Published in final edited form as:

Free Radic Biol Med. 2009 January 1; 46(1): 20–24. doi:10.1016/j.freeradbiomed.2008.09.018.

Association between haplotypes of manganese superoxide dismutase (SOD2), smoking, and lung cancer risk

Carla J. Gallagher^{1,2}, Kwangmi Ahn², Ashley L. Knipe², Anne-Marie Dyer², John P. Richie Jr.^{1,2}, Philip Lazarus^{1,2}, and Joshua E. Muscat²

¹ Cancer Prevention and Control Program, Penn State Cancer Institute, Penn State College of Medicine, Hershey, PA

² Departments of Pharmacology and Public Health Sciences, Penn State College of Medicine, Hershey, PA

Abstract

Tobacco smoke contains high concentrations of reactive oxygen species (ROS) that can damage DNA, proteins, and lipids. Manganese superoxide dismutase (SOD2) catalyzes the dismutation of superoxide radicals into hydrogen peroxide and protects against oxidative stress in lung tissues. Three tagSNPs were identified in one block of high linkage disequilibrium that spans the entire SOD2 gene and 5 kb promoter region. These tagSNPs, representing four haplotypes (TAA, TCA, TCG, CCG), were genotyped in 372 lung cancer cases and 605 controls. There was no association between the haplotype frequencies and overall lung cancer risk. The TCG haplotype (6% in controls) was significantly associated with a lower risk of lung cancer in light-smokers (< median pack-years; p-value = 0.02) but not in heavy smokers. In histologic-specific analysis, the TCG haplotype was significantly associated with a reduced risk of lung adenocarcinoma (odds ratio = 0.39, 95% CI 0.17–0.88), but an inverse association with squamous cell carcinoma was not significant. The association with adenocarcinoma was most apparent in light-smokers (haplotype-specific p-value = 0.005); none of the 61 case subjects with adenocarcinoma had the TCG allele. This study suggests that subjects with the SOD2 TCG haplotype may be at decreased risk for lung adenocarcinoma and that this association may depend upon smoking amount.

Introduction

Lung cancer is the leading cause of cancer mortality in the U.S. Tobacco smoking causes over 80% of all lung cancers (1), but only about 15% of lifetime smokers develop this malignancy (2). Heritable traits in the population likely explain why some smokers develop lung cancer (3–6). Tobacco smoke is a major source of reactive oxygen species (ROS) in the lungs, and ROS can damage DNA, proteins, and lipids, eventually leading to cancer. Numerous studies have shown that diets high in micronutrient antioxidants are associated with lower rates of lung cancer (7). Genetic variation in quenchers of ROS may therefore affect lung cancer risk.

Manganese superoxide dismutase (SOD2) catalyzes the dismutation of superoxide radicals (a major form of ROS) into hydrogen peroxide and plays an important protective role against oxidative stress in lung tissues (8–11). Due to their role in cellular metabolism, the

Please address correspondence to: Joshua E. Muscat, Ph.D., Department of Public Health Sciences, Penn State College of Medicine, 500 University Drive, H069, Hershey, PA; fax: (717) 531-0480; email: jmuscat@psu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

mitochondria are the major sites for the production of ROS. SOD2 is the only known superoxide scavenger in the mitochondria, and therefore may be important in the regulation of overall cellular ROS levels. The mitochondria plays an important role in apoptosis and cellular proliferation, and SOD2 may be directly involved in carcinogenesis as well by protecting against mitochondrial damage.

One coding polymorphism has been identified in SOD2 (Ala16Val). Previous lung cancer association studies of SOD2 focused on this SNP, and results were not consistent across various populations (12–16). Of the studies that evaluated the Ala16Val polymorphism, two studies did not detect an association with lung cancer risk in Asian populations (12,16), even after stratifying by histology, gender, and smoking (12,16). However, in a larger study, the Val/Val genotype was associated with an increase in lung cancer risk in Caucasians (14). The association was even stronger after adjusting for the effect of other genes (MPO, p53, XRCC1) (13,15). In particular, the greater risk for the double variant of p53 and XRCC1 in the SOD2 Val/Val genotype was specific only for adenocarcinoma risk and not squamous cell carcinoma (13).

To comprehensively evaluate the role of genetic variation between SOD2 and lung cancer risk, we used data from the International HapMap project (CEPH genotypes) to assess linkage disequilibrium and choose tagSNPs that represent the common genetic variation in people with European Ancestry for use in a haplotype association study.

Materials and Methods

Study Population

This hospital-based study was conducted at the H. Lee Moffitt Cancer Center (Tampa, FL) from 2000 to 2003, details have been described elsewhere (17). The current analysis includes white men and women with histologically-confirmed lung cancer and no past history of other tobacco-related cancers with available DNA ($n = 372$). White controls ($n = 605$ with available DNA) were randomly selected from thousands of community residents who underwent prostate-specific antigen testing, skin examinations, endoscopy, or mammography at the Lifetime Cancer Screening Center facility of the H. Lee Moffitt Cancer Center. Spiral computed tomography for lung cancer was not performed at the clinic. A list of control IDs was matched against the hospital patient database to identify and exclude subjects, if any, who might have developed cancer. Ninety-nine percent of the hospital patients and ninety-seven percent of the controls who were asked to participate in the study consented. All subjects signed a consent form approved by the institutional review board. A trained interviewer administered a structured questionnaire that obtained lifestyle and smoking history information including levels of education, occupation, year of smoking onset, current smoking status, number of cigarettes smoked per day, and years since quitting (for former smokers). The medical chart of the case subjects was reviewed to obtain diagnostic and pathology records.

SNP selection and genotyping

Genotypes for SNPs in SOD2 (and 5 kb promoter region) representing people with European Ancestry (CEPH), were downloaded from the International HapMap Project (18). Linkage disequilibrium (LD) in the SOD2 gene was determined using Haploview software (19). LD was estimated between all pairs of SNPs using the D' statistic. Haplotype block structure was determined using the Solid Spine of the LD option, with the block extended if pairwise D' between SNPs was greater than 0.80.

According to HapMap data, the coding region and promoter of the SOD2 gene are inherited together as one haplotype block of high LD (Figure 1). HapMap data was used to identify three

SNPs that tag the common haplotypes (>5%) of SOD2 (rs6917589, rs2758331, rs5746105; SNPbrowser 3.5, Applied Biosystems). DNA was extracted from oral buccal mucosa cells as previously described (17). TagSNPs were genotyped using pre-designed Taqman SNP genotyping assays and an Applied Biosystems 7900HT Fast Real-Time PCR System, according to manufacturer's instructions. SDS 2.2.2 software was used for automated calling of genotypes (Applied Biosystems, Foster City, CA). Laboratory personnel involved in genotyping were blinded to patient status. Approximately 10% of the samples were genotyped in duplicate by real-time PCR for all 3 SNP assays for quality control purposes.

Statistical Analysis

Hardy-Weinberg equilibrium was assessed by χ^2 tests. Haplo.stats software (20) was used to estimate haplotypes with the expectation maximization (EM) algorithm and to determine the haplotype association with lung cancer using the global score statistic. P-values were calculated by a 1,000-permutation algorithm, and a Hap-Score was calculated in which a positive sign shows a higher frequency in cases and a negative sign indicates a higher frequency in controls. Odds ratios and 95% confidence intervals were calculated using SAS statistical software (Cary, NC) and the SAS macro HAPPY. The associations were adjusted for known lung cancer confounders: age (continuous), sex (categorical – M, F), and pack-years of smoking (continuous). The lung cancer data set was also stratified by gender and smoking level (light and heavy smokers, stratified by the median pack-years among all smokers, 44 pack-years). Subgroup analyses were performed by lung histologic type using the entire control population as the comparison group. Rare haplotypes (<2%) were excluded from all analyses.

Results

The basic demographic characteristics of the lung cancer case and control subjects are shown in Table 1. About 41% of cases and 43% of controls were women, and the mean age of cases and controls was 64.4 ± 9.9 and 58.9 ± 10.1 , respectively. A higher percentage of cases than controls were current smokers (39% versus 17%). The mean pack-years of smoking was 56 for cases and 23 for controls. Twenty-three percent of cases and 39% of controls were college educated. The mean body mass index (BMI) of cases and controls was 26.8 ± 5.0 versus 27.5 ± 4.9 , respectively. Among the case group, the most frequent histology was adenocarcinoma (38%), followed by squamous cell carcinoma (24%).

One SNP, rs6917589, departed from Hardy-Weinberg equilibrium ($p = 0.022$) due to a shortage of heterozygotes, but rs2758331 and rs5746105 were consistent with Hardy-Weinberg equilibrium ($p > 0.05$). Of the 288 samples repeated for quality control purposes, all samples (100%) were concordant with the original genotype call. The minor allele frequencies of the tagSNPs in controls were 0.23 (rs6917589), 0.49 (rs2758331), and 0.29 (rs5746105).

The three tagSNPs distinguish the four common SOD2 haplotypes. The analysis results are summarized as the specific haplotype, the haplotype frequency in case and control subjects, the odds ratios and 95% confidence intervals, and a score for each haplotype (Hap-Score) in which a negative sign shows a higher proportion of this haplotype in control subjects and a positive sign indicates a higher proportion in the lung cancer case subjects. Also reported are haplotype-specific empiric p-values for significance of the Hap-Score. A global p-value is reported for the significance of the difference of all four haplotypes between the lung cancer case and control subjects.

There was no significant association between any of the SOD2 haplotypes and overall lung cancer risk (Table 2), after adjusting for age, sex, and pack-years. The odds ratio for the TCG haplotype compared to the most common haplotype was 0.80 (95% CI 0.50–1.29). When stratifying by the median pack-years among all smokers, the TCG haplotype was significantly

associated with decreased lung cancer risk in light-smokers. The haplotype occurred in 6.2% of controls, vs. 1.9% of cases (Haplotype-specific p-value = 0.018, OR = 0.33, 95% CI = 0.11–0.95; Table 3a). No association was observed in heavy-smokers. The association in light-smokers is consistent with the borderline association of the TCG haplotype observed in former smokers (P=0.068; data not shown). No other covariates, including BMI and education, modified the effect of the TCG haplotype on lung cancer risk as covariates in the overall model, and in stratified analysis. We examined the association between each tagSNP and lung cancer risk and found no association overall or in any of the subgroup analyses.

When examining the histologic-specific risks, the TCG haplotype was significantly associated with lung adenocarcinoma (p = 0.021, OR = 0.39, 95% CI = 0.17–0.88). The TCG haplotype was found in 6.2% of controls and 2.7% of adenocarcinoma cases (Table 3b). The TCG haplotype was also more common in controls than in cases with squamous cell carcinoma (4.5%) but this difference was not statistically significant. The other histologic types were too infrequent to test for statistical significance with reasonable precision. When stratified by pack-years of smoking in histologic subgroup analysis, the TCG haplotype was not present in any of the light-smoking adenocarcinoma cases (n=61) (p= 0.005; Table 3c). All stratified analyses were also adjusted for age, sex, and pack-years smoking. There were too few light smoking cases with squamous cell carcinoma to examine this association with precision.

Discussion

Because of the potential importance of SOD2 in lung cancer risk, we comprehensively tested the genetic variants of SOD2. Previous studies of SOD2 and lung cancer risk have focused on the (Ala16Val) polymorphism at codon 16. Two studies did not detect an association with lung cancer risk in Asian populations even after stratifying by histology, gender, and smoking (12,16). In a larger study, the Val/Val genotype was associated with an increase in lung cancer risk in Caucasians (14), which became even stronger when taking into account lung cancer risk alleles of other genes (MPO, p53, XRCC1) (13,15). The greater risk for the double variant of p53 and XRCC1 in the SOD2 Val/Val genotype was specific only for patients with adenocarcinoma (13).

Similar to the genotyping studies of Ala16Val polymorphism, our data indicates that genetic variation of SOD2 affects risk only for adenocarcinoma and not squamous cell carcinoma. However, the associated haplotype from our study (TCG) is not tagging the Ala16Val SNP; data from HapMap shows that Ala16Val is tagged by the most common haplotype (TAA), and indicates that the Ala16Val SNP is not associated with lung cancer risk in our data. In addition, data from HapMap shows that Ala16Val is in LD with the tagSNPs rs6917589 (D'=1) and rs2758331 (D'=1). These individual SNPs are not associated with lung cancer risk in our study, further indicating that the Ala16Val SNP is not associated with lung cancer risk in our data.

To identify potentially functional variants tagged by the TCG haplotype, the sequence databases were searched. The coding region of this gene was re-sequenced in 102 controls from the SNP500Cancer database (<http://snp500cancer.nci.nih.gov>) and 90 controls from the GeneSNPs database (<http://www.genome.utah.edu/genesnps>), so all common SNPs in the coding region are likely identified. Besides the Ala16Val SNP, no other coding SNPs with a frequency >1% were identified in these re-sequencing efforts. It is possible that the TCG haplotype tags for functional promoter or intron SNPs that alter SOD2 expression or splicing. Previously the SOD2 gene has been shown to produce at least two different mRNA products, one having a much larger 3' untranslated region (10,21), and recently Shao *et al* (22) identified a novel splice variant of SOD2, lacking exon 4, caused by an intron 3 polymorphism. In an analysis of the SOD2 promoter in the sequence databases, there exists 14 SNPs in the 1kb proximal promoter of SOD2, many of which alter putative transcription factor binding sites

when the minor allele is present (TESS, transcription element search software). However, the entire intronic and promoter regions of SOD2 have not been re-sequenced, so there are likely other variations in these regions that have not yet been discovered. Full sequencing and subsequent functional analysis of the promoter as well as intron region sequences must be performed in individuals homozygous for the TCG haplotype to determine if there are additional polymorphisms that are potentially important in SOD2-associated alterations in lung cancer risk.

It is uncertain why the effects of SOD2 genetic variants might be histologic-specific. Studies have shown that expression levels of SOD2 are significantly higher in adenocarcinoma than in squamous cell carcinoma for many cancer sites, including lung carcinoma (23–26). In addition, grape seed proanthocyanidin extract (GSPE), which provides protection against free radicals in vitro and in vivo, has been shown to demonstrate significant cytotoxicity towards lung adenocarcinoma cells (27), this has not been shown for other histologies of lung cancer. It is possible that free radical scavengers affect risk differentially by histology or cell origin.

The association of the SOD2 TCG haplotype with lung cancer risk was specific for light smokers. There is limited but conflicting data on whether smoking induces SOD2 expression. In a microarray study of 44 antioxidant related genes SOD2 did not display a difference in expression between smokers and nonsmokers in large airways of the lung (28). However, in a more recent study, expression of SOD2 was elevated in central bronchial epithelium and alveolar epithelium of smokers compared to nonsmokers (29). The protective effect of the TCG haplotype might be relatively more important in light than in heavy smokers if there is a saturation effect with smoking. Heavy smokers may be overwhelming their system with reactive oxygen species, where genetic variation may be unlikely to modify this effect.

The significant associations of SOD2 haplotypes with lung cancer risk occurred in subgroups. Since the TCG haplotype has a low frequency, the number of cases and controls with this haplotype in these subgroups is small and the associations could reflect a false positive result. These results need to be confirmed in larger populations of lung cancer cases and controls to gain increased power in subgroup analysis.

In conclusion, this study indicates that genetic variants of SOD2 may be protective against lung cancer risk in light smokers, particularly against lung adenocarcinoma. Further studies in independent cohorts are needed to determine if the current findings are not due to chance, and if so to identify the functional polymorphisms associated with this haplotype and examine their contribution to lung cancer risk.

References

1. Center for Disease Control and Prevention. Smoking and Health Database. 2004.
2. Spitz MR, Wei Q, Dong Q, Amos CI, Wu X. Genetic susceptibility to lung cancer: the role of DNA damage and repair. *Cancer Epidemiol Biomarkers Prev* 2003;12:689–98. [PubMed: 12917198]
3. Lorenzo Bermejo J, Hemminki K. Familial lung cancer and aggregation of smoking habits: a simulation of the effect of shared environmental factors on the familial risk of cancer. *Cancer Epidemiol Biomarkers Prev* 2005;14:1738–40. [PubMed: 16030110]
4. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer* 2002;99:260–6. [PubMed: 11979442]
5. Hemminki K, Lonnstedt I, Vaittinen P, Lichtenstein P. Estimation of genetic and environmental components in colorectal and lung cancer and melanoma. *Genet Epidemiol* 2001;20:107–116. [PubMed: 11119300]
6. Guo Z, Peng S, Jiang G. A genetic epidemiological study on lung cancer. *Zhonghua Yu Fang Yi Xue Za Zhi* 1996;30:154–6. [PubMed: 9208526]

7. El-Bayoumy, K.; Muscat, J.; Hoffmann, D.; Heber, D. *Nutritional Oncology*. New York: Elsevier Press; 2006. *Nutrition and Tobacco-Related Cancers*.
8. Fridovich I. The biology of oxygen radicals. *Science* 1978;201:875–80. [PubMed: 210504]
9. Halliwell B. Free radicals, antioxidants, and human disease: curiosity, cause, or consequence? *Lancet* 1994;344:721–4. [PubMed: 7915779]
10. Robinson BH. The role of manganese superoxide dismutase in health and disease. *J Inher Metab Dis* 1998;21:598–603. [PubMed: 9728339]
11. Zhu S, Manuel M, Tanaka S, Choe N, Kagan E, Matalon S. Contribution of reactive oxygen and nitrogen species to particulate-induced lung injury. *Environ Health Perspect* 1998;106(Suppl 5): 1157–63. [PubMed: 9788891]
12. Lin P, Hsueh YM, Ko JL, Liang YF, Tsai KJ, Chen CY. Analysis of NQO1, GSTP1, and MnSOD genetic polymorphisms on lung cancer risk in Taiwan. *Lung Cancer* 2003;40:123–9. [PubMed: 12711112]
13. Liu G, Zhou W, Park S, Wang LI, Miller DP, Wain JC, Lynch TJ, Su L, Christiani DC. The SOD2 Val/Val genotype enhances the risk of nonsmall cell lung carcinoma by p53 and XRCC1 polymorphisms. *Cancer* 2004;101:2802–8. [PubMed: 15534883]
14. Wang LI, Miller DP, Sai Y, Liu G, Su L, Wain JC, Lynch TJ, Christiani DC. Manganese superoxide dismutase alanine-to-valine polymorphism at codon 16 and lung cancer risk. *J Natl Cancer Inst* 2001;93:1818–21. [PubMed: 11734599]
15. Liu G, Zhou W, Wang LI, Park S, Miller DP, Xu LL, Wain JC, Lynch TJ, Su L, Christiani DC. MPO and SOD2 polymorphisms, gender, and the risk of non-small cell lung carcinoma. *Cancer Lett* 2004;214:69–79. [PubMed: 15331175]
16. Ho JC, Mak JC, Ho SP, Ip MS, Tsang KW, Lam WK, Chan-Yeung M. Manganese superoxide dismutase and catalase genetic polymorphisms, activity levels, and lung cancer risk in Chinese in Hong Kong. *J Thorac Oncol* 2006;1:648–53. [PubMed: 17409931]
17. Gallagher CJ, Muscat JE, Hicks AN, Zheng Y, Dyer AM, Chase GA, Richie J, Lazarus P. The UDP-glucuronosyltransferase 2B17 gene deletion polymorphism: sex-specific association with urinary 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol glucuronidation phenotype and risk for lung cancer. *Cancer Epidemiol Biomarkers Prev* 2007;16:823–8. [PubMed: 17416778]
18. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome. *Nature* 2005;437:1299–320. [PubMed: 16255080]
19. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–5. [PubMed: 15297300]
20. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002;70:425–34. [PubMed: 11791212]
21. Wan XS, Devalaraja MN, St Clair DK. Molecular structure and organization of the human manganese superoxide dismutase gene. *DNA Cell Biol* 1994;13:1127–36. [PubMed: 7702755]
22. Shao J, Chen L, Marrs B, Lee L, Huang H, Manton KG, Martin GM, Oshima J. SOD2 polymorphisms: unmasking the effect of polymorphism on splicing. *BMC Med Genet* 2007;8:7. [PubMed: 17331249]
23. Izutani R, Asano S, Imano M, Kuroda D, Kato M, Ohyanagi H. Expression of manganese superoxide dismutase in esophageal and gastric cancers. *J Gastroenterol* 1998;33:816–22. [PubMed: 9853553]
24. Kobayashi T, Matsumoto M, Iizuka H, Suzuki K, Taniguchi N. Superoxide dismutase in psoriasis, squamous cell carcinoma and basal cell epithelioma: an immunohistochemical study. *Br J Dermatol* 1991;124:555–9. [PubMed: 2064938]
25. Piyathilake CJ, Bell WC, Oelschlager DK, Heimburger DC, Grizzle WE. The pattern of expression of Mn and Cu-Zn superoxide dismutase varies among squamous cell cancers of the lung, larynx, and oral cavity. *Head Neck* 2002;24:859–67. [PubMed: 12211050]
26. Iizuka S, Taniguchi N, Makita A. Enzyme-linked immunosorbent assay for human manganese-containing superoxide dismutase and its content in lung cancer. *J Natl Cancer Inst* 1984;72:1043–9. [PubMed: 6585582]
27. Bagchi D, Bagchi M, Stohs S, Ray SD, Sen CK, Preuss HG. Cellular protection with proanthocyanidins derived from grape seeds. *Ann N Y Acad Sci* 2002;957:260–70. [PubMed: 12074978]

28. Hackett NR, Heguy A, Harvey BG, O'Connor TP, Luetlich K, Flieder DB, Kaplan R, Crystal RG. Variability of antioxidant-related gene expression in the airway epithelium of cigarette smokers. *Am J Respir Cell Mol Biol* 2003;29:331–43. [PubMed: 12702543]
29. Harju T, Kaarteenaho-Wiik R, Sirvio R, Paakko P, Crapo JD, Oury TD, Soini Y, Kinnula VL. Manganese superoxide dismutase is increased in the airways of smokers' lungs. *Eur Respir J* 2004;24:765–71. [PubMed: 15516670]

Figure 1. LD structure of SOD2 and 5 kb promoter. SOD2 HapMap genotypes from Caucasians (CEPH) were downloaded and LD was determined using Haploview. D' values are displayed in the squares (empty squares have a pairwise $D' = 1.0$). Red squares show high pairwise LD, gradually coloring down to white squares of low pairwise LD. Blue squares indicate high LD, but low significance. The black triangle indicates the single block of high LD comprising SOD2.

Table 1

Demographics of lung cancer cases and community controls, Tampa, Fla.

	cases N (%)	controls N (%)
Mean Age ^a	64.4 ± 9.9	58.9 ± 10.1
Women (%)	153 (41)	258 (43)
Mean BMI ^a	26.8 ± 5.0	27.5 ± 4.9
Current Smokers (%)	146 (39)	105 (17)
Pack-Years a	55.7 ± 38.3	23.3 ± 31.2
Years of education (%)		
< High School Degree	59 (16)	22 (4)
High School Degree	228 (61)	345 (57)
College Degree	56 (15)	146 (24)
Post-Graduate Degree	28 (8)	92 (15)
Histology (%)		
Adenocarcinoma	142 (38)	
Squamous Cell Carcinoma	87 (24)	
Non-Small Cell Carcinoma	66 (18)	
Small-Cell Carcinoma	36 (10)	
Large-Cell Carcinoma	27 (7)	
Mixed Histology/Other	13 (3)	
Genotype frequency		
rs6917589	192 (52), 137 (37), 20 (5)	359 (59), 189 (31), 41 (7)
rs2758331	94 (25), 190 (51), 79 (21)	153 (25) 300 (50), 144 (24)
rs5746105	168 (45) 158 (43), 27 (7)	311 (51), 225 (37), 58 (10)

^a Mean ± standard deviation

Table 2
Frequency distribution of the SOD2 haplotypes in cases and controls and lung cancer risk

Haplotype *	Case Frequency	Control Frequency	OR (95% CI) **	Hap-Score	Global P-value=0.073	
					Haplotype-Specific P-value	***
TAA	0.479	0.494	Reference	-1.17		0.273
TCG	0.05	0.062	0.80 (0.50-1.29)	-1.24		0.197
TCA	0.218	0.22	1.00 (0.76-1.31)	-0.64		0.496
CCG	0.254	0.224	1.11 (0.86-1.43)	0.79		0.403

* Haplotype of three tagSNPs in one LD block (rs6917589, rs2758331, rs5746105)
** OR were calculated comparing each haplotype to the most common haplotype (reference)
*** All P-values were calculated based on 1000 permutations and adjusted for age, sex, and pack-years smoking by comparing each haplotype to all others combined

Table 3
Subgroup and stratified haplotype association analysis in lung cancer cases and controls. A. light-smokers; B. adenocarcinoma cases; and C. light-smokers with adenocarcinoma

A.					
<i>Global P-value=0.011</i>				<i>***</i>	
Haplotype *	Case Frequency (%)	Control Frequency (%)	OR (95% CI) **	Hap-Score	Haplotype-Specific P-value ***
TAA	0.476	0.482	Reference	-0.82	0.407
TCG	0.019	0.062	0.33 (0.11–0.95)	-2.43	0.018
TCA	0.26	0.26	0.93 (0.60–1.44)	-0.82	0.417
CCG	0.245	0.195	1.21 (0.79–1.85)	0.87	0.394
B.					
<i>Global P-value=0.114</i>				<i>***</i>	
Haplotype *	Case Frequency (%)	Control Frequency (%)	OR (95% CI) **	Hap-Score	Haplotype-Specific P-value ***
TAA	0.487	0.494	Reference	0.12	0.89
TCG	0.027	0.062	0.39 (0.17–0.88)	-2.37	0.021
TCA	0.23	0.22	0.88 (0.61–1.27)	-0.71	0.489
CCG	0.256	0.224	1.09 (0.79–1.51)	1.01	0.306
C.					
<i>Global P-value=0.009</i>				<i>***</i>	
Haplotype *	Case Frequency (%)	Control Frequency (%)	OR (95% CI) **	Hap-Score	Haplotype-Specific P-value ***
TAA	0.527	0.482	Reference	0.58	0.557
TCG	0	0.062	NA ***	-2.71	0.005
TCA	0.218	0.26	0.61 (0.34–1.08)	-1.59	0.116
CCG	0.255	0.195	1.07 (0.65–1.75)	0.82	0.409

* Haplotype of three tagSNPs in one LD block (rs6917589, rs2758331, rs5746105)
** OR were calculated comparing each haplotype to the most common haplotype (reference)
*** All P-values were calculated based on 1000 permutations and adjusted for age, sex, and pack-years smoking by comparing each haplotype to all others combined
**** NA; OR not calculated due to no copies of this haplotype in cases