REVIEW

Tobacco, air pollution, environmental carcinogenesis, and thoughts on conquering strategies of lung cancer

Guangbiao Zhou
State Key Laboratory of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

ABSTRACT

Each year there will be an estimated 2.1 million new lung cancer cases and 1.8 million lung cancer deaths worldwide. Tobacco smoke is the No. 1 risk factors of lung cancer, accounting for > 85% lung cancer deaths. Air pollution, or haze, comprises ambient air pollution and household air pollution, which are reported to cause 252,000 and 304,000 lung cancer deaths each year, respectively. Tobacco smoke and haze (hereafter, smohaze) contain fine particles originated from insufficient combustion of biomass or coal, have quite similar carcinogens, and cause similar diseases. Smohaze exert hazardous effects on exposed populations, including induction of a large amount of mutations in the genome, alternative splicing of mRNAs, abnormalities in epigenomics, initiation of tumor-promoting chronic inflammation, and facilitating immune escape of transformed cells. Tackling smohaze and development of multi-targets-based preventive and therapeutic approaches targeting smohaze-induced carcinogenesis are the key to conquer lung cancer in the future.

KEYWORDS

Lung cancer; tobacco smoke; air pollution; smohaze; carcinogenesis

Introduction

Lung cancer is the most common cause of cancer death worldwide, with 2.1 million new cases and 1.8 million deaths predicted in 2018\(^1\). Lung cancer is also No. 1 cancer killer in China\(^2\) (Figure 1). Lung cancer is categorized by cell type into small-cell lung cancer (SCLC; accounting for 15% of all lung cancer cases) and non-small-cell lung cancer (NSCLC; 85%), whereas NSCLC can be divided into lung adenocarcinoma (LUAD; 40%), lung squamous cell carcinoma (LUSC; 30%), and large cell lung cancer (LCLC; 15%). The past two decades have witnessed enormous progress in development of targeted and immune therapies, which significantly improve the clinical outcome of the patients\(^3\). Targeted therapies include inhibitors of tyrosine kinases EGFR\(^4\), MET\(^5\), HER2\(^6\), and BRAF V600\(^7\), inhibitors of fusion proteins involved ALK\(^8\), ROSI\(^9\), RET\(^10,11\), and TRK\(^12\), antibodies against VEGF\(^13\), and others. Immune therapies include antibodies against cytotoxic T-cell lymphocyte-4 (CTLA-4)\(^14\), programmed cell death 1 (PD-1) receptor\(^15\), and PD-1 ligand 1 (PD-L1)\(^16\).

However, drug resistance will eventually occur in patients treated with these therapies\(^17\), leading to treatment failure of the patients. Currently, the 5-year overall survival of lung cancer of all stage combined is only 18.6%\(^18\) and 7%\(^19\), respectively. Therefore, new strategies remain urgent needs to conquer lung cancer. It is well-known that a large majority of patients suffering from lung cancer have their roots in...
unhealthy life style and exposure to environmental factors, hence much more attentions should be paid to the causes of lung cancer and related lung carcinogenesis when considering curative approaches for this deadly disease.

**Tobacco smoke: the No. 1 risk factor of lung cancer**

Tobacco smoke is the single biggest public health threat the world is currently facing. It is estimated that there are 1.1 billion smokers worldwide, and around 8 million people are killed by tobacco smoke each year globally. Among these deaths, more than 7 million are the result of direct tobacco use while around 1.2 million are the result of non-smokers being exposed to second-hand smoke\textsuperscript{20-22}. Tobacco use is responsible for approximately 22% of cancer deaths worldwide\textsuperscript{23}. Recent studies show that approximately 15% to 24% of lifetime smokers will get lung cancer\textsuperscript{24}, and tobacco smoking accounts for more than 85% of lung cancer worldwide\textsuperscript{25}.

During smoking, tobacco is heated to 880°C or higher, and more than 8,000 compounds have been identified in tobacco and tobacco smoke. More than 70 carcinogens and more than 20 lung carcinogens have been identified, which fall into group 1 (carcinogenic to humans), group 2A (probably carcinogenic to humans), and group 2B (possibly carcinogenic to humans) of carcinogen classifications of the International Agency for Research on Cancer (IARC) of the World Health Organization (WHO). Group 1 carcinogens found in tobacco smoke include polycyclic aromatic hydrocarbons [PAHs; e.g., benzo (a) pyrene, BaP], N-nitrosamines [e.g., N'-nitrosonornicotine; 4-(methyl nitrosamino)-1-(3-pyridyl)-1-butane, NNK], aromatic amines (e.g., 2-naphthylamine, 4-aminobiphenyl), heterocyclic aromatic amines [e.g., 2-amino-3-methylimidazo (4,5-f) quinoline], aldehydes (e.g., formaldehyde), volatile hydrocarbons (e.g., benzene), miscellaneous organic compounds (e.g., vinyl chloride, ethylene oxide), heavy metals and inorganic compounds [e.g., arsenic, beryllium, nickel, chromium (hexavalent), cadmium, radioisotope polonium-210], and others\textsuperscript{26,27}.

Cigarette smoking was linked to lung cancer in 1964 by the Surgeon General of U.S. Department of Health and Human Service in the first Surgeon General’s Report. This linking has had an enormous positive effect on public health, in that in U.S. male smoking prevalence has decreased from 33.3% to 16.5%\textsuperscript{24}. In China, male and female smoking prevalence has also began to decrease\textsuperscript{27}. Due to reduced tobacco use, lung cancer death rates in U.S. declined 45% from 1990 to 2015 among males and 19% from 2002 to 2015 among females\textsuperscript{1}. However, there will be an estimated 228,150 new lung cancer cases and 142,670 lung cancer deaths in 2019 in U.S.\textsuperscript{28}, and there will be an estimated 2,540,842 new lung cancer cases and 2,145,215 lung cancer deaths worldwide in 2025\textsuperscript{29}. Of note, no therapeutics so far are designed to target tobacco smoke-induced lung carcinogenesis, which remains to be elucidated.

**Air pollution represents the second risk factor of lung cancer**

Clean air is a basic requirement of life. Air pollution, or haze, is a diverse mixture of pollutants that originated from anthropogenic and natural sources, is comprised of particulate matters (PM), gases (e.g., sulfur oxides, carbon monoxide, ozone), organic compounds (e.g., PAHs; resin acids; anhydrous sugars; lignin pyrolysis products; hopanes), metals (e.g., lead, vanadium, and nickel), microbes, and others\textsuperscript{30-32}. The main PM in the environment are PM smaller than 10 μm in diameter (PM\textsubscript{10}) and PM smaller than 2.5 μm in diameter (PM\textsubscript{2.5}), and air pollution can be divided into ambient air pollution (AAP) and household air pollution (HAP), and is a global environmental health risk that affects the populations in developed and developing countries alike, and 91% of the world population is living in places where the WHO air quality guidelines levels (i.e. annual mean values of 20 μg/m\textsuperscript{3} for PM\textsubscript{10} and 10 μg/m\textsuperscript{3} for PM\textsubscript{2.5})\textsuperscript{33} are not met\textsuperscript{34}. WHO also sets guidelines for the protection of public health from risks due to a number of chemicals commonly present in indoor air [i.e. benzene, carbon monoxide, formaldehyde, naphthalene, nitrogen dioxide, polycyclic aromatic hydrocarbons [especially benzo (a) pyrene], radon, trichloroethylene and tetrachloroethylene]\textsuperscript{35}. AAP in both cities and rural areas is estimated to cause 4.2 million premature deaths worldwide each year\textsuperscript{34}, and 3.8 million people a year die prematurely from illness attributable to the HAP caused by the inefficient combustion of solid fuels and kerosene for cooking\textsuperscript{36}. It is estimated that there are 304,000\textsuperscript{37} and 252,000\textsuperscript{34} lung cancer deaths that are attributed to HAP and AAP each year worldwide, respectively.

Xuanwei (XW) City in Yunnan Province of China provides an example of the epidemiological association between PM\textsubscript{10}, PM\textsubscript{2.5} and lung cancer\textsuperscript{38-40}. This city has a large deposit of smoky coal and until the 1970s, local residents used smoky coal in unvented indoor firepits for domestic cooking and heating, all processes that release high concentrations of...
PM$_{10}$ and PM$_{2.5}$. These airborne particles contain high concentrations of PAHs including BaP and polar compounds that are highly mutagenic$^{40}$. Lung cancer incidence in XW is among the highest in China$^{38,40}$, and a reduction in lung cancer morbidity was noted in the 1990s after stove improvement in central XW, supporting the association between air pollution and lung cancer$^{41}$. These findings had been cited in the IARC monograph classifying HAP as “carcinogenic to humans (group 1)”$^{42}$. On the other hand, Raaschou-Nielsen et al.$^{43}$ showed that the risk of lung cancer rises by 18% for every increase of 5 $\mu$g/m$^3$ of PM$_{2.5}$ in the environment, and the risk increases by 22% for every increase of 10 $\mu$g/m$^3$ in PM$_{10}$. AAP has also been classified as group 1 carcinogen in humans by IARC$^{43}$. LUAD, LUSC, SCLC, and bronchiolo-alveolar carcinoma (BAC) represent the most frequently seen lung cancer subtypes in air pollution regions$^{38,44}$.

**Tobacco smoke and haze (smohaze) are responsible for more than 90% of lung cancer**

Tobacco smoke and haze bear many common characteristics in causing lung cancer (Table 1). For example, the fine particles in the two factors are PM$_{2.5}$, which are mainly originated from insufficient combustion of biomass or coal.

**Why smohaze**

Treating the two risk factors as one, smohaze, may result in some positive effects in prevention and treatment of related disease. First, the health hazards of tobacco smoke and haze are currently separately assessed, which may lead to inconsistent results. For example, it is estimated that > 85% (1,530,000) of the 1.8 million lung cancer deaths are

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
\textbf{PM} & \textbf{Origin of pollutants} & \textbf{Chemistry} & \textbf{Main carcinogens*} & \textbf{Exposure methods} & \textbf{Related diseases} & \textbf{Types of lung cancer} \\
\hline
\textbf{Tobacco smoke} & PM$_{2.5}$ & Burning of tobacco & Nicotine, N-Nitrosamines, PAHs, volatile compounds, heavy metals, aromatic amines, heterocyclic amines, etc. & PAHs (BaP), heterocyclic compounds, N-nitrosamines (N'-nitrosonornicotine, NNK), aromatic amines (2-naphthylamine, 4-aminoephenyl), heterocyclic aromatic amines, aldehydes (formaldehyde), phenolic compounds, volatile hydrocarbons, miscellaneous organic compounds (vinyl chloride, ethylene oxide), metals (As, Be, Ni, Cr, Cd, Lineless, etc.), etc. & Inhalation & Cancers, cardiovascular disease, diabetes, COPD, pneumonia, and others$^{22,22}$ & All types of lung cancer, in particular small cell lung and squamous cell carcinoma \\
\hline
\textbf{Haze} & PM$_{2.5}$ & Burning of coal, fossil fuel, biomass, etc. & Organic matter, nitrate, sulphate, ammonium, chloride, heavy metals, elemental carbon, etc. & PAHs (BaP), heterocyclic compounds, N-nitrosamines, aromatic amines, heterocyclic aromatic amines, aldehydes (formaldehyde), phenolic compounds, volatile hydrocarbons, miscellaneous organic compounds (vinyl chloride, ethylene oxide), metals (As, Be, Cr, Cd, Ni, Pb), etc.$^{20-32}$ & Inhalation & Respiratory infections, cancers, cardiovascular disease, COPD, asthma, and others$^{24,36}$ & LUAD, LUSC, SCLC, BAC$^{38,44}$ \\
\hline
\end{tabular}
\caption{Comparison of tobacco smoke and haze}
\end{table}

*For each class, only the group 1 carcinogens are listed in parentheses. As, arsenic; Be, beryllium; Ni, nickel; Pb, plumbum; NNK, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butane; Cr, chromium; Cd, cadmium; Lineless, radioisotope polonium-210.
attributed to tobacco smoke, and other reports show that 252,000 and 304,000 patients may die from AAP- and HAP-caused lung cancer, respectively. Obviously, deaths from these factors exceed the total lung cancer deaths worldwide. Under the concept of smohaze, the more accurate data will be obtained from the related agencies. Second, when considering preventive medicine, the two factors should be tackled simultaneously. Combination of the two would be helpful for the public health domain to develop more effective strategies to tackle the public health problems, in that anti-tobacco campaign and anti-haze efforts could be propagated simultaneously. Third, the emerging concept of smohaze may create additional opportunities for the elucidation of environmental lung carcinogenesis and development of effective preventive and therapeutic approaches to conquer lung cancer.

**Smohaze-induced lung carcinogenesis**

Efforts have been made to dissect tobacco smoke-induced tumorigenesis and air pollution in promoting lung cancer. These works show that smohaze exerts comprehensive effects on humans to trigger and promote lung cancer (Figure 2).

**Genomic mutations**

*Carcinogens and nucleotide substitutions*

As compared with counterpart normal controls, cancer genomes usually have six types of nucleotide changes, $C \rightarrow T/G \rightarrow A$, $C \rightarrow A/G \rightarrow T$, $C \rightarrow G/G \rightarrow C$, $A \rightarrow G/T \rightarrow C$, $A \rightarrow T/T \rightarrow A$, and $A \rightarrow C/T \rightarrow G$. Smohaze carcinogens can cause characteristic mutations (or signature) in the genome. For example, the $G \rightarrow T$ transversions have been described as a mutational fingerprint of tobacco smoke mutagens in smoking-associated lung cancers, and smokers more frequently show $G \rightarrow T$ transversions, whereas never-smokers have more $G \rightarrow A$ transitions$^{45-48}$. PAHs are the main carcinogens that produce $G \rightarrow T$ mutations$^{25,49}$. NNK may induce both $G \rightarrow T$ and $G \rightarrow A$ mutations$^{50}$, and 4-aminobiphenyl and 1,3-butadiene cause the $A \rightarrow T$ transversions$^{51-53}$. Other smohaze carcinogens also induce characteristic nucleotide substitutions in the genome$^{54}$. These mutations were firstly reported in genes such as $TP53$, $RAS$, and others$^{25,49,51-53}$. However, attentions should be paid to the fact that one type of nucleotide changes can be induced by different environmental factors or carcinogens, therefore the so-called “signature” is not stringently specific to one mutagen$^{54,55}$.

**Mutations in specific genes**

Husgafvel-Pursiainen et al.$^{56}$ showed that patients who smoke have a three-fold greater risk of $TP53$ mutations compared to those who do not smoke. Le Calvez et al.$^{57}$ showed that the rate of $TP53$ mutations increased from 47.5% in never-smokers to 77.4% in active smokers, and the risk of having a $TP53$ mutation was significantly proportional to the amount of tobacco consumed. $KRAS$ mutations are much more frequent in smokers, in that in active smokers and never-smokers the $KRAS$ mutation rates were 34% and 5%, respectively$^{58,59}$. $BRAF$ mutations are significantly more frequent in smokers (active or former)$^{60}$. On the contrary,
EGFR mutations and ALK rearrangements are much more frequent in never-smokers compared to active smokers\textsuperscript{58,59,61,62}. Barlesi et al.\textsuperscript{63} further reported significant differences between smokers and never-smokers for mutations in EGFR (4.5% vs. 36.3%), ALK (3.5% vs. 9.7%), KRAS (31.7% vs. 9.6%), BRAF (1.6% vs. 1.8%), and HER2 (0.2% vs. 3.8%), respectively (Table 2).

**Genomic mutations**

Alexandrov et al.\textsuperscript{64} analyzed somatic mutations and DNA methylation in 5,243 cancers of 17 types (including LUAD, LUSC, and SCLC) for which tobacco smoking confers an elevated risk. They showed that smoking is associated with increased mutation burdens of multiple distinct mutational signatures, which contribute to different extents in different cancers. One of these signatures, mainly found in cancers derived from tissues directly exposed to tobacco smoke, is attributable to misreplication of DNA damage caused by tobacco carcinogens. Others likely reflect indirect activation of DNA editing by APOBEC cytidine deaminases and of an endogenous clocklike mutational process. Smoking is associated with limited differences in methylation. Govindon et al.\textsuperscript{65} report the results of whole-genome and transcriptome sequencing of tumor and adjacent normal tissue samples from 17 NSCLCs. They identified 3,726 point mutations and more than 90 small insertions and deletions (indels) in the coding sequence, with an average mutation frequency more than 10-fold higher in smokers than in never-smokers. Deep digital sequencing revealed diverse clonality patterns in both never-smokers and smokers.

Only one work so far has looked at somatic genomic profiles in association with air pollution\textsuperscript{66}. In that work, we analyzed the somatic mutations of 164 NSCLCs from XW and control regions (CR) where smoky coal was not used. Whole genome sequencing revealed a mean of 289 somatic exonic mutations per tumor and the C:G→A:T substitutions in XW NSCLCs. Exome sequencing of 2010 genes showed that XW and CR NSCLCs had a mean of 68 and 22 mutated genes per tumor, respectively (P<0.0001). We found 167 genes (including CACNA1E, KRAS, MYH3, NRXN2, RB1CC1, RTL1, RYR2, TP53, XIRP2) which had significantly higher mutation frequencies in HPR than CR patients (Table 2), and mutations in most genes in XW NSCLCs differed from those in CR cases. The mutation rates of 70 genes (e.g., RYR2, MYH3, GPR144, CACNA1E) were associated with patients' lifetime BaP exposure. This study uncovers the mutation spectrum of air pollution-related lung cancers, and provides evidence for pollution exposure-genomic mutation relationship at a large scale.

Very few studies systematically dissect genomic mutations specifically occurred in “normal” tissues\textsuperscript{68,69} and their roles in lung carcinogenesis. We analyzed the genome of normal lung tissues and paired tumors of patients with LUAD\textsuperscript{70} and LUSC\textsuperscript{71}, and found genomic alterations in the “normal” lung tissues that can be verified by Sanger capillary sequencing. The C:G→T:A transitions, a signature of tobacco carcinogen N-methyl-N-nitro-N-nitrosoguanidine, were the predominant nucleotide changes in counterpart normal controls (CNCs) of LUAD and LUSC. The significance of these variations remains unclear. One possibility was that some of the mutated genes were pro-oncogenes (e.g., MUC4\textsuperscript{72} and CDC27\textsuperscript{73}) or tumor suppressors (e.g., NCOI\textsuperscript{74}), cells harboring these variations were in a “precancerous” stage, and accumulation of other mutations would result in transformation and development of malignant neoplasms. Secondly, many of the CNC altered genes were associated with immune response, DNA-damage response system, or other important signal pathways, which may facilitate avoiding immune destruction and other hallmarks to facilitate lung cancer\textsuperscript{70,71}. These possibilities warrant further investigation.

**Alternative splicing**

Alternative splicing, the process by which splice sites are differentially utilized to produce different mRNA isoforms, contributes to oncogenic activation in several types of

---

**Table 2** Mutation rates of selected genes (10% or more) in NSCLCs associated with smohaze or not

<table>
<thead>
<tr>
<th>Genes</th>
<th>Smohaze</th>
<th>Non-smohaze</th>
<th>P</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALK</td>
<td>3.5</td>
<td>9.7</td>
<td></td>
<td>\textsuperscript{63}</td>
</tr>
<tr>
<td>CACNA1E</td>
<td>19</td>
<td>1.2</td>
<td>0.0001</td>
<td>\textsuperscript{67}</td>
</tr>
<tr>
<td>EGFR</td>
<td>4.5</td>
<td>36</td>
<td>0.135</td>
<td>\textsuperscript{67}</td>
</tr>
<tr>
<td>KRAS</td>
<td>34</td>
<td>5</td>
<td></td>
<td>\textsuperscript{58,59}</td>
</tr>
<tr>
<td>MYH3</td>
<td>15.2</td>
<td>0.0</td>
<td>0.0002</td>
<td>\textsuperscript{67}</td>
</tr>
<tr>
<td>NRXN2</td>
<td>13.9</td>
<td>0.0</td>
<td>0.0004</td>
<td>\textsuperscript{67}</td>
</tr>
<tr>
<td>RB1CC1</td>
<td>13.9</td>
<td>1.2</td>
<td>0.0017</td>
<td>\textsuperscript{67}</td>
</tr>
<tr>
<td>RTL1</td>
<td>13.9</td>
<td>1.2</td>
<td>0.0017</td>
<td>\textsuperscript{67}</td>
</tr>
<tr>
<td>RYR2</td>
<td>29.1</td>
<td>8.2</td>
<td>0.0005</td>
<td>\textsuperscript{67}</td>
</tr>
<tr>
<td>TP53</td>
<td>77</td>
<td>47</td>
<td>&lt; 0.0001</td>
<td>\textsuperscript{57}</td>
</tr>
<tr>
<td>58</td>
<td>40</td>
<td>0.02</td>
<td></td>
<td>\textsuperscript{67}</td>
</tr>
<tr>
<td>XIRP2</td>
<td>19</td>
<td>3.5</td>
<td>0.0016</td>
<td>\textsuperscript{67}</td>
</tr>
</tbody>
</table>
Emerging evidence demonstrates that smohaze can induce alternative splicing of some critical genes to facilitate lung carcinogenesis. Weng et al. showed that of the 117 lung cancer tissue samples analyzed, 31 (26.5%) had splice variants for the MDM2 gene, with 26 samples bearing a splice variant lacking exons 3–11. Significant association was found between the frequency of alternative splicing and the smoking habits of the patients. 44.2% of the smoker patients had alternative MDM2 splice forms versus 16.2% of nonsmokers (P = 0.003). BaP and BPDE induced generation of MDM2 splicing products in H1355 LUAD cells. BPDE-induced MDM2 mRNA alternative splicing in H1355 cells may occur through the PI3K or MAPK pathway. We recently reported a splicing variant of Focal adhesion kinase (FAK), FAK6,7 that contains alternatively spliced exons of 18 bp (Box 6) and 21 bp (Box 7) on either side of codon for Y397 in 4 (4.4%) of 91 patients with NSCLC. Smokers had more FAK abnormalities than non-smokers. In TCGA RNA-seq data, Box 6/7-containing FAK variants were positive in 42 (8.3%) of 508 LUADs and 37 (7.4%) of 501 LUSCs, and current smokers had higher expression of Box 6/7 (+) FAK than reformed and never smokers. FAK6,7 promoted cell proliferation and migration, and exhibited increased autophosphorylation and was more sensitive to FAK inhibitor compared to wild type FAK. The effects of smohaze on mRNA splicing and splicing factors warrant further investigation.

**Less mutated genes that are critical to environmental lung carcinogenesis**

Cancer has been considered as a disease of the genome, and genomic mutations have been shown to be critical to tumorigenesis and served as targets for drug development. Some genes that are usually wild type also play crucial roles in smohaze-induced lung carcinogenesis.

**Aryl hydrocarbon receptor (AhR)**

AhR (Figure 3A) is a member of the basic helix-loop-helix–PER–ARNT–SIM (bHLH–PAS) subgroup of the bHLH superfamily of transcription factors. AhR is an environmental sensor integrating immune responses in health and disease. It can be activated by agonists such as 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD, dioxin) and BaP, and plays a critical role in endogenous ligand kynurenine-promoted and environmental carcinogen-induced tumorigenesis. A constitutively active AhR promotes hepatocarcinogenesis and induces stomach tumors in mice. Shimizu et al. investigated the response of AhR-deficient mice to BaP, and found that BaP induced subcutaneous tumors in AhR+/+ and AhR+/- mice. In contrast, no tumors were apparent in any of the AhR-deficient mice. We recently found that AhR mediated smohaze-induced PD-L1 expression on lung epithelial cells, and deficiency in AhR significantly suppresses BaP-induced lung cancer. AhR inhibitors alpha-naphthoflavone (ANF) and CH223191 exert significant antitumor activity in lung cancer mouse models. These results indicate that AhR is critical to smohaze-induced lung carcinogenesis, and represents an attractive therapeutic target.

**Other genes**

Smohaze may perturb the expression of some genes to...
facilitate lung carcinogenesis. NNK promotes migration and invasion of lung cancer cells through activation of c-Src/PKGII/FAK loop. Oncoprotein cancerous inhibitor of PP2A (CIP2A) was dramatically elevated in tumor samples compared to paratumor normal tissues of patients with NSCLC. CIP2A overexpression was associated with patients' smoking status, and chronic cigarette smoke exposure induced CIP2A expression in mice. Silencing CIP2A inhibited the proliferation and clonogenic activity of lung cancer cells. Smohaze may regulate the expression of some genes in an unexpected way. For example, we conducted a large-scale lethality screening in NSCLC cells to silence all the 1530 transcription factors and 696 ubiquitin pathway genes, and found that transcription factor Iroquois Homeobox 5 (IRX5) and E2 conjugase CDC34 were required for lung cancer cell proliferation. To our surprise, the expression of IRX5 was significantly higher in smoker patients than non-smoker cases, and BaP was able to upregulate IRX5 in lung epithelial cells. Silencing IRX5 significantly inhibited tumor growth in nude mice. We showed that CDC34 bound EGFR and competed with E3 ligase c-Cbl to inhibit the polyubiquitination and subsequent degradation of EGFR. In EGFR-L858R and EGFR-T790M/Del(exon 19)-driven lung cancer in mice, knockdown of CDC34 significantly inhibited tumor formation. CDC34 was elevated in tumor tissues in 67 of 102 (65.7%) NSCLCs, and smokers had much higher CDC34 than nonsmokers. These results indicate that further works should be done to identify critical genes induced or suppressed by smohaze to promote lung carcinogenesis.

**Epigenetic abnormalities**

**DNA methylation**

NNK induces DNA methyltransferase 1 (DNMT1) accumulation and tumor suppressor gene hypermethylation in mice and lung cancer patients. An epigenome wide association study was conducted using peripheral-blood DNA of 464 individuals (22 current smokers and 263 ex-smokers), and the results suggest the existence of dynamic, reversible site-specific methylation changes in response to cigarette smoking, which may contribute to the extended health risks associated with cigarette smoking. A meta-analyses of gene methylation and smoking behavior in NSCLCs showed that 7 hypermethylated genes (including CDKN2A, RASSF1, MGMT, RARB, DAPK, WIFI and FHIT) were significantly associated with the smoking behavior in NSCLC patients. CDKN2A hypermethylation was significantly associated with cigarette smoking in Japanese, Chinese and Americans, whereas RARB hypermethylation was associated with smoking status in Chinese patients. Jiang et al. analyzed genome-wide DNA methylation alterations in XW lung cancers, and obtained a comprehensive dataset of genome-wide CpG island methylation in air pollution-related lung cancers. BaP exposure induced multiple alterations in DNA methylation and in mRNA expressions of DNMTs and ten-11 translocation proteins; these alterations partially occurred in XW lung cancer. BaP-induced DKK2 and EN1 promoter hypermethylation and LPAR2 promoter hypomethylation led to down-regulation and up-regulation of the genes, respectively; the down-regulation of DKK2 and EN1 promoted cellular proliferation. Vitamin C and B6 reduced BaP-induced DNA methylation alterations.

**microRNAs (miRNAs)**

Schembri et al. examined whole-genome miRNA expression in bronchial airway epithelium from current and never smokers (n = 20) and found 28 miRNAs to be differentially expressed with the majority being down-regulated in smokers. These miRNAs contain potential binding sites for the differentially expressed mRNAs in their 3'-untranslated region (UTR). Among them, miR-218 expression was reduced in primary bronchial epithelium exposed to cigarette smoke condensate. Izzotti et al. analyzed the expression of 484 miRNAs in the lungs of rats exposed to environmental cigarette smoke (ECS), and found that ECS down-regulated 126 miRNAs (26.0%) at least 2-fold and 24 miRNAs more than 3-fold. The most remarkably down-regulated miRNAs belonged to the families of let-7, miR-10, miR-26, miR-30, miR-34, miR-99, miR-122, miR-123, miR-124, miR-125, miR-140, miR-145, miR-146, miR-191, miR-192, miR-219, miR-222, and miR-223, while miR-294 was up-regulated. They reported a strong parallelism in dysregulation of rodent microRNAs and their human homologues, which are often transcribed from genes localized in fragile sites deleted in lung cancer. Zhang et al. showed that cigarette smoke upregulates mir-25-3p maturation via N6-methyladenosine and inhibits PHLPP2 to activate AKT, leading to promotion of pancreatic cancer progression. We performed miRNA microarray analysis in NSCLCs from XW or CR, and found 13 down-regulated and 2 up-regulated miRNAs in XW NSCLCs. Among them, miR-144 was one of the most significantly down-regulated miRNAs. The expanded experiments showed that miR-144 was down-regulated in 45/51 (88.2%) XW NSCLCs and 34/54 (63%) CR NSCLCs (P = 0.016). MiR-144 interacted with the oncogene Zeb1 at 2 sites in its 3'-UTR, and a
decrease in miR-144 resulted in increased Zeb1 expression and an epithelial mesenchymal transition phenotype. Ectopic expression of miR-144 suppressed NSCLCs in vitro and in vivo by targeting Zeb1. Smohaze can also perturb the expression of other miRNAs to activate oncogenes, suppress tumor suppressors, promote cell proliferation, angiogenesis, and cell cycle, and inhibit apoptosis, to facilitate carcinogenesis.\(^{98,99}\)

**Long noncoding RNAs (lncRNAs)**

lncRNAs are a group of non-coding RNAs consisting of > 200 nucleotides and having no or low translational potential.\(^{100}\) Cigarette smoke or BaP can induce the expression of several lncRNAs, such as lncRNA–1 (SCAL1), DQ786227, and LOC728228, in cells.\(^{101-104}\) We screened for abnormal lncRNAs in XW lung cancers and reported that XW patients had much more dysregulated lncRNAs than patients from CR. The lncRNA CAR intergenic 10 (CAR10) was up-regulated in 39/62 (62.9%) of the XW patients, which was much higher than in patients from CR (32/86, 37.2%; \(P = 0.002\)). PAH compound dibenz (a, h) anthracene (DbA) up-regulated CAR10 by increasing the expression of transcription factor FoxF2. CAR10 bound and stabilized transcription factor Y-box-binding protein 1 (YB-1), leading to up-regulation of EGFR and proliferation of lung cancer cells. Knockdown of CAR10 inhibited cell growth in vitro and tumor growth in vivo.\(^{105}\) These results demonstrate the role of lncRNAs in environmental lung carcinogenesis.

**Chronic cancer-promoting inflammation**

Tobacco smoke induces pulmonary inflammation, which is believed to play a role in progressive lung destruction in COPD.\(^{106,107}\) Tobacco smoke induces production of cytokines and chemokines, and chronic inflammation may contribute to tumor initiation/promotion through the production of reactive oxygen and nitrogen species that contribute to DNA damage and induction of oncogenic mutations.\(^{106,108,109}\) Takahashi et al.\(^{110}\) demonstrated that repetitive exposure to tobacco smoke promotes tumor development both in carcinogen-treated mice and in transgenic mice undergoing sporadic K-ras activation in lung epithelial cells. Tumor promotion is due to induction of inflammation that results in enhanced pneumocyte proliferation and is abrogated by IKKβ ablation in myeloid cells or inactivation of JNK1.\(^{110}\) We systematically screened for clinically relevant inflammatory factors that are critical for carcinogenesis, and reported that a chemokine CCL20 was significantly up-regulated by NNK. In 78/173 (45.1%) patients the expression of CCL20 was higher in tumor samples than their adjacent normal lung tissues; CCL20 was up-regulated in 48/92 (52.2%) smoker and 29/78 (37.2%) non-smoker patients (\(P = 0.05\)), and high CCL20 was associated with poor prognosis. Anti-inflammation drug dexamethasone inhibited NNK-induced CCL20 production and suppressed lung cancer in vitro and in vivo.\(^{111}\) We screened for abnormal inflammatory factors in NSCLCs from XW and CR, and found that a chemokine CXCL13 was overexpressed in 63/70 (90%) of XW NSCLCs and 44/71 (62%) of smoker and 27/60 (45%) of non-smoker CR patients. CXCL13 overexpression was associated with the XW region and cigarette smoke. The smohaze carcinogen BaP induced AhR-mediated CXCL13 production in lung epithelial cells and in mice prior to development of detectable lung cancer (Figure 3B). Deficiency in Cxcl13 or its receptor, Cxcr5, attenuated BaP-induced lung cancer in mice, demonstrating CXCL13’s critical role in PAH-induced lung carcinogenesis.\(^{112}\) In a nested case-control study (\(n = 526\) lung cancer patients and \(n = 592\) control subjects) measuring serum levels of 77 inflammation markers, CXCL13 and C-reactive protein (CRP), CCL22, and IL-1RA provided good separation in 10-year lung cancer cumulative risks among former smokers and current smokers even after adjustment for smoking.\(^{113}\) Lung inflammation caused by tobacco smoke exposure also converted disseminated, dormant cancer cells to aggressively growing metastases.\(^{114}\)

**Immune escape**

Whether tobacco carcinogens confer the exposed cells immune escape to initiate carcinogenesis, and why smoker patients response better to immunotherapies than non-smokers,\(^{14,15,115}\) remain poorly understood. We reported that cigarette smoke and carcinogen BaP induced PD-L1 expression on lung epithelial cells in vitro and in vivo, which is mediated by AhR (Figure 3B). Anti-PD-L1 antibody or deficiency in AhR significantly suppresses BaP-induced lung cancer. In 37 patients treated with anti-PD-1 antibody pembrolizumab, 13/16 (81.3%) patients who achieve partial response or stable disease express high levels, whereas 12/16 (75%) patients with progression disease exhibit low levels, of AhR in tumor tissues. AhR inhibitors exert significant antitumor activity and synergize with anti-PD-L1 antibody in lung cancer mouse models. These results demonstrate that tobacco smoke induces lung epithelial cells escape from adaptive immunity to promote tumorigenesis, and AhR predicts responses to immunotherapy and represent an attractive therapeutic target. The expression of CTLA-4 was
analyzed in 909 NSCLC patients, and the results showed that CTLA-4 expression was significantly higher in LUSC and current/former smokers. Cigarette smoke also induces lung inflammation and formation of neutrophil extracellular traps, which awaken dormant cancer cells by the neutrophil elastase and matrix metalloproteinase 9. A recent study showed that carcinogenesis in the lung involves a dynamic co-evolution of pre-invasive bronchial cells and the immune response/immune escape through immune checkpoints and suppressive interleukins from high-grade pre-invasive lesions.

The anti-lung cancer strategies

To tame lung cancer, WHO and many countries have launched measures to monitor tobacco use and prevention policies, protect people from tobacco use, offer help to quit tobacco use, warn about the dangers of tobacco, enforce bans on tobacco advertising, promotion and sponsorship, and raise taxes on tobacco. China and other countries have been tackling the health effects of air pollution. Tremendous efforts have been made to develop preventive and therapeutic approaches. However, the cancer incidence is still high and the 5-year overall survival rate of lung cancer remains dismally low.

Rethinking of the current anti-lung cancer strategies may uncover the following limitations for further improvement. First, smohaze-induced lung carcinogenesis remains to be elucidated. Though numerous studies have been performed to dissect environmental tumorigenesis, many of the works were conducted at cellular and animal models, the clinical relevance of the results should be further tested. Second, no druggable target for smohaze-induced lung cancer was identified and no drug was developed to target smohaze-induced lung carcinogenesis. This represents a major limitation of lung cancer studies in the past 5 decades. Third, while great efforts have been made to characterize genomic mutations, those critical genes that rarely mutate should be unveiled. Thousands of lung cancer genomes have been sequenced in the past decade and a large amount of somatic abnormalities have been reported, but only a few mutated genes have been shown to be the “driver” mutations, suggesting that some less mutated genes should play critical roles in smohaze-induced lung carcinogenesis. AhR represents one of this kind of critical genes, and other critical genes should be uncovered by systematic studies. Of course, the less mutated genes usually play important roles in physiological settings, targeting these genes may possibly induce side effects in the patients. Hence, efforts should be made to scrutinize the requirement of the genes by cancerous and normal tissues, and molecules that are more critical to tumor cells than normal cells could be appropriate targets exemplified by proteasome in multiple myeloma. Fourth, the current single target-oriented treatment regimens will eventually fail due to development of drug resistance. Inhibition of EGFR gain-of-function mutations marks a revolution in the history of lung cancer treatment, and while resistance occurred the second, third, and even fourth generations of EGFR inhibitors were developed, though drug resistance may develop again. This single target-based passive strategy could be improved by combinatorial treatment regimens comprised of different components targeting different targets, exemplified by the traditional Chinese medicine formulae that usually contain compounds of “Jun” (Emperor), “Chen” (Minister), “Zuo” (Assistant), and “Shi” (Delivering Servant) to reach synergistic anti-cancer effects.

Perspectives

Up to 90% of cancer cases are caused by environmental factors or lifestyle. The fact that > 90% of lung cancer deaths are caused by smohaze clearly demonstrates the smohaze-induced lung carcinogenesis as the key to develop effective preventive and therapeutic strategies to tame lung cancer, given that the hazardous health effects of smohaze will not be avoidable for billions of people in the near future. The discoveries that deficiency in AhR significantly inhibited BaP-induced lung and skin cancers indicate that inhibition of AhR may exert beneficial effects in prevention and treatment of lung cancer. Moreover, deficiency in CXCL13 or suppression of PD-L1, both of which are AhR target genes, inhibited BaP-induced lung cancer, further confirming the critical roles of AhR in smohaze-induced lung carcinogenesis. With uncovering of more and more critical molecules in smohaze-induced lung carcinogenesis, more targets will be provided to develop more effective preventive and therapeutic approaches. Based on these targets, vaccines and anti-lung cancer formulae can be developed, and early diagnosis and early treatment can be achieved. I believe that these efforts will eventually result in conquering of lung cancer in the future.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Grant No. 2016YFC0905501), the National Natural Science Funds for
Distinguished Young Scholar (Grant No. 81425025), the Key Project of the National Natural Science Foundation of China (Grant No. 81830093), the CAMS Innovation Fund for Medical Sciences (Grant No. CIFMS; 2019-I2M-1-003), and the National Natural Science Foundation of China (Grant No. 81672765).

**Conflict of interest statement**

No potential conflicts of interest are disclosed.

**References**


29. International Agency for Research on Cancer of World Health


50. Ronai ZA, Gradia S, Peterson LA, Hecht SS. G to a transitions and g to t transversions in codon 12 of the K-ras oncogene isolated from mouse lung tumors induced by 4-(methylthio)nitrosamino)-1-(3-pyridyl)-1-butanone (NNK) and related DNA methylating and pyridylloxobutylating agents. Carcinogenesis. 1993; 14: 2419-22.


by enhancing expression of the oncogene CIP2A. Am J Respir Cell Mol Biol. 2018; 59: 695-705.


