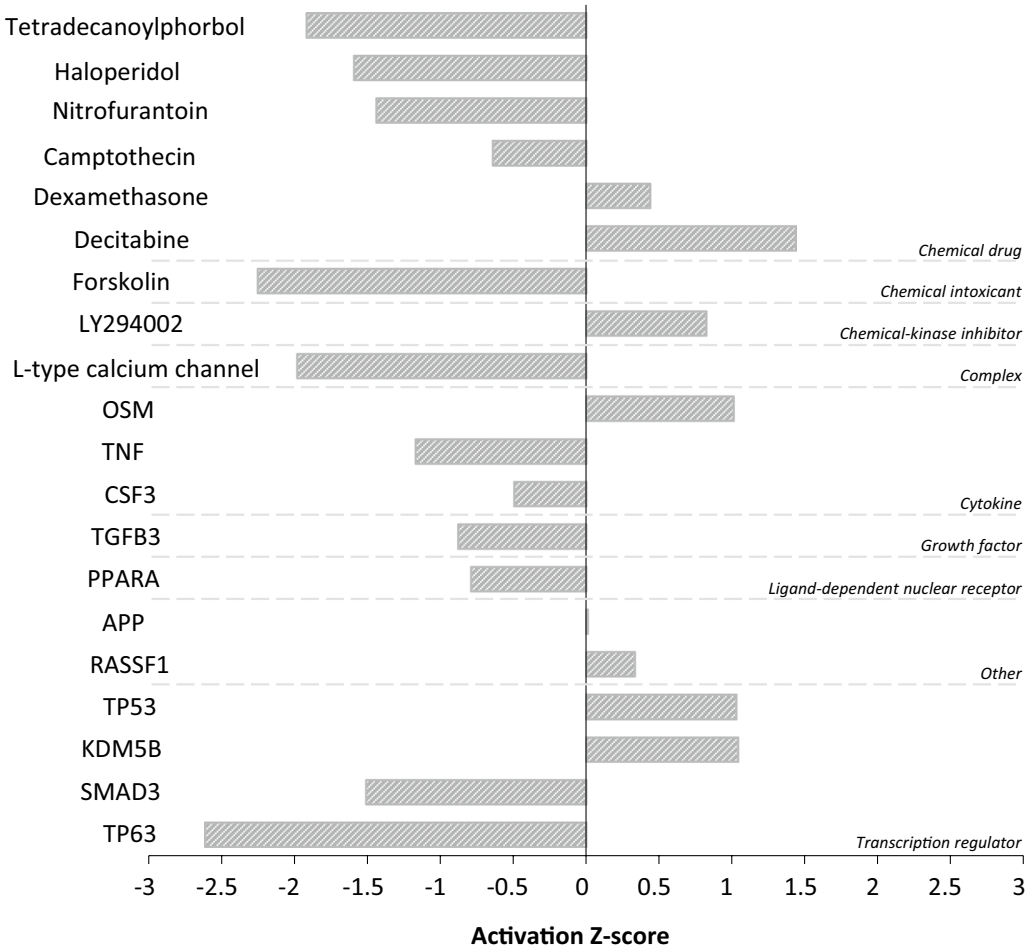

Fig. 4 Top upstream regulators predicted by Ingenuity Pathway Analysis in smoker buccal mucosa. A positive z score indicates activation, while a negative z score indicates inhibition

as the drug dexamethasone, among other molecules (Table 3). Figure 4 illustrates the data in Table 3 and emphasizes the predicted activation status of the top upstream regulators as revealed by IPA. As can be seen from Fig. 4, the most inhibited upstream regulator in smoker buccal mucosa is predicted to be the TP63 protein.

Dexamethasone was predicted to be a top upstream regulator and affected a total of 78 genes via indirect interactions (Fig. 5A). Likewise, microRNA-8 (miR-8) was found by IPA to be among the top upstream regulators to be activated, as miR-8 targeted 7 of the DE genes between smokers and never smokers (Fig. 5B). Of those genes, 5 (*CCND2*, *ITGAV*, *QKI*, *RPS6KB1*, and *SMAD2*) were under-expressed and 2 (*BMP2* and *CLDN3*) were over-expressed.

Further analysis of the top upstream regulator proteins resulted in the construction of gene–gene (Fig. 6) and protein–protein (Fig. 7) interaction networks. Figure 6 shows that the 36.04% of the top upstream regulator proteins were predicted to have interactions with one another, 26.19% have shared protein domains, and 22.85% were co-expressed. Similarly, Fig. 7 shows that the TP53 and TNF proteins had the highest number of interactions with the other top upstream regulator proteins.

Enriched biological pathways

The most significant canonical pathway was identified as the nitric oxide and reactive oxygen production at macrophages (Table 4).

Correlation of smoker buccal mucosa with other diseases

The DE genes in smoker buccal mucosa are significantly associated with cancer and organismal injury, among other diseases (Table 5).

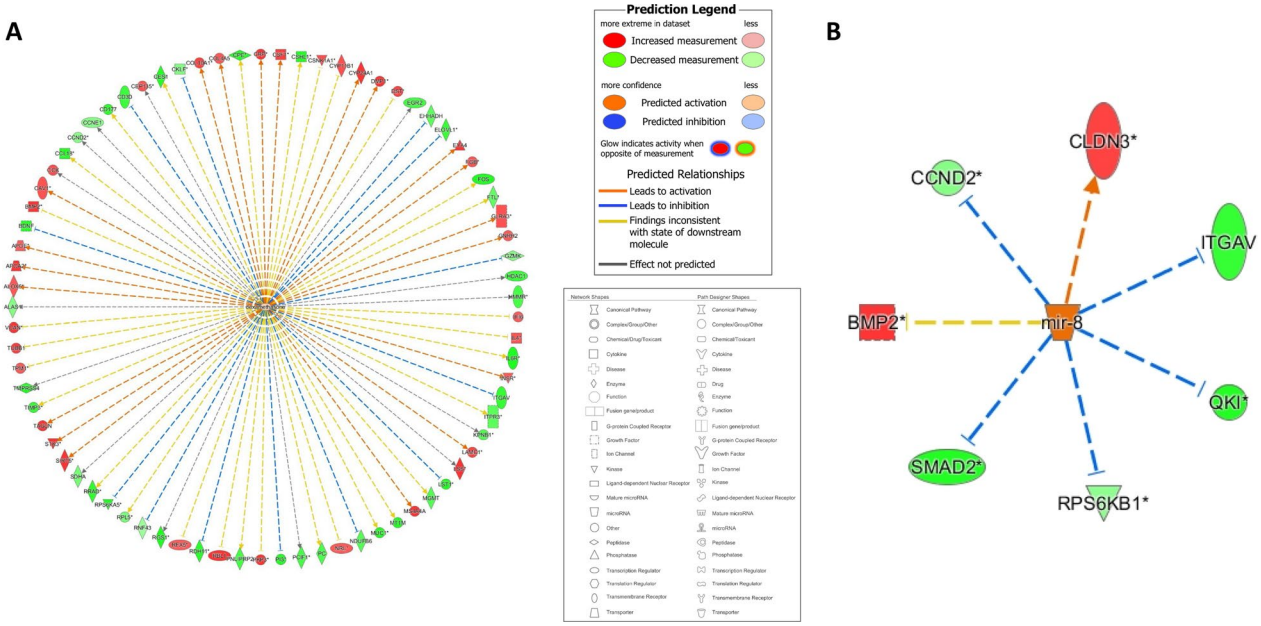

Fig. 5 Upstream regulators dexamethasone and miR-8. **A** Chemical drug dexamethasone is predicted to be activated in smoker buccal mucosa with P value = 1.10×10^{-06} and Z score = 0.438. **B** miR-8 is predicted to be activated in smoker buccal mucosa with *P* value = 8.90×10^{-02} and Z score = 2.1. Different shapes represent the molecular class of the protein. Red and green indicate upregulation and downregulation, respectively, while blue and orange indicate inhibition and activation, respectively. A solid line indicates a direct interaction, a dashed line indicates an indirect interaction, and a dotted line indicates inferred correlation from machine-based learning. An asterisk indicates that multiple identifiers in the dataset file map to a single gene or chemical in the Global Molecular Network

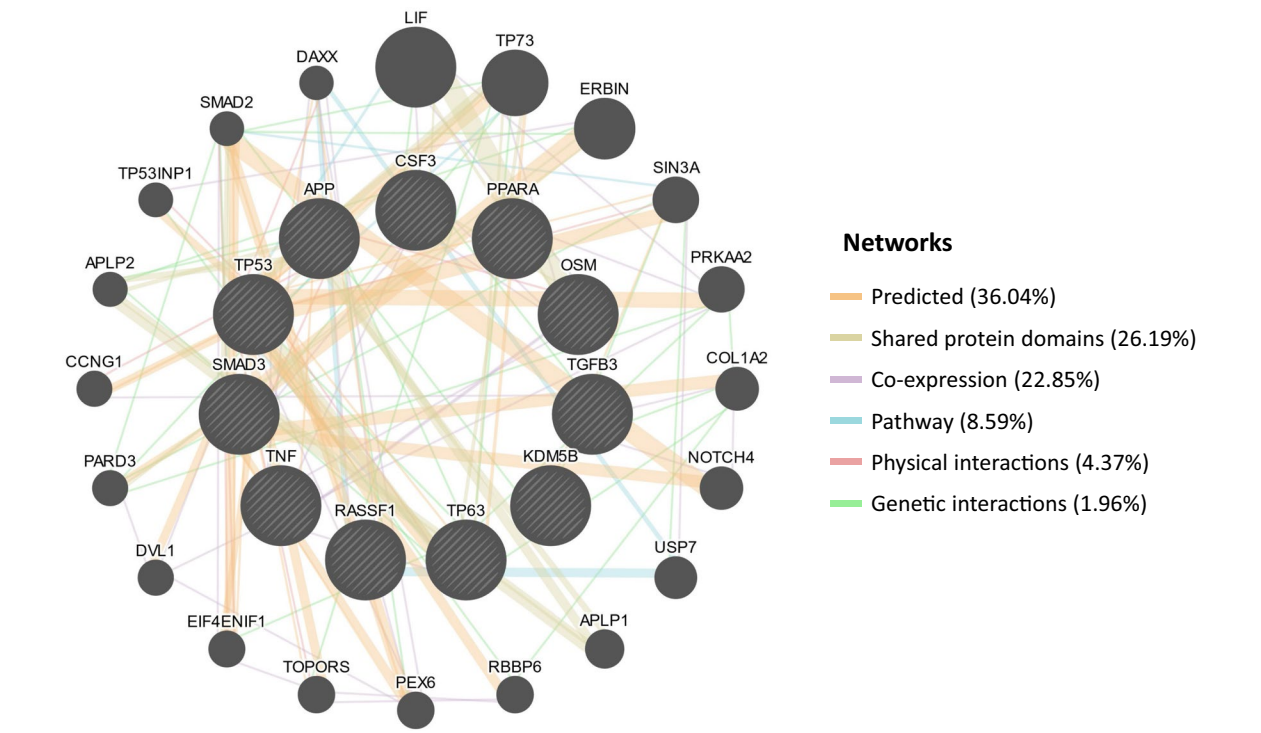

Fig. 6 Construction of a gene–gene interaction network of the upstream regulators with the most significant differential expression. Black circles with white stripes indicate genes that were entered as query terms, while black circles indicate the associated genes. The size of the circle corresponds with the number of correlations with other genes in the network

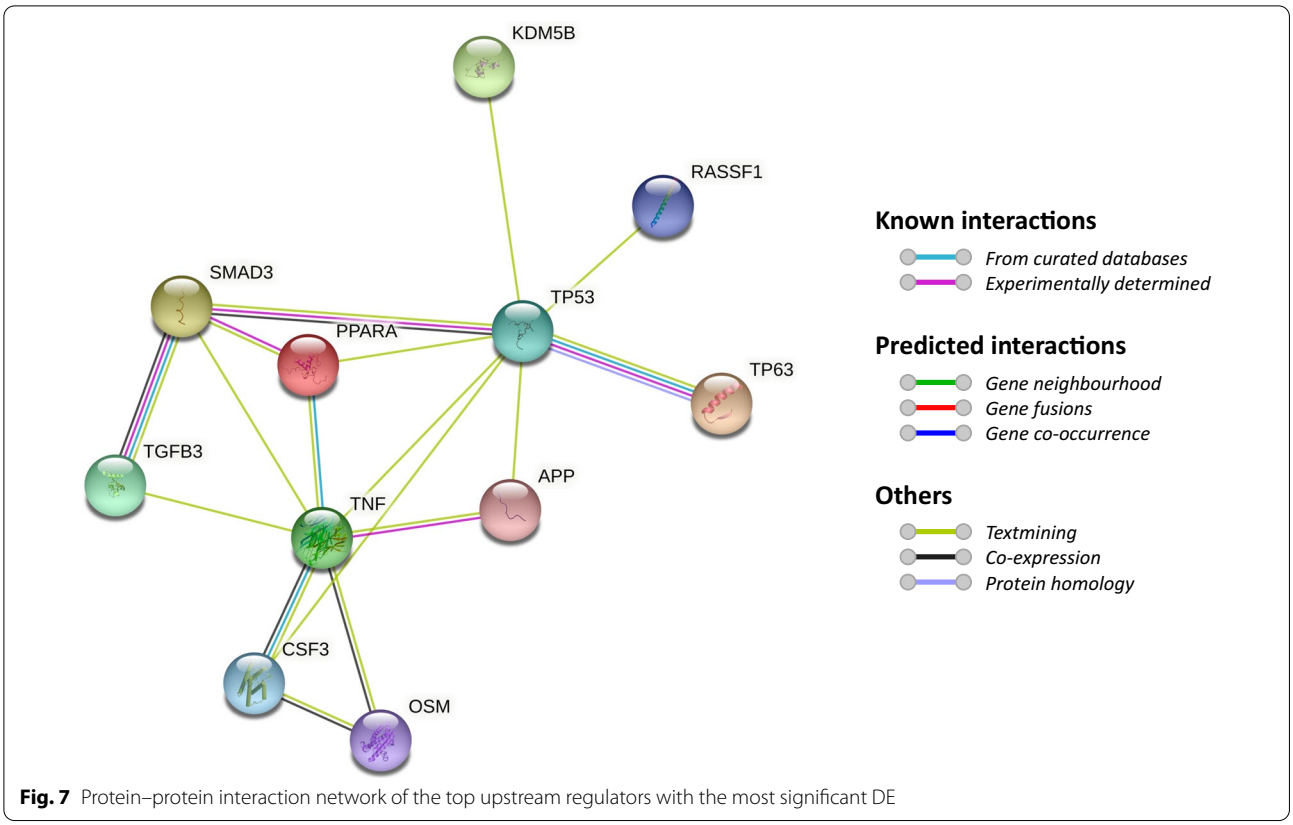

Fig. 7 Protein–protein interaction network of the top upstream regulators with the most significant DE

Table 4 Top five canonical pathways revealed by ingenuity pathway analysis

Pathway	*p* value	−log(*p* value)	Z score
Nitric oxide and reactive oxygen production at macrophages	0.00124	2.9	−1.26
Retinol biosynthesis pathways	0.00159	2.7	−1.3
Thyroid tumor signaling pathways	0.00441	2.3	−0.81
Insulin signaling pathways	0.00860	2.0	−2.11
Macrophages/fibroblasts and endothelial cells in rheumatoid arthritis	0.00895	2.0	N/A

Table 5 Top five diseases or disorders revealed by Ingenuity Pathway Analysis

Disease	Number of molecules	*p* value
Organismal injury/abnormalities	436	5.00×10^{-04} to 6.75×10^{-23}
Cancer	432	5.00×10^{-04} to 6.75×10^{-23}
Gastrointestinal disease	393	4.66×10^{-04} to 1.01×10^{-14}
Endocrine system disorders	385	4.99×10^{-04} to 1.25×10^{-18}
Reproductive system disease	311	4.99×10^{-04} to 7.24×10^{-13}

Pathway and functional enrichment analysis

Figure 8 illustrates the most over-represented biological processes in smoker buccal mucosa. Interestingly, craniosynostosis and fibroid tumors were revealed to be the topmost significantly over-represented biological processes.

Figure 9 shows the results of signaling network analysis of the 459 significantly DE genes, with the *SMAD2* gene having the most interactions. *SMAD2* is directly downregulated by the *CTDSPL* and *SKIL* genes and indirectly upregulated by the *BMP2* gene.

Discussion

The most significantly differentially expressed (DE) protein-coding genes in smoker buccal mucosa were the *CHD5*, *QKI*, *BATF3,* and *IL6R* genes, which have previously reported associations with smoking and related diseases.

The *CHD5* gene, which is a tumor suppressor gene that is preferentially expressed in the nervous system and testis, was significantly upregulated in smoker buccal mucosa [20, 21]. *CHD5* is believed to serve as a master regulator in tumor-suppressive networks, and *CHD5* expression levels are strongly associated with the

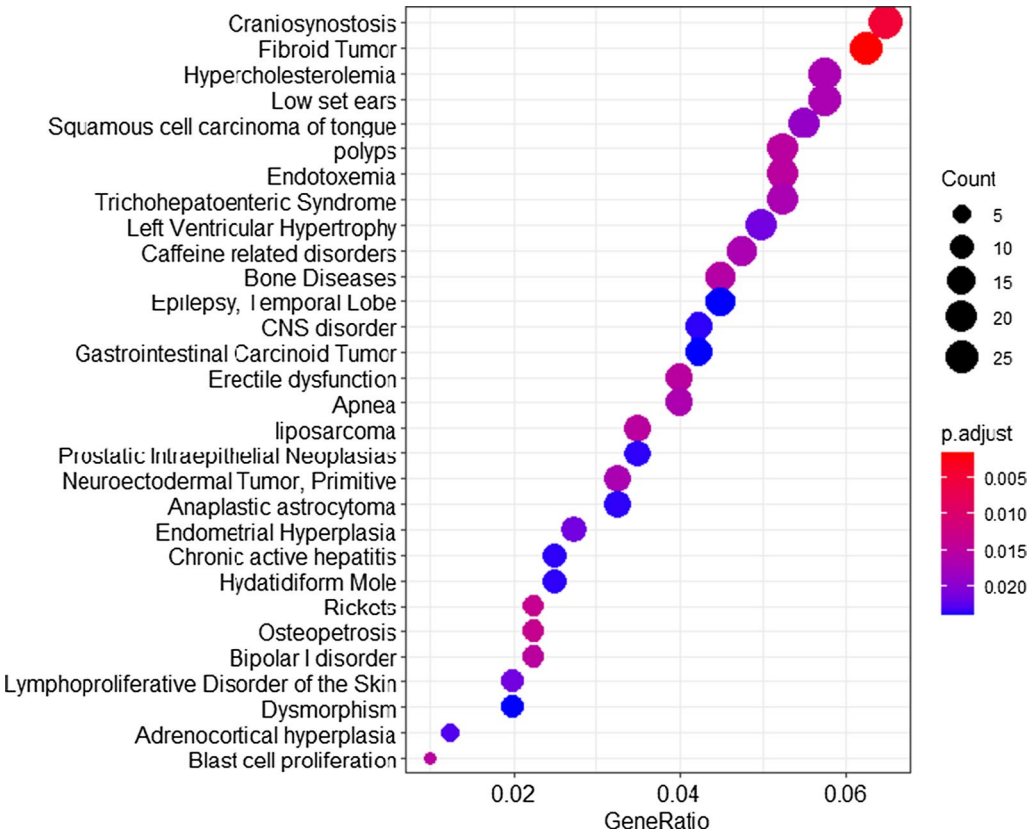

Fig. 8 Over-representation analysis of the most significantly differentially expressed genes ($p < 0.05$). The size of the circle corresponds with the number of genes that are mapped to a biological process, while a redder color indicates increased significance

prognosis of several cancers, including hepatocellular carcinoma and non-small cell lung cancer [20, 22–24]. One study found that a rare *CHD5* variant, rs12564469-rs9434711, contributed to the risk of hepatocellular carcinoma, a risk effect which was statistically significant in alcohol drinkers but not smokers [25].

The *QKI* gene contributes to a number of human diseases, including cancers, myelin disorders, and schizophrenia, and it is a critical regulator of alternative splicing in cardiac myofibrillogenesis and contractile function [26]. *QKI* has also been identified as a master regulator of alternative splicing in human lung cancer cell lines, but no significant statistical association was found between *QKI* expression and smoking status in lung tumors [27, 28]. Moreover, *QKI* was identified as a significantly altered gene in the ciliated epithelial cells of lungs affected by chronic obstructive pulmonary disease (COPD), a disease that is primarily caused by tobacco smoking [29].

The *BATF3* gene belongs to the AP-1 transcription factor family, whose members respond to a range of pathological and physiological stimuli by mediating gene expression [30]. BATF3 controls the differentiation of dendritic cells, inhibits the differentiation of regulatory T cells, and critically regulates the development of memory T cells [31, 32]. *BATF3* expression in the lungs was necessary in order to induce protection against allergic airway inflammation through tolerization with *Helicobacter pylori* extract [33]. Moreover, the acute inhalation of electronic cigarette smoke by healthy never smokers led to the significant upregulation of *BATF3*, among other genes that play a role in promoting tumorigenesis [34].

The *IL6R* gene is a pleiotropic regulator of both acquired and innate immune responses, and it is believed to be expressed in the lungs [35]. There have been conflicting findings regarding the benefits of anti-IL-6R therapy for COVID-19-induced acute respiratory distress syndrome [36, 37]. In the context of smoking, exposure to cigarette smoke led to increased *IL6R* mRNA levels in primary bronchial epithelial cell lines [38]. Moreover, a certain *IL6R* haplotype (rs6684439-rs7549250-rs4129267-rs10752641-rs407239) has been associated with a lower COPD risk in a Mexican Mestizo population, while the IL6R variant Asp358Ala did not show any association with COPD [39, 40].

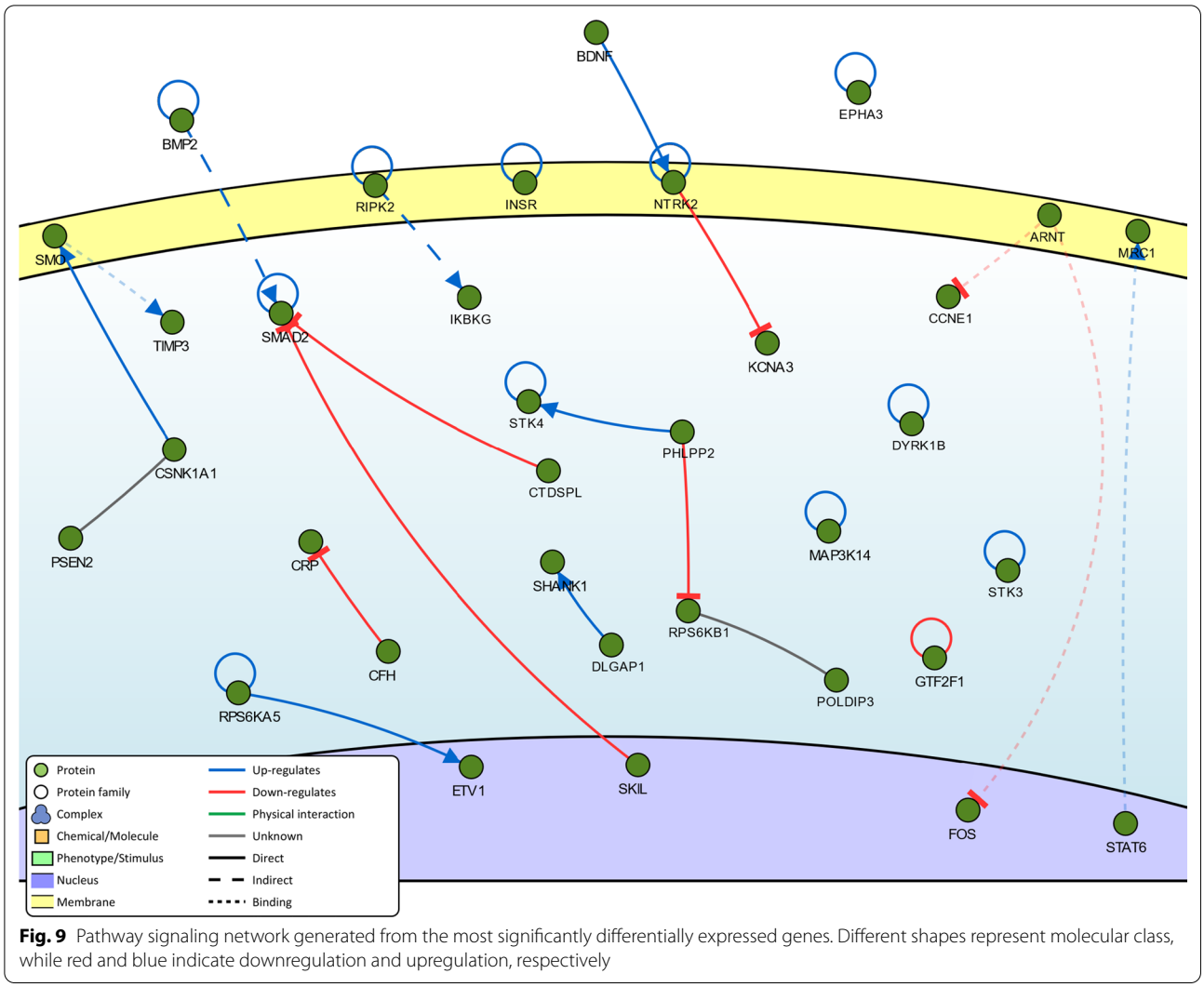

Fig. 9 Pathway signaling network generated from the most significantly differentially expressed genes. Different shapes represent molecular class, while red and blue indicate downregulation and upregulation, respectively

Pseudogene expression was also altered in smoker buccal mucosa, most notably in the upregulation of *FMO6P*, *ZNF259P1*, and *ZNF702P* and the downregulation of *ALDOAP2* and *PNLIPRP2*. *FMO6P* has significant sequence homology with the *FMO3* gene, the latter of which functions to metabolize a small amount of nicotine [41]. A single nucleotide variation in the *FMO6P* pseudogene, rs6608453, was associated with nicotine dependence in African Americans [42]. Likewise, *ALDOAP2* was over-expressed in both healthy and non-healthy smokers compared to non-smokers, while exposure to cigarette smoke resulted in the upregulation of the *PNLIPRP2* polymorphic pseudogene in a murine model [43, 44]. In contrast, *ZNF259P1* and *ZNF702P* did not have previously reported associations with smoking. *ZNF259P1* was significantly correlated with the tumor size of primary lung adenocarcinomas, while *ZNF702P* was found to be upregulated after *BCL2L10* knockdown in two ovarian cell lines [45, 46].

Analysis of upstream regulators revealed that the tumor protein 53 (*TP53*) gene was the most significantly DE regulator in smoker buccal mucosa. *TP53* contains cellular proliferation by guarding against genomic mutation, and *TP53* mutations are among the most common genetic alterations in human cancers [47]. Tobacco smoking is known to influence *TP53* mutation patterns and frequencies in lung cancer and urothelial cell carcinoma patients [48, 49]. In fact, a large proportion of *TP53* mutations in the lung cancers of smokers were G → T transversions, a primary mutagenic signature that is caused by DNA damage from tobacco smoke [50].

The most significant canonical pathway identified by IPA was the "nitric oxide and reactive oxygen production at macrophages". Nitric oxide and reactive oxygen

species are essential for maintaining redox balance, but they also act in pathological processes [51]. Tobacco smoke contains large numbers of free radicals, including nitric oxide and reactive oxygen species (ROS), that cause oxidative stress on the cellular and sub-cellular levels [52, 53]. In turn, smoking-induced oxidative stress activates inflammatory response pathways that produce endogenous ROS at the site of oxidative stress, potentially causing further oxidative damage to that site [53]. Smoking also reduces the production of nitric oxide while also elevating the production of ROS in endothelial cells [54, 55]. Smoking-induced ROS production is especially concerning as it may contribute to the progression of endometrial adenocarcinoma [56].

Among the DE genes, those associated with craniosynostosis and fibroid tumors were over-represented in smoker buccal mucosa.

Craniosynostosis, which is caused by the premature fusion of cranial sutures, is the second-most common cranio-facial anomaly [57]. Smoking during pregnancy was associated with an increased risk of craniosynostosis, while exposure to secondhand smoke modestly increased the risk of this birth defect [58]. Maternal smoking impacts cranio-facial development by acting upon variant alleles of the transforming growth factor alpha (TGF-α) gene, and genetic variation of the TGF-α gene is associated with increased risk of cranio-facial defects [59, 60].

Fibroid tumors are non-cancerous growths that develop inside or on the uterus and are the most common type of pelvic tumor detected in women [61]. Previous studies that investigated the impact of smoking on fibroid tumors yielded conflicting results. Earlier studies suggested that smoking had a protective effect against fibroid tumors, but subsequent studies have shown either a negative effect or no relationship at all [61, 62]. It is worthwhile to note that smoking has been shown to have an anti-estrogenic effect in women, resulting in an earlier natural menopause as well as protective associations with the risk of estrogen-related cancers [63, 64].

Pathway network analysis revealed that the *SMAD2* gene had the highest number of interactions with other DE genes, and it was also a target of miR-8. *SMAD3* was predicted by IPA to be an inhibited upstream regulator. The SMAD Family Member 2 (*SMAD2*) gene encodes for a protein that is vital for early development, and *SMAD2* mutations were associated with complex cranio-facial defects in a murine model [65]. SMAD2, SMAD3, and SMAD4 mediate the signal transduction of transforming growth factor-β (TGF-β) superfamily members, the latter of which induce a range of effects that involve cellular differentiation, proliferation, migration, and apoptosis [66].

The present study is affected by a few limitations. The sample size was relatively small, and the patient samples differed in terms of sex and race, which could confound the interpretation of the genetic variation. Additionally, several differentially expressed genes in smoker buccal mucosa were uncharacterized or unmapped to pathways, meaning that their effects are not considered in the current analysis.

Conclusion

The current findings signify the importance of inflammatory response and oxidative stress as a major component of smoking-induced tissue injury. Most significantly, nitric oxide-related inflammation stands as one of the canonical pathways underlying genetic and molecular pathways changes coupled with exposure to cigarette smoke. Future lines of research should focus on validating the results of the current study in a larger population to ascertain potential therapeutic targets in the context of smoking-induced damage.

Abbreviations

Chr: Chromosome; COPD: Chronic obstructive pulmonary disease; COVID-19: Coronavirus disease 2019; DE: Differentially expressed; DNA: Deoxyribonucleic acid; GEO: Gene Expression Omnibus; IPA: Ingenuity Pathway Analysis; NCBI: National Center for Biotechnology Information; RNA: Ribonucleic acid; ROS: Reactive oxygen species; SIGNOR 2.0: SIGnaling Network Open Resource 2.0.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43042-022-00268-y.

Additional file 1. List of DE genes downloaded from GEO2R. A complete list of gene names and their abbreviations are delivered in the supplementary data files.

Acknowledgements

The authors would like to acknowledge the efforts of the University of Petra's Faculty of Pharmacy and Medical Sciences, its dean, and its Department Head of Pharmacology and Biomedical Sciences.

Authors' contributions

AK was involved in conceptualization, writing (review & editing), formal analysis, validation, and visualization. BK and LS were involved in formal analysis, methodology, and writing (original draft). AT and TQ were involved in visualization and writing (review & editing). All authors have read and approved the manuscript.

Funding

This study was not funded.

Availability of data and materials

The current report utilized a previously published dataset for the analysis. The dataset used in this work was acquired from The National Center for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO) depository (accession number GSE8987).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Pharmacology and Biomedical Sciences, Faculty of Pharmacy and Medical Sciences, University of Petra, Amman, Jordan. [2]Department of Pharmaceutical Medicinal Chemistry and Pharmacognosy, Faculty of Pharmacy and Medical Sciences, University of Petra, Amman, Jordan. [3]Department of Medical Laboratory Sciences, Pharmacological and Diagnostic Research Centre, Faculty of Allied Medical Sciences, Al-Ahliyya Amman University, Amman, Jordan. [4]Independent Researcher, Doha, Qatar.

Received: 10 September 2021 Accepted: 16 February 2022
Published online: 17 March 2022

References

1. Glantz S, Gonzalez M (2012) Effective tobacco control is key to rapid progress in reduction of non-communicable diseases. The Lancet 379:1269–1271. https://doi.org/10.1016/S0140-6736(11)60615-6
2. Beaglehole R, Bonita R, Horton R, Adams C, Alleyne G, Asaria P et al (2011) Priority actions for the non-communicable disease crisis. Lancet 377:1438–1447. https://doi.org/10.1016/S0140-6736(11)60393-0
3. Reitsma MB, Kendrick PJ, Ababneh E, Abbafati C, Abbasi-Kangevari M, Abdoli A et al (2021) Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. The Lancet 397:2337–2360. https://doi.org/10.1016/S0140-6736(21)01169-7
4. Bade BC, Dela Cruz CS (2020) Lung cancer 2020: epidemiology, etiology, and prevention. Clin Chest Med 41:1–24. https://doi.org/10.1016/j.ccm.2019.10.001
5. Talhout R, Schulz T, Florek E, Van Benthem J, Wester P, Opperhuizen A (2011) Hazardous compounds in tobacco smoke. Int J Environ Res Public Health 8:613–628. https://doi.org/10.3390/ijerph8020613
6. Gümüş ZH, Du B, Kacker A, Boyle JO, Bocker JM, Mukherjee P et al (2008) Effects of tobacco smoke on gene expression and cellular pathways in a cellular model of oral leukoplakia. Cancer Prev Res (Phila) 1:100–111. https://doi.org/10.1158/1940-6207.CAPR-08-0007
7. Hecht SS (2006) Cigarette smoking: cancer risks, carcinogens, and mechanisms. Langenbecks Arch Surg 391:603–613. https://doi.org/10.1007/s00423-006-0111-z
8. Peppone LJ, Mustian KM, Morrow GR, Dozier AM, Ossip DJ, Janelsins MC et al (2011) The effect of cigarette smoking on cancer treatment-related side effects. Oncologist 16:1784–1792. https://doi.org/10.1634/theoncologist.2011-0169
9. Kopa PN, Pawliczak R (2018) Effect of smoking on gene expression profile – overall mechanism, impact on respiratory system function, and reference to electronic cigarettes. Toxicol Mech Methods 28:397–409. https://doi.org/10.1080/15376516.2018.1461289
10. Na H-K, Kim M, Chang S-S, Kim S-Y, Park JY, Chung MW et al (2015) Tobacco smoking-response genes in blood and buccal cells. Toxicol Lett 232:429–437. https://doi.org/10.1016/j.toxlet.2014.10.005
11. Tsai P-C, Glastonbury CA, Eliot MN, Bollepalli S, Yet I, Castillo-Fernandez JE et al (2018) Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health. Clin Epigenetics 10:126. https://doi.org/10.1186/s13148-018-0558-0
12. Cai G, Bossé Y, Xiao F, Kheradmand F, Amos CI (2020) Tobacco smoking increases the lung gene expression of ACE2, the receptor of SARS-CoV-2. Am J Respir Crit Care Med 201:1557–1559. https://doi.org/10.1164/rccm.202003-0693LE
13. Hijazi K, Malyszko B, Steiling K, Xiao X, Liu G, Alekseyev YO et al (2019) Tobacco-related alterations in airway gene expression are rapidly reversed within weeks following smoking-cessation. Sci Rep 9:6978. https://doi.org/10.1038/s41598-019-43295-3
14. Huan T, Joehanes R, Schurmann C, Schramm K, Pilling LC, Peters MJ et al (2016) A whole-blood transcriptome meta-analysis identifies gene expression signatures of cigarette smoking. Hum Mol Genet 25:4611–4623. https://doi.org/10.1093/hmg/ddw288
15. Sridhar S, Schembri F, Zeskind J, Shah V, Gustafson AM, Steiling K et al (2008) Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. BMC Genomics 9:259. https://doi.org/10.1186/1471-2164-9-259
16. Blighe K, Rana S, Lewis M (2018) EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling. https://github.com/kevinblighe/EnhancedVolcano
17. Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS J Integr Biol 16:284–287. https://doi.org/10.1089/omi.2011.0118
18. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z et al (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation. https://doi.org/10.1016/j.xinn.2021.100141
19. Licata L, Lo Surdo P, Iannuccelli M, Palma A, Micarelli E, Perfetto L (2020) SIGNOR 2.0, the SIGnaling network open resource 2.0 2019 update. Nucleic Acids Res 48:D504–D510. https://doi.org/10.1093/nar/gkz949
20. Kolla V, Zhuang T, Higashi M, Naraparaju K, Brodeur GM (2014) Role of CHD5 in human cancers: 10 years later. Cancer Res 74:652–658. https://doi.org/10.1158/0008-5472.CAN-13-3056
21. Thompson PM, Gotoh T, Kok M, White PS, Brodeur GM (2003) CHD5, a new member of the chromodomain gene family, is preferentially expressed in the nervous system. Oncogene 22:1002–1011. https://doi.org/10.1038/sj.onc.1206211
22. Zhao R, Wang N, Huang H, Ma W, Yan Q (2014) CHD5 a tumour suppressor is epigenetically silenced in hepatocellular carcinoma. Liver Int 34:e151-160. https://doi.org/10.1111/liv.12503
23. Zhu X, Kong Q, Xie L, Chen Z, Li H, Zhu Z et al (2017) The single-nucleotide polymorphisms in CHD5 affect the prognosis of patients with hepatocellular carcinoma. Oncotarget 9:13222–13230
24. Baykara O, Tansarikaya M, Bulut P, Demirkaya A, Buyru N (2017) CHD5 is a potential tumor suppressor in non small cell lung cancer (NSCLC). Gene 618:65–68. https://doi.org/10.1016/j.gene.2017.04.010
25. Xiao Q, Chen L, Luo H, Li H, Kong Q, Jiao F et al (2018) A rare CHD5 haplotype and its interactions with environmental factors predicting hepatocellular carcinoma risk. BMC Cancer 18:658. https://doi.org/10.1186/s12885-018-4551-y
26. Chen X, Liu Y, Xu C, Ba L, Liu Z, Li X et al (2021) QKI is a critical pre-mRNA alternative splicing regulator of cardiac myofibrillogenesis and contractile function. Nat Commun 12:89. https://doi.org/10.1038/s41467-020-20327-5
27. Zong F-Y, Fu X, Wei W-J, Luo Y-G, Heiner M, Cao L-J et al (2014) The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. PLoS Genet 10:e1004289. https://doi.org/10.1371/journal.pgen.1004289
28. de Miguel FJ, Pajares MJ, Martínez-Terroba E, Ajona D, Morales X, Sharma RD et al (2016) A large-scale analysis of alternative splicing reveals a key role of QKI in lung cancer. Mol Oncol 10:1437–1449. https://doi.org/10.1016/j.molonc.2016.08.001
29. Li X, Noell G, Tabib T, Gregory AD, Trejo Bittar HE, Vats R et al (2021) Single cell RNA sequencing identifies IGFBP5 and QKI as ciliated epithelial cell genes associated with severe COPD. Respir Res 22:100. https://doi.org/10.1186/s12931-021-01675-2
30. Hess J, Angel P, Schorpp-Kistner M (2004) AP-1 subunits: quarrel and harmony among siblings. J Cell Sci 117:5965–5973. https://doi.org/10.1242/jcs.01589
31. Lee W, Kim HS, Hwang SS, Lee GR (2017) The transcription factor Batf3 inhibits the differentiation of regulatory T cells in the periphery. Exp Mol Med 49:e393–e393. https://doi.org/10.1038/emm.2017.157
32. Qiu Z, Khairallah C, Romanov G, Sheridan BS (2020) Cutting edge: Batf3 expression by CD8 T cells critically regulates the development of memory populations. J Immunol. https://doi.org/10.4049/jimmunol.2000228

33. Engler DB, Reuter S, van Wijck Y, Urban S, Kyburz A, Maxeiner J et al (2014) Effective treatment of allergic airway inflammation with Helicobacter pylori immunomodulators requires BATF3-dependent dendritic cells and IL-10. PNAS 111:11810–11815. https://doi.org/10.1073/pnas.1410579111
34. Staudt MR, Salit J, Kaner RJ, Hollmann C, Crystal RG (2018) Altered lung biology of healthy never smokers following acute inhalation of E-cigarettes. Respir Res 19:78. https://doi.org/10.1186/s12931-018-0778-z
35. Wilk JB, Walter RE, Laramie JM, Gottlieb DJ, O'Connor GT (2007) Framingham Heart Study genome-wide association: results for pulmonary function measures. BMC Med Genet 8:S8. https://doi.org/10.1186/1471-2350-8-S1-S8
36. Garbers C, Rose-John S (2021) Genetic IL-6R variants and therapeutic inhibition of IL-6 receptor signalling in COVID-19. Lancet Rheumatol 3:e96–e97. https://doi.org/10.1016/S2665-9913(20)30416-1
37. Buonaguro FM, Puzanov I, Ascierto PA (2020) Anti-IL6R role in treatment of COVID-19-related ARDS. J Transl Med 18:165. https://doi.org/10.1186/s12967-020-02333-9
38. Stolarczyk M, Amatngalim GD, Yu X, Veltman M, Hiemstra PS, Scholte BJ (2016) ADAM17 and EGFR regulate IL-6 receptor and amphiregulin mRNA expression and release in cigarette smoke-exposed primary bronchial epithelial cells from patients with chronic obstructive pulmonary disease (COPD). Physiol Rep 4:12878
39. Pérez-Rubio G, Silva-Zolezzi I, Fernández-López JC, Camarena Á, Velázquez-Uncal M, Morales-Mandujano F et al (2016) Genetic variants in IL6R and ADAM19 are associated with COPD severity in a Mexican mestizo population. COPD J Chronic Obstruct Pulm Dis 13:610–615. https://doi.org/10.3109/15412555.2016.1161017
40. Farahi N, Paige E, Balla J, Prudence E, Ferreira RC, Southwood M et al (2017) Neutrophil-mediated IL-6 receptor trans-signaling and the risk of chronic obstructive pulmonary disease and asthma. Hum Mol Genet 26:1584–1596. https://doi.org/10.1093/hmg/ddx053
41. Legaz I, Pérez-Cárceles MD, de la Calle I, Arjona F, Roca M, Cejudo P et al (2019) Genetic susceptibility to nicotine and/or alcohol addiction: a systematic review. Toxin Rev. https://doi.org/10.1080/15569543.2019.1619085
42. Zhang T-X, Saccone NL, Bierut LJ, Rice JP (2017) Targeted sequencing identifies genetic polymorphisms of flavin-containing monooxygenase genes contributing to susceptibility of nicotine dependence in European American and African American. Brain Behavior 7:e00651. https://doi.org/10.1002/brb3.651
43. Boué S, Tarasov K, Jänis M, Lebrun S, Hurme R, Schlage W et al (2012) Modulation of atherogenic lipidome by cigarette smoke in apolipoprotein E-deficient mice. Atherosclerosis 225:328–334. https://doi.org/10.1016/j.atherosclerosis.2012.09.032
44. Wang DC, Shi L, Zhu Z, Gao D, Zhang Y (2017) Genomic mechanisms of transformation from chronic obstructive pulmonary disease to lung cancer. Semin Cancer Biol 42:52–59. https://doi.org/10.1016/j.semcancer.2016.11.001
45. Lee S-Y, Kwon J, Lee K-A (2021) Bcl2l10 induces metabolic alterations in ovarian cancer cells by regulating the TCA cycle enzymes SDHD and IDH1. Oncol Rep 45:1–15. https://doi.org/10.3892/or.2021.7998
46. Kerkentzes K, Lagani V, Tsamardinos I, Vyberg M, Røe OD (2014) Hidden treasures in "ancient" microarrays: gene-expression portrays biology and potential resistance pathways of major lung cancer subtypes and normal tissue. Front Oncol 4:251. https://doi.org/10.3389/fonc.2014.00251
47. Olivier M, Hollstein M, Hainaut P (2010) TP53 mutations in human cancers: origins, consequences, and clinical use. Cold Spring Harb Perspect Biol 2:a001008. https://doi.org/10.1101/cshperspect.a001008
48. Halvorsen AR, Silwal-Pandit L, Meza-Zepeda LA, Vodak D, Vu P, Sagerup C et al (2016) TP53 mutation spectrum in smokers and never smoking lung cancer patients. Front Genet 7:85. https://doi.org/10.3389/fgene.2016.00085
49. Wallerand H, Bakkar AA, de Medina SGD, Pairon J-C, Yang Y-C, Vordos D et al (2005) Mutations in TP53, but not FGFR3, in urothelial cell carcinoma of the bladder are influenced by smoking: contribution of exogenous versus endogenous carcinogens. Carcinogenesis 26:177–184. https://doi.org/10.1093/carcin/bgh275
50. Hainaut P, Olivier M, Pfeifer GP (2001) TP53 mutation spectrum in lung cancers and mutagenic signature of components of tobacco smoke: lessons from the IARC TP53 mutation database. Mutagenesis 16:551–553. https://doi.org/10.1093/mutage/16.6.551
51. Pires KMP, Lanzetti M, Rueff-Barroso CR, Castro P, Abrahão A, Koatz VLG et al (2012) Oxidative damage in alveolar macrophages exposed to cigarette smoke extract and participation of nitric oxide in redox balance. Toxicol In Vitro 26:791–798. https://doi.org/10.1016/j.tiv.2012.05.011
52. Ozguner F, Koyu A, Cesur G (2005) Active smoking causes oxidative stress and decreases blood melatonin levels. Toxicol Ind Health 21:21–26. https://doi.org/10.1191/0748233705th211oa
53. Caliri AW, Tommasi S, Besaratinia A (2021) Relationships among smoking, oxidative stress, inflammation, macromolecular damage, and cancer. Mutat Res Rev Mutat Res 787:108365. https://doi.org/10.1016/j.mrrev.2021.108365
54. Venardos K, Zhang W-Z, Chin-Dusting J, Kaye DM (2006) Cigarette smoke adversely influences nitric oxide bioavailability by effects on L-arginine transport and oxidative stress in endothelial cells. J Mol Cell Cardiol 41:748. https://doi.org/10.1016/j.yjmcc.2006.06.057
55. Barua RS, Ambrose JA, Srivastava S, DeVoe MC, Eales-Reynolds L-J (2003) Reactive oxygen species are involved in smoking-induced dysfunction of nitric oxide biosynthesis and upregulation of endothelial nitric oxide synthase: an in vitro demonstration in human coronary artery endothelial cells. Circulation 107:2342–2347. https://doi.org/10.1161/01.CIR.0000066691.52789.BE
56. Kim S-M, Hwang K-A, Choi D-W, Choi K-C (2018) The cigarette smoke components induced the cell proliferation and epithelial to mesenchymal transition via production of reactive oxygen species in endometrial adenocarcinoma cells. Food Chem Toxicol 121:657–665. https://doi.org/10.1016/j.fct.2018.09.023
57. Kutkowska-Kaźmierczak A, Gos M, Obersztyn E (2018) Craniosynostosis as a clinical and diagnostic problem: molecular pathology and genetic counseling. J Appl Genet 59:133–147. https://doi.org/10.1007/s13353-017-0423-4
58. Carmichael SL, Ma C, Rasmussen SA, Honein MA, Lammer EJ, Shaw GM (2008) Craniosynostosis and maternal smoking. Birth Defects Res A 82:78–85. https://doi.org/10.1002/bdra.20426
59. Ardinger HH, Buetow KH, Bell GI, Bardach J, VanDemark DR, Murray JC (1989) Association of genetic variation of the transforming growth factor-alpha gene with cleft lip and palate. Am J Hum Genet 45:348–353
60. Durham EL, Howie RN, Cray JJ (2017) Gene/environment interactions in craniosynostosis: a brief review. Orthod Craniofac Res 20:8–11. https://doi.org/10.1111/ocr.12153
61. Pavone D, Clemenza S, Sorbi F, Fambrini M, Petraglia F (2018) Epidemiology and risk factors of uterine fibroids. Best Pract Res Clin Obstet Gynaecol 46:3–11. https://doi.org/10.1016/j.bpobgyn.2017.09.004
62. Wong JYY, Chang P-Y, Gold EB, Johnson WO, Lee JS (2016) Environmental tobacco smoke and risk of late-diagnosis incident fibroids in the Study of Women's Health across the Nation (SWAN). Fertil Steril 106:1157–1164. https://doi.org/10.1016/j.fertnstert.2016.06.025
63. Baron JA, La Vecchia C, Levi F (1990) The antiestrogenic effect of cigarette smoking in women. Am J Obstet Gynecol 162:502–514. https://doi.org/10.1016/0002-9378(90)90420-C
64. Baron JA, Nichols HB, Anderson C, Safe S (2021) Cigarette smoking and estrogen-related cancer. Cancer Epidemiol Biomarkers Prev 30:1462–1471. https://doi.org/10.1158/1055-9965.EPI-20-1803
65. Nomura M, Li E (1998) Smad2 role in mesoderm formation, left–right patterning and craniofacial development. Nature 393:786–790. https://doi.org/10.1038/31693
66. Nakao A, Imamura T, Souchelnytskyi S, Kawabata M, Ishisaki A, Oeda E et al (1997) TGF-beta receptor-mediated signalling through Smad2, Smad3 and Smad4. EMBO J 16:5353–5362. https://doi.org/10.1093/emboj/16.17.5353

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.