frontiers
in Public Health

Genomic and Bioinformatics Approaches for Analysis of Genes Associated With Cancer Risks Following Exposure to Tobacco Smoking

Mohammed A. I. Al-Obaide[1]*, Buthainah A. Ibrahim[2], Saif Al-Humaish[3] and Abdel-Salam G. Abdel-Salam[4]

[1] Department of Biomedical Science, School of Pharmacy, Texas Tech University Health Science Center, Amarillo, TX, United States, [2] Department of Physics, Diyala University, Baquba, Iraq, [3] Biomedica, LLC, Sterling Heights, MI, United States, [4] Department of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, Doha, Qatar

Cancer is a significant health problem in the Middle East and global population. It is well established that there is a direct link between tobacco smoking and cancer, which will continue to pose a significant threat to human health. The impact of long-term exposure to tobacco smoke on the risk of cancer encouraged the study of biomarkers for vulnerable individuals to tobacco smoking, especially children, who are more susceptible than adults to the action of environmental carcinogens. The carcinogens in tobacco smoke condensate induce DNA damage and play a significant role in determining the health and well-being of smokers, non-smoker, and primarily children. Cancer is a result of genomic and epigenomic malfunctions that lead to an initial premalignant condition. Although premalignancy genetic cascade is a much-delayed process, it will end with adverse health consequences. In addition to the DNA damage and mutations, tobacco smoke can cause changes in the DNA methylation and gene expression associated with cancer. The genetic events hint on the possible use of genomic–epigenomic changes in genes related to cancer, in predicting cancer risks associated with exposure to tobacco smoking. Bioinformatics provides indispensable tools to identify the cascade of expressed genes in active smokers and non-smokers and could assist the development of a framework to manage this cascade of events linked with the evolvement of disease including cancer. The aim of this mini review is to cognize the essential genomic processes and health risks associated with tobacco smoking and the implications of bioinformatics in cancer prediction, prevention, and intervention.

Keywords: tobacco smoking, bioinformatics, cancer, single-nucleotide polymorphisms, DNA methylation, carcinogenic chemicals, health risk

INTRODUCTION

Direct involvement of chemicals in the tobacco smoke with cancer has been shown since the 1970s and considered as pleiotropic carcinogens that cause DNA damage (1–3). Although, tobacco smoke associated with many types of cancer, it is linked primarily to lung cancer in 90% of men and 70–80% of lung cancer in women (4). Furthermore, tobacco smoke causes more than 13 types of cancer in men and women (5–11). The main consequences of long-term exposure to tobacco smoke associated with

the premalignancy mechanism, which is a considerably delayed process, sometimes taking several years (12, 13). Accordingly, cancer surveillance plays an essential role in cancer prevention and intervention of various socioeconomic aspects associated with cancer attributed to tobacco smoking. The health risk of tobacco smoke is of utmost importance to assess in children and non-smokers. These health issues require sophisticated genomic assays besides the clinical procedures (14–16). The incidence of disease and cancer risks from cigarette smoking among male and female smokers, increased in the Middle East and worldwide over most of the twentieth century, and continued in the twenty-first century as compared with persons who have never smoked (5, 6, 17–19).

In this study, we will highlight genomic and molecular features, which can be used to develop approaches to predict cancer after long-term tobacco smoke exposure. Most importantly, the influence of tobacco smoke on children and non-smokers. Unfortunately, the negative impact of tobacco smoke on the health of non-smokers is a result of involuntary exposure to tobacco smoke. It is, therefore, crucial that the outcome of research efforts in this field go hand in hand with efforts to develop practical informatics and bioinformatics programs to face the challenges through good practices and to strengthen public awareness and understanding of hidden dangers of tobacco smoking. Thus, research activities in this area will participate in creating more efficient cancer prevention program and find novel molecular methods for early detection of premalignancy lesions caused by tobacco smoke carcinogens.

CARCINOGENIC CHEMICALS IN TOBACCO SMOKE

Tobacco smoke constitutes more than 4,000 compounds, of which over 70 with carcinogenic activities, such as polycyclic aromatic hydrocarbons (PAHs), nitrosamines, aromatic amines, and others (6, 20, 21). The most critical nitrosamines carcinogens in tobacco smoke are N-dimethylnitrosamine (DMN, also known as N-nitrosodimethylamine), 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), and N'-nitrosonornicotine (NNN) (1–3). N-nitrosamines and tobacco-specific nitrosamines (TSNAs) can attack DNA at the O6-position of guanine forming DNA adducts O6-methylguanine and O(6)-[4-oxo-4-(3-pyridyl) butyl]guanine (O6pobG). Both are mutagenic lesions (22), if left unrepaired would pair with thymine and introduce G:C to A:T transition upon DNA replication. Another study showed no reduction observed in tobacco carcinogenic nitrosamines associated with cancer (23). The study revealed that the sum of the two potent carcinogens, NNK and NNN in cigarette filler averaged 2.54 ± 1.05 µg/g tobacco, which is identical to the results of these two carcinogens reported for the tobacco of a U.S. filtered cigarette in 1979. Tobacco smoke carcinogenic chemicals can cause DNA damage and can give rise to mutations and single-nucleotide polymorphisms (SNPs) in genes and regulatory sequences associated with various types of cancers. DNA damage caused by tobacco smoke carcinogens is related to cancer development (2) and found associated with lung cancer (4). Interestingly, though tobacco use causes more than 13 types of cancer in men and women (5–11), it is the leading preventable cause of cancer (6). **Figure 1** summarizes the events followed exposure to tobacco smoke and development of cancer.

CONSEQUENCES OF ENVIRONMENTAL TOBACCO SMOKE (ETS)

The health risks because of inhaling tobacco smoke are not limited to tobacco smokers. Because cancer risks owing to tobacco smoke inhalation can extend to non-smokers especially children, who involuntarily inhale ETS at home, at work, or in public places (24). More recent studies showed the non-smokers represent a significant percentage of lung cancer patients, which is estimated approximately 25% of all lung cancer cases indicating the deadly effect of passive and second-hand smoking (25). Further, children and non-smokers exposed to third-hand smoke (THS), which is residual tobacco smoke deposited on surfaces and dust and linked
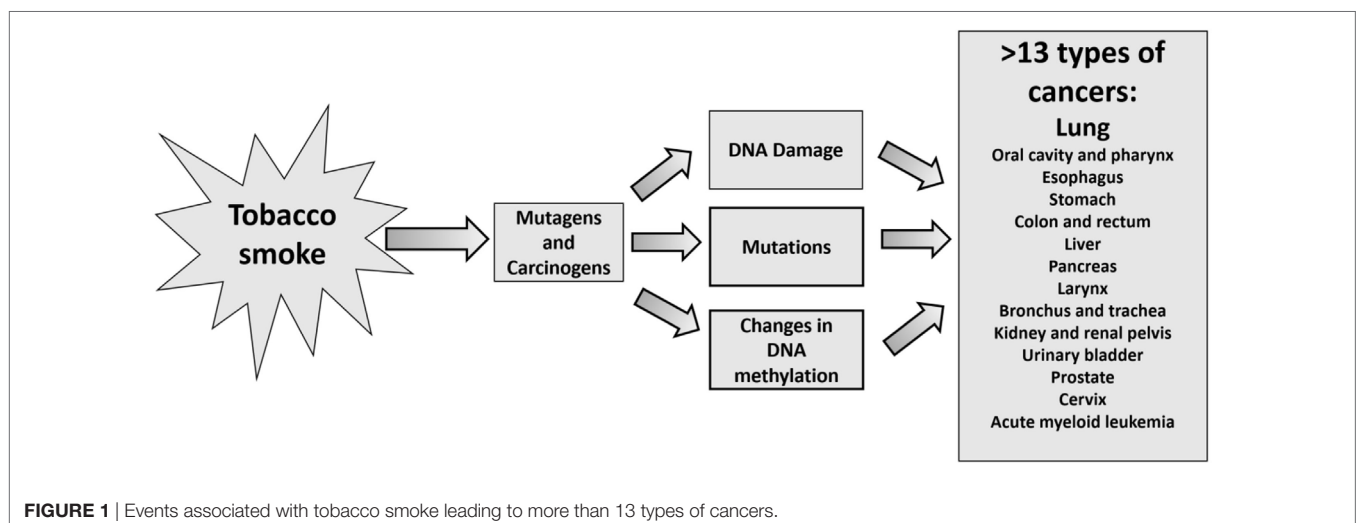

FIGURE 1 | Events associated with tobacco smoke leading to more than 13 types of cancers.

to higher cancer risk incidence in non-smokers (26). The study showed the likely cancer risk by age group *via* dermal exposure to carcinogen *N*-nitrosamines and TSNAs measured in house dust samples. The calculated TSNAs risks surpassed the upper-bound risk advocated by the United States Environmental Protection Agency (EPA/U.S. EPA).

The consequences of risks following exposure of children and non-smoker, the main victims of the ETS and THS, to TSNAs are not limited to the ability of TSNAs to induce DNA damage and play a significant role in determining the health and well-being of children. But extended to the association of mutagenic, carcinogenic and DNA-damaging effects of various chemicals in tobacco smoke inhalation associated with the premalignancy mechanism (12, 13), which is a much-delayed process, sometimes taking as long as 30 years. Although there is some controversial association between parental smoking during pregnancy and risk of childhood tumors (27), recent studies have reported positive associations between paternal smoking during pregnancy and childhood brain tumor risk (28, 29) and childhood acute lymphoblastic and myeloid leukemia (30).

TOBACCO SMOKE, SNPs, AND CANCER

Mutations are considered the primary sources of SNPs in genes leading to the formation of genetic variants linked to the development of cancer following exposure to environmental risk factors (31–33). Many SNPs located within the regulatory regions of the genes, which may influence the expression of the genes (34, 35). SNPs of glutathione S-transferase (*GST*) and *TP53* are examples. The GST enzymes contribute to the activation and inactivation of oxidative metabolites of carcinogenic agents correlated with cancer (36). The identified polymorphisms in GST genes associated with cancer susceptibility and increased the risk resulting from exposure to tobacco smoke and decreased the ability of GST enzymes to detoxify the carcinogens (28, 36). The TP53 gene is a tumor suppressor gene and involved in several vital cellular functions (37, 38). The *TP53* gene harbors high frequency of functional SNPs, which may alter P53 protein function (39). Studies reported the association of several TP53 SNPs with risk of cancer. Mutations in the *TP53* tumor suppressor gene implicated in different cancer types, including lung cancer (40–42). Multifactor dimensionality reduction analysis used and provided evidence on interactions among various *P53* SNPs and tobacco habit (43).

IMPACT OF TOBACCO SMOKE ON GENOMIC DNA METHYLATION

DNA methylation plays a significant role in silencing gene expression, and changes in genomic DNA methylation patterns from the natural state will lead to the development of cancers (44, 45) (**Figure 1**). Tobacco smoke causes abnormal DNA methylation profiles in the genome of smokers and involuntarily exposed non-smokers. Evidence of significant differences observed in the degree of site-specific methylation in each of the 22 autosomes as a function of tobacco smoking (46), suggesting DNA methylation aberration is another criterion associated with the prediction of

cancer risk. The epigenetic changes observed at CpGs in blood. Su et al. (47) identified novel CpGs associated with current smoking, pack-years, duration. The authors were able to show unique profiles of smoking-associated DNA methylation and gene expression among immune cell types. Another study showed from CpG site methylation that tobacco smoking differently influences cell types of the innate and adaptive immune system-indications (48). The link between lung cancer-related genes and tobacco smoking was confirmed through identifications of 13 novel CpG with a decrease of DNA methylation compared to never smoking sites located in 8 genes: *KLF6*, *STK32A*, *TERT*, *MSH5*, *ACTA2*, *GATA3*, *VTI1A*, and *CHRNA5* (49).

BIOINFORMATICS APPLICATIONS IN TOBACCO SMOKE-ASSOCIATED CANCER GENES

There are several definitions of "Bioinformatics," in this article we use NIH working definition of bioinformatics "Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data." The U.S. National Library of Medicine gives a similar definition "The organization and analysis of biological and related information, usually involving the use of computers to develop databases, retrieval mechanisms, and data analysis tools, especially in the fields of molecular biology, structural biology, and genetics." Bioinformatics applies principles of information sciences and technologies to make the vast, diverse, and complex life sciences data more understandable and accessible. Thus, it is the bridge between domain users (i.e., researchers) to the software needed to support the workflow of their domain. Here, we show an example that used bioinformatics to develop a knowledgebase for addiction-related genes (KARG) database (50). KARG is a cross-platform database links genes and chromosome regions and pathways to addiction. An important accomplishment of the database identification of pathway for addiction chemicals. The KARG database identified addiction pathway of nicotine, which is a result of the neuroactive ligand-receptor interaction.

Furthermore, bioinformatics databases can provide indispensable tools to assist identification of changes in the expression of genes associated with nicotine addiction in active smokers and never smokers (see Discussion). Also, the bioinformatics databases can provide a framework for scientists to manage the deluge of data and could help to discover new genes associated with tobacco smoking addiction and initiation. The data from NCBI-Gene and UniProt were used to show the expression and health risks of tobacco smoking associated with nine genes related to the initiation of smoking and nicotine dependence, and four genes linked to lung cancer (**Table 1**). The table also shows the potential to discover a new gene, *KMO* that can help in tobacco smoking cessation. Modulating *KMO* gene expression could be a useful tactic for the treatment of tobacco dependence (51). Collectively, the data show the bioinformatic methods help to investigate the genes' involvement in tobacco smoking initiation and dependence, cancer and other diseases and therapy.

TABLE 1 | Examples of genes associated with health risks of tobacco smoking.

Gene	Protein	Smoking trait	Disease	Gene expression	Reference
CHRNA3	Neuronal acetylcholine receptor subunit alpha-3	Initiation, dependence	Lung cancer, obstructive pulmonary	0.197 ± 0.088^a	(52–55)
CHRNA5	Neuronal acetylcholine receptor subunit alpha-5	Initiation	Obstructive pulmonary, lung and colorectal cancer	0.234 ± 0.194^a	(52, 55–57)
CHRNB3	Neuronal acetylcholine receptor subunit beta-3	Dependence	Esophageal adenocarcinoma	0.132 ± 0.071^a	(55, 58, 59)
CHRNB4	Neuronal acetylcholine receptor subunit beta-4	Dependence	Lung cancer	0.022 ± 0.016^a	(55, 60)
CYP2A6	Cytochrome P450 2A6	Dependence	Lung cancer	542 ± 278^b	(55, 61–63)
NTRK2	BDNF/NT-3 growth factors receptor	Nicotine dependence	Mood disorders	64.619 ± 20.74^a	(52, 64)
GABARAP	Gamma-aminobutyric acid receptor-associated protein	Nicotine dependence	–	148.01 ± 24.08^a	(52, 65)
ANKK1	Ankyrin repeat and protein kinase domain-containing protein 1	Nicotine dependence	Attention deficit hyperactivity disorder	0.01 ± 0.01^a	(52, 66)
KMO	Kynurenine 3-monooxygenase	Addiction therapy	Huntington disease, schizophrenia	1.976 ± 0.559^a	(51, 52)

The data extracted from NCBI-Genes and proteins, and UniProt databases. The measurements of expression of genes presented as mean Reads Per Kilobase Million in the ^{a}brain and ^{b}liver.

DISCUSSION

The bioinformatics methods used in the analyses of genomic sequences and gene expression profiles of genes and regulatory elements can identify the changes that arise because of exposures to tobacco smoke. The induced changes in tobacco smoke cause mutations and form SNPs, and malfunction of regulatory modes of gene expressions (36, 49). Many chemicals and carcinogens identified in tobacco smoke are linked to several human malignancies and other diseases (**Table 1**). Tobacco smoking can cause cancer, because of N-nitrosamines and TSNAs (22, 23), recent studies showed that nicotine is associated with the development of cancer (67). This finding exacerbates the health risks of nicotine addiction and suggests its involvement in worsening and relapse of cancer, considering the formation of TSNAs from nicotine in the body (67–69).

The data from NCBI-Gene and Uniprot showed the link between tobacco smoking and health risk of eight genes (**Table 1**). Although GABARAP showed no health risk, its involvement in nicotine addiction suggested the potential formation of TSNAs from nicotine and development of cancer (67–69). The table also shows the potential to discover the new gene KMO that can help in tobacco smoking cessation. Modulating KMO gene expression could be a useful tactic for the treatment of tobacco dependence (50). The expression KMO is reduced in schizophrenia and in bipolar patients with a history of psychosis (50), and two KMO SNPs were found frequently in schizophrenia patients compared with healthy controls (51).

Integrated bioinformatics approaches were used to show the differentially expressed genes (DEGs) in smokers and non-smokers. Cao et al. (70) used in their analysis of DEGs: gene ontology, pathway enrichment analyses of DEGs, protein–protein interaction network, and transcriptional regulatory network along with miRNA-target regulatory network construction. The authors showed DEGs, for example, CYP1A1, CYP1B1, YWHAZ, and PTPRD, may have the potential to be used as biomarkers and therapeutic targets in tobacco smoke-related pathological changes. Bioinformatic methods also showed applications in investigating the impact of tobacco smoke on regulatory factors associated with gene expression in various ways, for example, an aberration in DNA methylation (46–49, 71, 72), transcriptional binding sites (73), and noncoding RNA (70, 74, 75).

Another topic requires consideration is the correlation between medical informatics and bioinformatics. In their article, Wyatt and Liu suggested that medical informatics include bioinformatics, clinical informatics, consumer health informatics (CHI), and public health informatics (76). Bio and medical informatics have applications in various fields specifically, genomic medicine. Generally, their contributions were in areas such as knowledge-based systems, database design, data mining, genomic sequences, and structure analysis (77). Of interest, CHI, which is interacting topic with bioinformatics. In their report, Gibbons et al. (78) showed the outcomes related to the use of tobacco that included self-management, program adherence, and change in health behaviors. The outcome indicates that CHI applications had statistically significant effects on health.

Finally, the main enigma of health risk associated with tobacco smoking is not restricted to cancer but it is also linked to other noncommunicable diseases (NCDs), cardiovascular diseases, chronic respiratory, and diabetes (79, 80). Tobacco smokers with diabetes have higher risks for serious cardiovascular complications (81). The NCDs, are the major reason for mortality in the world that represent 63% of all annual deaths, over 36 million people each year. Intriguingly, nearly 80% of all NCDs deaths take place in low and middle-income countries (80). The impact of the sociodemographic and psychosocial factors associated with men and women is apparent in emerging adulthood who are prone to the tobacco smoking habit. Nowadays, although smoking prevalence and cigarette consumption have declined in some countries, smoking prevalence among women is increasing (20, 82–87).

In conclusion, tobacco smoke is a widespread carcinogen; which contains many chemicals known to cause cancer, including PAHs, nitrosamines, and aromatic amines. It can cause DNA damage, mutations, and changes in DNA methylation. A considerable percentage of lung cancer is attributable to ETS carcinogens, besides more than 13 other types of cancer and NCDs. Applying bioinformatics, medical informatics and genomic for cancer surveillance in smokers and non-smokers could be utilized in cancer prediction, prevention, and intervention.

AUTHOR CONTRIBUTIONS

The authors have made equal contributions in the writing and revising of this mini review.

REFERENCES

1. Hecht SS. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat Rev Cancer* (2003) 3(10):733–44. doi:10.1038/nrc1190

2. Hecht SS. Progress and challenges in selected areas of tobacco carcinogenesis. *Chem Res Toxicol* (2008) 21(1):160–71. doi:10.1021/tx7002068

3. Hecht SS, Hoffmann D. Tobacco-specific nitrosamines, an important group of carcinogens in tobacco and tobacco smoke. *Carcinogenesis* (1988) 9:875–84. doi:10.1093/carcin/9.6.875

4. Walser T, Cui X, Yanagawa J, Lee JM, Heinrich E, Lee G, et al. Smoking and lung cancer: the role of inflammation. *Proc Am Thorac Soc* (2008) 5:811–5. doi:10.1513/pats.200809-100TH

5. Ibrahim WH, Rasul KI, Khinji A, Ahmed MS, Bener A. Clinical and epidemiological characteristics of lung cancer cases in Qatar. *East Mediterr Health J* (2010) 16(2):166–70.

6. Henley SJ, Thomas CC, Sharapova SR, Momin B, Massetti GM, Winn DM, et al. Vital signs: disparities in tobacco-related cancer incidence and mortality – United States, 2004–2013. *MMWR Morb Mortal Wkly Rep* (2016) 65:1212–8. doi:10.15585/mmwr.mm6544a3

7. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. *The Health Consequences of Smoking-50 Years of Progress: A Report of the Surgeon General*. Atlanta, GA: Centers for Disease Control and Prevention (US) (2014).

8. Working Group on the Evaluation of Carcinogenic Risks to Humans, International Agency for Research on Cancer. IARC monographs on the evaluation of carcinogenic risks to humans. *Personal Habits and Indoor Combustions*. (Vol. 100E), Lyon, France: International Agency for Research on Cancer (2012). p. 1–598.

9. Henley SJ, Thun MJ. Chapter 4: Health consequences of smokeless tobacco use. In: Hatsukami DK, Zeller M, Gupta P, Parascandola M, Asma S, editors. *Smokeless Tobacco and Public Health: A Global Perspective*. Bethesda, MD: US Department of Health and Human Services, CDC, National Institutes of Health, National Cancer Institute. NIH Publication No. 14-7983 (2014). p. 1–558.

10. National Cancer Institute. Smoking and tobacco control monograph 9. *Cigars: Health Effects and Trends*. Bethesda, MD: US Department of Health and Human Services, National Institutes of Health, National Cancer Institute (1998).

11. US Department of Health and Human Services. *The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General*. Atlanta, GA: US Department of Health and Human Services, CDC (2006).

12. Jackson PE, Kuang SY, Wang JB, Strickland PT, Muñoz A, Kensler TW, et al. Prospective detection of codon 249 mutations in plasma of hepatocellular carcinoma patients. *Carcinogenesis* (2003) 24(10):1657–63. doi:10.1093/carcin/bgg101

13. Antony J, Gopalan V, Smith RA, Lam AK. Carcinoma ex pleomorphic adenoma: a comprehensive review of clinical, pathological and molecular data. *Head Neck Pathol* (2012) 6(1):1–9. doi:10.1007/s12105-011-0281-z

14. De S, Ganesan S. Looking beyond drivers and passengers in cancer genome sequencing data. *Ann Oncol* (2017) 28(5):938–45. doi:10.1093/annonc/mdw677

15. Dimitrakopoulos CM, Beerenwinkel N. Computational approaches for the identification of cancer genes and pathways. *Wiley Interdiscip Rev Syst Biol Med* (2017) 9:e1364. doi:10.1002/wsbm.1364

16. ASGE Standards of Practice Committee, Evans JA, Chandrasekhara V, Chathadi KV, Decker GA, Early DS, et al. The role of endoscopy in the management of premalignant and malignant conditions of the stomach. *Gastrointest Endosc* (2015) 82(1):1–8. doi:10.1016/j.gie.2015.03.1967

17. Poirier AE, Grundy A, Khandwala F, Tamminen S, Friedenreich CM, Brenner DR. Cancer incidence attributable to tobacco in Alberta, Canada, in 2012. *CMAJ Open* (2016) 4(4):E578–87. doi:10.9778/cmajo.20150069

18. Galceran J, Ameijide A, Carulla M, Mateos A, Quirós JR, Rojas D, et al. Cancer incidence in Spain, 2015. *Clin Transl Oncol* (2017) 19(7):799–825. doi:10.1007/s12094-016-1607-9

19. Thun MJ, Carter BD, Feskanich D, Freedman ND, Prentice R, Lopez AD, et al. 50-year trends in smoking-related mortality in the United States. *N Engl J Med* (2013) 368(4):351–64. doi:10.1056/NEJMsa1211127

20. Taioli E. Gene-environment interaction in tobacco-related cancers. *Carcinogenesis* (2008) 29:1467–74. doi:10.1093/carcin/bgn062

21. Zienolddiny S, Skaug V. Single nucleotide polymorphisms as susceptibility, prognostic, and therapeutic markers of nonsmall cell lung cancer. *Lung Cancer (Auckl)* (2011) 3:1–14. doi:10.2147/LCTT.S13256

22. Christmann M, Kaina B. O(6)-methylguanine-DNA methyltransferase (MGMT): impact on cancer risk in response to tobacco smoke. *Mutat Res* (2012) 736:64–74. doi:10.1016/j.mrfmmm.2011.06.004

23. Stepanov I, Knezevich A, Zhang L, Watson C, Hatsukami DK, Hecht SS. Carcinogenic tobacco-specific N-nitrosamines in US cigarettes: three decades of remarkable neglect by the tobacco industry. *Tob Control* (2012) 21(1):44–8. doi:10.1136/tc.2010.042192

24. Tobacco Control Monograph Series. *Monograph 10: Health Effects of Exposure to Environmental Tobacco Smoke. Broad review of Environmental Tobacco Smoke*. USA: The National Cancer Institute established the Tobacco Control Monograph series, National Cancer Institute (1999).

25. Pallis AG, Syrigos KN. Lung cancer in never smokers: disease characteristics and risk factors. *Crit Rev Oncol Hematol* (2013) 88(3):494–503. doi:10.1016/j.critrevonc.2013.06.011

26. Ramírez N, Özel MZ, Lewis AC, Marcé RM, Borrull F, Hamilton JF. Exposure to nitrosamines in thirdhand tobacco smoke increases cancer risk in non-smokers. *Environ Int* (2014) 71:139–47. doi:10.1016/j.envint.2014.06.012

27. Huang Y, Huang J, Lan H, Zhao GY, Huang CZ. A meta-analysis of parental smoking and the risk of childhood brain tumors. *PLoS One* (2014) 9(7):e102910. doi:10.1371/journal.pone.0102910

28. Barrington-Trimis JL, Nielsen SS, Preston-Martin S, James Gauderman W, Holly EA, Farin FM, et al. Parental smoking and risk of childhood brain tumors by functional polymorphisms in polycyclic aromatic hydrocarbon metabolism genes. *PLoS One* (2013) 8(11):e79110. doi:10.1371/journal.pone.0079110

29. Plichart M, Menegaux F, Lacour B, Hartmann O, Frappaz D, Doz F, et al. Parental smoking, maternal alcohol, coffee and tea consumption during pregnancy and childhood malignant central nervous system tumours: the ESCALE study (SFCE). *Eur J Cancer Prev* (2008) 17(4):376–83. doi:10.1097/CEJ.0b013e3282f75e6f

30. Metayer C, Zhang L, Wiemels JL, Bartley K, Schiffman J, Ma X, et al. Tobacco smoke exposure and the risk of childhood acute lymphoblastic and myeloid leukemias by cytogenetic subtype. *Cancer Epidemiol Biomarkers Prev* (2013) 22(9):1600–11. doi:10.1158/1055-9965.EPI-13-0350

31. 1000 Genomes Project Consortium; Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* (2012) 491(7422):56–65. doi:10.1038/nature11632

32. Park JH, Gail MH, Greene MH, Chatterjee N. Potential usefulness of single nucleotide polymorphisms to identify persons at high cancer risk: an evaluation of seven common cancers. *J Clin Oncol* (2012) 30(17):2157–62. doi:10.1200/JCO.2011.40.1943

33. Van den Broeck T, Joniau S, Clinckemalie L, Helsen C, Prekovic S, Spans L, et al. The role of single nucleotide polymorphisms in predicting prostate cancer risk and therapeutic decision making. *Biomed Res Int* (2014) 2014:627510. doi:10.1155/2014/627510

34. Faber K, Glatting KH, Mueller PJ, Risch A, Hotz-Wagenblatt A. Genome-wide prediction of splice-modifying SNPs in human genes using a new analysis pipeline called AASsites. *BMC Bioinformatics* (2011) 12(Suppl 4):S2. doi:10.1186/1471-2105-12-S4-S2

35. Guo Y, Jamison DC. The distribution of SNPs in human gene regulatory regions. *BMC Genomics* (2005) 6:140. doi:10.1186/1471-2164-6-140

36. Lavender NA, Benford ML, VanCleave TT, Brock GN, Kittles RA, Moore JH, et al. Examination of polymorphic glutathione S-transferase (GST) genes, tobacco smoking and prostate cancer risk among men of African descent: a case-control study. *BMC Cancer* (2009) 9:397. doi:10.1186/1471-2407-9-397

37. Whibley C, Pharoah PD, Hollstein M. p53 polymorphisms: cancer implications. *Nat Rev Cancer* (2009) 9(2):95–107. doi:10.1038/nrc2584

38. Li Y, Chang SC, Niu R, Liu L, Crabtree-Ide CR, Zhao B, et al. TP53 genetic polymorphisms, interactions with lifestyle factors and lung cancer risk: a case control study in a Chinese population. *BMC Cancer* (2013) 13:607. doi:10.1186/1471-2407-13-607

39. Grochola LF, Zeron-Medina J, Meriaux S, Bond GL. Single-nucleotide polymorphisms in the p53 signaling pathway. *Cold Spring Harb Perspect Biol* (2010) 2(5):a001032. doi:10.1101/cshperspect.a001032

40. Yan L, Zhang D, Chen C, Mao Y, Xie Y, Li Y, et al. TP53 Arg72Pro polymorphism and lung cancer risk: a meta-analysis. *Int J Cancer* (2009) 125(12):2903–11. doi:10.1002/ijc.24603

41. Francisco G, Menezes PR, Eluf-Neto J, Chammas R. Arg72Pro TP53 polymorphism and cancer susceptibility: a comprehensive meta-analysis of 302 case-control studies. *Int J Cancer* (2011) 129(4):920–30. doi:10.1002/ijc.25710

42. Bellini MF, Cadamuro AC, Succi M, Proença MA, Silva AE. Alterations of the TP53 gene in gastric and esophageal carcinogenesis. *J Biomed Biotechnol* (2012) 2012:891961. doi:10.1155/2012/891961

43. Kotnis A, Namkung J, Kannan S, Jayakrupakar N, Park T, Sarin R, et al. Multiple pathway-based genetic variations associated with tobacco related multiple primary neoplasms. *PLoS One* (2012) 7(1):e30013. doi:10.1371/journal.pone.0030013

44. Ibrahim MA. Advances in genomic DNA methylation analysis. *Biotechnology* (2010) 9(4):459–68. doi:10.3923/biotech.2010.459.468

45. Ibrahim MA. Perspective of DNA methylation in cancer research. *Int J Cancer Res* (2010) 6(4):188–201. doi:10.3923/ijcr.2010.188.201

46. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One* (2013) 8(5):e63812. doi:10.1371/journal.pone.0063812

47. Su D, Wang X, Campbell MR, Porter DK, Pittman GS, Bennett BD, et al. Distinct epigenetic effects of tobacco smoking in whole blood and among leukocyte subtypes. *PLoS One* (2016) 11(12):e0166486. doi:10.1371/journal.pone.0166486

48. Bauer M, Fink B, Thürmann L, Eszlinger M, Herberth G, Lehmann I. Tobacco smoking differently influences cell types of the innate and adaptive immune system-indications from CpG site methylation. *Clin Epigenetics* (2016) 7:83. doi:10.1186/s13148-016-0249-7

49. Gao X, Zhang Y, Breitling LP, Brenner H. Tobacco smoking and methylation of genes related to lung cancer development. *Oncotarget* (2016) 7(37):59017–28. doi:10.18632/oncotarget.10007

50. Li CY, Mao X, Wei L. Genes and (common) pathways underlying drug addiction. *PLoS Comput Biol* (2008) 4(1):e2. doi:10.1371/journal.pcbi.0040002

51. Secci ME, Auber A, Panlilio LV, Redhi GH, Thorndike EB, Schindler CW, et al. Attenuating nicotine reinforcement and relapse by enhancing endogenous brain levels of kynurenic acid in rats and squirrel monkeys. *Neuropsychopharmacology* (2017) 42(8):1619–29. doi:10.1038/npp.2017.21

52. Korytina GF, Akhmadishina LZ, Viktorova EV, Kochetova OV, Viktorova TV. IREB2, CHRNA5, CHRNA3, FAM13A & hedgehog interacting protein genes polymorphisms & risk of chronic obstructive pulmonary disease in Tatar population from Russia. *Indian J Med Res* (2016) 44(6):865–76. doi:10.4103/ijmr.IJMR_1233_14

53. Qu H, Wang K, Dong W, Shen H, Wang Y, Liu Q, et al. Association between two CHRNA3 variants and susceptibility of lung cancer: a meta-analysis. *Sci Rep* (2016) 6:20149. doi:10.1038/srep20149

54. Wen L, Jiang K, Yuan W, Cui W, Li MD. Contribution of variants in CHRNA5/A3/B4 gene cluster on chromosome 15 to tobacco smoking: from genetic association to mechanism. *Mol Neurobiol* (2016) 53(1):472–84. doi:10.1007/s12035-014-8997-x

55. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics* (2014) 13(2):397–406. doi:10.1074/mcp.M113.035600

56. Wang Y, Peng X, Zhu L, Hu L, Song Y. Genetic variants of CHRNA5-A3 and CHRNB3-A6 predict survival of patients with advanced non-small cell lung cancer. *Oncotarget* (2016) 7(18):26436–43. doi:10.18632/oncotarget.8510

57. Pérez-Rubio G, Pérez-Rodríguez ME, Fernández-López JC, Ramírez-Venegas A, García-Colunga J, Ávila-Moreno F, et al. SNPs in NRXN1 and CHRNA5 are associated to smoking and regulation of GABAergic and glutamatergic pathways. *Pharmacogenomics* (2016) 17(10):1145–58. doi:10.2217/pgs-2016-0020

58. Cui WY, Wang S, Yang J, Yi SG, Yoon D, Kim YJ, et al. Significant association of CHRNB3 variants with nicotine dependence in multiple ethnic populations. *Mol Psychiatry* (2013) 18(11):1149–51. doi:10.1038/mp.2012.190

59. Song Y, Wang Y, Xu L, Ma J, Chen E, Zang R, et al. A genetic variant in CHRNB3-CHRNA6 increases risk of esophageal squamous cell carcinoma in Chinese populations. *Carcinogenesis* (2015) 36(5):538–42. doi:10.1093/carcin/bgv019

60. Zhang Y, Jiang M, Li Q, Liang W, He Q, Chen W, et al. Chromosome 15q25 (CHRNA3-CHRNB4) variation indirectly impacts lung cancer risk in Chinese males. *PLoS One* (2016) 11(3):e0149946. doi:10.1371/journal.pone.0149946

61. Ito T, Tsuji M, Mori Y, Kanda H, Hidaka T, Kakamu T, et al. Effect of cyp2a6*4 genetic polymorphisms on smoking behaviors and nicotine dependence in a general population of Japanese men. *Fukushima J Med Sci* (2015) 61(2):125–30. doi:10.5387/fms.2015-14

62. Pérez-Rubio G, López-Flores LA, Ramírez-Venegas A, Noé-Díaz V, García-Gómez L, Ambrocio-Ortiz E, et al. Genetic polymorphisms in CYP2A6 are associated with a risk of cigarette smoking and predispose to smoking at younger ages. *Gene* (2017) 628:205–10. doi:10.1016/j.gene.2017.07.051

63. Park SL, Murphy SE, Wilkens LR, Stram DO, Hecht SS, Le Marchand L. Association of CYP2A6 activity with lung cancer incidence in smokers: the multiethnic cohort study. *PLoS One* (2017) 12(5):e0178435. doi:10.1371/journal.pone.0178435

64. Beuten J, Ma JZ, Payne TJ, Dupont RT, Lou XY, Crews KM, et al. Association of specific haplotypes of neurotrophic tyrosine kinase receptor 2 gene (NTRK2) with vulnerability to nicotine dependence in African-Americans and European-Americans. *Biol Psychiatry* (2007) 61(1):48–55. doi:10.1016/j.biopsych.2006.02.023

65. Lou XY, Ma JZ, Sun D, Payne TJ, Li MD. Fine mapping of a linkage region on chromosome 17p13 reveals that GABARAP and DLG4 are associated with vulnerability to nicotine dependence in European-Americans. *Human Mol Genet* (2007) 16(2):142–53. doi:10.1093/hmg/ddl450

66. Gelernter J, Yu Y, Weiss R, Brady K, Panhuysen C, Yang B-Z, et al. Haplotype spanning TTC12 and ANNKK1, flanked by the DRD2 and NCAM1 loci, is strongly associated to nicotine dependence in two distinct American populations. *Human Mol Genet* (2006) 15(24):3498–507. doi:10.1093/hmg/ddl426

67. Sanner T, Grimsrud TK. Nicotine: carcinogenicity and effects on response to cancer treatment – a review. *Front Oncol* (2015) 5:196. doi:10.3389/fonc.2015.00196

68. Schaal C, Chellappan SP. Nicotine-mediated cell proliferation and tumor progression in smoking-related cancers. *Mol Cancer Res* (2014) 12:14–23.1. doi:10.1158/1541-7786.MCR-13-0541

69. Grando SA. Connections of nicotine to cancer. *Nat Rev Cancer* (2014) 14:419–29. doi:10.1038/nrc3725

70. Cao C, Chen J, Lyu C, Yu J, Zhao W, Wang Y, et al. Bioinformatics analysis of the effects of tobacco smoke on gene expression. *PLoS One* (2015) 10(12):e0143377. doi:10.1371/journal.pone.0143377

71. Lee KW, Pausova Z. Cigarette smoking and DNA methylation. *Front Genet* (2013) 4:132. doi:10.3389/fgene.2013.00132

72. Freeman JR, Chu S, Hsu T, Huang YT. Epigenome-wide association study of smoking and DNA methylation in non-small cell lung neoplasms. *Oncotarget* (2016) 7(43):69579–91. doi:10.18632/oncotarget.11831

73. Wang X, Chorley BN, Pittman GS, Kleeberger SR, Brothers J II, Liu G, et al. Genetic variation and antioxidant response gene expression in the bronchial airway epithelium of smokers at risk for lung cancer. *PLoS One* (2010) 5(8):e11934. doi:10.1371/journal.pone.0011934

74. Metzler-Guillemain C, Victorero G, Lepoivre C, Bergon A, Yammine M, Perrin J, et al. Sperm mRNAs and microRNAs as candidate markers for the impact of toxicants on human spermatogenesis: an application to tobacco smoking. *Syst Biol Reprod Med* (2015) 61(3):139–49. doi:10.3109/19396368.2015.1022835

75. Kappil M, Chen J. Environmental exposures in utero and microRNA. *Curr Opin Pediatr* (2014) 26(2):243–51. doi:10.1097/MOP.0000000000000073

76. Wyatt JC, Liu JL. Basic concepts in medical informatics. *J Epidemiol Community Health* (2002) 56(11):808–12. doi:10.1136/jech.56.11.808

77. Maojo V, Kulikowski CA. Bioinformatics and medical informatics: collaborations on the road to genomic medicine? *J Am Med Inform Assoc* (2003) 10(6):515–22. doi:10.1197/jamia.M1305

78. Gibbons MC, Wilson RF, Samal L, Lehmann CU, Dickersin K, Lehmann HP, et al. *Impact of Consumer Health Informatics Applications. Evidence Report/Technology Assessment No. 188*. (Prepared by Johns Hopkins University Evidence-Based Practice Center under Contract No. HHSA 290-2007-10061-I). AHRQ Publication No. 09(10)-E019. Rockville, MD: Agency for Healthcare Research and Quality (2009).

79. Thakur JS, Garg R, Narain JP, Menabde N. Tobacco use: a major risk factor for non communicable diseases in South-East Asia region. *Indian J Public Health* (2011) 55(3):155–60. doi:10.4103/0019-557X.89943

80. The U.S. Government and Global Non-Communicable Disease Efforts. *Fact Sheet*. (2017). Available from: www.kff.org

81. Smoking and Diabetes. *Centers for Disease Control and Prevention*. (2018). Available from: https://www.cdc.gov/tobacco/campaign/tips/diseases/diabetes.html (accessed December 10, 2017).

82. *WHO Global Report on Trends in Prevalence of Tobacco Smoking/2015*. NLM Classification: WM 290. World Health Organization (2015). p. 1–359.

83. Centers for Disease Control and Prevention (CDC). *Office on Smoking and Health (OSH) – Global Tobacco Surveillance System (GTSS) – Global Adult Tobacco Survey (GATS)*. (2016). Available from: https://www.cdc.gov/tobacco/global/gtss/gtssdata/index.html

84. Li X, Holahan CK, Holahan CJ. Sociodemographic and psychological characteristics of very light smoking among women in emerging adulthood, national survey of drug use and health, 2011. *Prev Chronic Dis* (2015) 12:E111. doi:10.5888/pcd12.140547

85. Cloonan SM, Glass K, Laucho-Contreras ME, Bhashyam AR, Cervo M, Pabón MA, et al. Mitochondrial iron chelation ameliorates cigarette smoke-induced bronchitis and emphysema in mice. *Nat Med* (2016) 22(2):163–74. doi:10.1038/nm.4021

86. U.S. Department of Health & Human Services. *The Health Consequences of Smoking: Nicotine and Addiction. A Report of the Surgeon General*. Washington, DC: USDHHS. Report No.: DHHS Publication No. (CDC) 88–8406 (1988). p. 1–643.

87. *10 Facts on Noncommunicable Diseases*. WHO (2018). Available from: http://www.who.int/features/factfiles/noncommunicable_diseases/en/ (accessed December 10, 2017).

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.