

Genome-Wide Analysis of DNA Methylation and Cigarette Smoking in Chinese

Xiaoyan Zhu, Jun Li, Siyun Deng, Kuai Yu, Xuezhen Liu,
Qifei Deng, Huizhen Sun, Xiaomin Zhang, Meian He, Huan Guo,
Weihong Chen, Jing Yuan, Bing Zhang, Dan Kuang,
Xiaosheng He, Yansen Bai, Xu Han, Bing Liu, Xiaoliang Li,
Liangle Yang, Haijing Jiang, Yizhi Zhang, Jie Hu,
Longxian Cheng, Xiaoting Luo, Wenhua Mei, Zhiming Zhou,
Shunchang Sun, Liyun Zhang, Chuanyao Liu, Yanjun Guo,
Zhihong Zhang, Frank B. Hu, Liming Liang, and Tangchun Wu

<http://dx.doi.org/10.1289/ehp.1509834>

Received: 17 February 2015

Accepted: 22 December 2015

Advance Publication: 12 January 2016

Note to readers with disabilities: *EHP* will provide a [508-conformant](#) version of this article upon final publication. If you require a 508-conformant version before then, please contact ehp508@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

Genome-Wide Analysis of DNA Methylation and Cigarette Smoking in Chinese

Xiaoyan Zhu^{1*}, Jun Li^{1,2*}, Siyun Deng¹, Kuai Yu¹, Xuezheng Liu¹, Qifei Deng¹, Huizhen Sun¹, Xiaomin Zhang¹, Meian He¹, Huan Guo¹, Weihong Chen¹, Jing Yuan¹, Bing Zhang¹, Dan Kuang¹, Xiaosheng He¹, Yansen Bai¹, Xu Han¹, Bing Liu¹, Xiaoliang Li¹, Liangle Yang¹, Haijing Jiang¹, Yizhi Zhang¹, Jie Hu¹, Longxian Cheng³, Xiaoting Luo⁴, Wenhua Mei⁴, Zhiming Zhou⁵, Shunchang Sun⁵, Liyun Zhang⁶, Chuanyao Liu¹, Yanjun Guo¹, Zhihong Zhang¹, Frank B. Hu^{2,7,8}, Liming Liang^{2,9}, and Tangchun Wu¹

¹Department of Occupational and Environmental Health and Ministry of Education Key Lab for Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; ²Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; ³Department of Cardiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; ⁴Department of Cardiology, People's Hospital of Zhuhai, Zhuhai, China; ⁵Department of Cardiology, Bao'an Hospital, Shenzhen, China; ⁶Department of Cardiology, Wuhan Central Hospital, Wuhan, China; ⁷Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; ⁸Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; ⁹Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. *These authors contributed equally to this work.

Address correspondence to Tangchun Wu, Department of Occupational and Environmental Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, 13 Hangkong Rd, Wuhan 430030, Hubei, China. Telephone: 86-27-83692347. Fax: 86-27-83692560. E-mail: wut@mails.tjmu.edu.cn

Running title: Smoking related methylation changes in Chinese

Acknowledgments: We would like to thank all individuals for participating in the present study. This study is supported by the National Key Basic Research and Development Program (973 project, grant 2011CB503806), the Natural National Scientific Foundation of China (81230069) and the China Medical Board of New York.

Competing financial interests: The authors have no actual or potential competing financial interests.

Abstract

Background: Smoking is a risk factor for many human diseases. DNA methylation has been related to smoking, but genome-wide methylation data on smoking in Chinese is limited.

Objectives: We aimed to investigate epigenome-wide methylation in relation to smoking in Chinese.

Methods: We measured the methylation levels at >485,000 CpGs of blood leukocytes using HumanMethylation450 BeadChip and conducted a genome-wide meta-analysis of smoking in a total of 596 Chinese participants. For the smoking-related CpGs, we further evaluated their associations with internal polycyclic aromatic hydrocarbons (PAHs) biomarkers and their correlations with the expression of corresponding genes.

Results: We identified 318 CpGs whose methylation levels were associated with smoking at genome-wide significance level (*false discovery rate* < 0.05), among which 161 CpGs annotated to 123 genes were not associated with smoking in recent studies of Europeans and African Americans. Of these smoking-related CpGs, methylation levels at 80 CpGs showed significant correlations with the expression of corresponding genes (including *RUNX3*, *IL6R*, *PTAFR*, *ANKRD11*, *CEP135* and *CDH23*), and methylation at 15 CpGs were significantly associated with urinary 2-hydroxynaphthalene, the most representative internal monohydroxy-PAHs biomarker for smoking.

Conclusion: We identified DNA methylation markers associated with smoking in Chinese populations, including some that also were correlated with gene expression. Exposure to naphthalene, a by-product of tobacco smoke, may contribute to smoking-related methylation.

Introduction

Tobacco kills nearly 6 million people per year on account of direct tobacco use or indirect smoke exposure (WHO 2014). Cigarette smoking, the primary method of tobacco consumption, is a major cause of preventable diseases (including cardiovascular diseases, respiratory diseases, and cancers) (Cunningham et al 2014; Rea et al 2002; Sosnowski and Przewozniak 2014) and mortality (Ezzati and Lopez; Mathers and Loncar 2006). Various human carcinogens have been identified in cigarette smoke, including polycyclic aromatic hydrocarbons (PAHs) (IARC 2004; CDC 2010; Rodgman et al. 2000). Although the adverse health effects of smoking are well acknowledged, less is known about its underlying mechanisms of toxicity, especially at the molecular level.

DNA methylation is an epigenetic modification of the genome that is involved in regulating gene expression and genome stability (Lee and Pausova 2013). Methylation status could be modified by both genetic and environmental factors, and it could integrate the effects of both gene and environment on a phenotype or disease (Feil and Fraga 2011; Schadt 2009). Previous studies using targeted approaches (global methylation and candidate gene methylation) have established potential links between smoking and DNA methylation (Furniss et al. 2008; Philibert et al. 2010; Smith et al. 2007), but it was not until the wide applications of genome-wide methylation technologies that hundreds of smoking-related methylation markers were discovered and their relations with smoking-related diseases were evaluated (Besingi and Johansson 2014; Breitling et al. 2011; Elliott et al. 2014; Harlid et al. 2014; Joubert et al. 2012; Markunas et al. 2014; Shenker et al. 2013; Sun et al. 2013; Zeilinger et al. 2013). However,

previous genome-wide methylation analyses of smoking were conducted in Europeans (Guida et al., 2015; Shenker et al. 2013; Zeilinger et al. 2013) and African Americans (Dogan et al 2014; Philibert et al. 2013; Sun et al. 2013), and populations of mid-income countries such as China, the biggest cigarette producer and customer in the world, have not been evaluated.

To investigate epigenome-wide methylation alterations in relation to cigarette smoking in Chinese population, we measured DNA methylation levels at >485,000 CpGs of peripheral blood leukocytes and conducted a genome-wide meta-analysis of smoking in a total of 596 Chinese participants. For the smoking-related CpGs, we further investigated their correlations with the expression of annotated genes and their associations with urinary monohydroxy-PAHs (OH-PAHs) metabolites.

Methods

Study subjects

In the present study, the genome-wide methylation analysis of smoking was conducted in 596 Chinese participants selected from the Coke Oven Cohort, acute coronary syndrome (ACS) patients from Wuhan and Guangdong, China, and the Wuhan-Zhuhai (WHZH) Cohort (see Supplemental Material, Figure S1, for a flowchart of the study).

The Coke Oven Cohort

A total of 1,628 coke oven workers (COW) were recruited from a coke oven plant in Wuhan, China in 2010 (Li et al. 2012). We included 144 workers in the present study based on the following criteria: 1) donated blood and urine samples; 2) had baseline total urinary OH-PAHs

(ΣOH-PAHs) levels in the high tertile; 3) had worked in the plant for more than 5 years; 4) had no self-reported diseases or discomfort; 5) had no fever or infectious conditions within 2 weeks; 6) did not take prescribed medicine in the past month; and 7) had a body mass index (BMI) of 18.0–30.5. After quality controls of methylation and genotyping data, 137 individuals (abbreviated as COW-1) retained in the present study.

Acute coronary syndrome (ACS) patients

The present study also included 103 clinically confirmed ACS patients from Wuhan, China (recruited in Union Hospital and Wuhan Central Hospital) and 103 ACS patients from Guangdong, China (recruited in Bao'an hospital and Peoples' Hospital of Zhuhai). Patients were 1) diagnosed with acute myocardial infarction or unstable angina pectoris by professional clinicians; 2) were not complicated with congenital heart diseases, cardiomyopathy, autoimmune diseases, acute infections, tuberculosis, chronic obstructive pulmonary diseases, diabetes mellitus, severe kidney or liver diseases, hyperthyroidism, or malignant neoplasms; and 3) donated blood samples at the earliest convenient time of the first day of admission. We included 101 patients from Wuhan (abbreviated as ACS-1) and 97 patients from Guangdong (abbreviated as ACS-2) who passed quality controls of both methylation data and genotyping data in the present analysis.

The Wuhan-Zhuhai (WHZH) Cohort

The WHZH Cohort is a community-based cohort established in 2011 with 4812 individuals (3053 from Wuhan and 1759 from Zhuhai, respectively) recruited at baseline (Song et al. 2014).

From all participants who 1) had no acute or chronic diseases or any kind of discomfort; 2) had no sign of abnormalities in clinical examinations; 3) had no fever or infectious conditions within 2 weeks; 4) did not take prescribed medicine in the past month and 5) donated both blood and urine samples, a total of 180 Wuhan residents were sampled as healthy controls for the ACS patients in Wuhan (matched on age, gender and BMI, n=103) and/or healthy and low PAHs-exposed controls for COWs in Wuhan (matched for age, gender and BMI, and had urinary Σ OH-PAHs at low tertile, n=144; ACS patients and COWs shared 64 controls), whereas 103 Guangdong residents were sampled as healthy controls for ACS patients from Guangdong (matched for age, gender and BMI). We included 162 Wuhan residents and 99 Guangdong residents (abbreviated as WHZH) who passed quality controls of both methylation and genotyping data in the present analysis.

Subjects for investigating methylation-expression correlations

To investigate the correlation between DNA methylation and gene expression, we recruited 144 individuals who participated in the regular health examinations at Health Examination Center of Dongfeng Central Hospital (Dongfeng Motor Corporation and Hubei University of Medicine) in Shiyan, China during April and May in 2015. The selected participants met the following criteria: 1) were 20 to 70 years of age; 2) had no self-reported diseases or discomfort; 3) had no fever or infectious conditions within 2 weeks; 4) took no prescribed medicine in the past month; and 5) donated both blood and urine samples. The methylation and expression data of all the 144 subjects passed quality control criteria and were included in the present analysis (abbreviated as SY).

Our study was approved by the Ethics Committee of Tongji Medical College and informed written consent was obtained from each participant. We require all participants to have a bland diet and fast for at least 12 hours before donating blood samples. Biological samples from all study panels were collected as per the same protocol and stored under similar conditions.

Laboratory Assays

Illumina HumanMethylation450 BeadChip

Genomic DNA of whole blood was extracted using BioTeke Whole Blood DNA Extraction Kit (BioTeke) and then stored at -80°C. One thousand nanogram of each sample was bisulfite converted using the Zymo EZ DNA Methylation kit (Zymo Research) according to the manufacturer's instructions, and was diluted at 60ng/ul. DNA methylation contents at > 485,000 CpGs were assayed using HumanMethylation450 BeadChip (Illumina) with 4 ul bisulfite converted samples.

HumanHT-12 v4 Expression BeadChip

Leukocytes were isolated from whole blood immediately after blood collection and total RNA of blood leukocytes was isolated using TRIZOL LS solution (Invitrogen) as per manufacturer's instructions. Gene expressions were profiled by a commercial company (ETMD, Beijing, China) using HumanHT-12 v4 Expression BeadChip according to standard protocols from Illumina. We acquired raw expression values using GenomeStudio (Illumina) and normalized the expression data using quantile-quantile normalization with "beadarray" package (Dunning et al. 2007) in R 3.1.2 (R Core Team. 2013). All unexpressed signals were assigned as 0 before analysis.

Urinary creatinine and OH-PAHs measurement

The urinary measures of creatinine and 12 OH-PAH metabolites in the WHZH cohort (Song et al. 2014) and the Cohort of Coke Oven Workers (Deng et al. 2014; Li et al. 2012) have been previously reported. All urine samples were collected in sterile conical tubes and were stored at -20°C until laboratory assays. The identification and quantification of PAH metabolites were based on retention time, mass-to-charge ratio and peak area using a linear regression curve obtained from separate internal standard solutions. Among the 12 urinary OH-PAH metabolites, 10 noncarcinogenic metabolites, including 1-hydroxynaphthalene, 2-hydroxynaphthalene, 2-hydroxyfluorene, 9-hydroxyfluorene, 1-hydroxyphenanthrene, 2-hydroxyphenanthrene, 3-hydroxyphenanthrene, 4-hydroxyphenanthrene, 9-hydroxyphenanthrene, and 1-hydroxypyrene were above the limits of quantification (LOQ) and were hence included in the present analysis, whereas the other 2 carcinogenic metabolites, i.e. 6-hydroxy chrysene and 3-hydroxy benzo[a]pyrene, were below the LOQ (Deng et al. 2014; Li et al. 2012) and therefore were not used in the present analysis. The OH-PAHs levels were calibrated by urinary creatinine and presented as micromoles per millimole creatinine.

Quality controls of genome-wide data

We randomized sample pairs of cases (disease or exposed group) and matched controls across different plates and beadchips to minimize batch effects. We used minfi package (Aryee et al. 2014) to preprocess the IDAT files. Signal outliers were identified by multidimensional scaling (MDS) analysis. We examined potential sample mix-ups by matching genotypes of the 65 single nucleotide polymorphisms (SNPs) on the Methylation450k Beadchips with the genotypes of the

same SNPs obtained from the genome wide association study (GWAS) data. Methylation probes were excluded if they: 1) were the 65 SNPs probes; 2) had missing rate $> 20\%$ across samples (missing was defined as for a probe of a certain sample, detection p value > 0.01 or bead counts < 3); or 3) potentially contain or extend on SNPs with MAF > 0.05 in 1000 Genomes Project 20110521 release for ASN population, or cross-hybridized to other genomic locations (41296 probes). Samples were excluded if they: 1) were MDS outliers; 2) were mix-up samples; 3) had missing rate > 0.05 across probes; or 4) failed GWAS quality controls, including unexpected duplicates or relatives (in IBD analysis, $PI_HAT > 0.185$), gender discrepancies, heterozygosity outliers, or individual call rate < 0.98 . After filtering, methylation values at 431,369 CpGs were normalized using dasen method in watermelon package (Pidsley et al. 2013). Methylation values with detection p value > 0.01 or bead counts < 3 were assigned as NA before further analysis.

Statistical analysis

Genome-wide analyses of smoking

Participants who had smoked an average of > 1 cigarette/day over the previous 6 months were defined as current smokers; participants who stopped smoking for > 6 months were defined as former smokers; and participants who never smoked for lifetime were classed as never smokers. Individuals who drank > 1 time/week over the previous 6 months were defined as current drinkers; individuals who stopped drinking for > 6 months were defined as former drinkers; and individuals who never had liquor were defined as never drinkers. Surrogate variable analysis (SVA) was conducted separately in each panel before genome-wide association analyses using

SVA package (Leek et al. 2012, <https://www.bioconductor.org/packages/3.3/bioc/html/sva.html>). Variables used in the SVA included smoking status (coded as 0, 1, 2 for never, former and current smokers, respectively), age (years; as a continuous variable), gender (coded as 1 and 2 for male and female, respectively), drinking status (coded as 0, 1, 2 for never, former and current drinkers, respectively), and BMI (kg/m^2 ; as a continuous variable). Surrogate variables (SVs) could capture major unknown variations of the genome-wide data that cannot be explained by included variables. Association analyses were performed separately in each panel using linear regression models, with inverse-normal transformed (INT) methylation beta-values included as dependent variables, and smoking status, age, gender, drinking status, BMI, and SVs included as independent variables. In the analyses of COW and WHZH, $\Sigma\text{OH-PAHs}$ were also included in the models as covariates since $\Sigma\text{OH-PAHs}$ were considered in sample selection in the two panels. Results from all the 4 panels were combined using a fixed effect meta-analysis, with a sample-size weighted method to get p values and an inverse-variance weighted method to obtain estimates of effect size. The significance threshold for the genome-wide meta-analysis was *false discovery rate* (FDR) < 0.05 . The analyses were implemented in R 3.1.2.

Correlation between CpGs and gene expression

CpGs and expression probes were paired based on annotation files provided by Illumina, which provide information on genomic locations and gene annotations for both expression probes and CpGs probes. Linear regressions, of which dependent variables were inverse-normal transformed expression values and independent variables were methylation values, age, and gender, were used to estimate associations between methylation and expression. For each CpG,

the significance threshold was defined as 0.05/number of expression probes of the corresponding gene.

Urinary PAHs metabolites and smoking-related methylation alterations

We evaluated which urinary OH-PAHs could be used as representative biomarkers of smoking exposure by calculating the contribution of smoking to each OH-PAHs metabolite (defined by the difference of R^2 between the models with and without smoking status; other covariates were age, drinking status, BMI, occupation, geographical regions and beadchip operation date; geographical regions were coded as 1 and 2 for Wuhan and Guangdong, respectively) using linear regression models in males from the WHZH cohort. The association between methylation values of the smoking-related CpGs and urinary 2-hydroxynaphthalene levels were separately analyzed in males from the WHZH cohort and the Coke Oven Cohort. Mediation analysis was performed to evaluate if 2-hydroxynaphthalene showed mediation effect of smoking on methylation alterations in males from the WHZH cohort with adjustment for age, drinking status, BMI, occupation, differential leukocyte proportions, geographical regions and beadchip operation date (Valeri and Vanderweele 2013). The association analyses were conducted in R 3.1.2 and the mediation analyses were performed in SAS 9.2.

Results

Basic characteristics of the subjects

The genome-wide meta-analysis contained a total of 596 participants recruited from China, including 137 coke oven workers (107 males; mean age = 46.51), 198 ACS patients (including

101 from Wuhan with 81 males and a mean age of 58.96 years; and 97 from Guangdong with 78 males and a mean age of 59.37 years), and 261 community residents from the WHZH cohort (206 males, mean age = 53.84). The characteristics of the study populations are summarized in Table 1.

Genome-wide analysis of DNA methylation and smoking

In our genome-wide methylation meta-analysis, we identified 318 CpGs whose methylation levels were associated with smoking at genome-wide significance level ($FDR < 0.05$, Figure 1). Of these, 161 CpGs annotated to 123 genes were not reported to be significantly associated with smoking in previous genome-wide studies of methylation and smoking in Europeans (Guida et al. 2015; Shenker et al. 2013; Zeilinger et al. 2013) or in African Americans (Dogan et al 2014; Philibert et al. 2013; Sun et al. 2013) (Supplemental Material, Table S1). The association results for the top 40 smoking-related CpGs ($FDR < 0.01$) were presented in Table 2, and the association results of the 318 smoking-related CpGs in each panel were presented in Supplemental Material, Table S2. For most of the 318 CpGs, we also observed a gradational alteration trend of the methylation levels from never, former to current smokers; the methylation alterations from current smokers to nonsmokers were larger than alterations from former smokers to nonsmokers (see Supplemental Material, Table S3).

Correlations with the expression of annotated genes

We further investigated whether the methylation values of the smoking-related CpGs were correlated with the expression of corresponding genes in an independent set of 144 healthy

individuals whose methylome and gene-expression profiles were both measured (Table 1). Seventy-seven of the 318 smoking-related CpGs were excluded from the analysis either because no expression probes were designed for the genes or because of the low expression rate in blood leukocytes. Of the remaining 241 CpGs (a total of 414 CpG-expression probe pairs) that had qualified expression data for the annotated genes, we observed that methylation levels at 80 CpGs were associated with the expression of their corresponding genes ($p < 0.05/\text{number of expression probes of the corresponding gene}$; e.g. on the body of *RUNX3*, $p = 1.57 \times 10^{-7}$ for cg10951873 and ILMN_1787461; on the body of *IL6R*, $p = 1.98 \times 10^{-9}$ for cg09257526 and ILMN_1696394, $p = 5.61 \times 10^{-6}$ for cg09257526 and ILMN_1754753; within 1500 bps from transcription start site of *CEP135*, $p = 1.82 \times 10^{-2}$ for cg26542660 and ILMN_1693766; on the body of *CDH23*, $p = 9.45 \times 10^{-3}$ for cg10750182 and ILMN_1779934; within 1500 bps from transcription start site of *PTAFR*, $p = 2.07 \times 10^{-16}$ for cg20460771 and ILMN_1746836; in the 5' untranslated regions of *ANKRD11*, $p = 1.03 \times 10^{-8}$ for cg01107178 and ILMN_2108709) (see Supplemental Material, Table S4).

Associations of smoking-related CpGs and urinary 2-hydroxynaphthalene

Considering that the main smokers in our study were males (98.52%) and to avoid effects due to occupational exposures, the analysis were mainly conducted in males from the WHZH Cohort. We first tested which OH-PAHs metabolite was the most representative biomarker for smoking. We observed that smoking could account for 18.0% of the variation of urinary 2-hydroxynaphthalene, larger than the variations it explained for the other 9 OH-PAHs metabolites (see Supplemental Material, Table S5).

We then assessed the association between methylation levels at the 318 smoking-related CpGs and urinary 2-hydroxynaphthalene levels (Supplemental Material, Table S6), and found 15 significant associations after Bonferroni corrections ($p < 1.57 \times 10^{-4}$) (Figure 2). When restricting the analysis only in nonsmokers, these associations were greatly attenuated (see Supplemental Material, Table S6 and Figure 2), suggesting that the correlations between DNA methylation and urinary 2-hydroxynaphthalene were mainly attributed to smoking. We further investigated whether the 2-hydroxynaphthalene could be a mediator of these smoking-induced methylation alterations, and found that among the 15 CpGs associated with 2-hydroxynaphthalene, the smoking-related methylation variation at 12 CpGs (including cg05575921, cg23916896, cg24090911 and cg26703534 on *AHRR*) might be partially mediated by their associations with urinary 2-hydroxynaphthalene levels ($p < 0.05$) (Table 3).

Even though subjects from the Coke Oven Cohort had occupational PAHs exposures, similar association patterns between smoking, 2-hydroxynaphthalene and the methylation at these CpGs were observed in male subjects from the Coke Oven Cohort after adjustment for 1-hydroxypyrene, an occupational exposure marker for coke oven workers (see Supplemental Material, Table S6 and Figure 2).

Discussion

In the present study, we identified 318 smoking-related CpGs through a genome-wide meta-analysis of DNA methylation in several Chinese populations. Among the identified CpGs, 161 annotated to 123 genes were not associated with smoking in recent studies of Europeans (Guida et al. 2015; Shenker et al. 2013; Zeilinger et al. 2013) or African Americans (Dogan et al

2014; Philibert et al. 2013; Sun et al. 2013). We also observed that methylation levels at some smoking-related CpGs might affect the expression of corresponding genes, and some smoking-related methylation alterations might be partly mediated by exposure to naphthalene.

Although China is the largest consumer and producer of tobacco in the world (Gu et al. 2009), genome-wide methylation studies of smoking have not been conducted in Chinese populations. The present study identified 318 smoking-related CpGs in Chinese; 157 of them have been reported by the previous methylation studies, suggesting that smoking-related methylation alterations were mainly consistent across Chinese and Western populations; the other 161 CpGs that have not been prior reported in Europeans or African Americans might suggest novel smoking-related sites or sites specific to Chinese, which calls for replication by further studies among Chinese populations. Most of the identified loci were annotated on genes that are involved in metabolism of smoking-released chemicals (e.g. *AHRR* is a repressor of nuclear receptor for aryl hydrocarbons that is involved in xenobiotic metabolism (Shenker et al. 2013)) or might be involved in smoking-related health effects (e.g. methylation of *F2RL3* mediated the detrimental impact of smoking and was related to the mortality of coronary heart disease (Breitling et al. 2012; Zhang et al. 2014)).

DNA methylation might be a potential link between smoking and human diseases. In the present study, the smoking-related methylation changes on *RUNX3*, *IL6R*, *PTAFR*, *ANKRD11* (cardiovascular related genes) and *CEP135*, *CDH23* (cancer related genes) corresponded to increased gene expression levels. *RUNX3* encodes a member of the runt domain-containing family of transcription factors, which might play important functions in innate and adaptive

immune cell types and might be associated with several inflammatory-related diseases (Lotem et al. 2015). Interleukin 6 is a cytokine with vital roles in inflammatory responses, and its dysregulation has been implicated in many health problems (Ferreira et al. 2013). *PTAFR* encodes a receptor for platelet-activating factor (PAF) that plays a significant role in pro-inflammatory processes (Ninio et al. 2004). Besides its critical role in hemostasis and thrombosis, platelets are also involved in regulating inflammatory and immune responses (Von Hundelshausen and Weber 2007). *ANKRD11* might be involved in apoptosis pathways (e.g. p53 signaling) (Lim et al. 2012; Neilsen et al. 2008), which have been reported to play key roles in the pathogenesis of cardiovascular diseases (Lee and Gustafsson 2009) It has been speculated that smoking-induced abnormal physiological processes might be important mechanisms in the development of cardiovascular diseases (Frostegard 2013). Our findings that smoking was associated with methylation of cardiovascular related genes which were correlated with the corresponding expression levels suggested that, DNA methylation might contribute to the disease progression through immune reactions, inflammation responses and apoptosis induced by smoking.

CEP135 encodes a centrosomal protein that acts as a scaffolding protein during early centriole biogenesis (Kim et al. 2008). Centrosomes play crucial roles in many processes (including organizing mitotic spindle poles) and centrosome aberrations (Nigg 2002) were included in many human tumors (Rusan and Peifer 2007). Notably, antimetabolic compounds have been identified in tobacco-smoke condensate, and smoking could induce mitotic abnormalities (Qiao et al. 2003; Vogt 2004). *CDH23* encodes cadherin 23 that acts as a mediator in

intercellular junctions, cellular differentiation and cell migration (Agarwal 2014). Previous studies demonstrated that *CDH23* was up-regulated in breast cancer tissues and was involved in metastatic processes (Binai et al. 2013). Recent evidence suggested that active smoking played a potentially casual role for breast cancer (Reynolds et al. 2009). Smoking is also a well-established cause for many cancers (e.g. cancers of lung, colon, and stomach) (Gandini et al. 2008). Therefore, it is possible that methylation alterations might be potential mechanisms of smoking-induced adverse effects and cancers.

Cigarette smoking is one of the major sources of PAHs exposures, especially for naphthalene (Ding et al. 2005; Jacob et al. 2013). We estimated that cigarette smoking accounted for 18.0% of the variation in urinary 2-hydroxynaphthalene among males in the WHZH cohort, which supports further investigation of urinary 2-hydroxynaphthalene as a possible biomarker of internal exposure to smoking-sourced PAHs. Smoking-related methylation alterations of *AHRR* might be caused by the exposure to PAHs (Shenker et al. 2013). *AHRR* encodes a repressor of the aryl hydrocarbon receptor (AhR) (Harlid et al. 2014). Previous studies suggested that the AhR pathway was important in the metabolism of various xenobiotics including PAHs (Zeilinger et al. 2013) and was modified in responses to exposure to smoking (Besingi and Johansson 2014). Our present data suggested that smoking-released naphthalene might alter the AhR pathway by changing the methylation levels of the vital genes in the AhR pathway.

Different cells or tissues have distinct DNA methylation signatures (Ohgane et al. 2008). Peripheral blood DNA is reasonable for the present study for two reasons. First, peripheral blood

was an important carrier for many xenobiotics absorbed into human bodies (Barr et al. 2007); peripheral blood cells contact directly with the internal forms of the xenobiotics and react to them (Bonassi et al. 2007). Second, blood samples are the most convenient in collection in large-scale studies and using blood cells can keep our results comparable to other studies. A limitation of blood leukocytes as the source of DNA for methylation analyses is that methylation varies among leukocyte subtypes, and the distribution of leukocyte subtypes may vary in association with exposure, thus resulting in potential confounding of associations between exposures and methylation (Reinius et al. 2012). As suggested by a previous study that factor-based 'batch' correction methodology like surrogate variable analysis is not only able to control for batch effects but can also empirically estimate and control for cell-type compositions (Jaffe and Irizarry 2014), we adopted surrogate variables in our genome-wide methylation analyses to limit effects of batch and cellular compositions simultaneously. When investigating associations between the smoking-related CpGs and urinary 2-hydroxynaphthalene, we adjusted for differential white blood cell proportions in the analysis models. However, we cannot rule out the potential for residual confounding related to leukocyte subtypes variations or other factors. In addition, we could not establish the temporal relation between smoking and methylation given the cross-sectional study design.

Conclusions

On the basis of a genome-wide methylation analysis of smoking in Chinese, we identified 318 smoking-related CpGs, among which 161 CpGs annotated to 123 genes have not been previously reported in Europeans and African Americans. Some smoking-related CpGs might

have a role in gene regulation. We also found that naphthalene might be one of the smoking-released chemicals inducing the methylation alterations we observed for smoking. Additional studies are needed to replicate our findings and determine their potential relevance to health outcomes, and to elucidate underlying mechanisms linking smoking and methylation.

References

- Agarwal SK. 2014. Integrins and cadherins as therapeutic targets in fibrosis. *Front Pharmacol* 5:131; doi: 10.3389/fphar.2014.00131 [Online 3 June 2014]
- Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. 2014. Minfi: A flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30:1363-1369.
- Barr DB, Bishop A, Needham LL. 2007. Concentrations of xenobiotic chemicals in the maternal-fetal unit. *Reprod Toxicol* 23:260-266.
- Besingi W, Johansson A. 2014. Smoke-related DNA methylation changes in the etiology of human disease. *Hum Mol Genet* 23:2290-2297.
- Binai NA, Carra G, Lower J, Lower R, Wessler S. 2013. Differential gene expression in ER α -positive and ER α -negative breast cancer cells upon leptin stimulation. *Endocrine* 44:496-503.
- Bonassi S, Znaor A, Ceppi M, Lando C, Chang WP, Holland N, et al. 2007. An increased micronucleus frequency in peripheral blood lymphocytes predicts the risk of cancer in humans. *Carcinogenesis* 28:625-631.
- Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. 2011. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet* 88:450-457.
- Breitling LP, Salzman K, Rothenbacher D, Burwinkel B, Brenner H. 2012. Smoking, *F2RL3* methylation, and prognosis in stable coronary heart disease. *Eur Heart J* 33:2841-2848.
- CDC (Centers for Disease Control and Prevention). 2010. 2010 surgeon general's report-how tobacco smoke causes disease: The biology and behavioral basis for smoking-attributable disease. Atlanta, GA: Centers for Disease Control and Prevention. Available: http://www.cdc.gov/tobacco/data_statistics/sgr/2010/ [accessed 13 September 2014].

- Cunningham TJ, Ford ES, Rolle IV, Wheaton AG, Croft JB. 2014. Associations of self-reported cigarette smoking with chronic obstructive pulmonary disease and co-morbid chronic conditions in the United States. *COPD*; doi:10.3109/15412555.2014.949001 [Online 10 September 2014].
- Deng Q, Huang S, Zhang X, Zhang W, Feng J, Wang T, et al. 2014. Plasma microRNA expression and micronuclei frequency in workers exposed to polycyclic aromatic hydrocarbons. *Environ Health Perspect* 122:719-725.
- Ding YS, Trommel JS, Yan XJ, Ashley D, Watson CH. 2005. Determination of 14 polycyclic aromatic hydrocarbons in mainstream smoke from domestic cigarettes. *Environ Sci Technol* 39:471-478.
- Dogan MV, Shields B, Cutrona C, Gao L, Gibbons FX, Simons R, et al. 2014. The effect of smoking on DNA methylation of peripheral blood mononuclear cells from african american women. *BMC Genomics* 15:151; doi: 10.1186/1471-2164-15-151 [Online 3 February 2014].
- Dunning MJ, Smith ML, Ritchie ME, Tavaré S. 2007. Beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* 23:2183-2184.
- Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, et al. 2014. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenetics* 6:4; doi:10.1186/1868-7083-6-4 [Online 3 February 2014].
- Ezzati M, Lopez AD. 2003. Estimates of global mortality attributable to smoking in 2000. *Lancet* 362:847-852.
- Feil R, Fraga MF. 2011. Epigenetics and the environment: Emerging patterns and implications. *Nat Rev Genet* 13:97-109.
- Ferreira RC, Freitag DF, Cutler AJ, Howson JM, Rainbow DB, Smyth DJ, et al. 2013. Functional *IL6R* 358Ala allele impairs classical IL-6 receptor signaling and influences risk of diverse inflammatory diseases. *PLoS Genet* 9:e1003444; doi:10.1371/journal.pgen.1003444 [Online 4 April 2013].
- Frostegard J. 2013. Immunity, atherosclerosis and cardiovascular disease. *BMC Med* 11:117.

- Furniss CS, Marsit CJ, Houseman EA, Eddy K, Kelsey KT. 2008. Line region hypomethylation is associated with lifestyle and differs by human papillomavirus status in head and neck squamous cell carcinomas. *Cancer Epidemiol Biomarkers Prev* 17:966-971.
- Gandini S, Botteri E, Iodice S, Boniol M, Lowenfels AB, Maisonneuve P, et al. 2008. Tobacco smoking and cancer: a meta-analysis. *Int J Cancer* 122:155-164.
- Gu D, Kelly TN, Wu X, Chen J, Samet JM, Huang JF, et al. 2009. Mortality attributable to smoking in China. *N Engl J Med* 360:150-159.
- Guida F, Sandanger TM, Castagne R, Campanella G, Polidoro S, Palli D, et al. 2015. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum Mol Genet* 24:2349-2359.
- Harlid S, Xu Z, Panduri V, Sandler DP, Taylor JA. 2014. CpG sites associated with cigarette smoking: Analysis of epigenome-wide data from the Sister Study. *Environ Health Perspect* 122:673-678.
- IARC (International Agency for Research on Cancer). 2004. Tobacco smoke and involuntary smoking. *IARC Monogr Eval Carcinog Risks Hum.* 83:1-1438.
- Jacob P, Abu Raddaha AH, Dempsey D, Havel C, Peng M, Yu L, et al. 2013. Comparison of nicotine and carcinogen exposure with water pipe and cigarette smoking. *Cancer Epidemiol Biomarkers Prev* 22:765-772.
- Jaffe AE, Irizarry RA. 2014. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 15:R31; doi:10.1186/gb-2014-15-2-r31 [Online 4 February 2014].
- Joubert BR, Haberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 2012. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect* 120:1425-1431.
- Kim K, Lee S, Chang J, Rhee K. 2008. A novel function of CEP135 as a platform protein of C-NAP1 for its centriolar localization. *Exp Cell Res* 314:3692-3700.
- Lee KW, Pausova Z. 2013. Cigarette smoking and DNA methylation. *Front Genet* 4:132; doi:10.3389/fgene.2013.00132 [Online 17 July 2013].

- Lee Y, Gustafsson AB. 2009. Role of apoptosis in cardiovascular disease. *Apoptosis* 14:536-548.
- Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28:882-883.
- Li X, Feng Y, Deng H, Zhang W, Kuang D, Deng Q, et al. 2012. The dose-response decrease in heart rate variability: Any association with the metabolites of polycyclic aromatic hydrocarbons in coke oven workers? *PLoS One* 7:e44562; doi:10.1371/journal.pone.0044562 [Online 14 September 2012].
- Lim SP, Wong NC, Suetani RJ, Ho K, Ng JL, Neilsen PM, et al. 2012. Specific-site methylation of tumour suppressor *ANKRD11* in breast cancer. *Eur J Cancer* 48:3300-3309.
- Lotem J, Levanon D, Negreanu V, Bauer O, Hantisteanu S, Dicken J, et al. 2015. Runx3 at the interface of immunity, inflammation and cancer. *Biochim Biophys Acta* 1855:131-143.
- Markunas CA, Xu Z, Harlid S, Wade PA, Lie RT, Taylor JA, et al. 2014. Identification of DNA methylation changes in newborns related to maternal smoking during pregnancy. *Environ Health Perspect* 122:1147-1153.
- Mathers CD, Loncar D. 2006. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 3:e442; doi:10.1371/journal.pmed.0030442 [Online 28 November 2006].
- Neilsen PM, Cheney KM, Li CW, Chen JD, Cawrse JE, Schulz RB, et al. 2008. Identification of ANKRD11 as a p53 coactivator. *J Cell Sci* 121:3541-3552.
- Nigg EA. 2002. Centrosome aberrations: cause or consequence of cancer progression? *Nat Rev Cancer* 2:815-825.
- Ninio E, Tregouet D, Carrier JL, Stengel D, Bickel C, Perret C, et al. 2004. Platelet-activating factor-acetylhydrolase and PAF-receptor gene haplotypes in relation to future cardiovascular event in patients with coronary artery disease. *Hum Mol Genet* 13:1341-1351.
- Ohgane J, Yagi S, Shiota K. 2008. Epigenetics: The DNA methylation profile of tissue-dependent and differentially methylated regions in cells. *Placenta* 29 Suppl A:S29-35.

- Philibert RA, Beach SR, Gunter TD, Brody GH, Madan A, Gerrard M. 2010. The effect of smoking on maoa promoter methylation in DNA prepared from lymphoblasts and whole blood. *Am J Med Genet B Neuropsychiatr Genet* 153B:619-628.
- Philibert RA, Beach SR, Lei MK, Brody GH. 2013. Changes in DNA methylation at the aryl hydrocarbon receptor repressor may be a new biomarker for smoking. *Clin Epigenetics* 5:19; doi: 10.1186/1868-7083-5-19. [Online 10 February 2014].
- Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. 2013. A data-driven approach to preprocessing Illumina 450k methylation array data. *BMC Genomics* 14:293; doi:10.1186/1471-2164-14-293 [Online 1 May 2013].
- Qiao D, Seidler FJ, Violin JD, Slotkin TA. 2003. Nicotine is a developmental neurotoxicant and neuroprotectant: stage-selective inhibition of DNA synthesis coincident with shielding from effects of chlorpyrifos. *Brain Res Dev Brain Res* 147:183-190.
- R Core Team. 2013. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available: <http://www.R-project.org/>. [accessed 5 July 2014]
- Rea TD, Heckbert SR, Kaplan RC, Smith NL, Lemaitre RN, Psaty BM. 2002. Smoking status and risk for recurrent coronary events after myocardial infarction. *Ann Intern Med* 137:494-500.
- Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, et al. 2012. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One* 7:e41361; doi:10.1371/journal.pone.0041361 [Online 25 July 2012].
- Reynolds P, Goldberg D, Hurley S, Nelson DO, Largent J, Henderson KD, et al. 2009. Passive smoking and risk of breast cancer in the California teachers study. *Cancer Epidemiol Biomarkers Prev* 18:3389-3398.
- Rodgman A, Smith CJ, Perfetti TA. 2000. The composition of cigarette smoke: A retrospective, with emphasis on polycyclic components. *Hum Exp Toxicol* 19:573-595.
- Rusan NM, Peifer M. 2007. A role for a novel centrosome cycle in asymmetric cell division. *J Cell Biol* 177:13-20.

- Schadt EE. 2009. Molecular networks as sensors and drivers of common human diseases. *Nature* 461:218-223.
- Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. 2013. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum Mol Genet* 22:843-851.
- Smith IM, Mydlarz WK, Mithani SK, Califano JA. 2007. DNA global hypomethylation in squamous cell head and neck cancer associated with smoking, alcohol consumption and stage. *Int J Cancer* 121:1724-1728.
- Song Y, Hou J, Huang X, Zhang X, Tan A, Rong Y, et al. 2014. The Wuhan-Zhuhai (WHZH) cohort study of environmental air particulate matter and the pathogenesis of cardiopulmonary diseases: study design, methods and baseline characteristics of the cohort. *BMC Public Health* 14:994; doi:10.1186/1471-2458-14-994 [Online 24 September 2014].
- Sosnowski R, Przewozniak K. 2014. The role of the urologist in smoking cessation: Why is it important? *Urol Oncol* 33:30-39.
- Sun YV, Smith AK, Conneely KN, Chang Q, Li W, Lazarus A, et al. 2013. Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans. *Hum Genet* 132:1027-1037.
- Valeri L, Vanderweele TJ. 2013. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with sas and spss macros. *Psychol Methods* 18:137-150.
- Vogt IC. 2004. Maternal smoking, intrauterine growth restriction, and placental apoptosis. *Pediatr Dev Pathol* 7:433-442.
- Von Hundelshausen P, Weber C. 2007. Platelets as immune cells: bridging inflammation and cardiovascular disease. *Circ Res* 100:27-40.
- WHO (World Health Organization). 2014. Tobacco. Available: <http://www.who.int/mediacentre/factsheets/fs339/en/> [accessed 26 August 2014].

Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. 2013. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One* 8:e63812; doi:10.1371/journal.pone.0063812 [Online 17 May 2013].

Zhang Y, Yang R, Burwinkel B, Breitling LP, Holleczeck B, Schottker B, et al. 2014. *F2RL3* methylation in blood DNA is a strong predictor of mortality. *Int J Epidemiol* 43:1215-1225.

Table 1. Characteristics of the study participants (*n* or mean \pm SD).

Characteristics	Genome-wide meta-analysis of methylation and smoking				Dataset for methylation-expression association analysis
	COW-1 (<i>n</i> = 137)	ACS-1 (<i>n</i> = 101)	ACS-2 (<i>n</i> = 97)	WHZH (<i>n</i> = 261)	SY (<i>n</i> = 144)
Age, years	46.51 \pm 8.91	58.96 \pm 10.20	59.37 \pm 11.47	53.84 \pm 13.06	41.31 \pm 10.27
Male, N (%)	107 (78.1%)	81 (80.2%)	78 (80.4%)	206 (78.9%)	107 (74.3%)
Smoking status, current/ex/no	84/3/50	40/23/38	40/13/44	109/25/127	45/2/97
Drinking status, current/ex/no	49/4/84	23/0/78	20/0/77	81/5/175	55/1/88
Body mass index, kg/m ²	23.47 \pm 2.72	24.84 \pm 2.80	22.97 \pm 2.43	23.37 \pm 2.76	24.22 \pm 2.67
White blood count, 10 ⁹ /L	6.71 \pm 1.51	7.31 \pm 2.44	10.52 \pm 3.82	6.00 \pm 1.57	5.97 \pm 1.40
Neutrophil proportions, %	59.62 \pm 7.26	63.38 \pm 10.50	72.50 \pm 13.04	55.83 \pm 8.33	57.97 \pm 8.18
Lymphocyte proportions, %	36.83 \pm 7.11	27.63 \pm 9.73	19.07 \pm 11.46	37.38 \pm 8.17	33.68 \pm 7.52
Intermediate cell ^a proportions, %	3.55 \pm 1.16	8.97 \pm 3.08	8.44 \pm 4.07	6.85 \pm 3.75	8.35 \pm 2.71

Abbreviations: COW-1, participants from the Coke Oven Cohort; ACS-1, ACS patients from Wuhan; ACS-2, ACS patients from Guangdong; WHZH, residents selected from the WHZH Cohort; SY, individuals who attended regular health examinations in Shiyan.

^a Intermediate cells were defined as the sum of monocytes, eosinophils and basophils.

Table 2. The 40 CpGs associated with cigarette smoking in the genome-wide meta-analysis ($FDR < 0.01$).

Chr	Position	Gene	Relation to gene	CpG	Effect (s.e.) ^a	<i>p</i>	<i>FDR</i>
1	11908164	<i>NPPA</i>	TSS1500	cg05396397	0.257 (0.047)	1.02E-07	6.97E-04
1	19717337	<i>CAPZB</i>	Body	cg07573717	-0.104 (0.020)	1.00E-07	6.97E-04
1	21617442	<i>ECE1</i>	TSS1500	cg26348226	-0.143 (0.030)	2.18E-06	0.007
1	42367407	<i>HIVEP3</i>	5'UTR	cg14663208	0.193 (0.035)	1.63E-07	9.99E-04
1	154299179	<i>ATP8B2</i>	TSS1500	cg06811467	-0.126 (0.021)	1.29E-08	1.22E-04
1	154379696	<i>IL6R</i>	Body	cg09257526	-0.112 (0.020)	5.59E-08	4.16E-04
2	11969958	-		cg02560388	-0.180 (0.037)	1.60E-06	0.006
2	176987918	<i>HOXD9</i>	1stExon	cg22674699	0.216 (0.043)	7.15E-07	0.003
2	231790037	<i>GPR55</i>	TSS200	cg16382047	-0.128 (0.026)	1.31E-06	0.005
3	99792561	<i>C3orf26</i>	Body	cg15554421	-0.159 (0.031)	5.94E-07	0.003
4	56813860	<i>CEP135</i>	TSS1500	cg26542660	-0.150 (0.026)	1.57E-08	1.41E-04
4	95679705	<i>BMPR1B</i>	5'UTR	cg09156233	0.198 (0.042)	3.15E-06	0.010
5	146614298	<i>STK32A</i>	TSS1500	cg09088988	0.177 (0.035)	5.57E-07	0.003
6	46702983	<i>PLA2G7</i>	1stExon	cg18630040	0.196 (0.038)	6.68E-07	0.003
7	1102177	<i>C7orf50</i>	Body	cg15693483	-0.161 (0.027)	3.33E-09	3.79E-05
7	147065665	<i>MIR54814</i>	Body	cg15700587	0.181 (0.034)	1.81E-07	0.001
7	158937969	<i>VIPR2</i>	TSS1500	cg23572908	0.239 (0.048)	1.16E-06	0.004
9	127054428	<i>NEK6</i>	TSS1500	cg14556677	-0.111 (0.019)	7.99E-09	8.02E-05
10	49892930	<i>WDFY4</i>	TSS1500	cg15164194	-0.091 (0.020)	1.81E-06	0.006
10	116298339	<i>ABLIM1</i>	Body	cg07978738	-0.145 (0.028)	3.62E-07	0.002
10	128994432	<i>DOCK1</i>	Body	cg03242819	0.229 (0.044)	3.41E-07	0.002
11	65201834	-		cg10416861	0.216 (0.043)	8.67E-07	0.004
11	65550444	-		cg09419102	-0.141 (0.029)	1.60E-06	0.006

11	122709551	<i>CRTAM</i>	Body	cg22512531	0.138 (0.026)	4.70E-07	0.002
12	7055657	<i>PTPN6</i>	TSS200	cg23193870	-0.204 (0.042)	1.84E-06	0.006
14	78051204	<i>SPTLC2</i>	Body	cg14544289	-0.199 (0.041)	3.06E-06	0.010
14	89933549	<i>FOXN3</i>	5'UTR	cg13679772	0.149 (0.031)	1.94E-06	0.007
14	106331803	-		cg14387626	-0.142 (0.026)	4.34E-08	3.47E-04
14	106354912	-		cg27113548	-0.284 (0.047)	4.60E-09	5.09E-05
15	99194021	<i>IGF1R</i>	Body	cg07779120	0.251 (0.046)	2.35E-07	0.001
17	4923126	<i>KIF1C</i>	Body	cg03877174	0.154 (0.030)	7.12E-07	0.003
17	9921982	<i>GAS7</i>	Body	cg02018337	-0.122 (0.024)	4.43E-07	0.002
17	27050723	<i>RPL23A</i>	Body	cg18150958	-0.229 (0.046)	8.85E-07	0.004
17	27401793	<i>TIAF1</i>	5'UTR	cg18960216	0.181 (0.034)	1.97E-07	0.001
17	56082867	<i>SFRS1</i>	3'UTR	cg08591265	-0.152 (0.032)	1.42E-06	0.005
18	76739409	<i>SALL3</i>	TSS1500	cg05080154	0.212 (0.038)	4.87E-08	3.75E-04
19	40919465	<i>PRX</i>	TSS200	cg01447828	0.233 (0.044)	1.54E-07	9.63E-04
19	53758055	<i>ZNF677</i>	5'UTR	cg03217253	0.216 (0.042)	3.10E-07	0.002
20	43438809	<i>RIMS4</i>	Body	cg15207742	0.212 (0.043)	1.40E-06	0.005
22	20792535	<i>SCARF2</i>	TSS1500	cg14785479	0.123 (0.025)	1.10E-06	0.004

Abbreviations: Chr, Chromosome; *FDR*, false discovery rate; TSS200, within 200 bps from transcription start site; TSS1500, within 1500 bps from transcription start site; Body, gene body; UTR, untranslated regions.

^a Estimates were calculated based on inverse-normal transformed methylation values. Fixed effect meta-analysis was used with a sample size-weighted method to get *p* values, and an inverse-variance weighted method to get estimates of effect size.

Table 3. Mediation analysis of 15 significant CpGs whose methylation levels were correlated with urinary 2-hydroxynaphthalene in males from the WHZH Cohort.

CpG	Gene	Relations between smoking, 2-hydroxynaphthalene, and methylation			
		Total effect of smoking on methylation ^a		Mediating effect of 2-hydroxynaphthalene in associations between smoking and methylation ^a	
		Total effect (s.e.)	<i>p</i>	Indirect effect (s.e.)	<i>p</i>
cg05575921	<i>AHRR</i>	-0.667 (0.060)	2.17E-28	-0.084 (0.032)	0.008
cg21161138	<i>AHRR</i>	-0.305 (0.074)	<1.00E-05	-0.134 (0.041)	0.059
cg23916896	<i>AHRR</i>	-0.085 (0.078)	<1.00E-05	-0.136 (0.042)	0.017
cg23576855	<i>AHRR</i>	-0.655 (0.059)	<1.00E-05	-0.056 (0.030)	0.100
cg24090911	<i>AHRR</i>	-0.203 (0.074)	1.81E-10	-0.124 (0.040)	0.009
cg25648203	<i>AHRR</i>	-0.360 (0.071)	<1.00E-05	-0.089 (0.037)	0.088
cg26703534	<i>AHRR</i>	-0.428 (0.067)	1.10E-17	-0.092 (0.035)	0.008
cg01940273	-	-0.542 (0.063)	<1.00E-05	-0.088 (0.033)	0.016
cg03636183	<i>F2RL3</i>	-0.464 (0.068)	3.70E-05	-0.059 (0.034)	0.001
cg05284742	<i>ITPK1</i>	-0.273 (0.070)	0.273	-0.084 (0.036)	0.001
cg05951221	-	-0.430 (0.068)	1.39E-04	-0.057 (0.035)	0.015
cg09022230	<i>TNRC18</i>	-0.282 (0.073)	1.24E-04	-0.103 (0.039)	0.008
cg09935388	<i>GFII</i>	-0.274 (0.072)	0.006	-0.091 (0.037)	0.002
cg21322436	<i>CNTNAP2</i>	-0.284 (0.073)	9.90E-05	-0.102 (0.038)	0.008
cg21566642	-	-0.534 (0.067)	1.05E-04	-0.083 (0.035)	0.021

Methylation values were inverse-normal transformed and concentrations of urinary 2-hydroxynaphthalene were ln-transformed.

^aAdjusting for age, drinking status, BMI, occupation, differential leukocyte proportions, geographical regions and beadchip operation date in a mediation macro in SAS 9.2 : %macro mediation (data=, yvar=, avar=, mvar=, cvar=, a0=, a1=, m=, nc=, yreg=, mreg=, Interaction=, casecontrol= false, output= reduced, c=, boot=) (Valeri and Vanderweele 2013).

Figure Legends

Figure 1. Manhattan plot and Q-Q plot of the p values of the associations between methylation and cigarette smoking in the genome-wide meta-analysis. In the Manhattan plot, the x-axis indicates genomic locations of the CpGs; the y-axis indicates $-\log_{10}(p\text{-values})$ of the associations; and the red line indicates the $-\log(p\text{-value})$ at $FDR = 0.05$. In the Q-Q plot, the x-axis shows the expected $-\log_{10}(p\text{-values})$ whereas the y-axis indicates the observed $-\log_{10}(p\text{-values})$.

Figure 2. Associations of the 15 smoking-related CpGs and urinary 2-hydroxynaphthalene levels in males from the WHZH Cohort and the Coke Oven Cohort.

Figure 1.

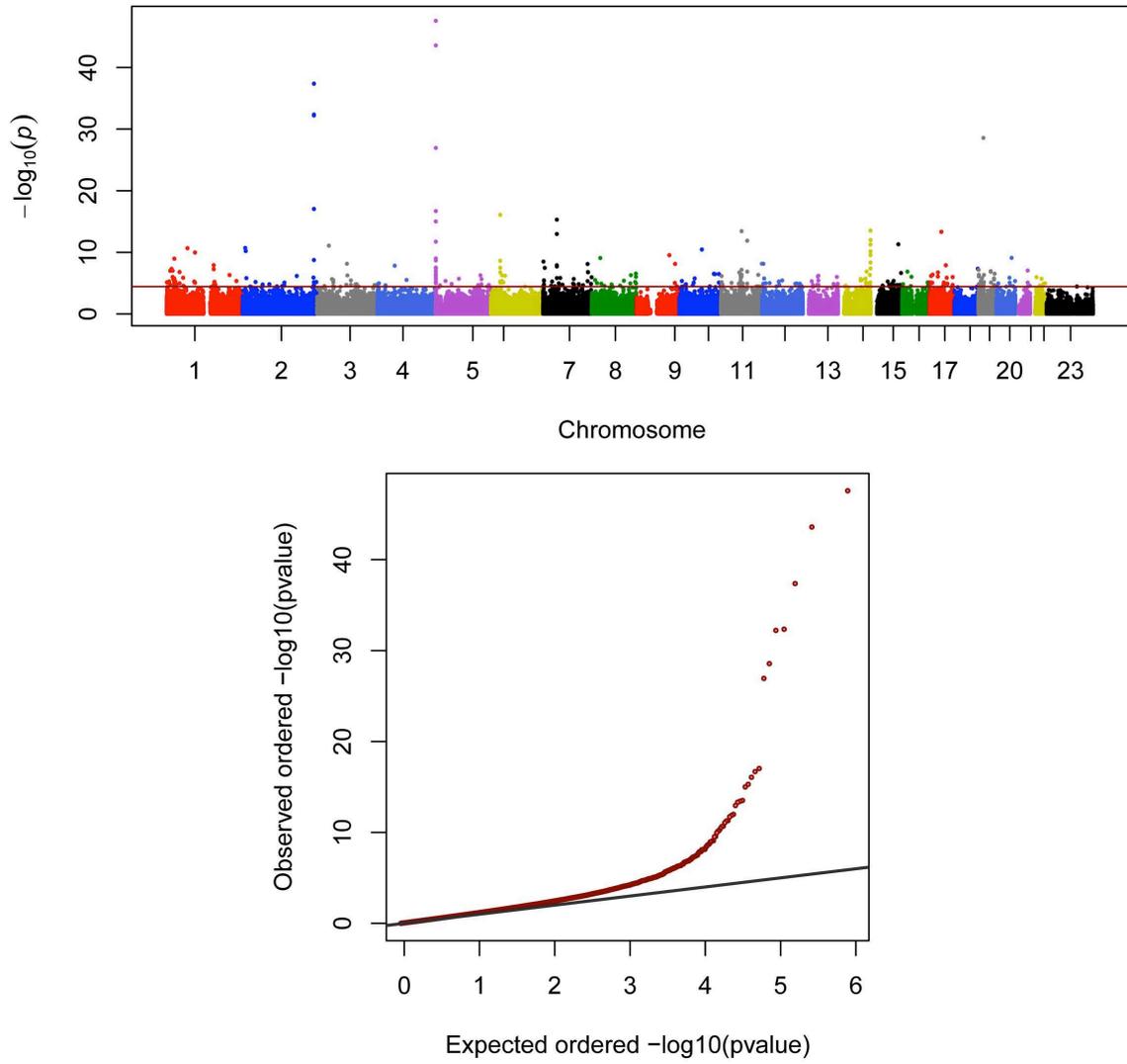


Figure 2.

