Associations between variants of the 8q24 chromosome and nine smoking-related cancer sites

Sungshim Lani Park[1], Shen-Chih Chang[1], Lin Cai[2], Carlos Cordon-Cardo[3], Bao-Guo Ding[4], Sander Greenland[1,5], Shehnaz K. Hussain[1,6], Qingwu Jiang[7], Simin Liu[1,6], Ming-Lan Lu[8], Jenny T. Mao[9], Hal Morgenstern[10], Li-Na Mu[11], Leslie J. Ng[1], Allan Pantuck[12], Jianyu Rao[1,13], Victor E. Reuter[8], Donald P. Tashkin[9], Nai-Chieh Y. You[1], Can-Qing Yu[1,14], Shun-Zhang Yu[7], Jin-Kou Zhao[15], Arie Belldegrun[12], Zuo-Feng Zhang[1,6]

Affiliations of Authors:

[1]Department of Epidemiology, University of California, Los Angeles (UCLA) School of Public Health, Los Angeles, CA, USA; [2]Department of Epidemiology, School of Public Health, Fujian Medical University, Fuzhou, China; [3]Herbert Irwing Comprehensive Cancer Center, Columbia University, New York, NY; [4]Taixing City Center for Disease Prevention and Control (CDC), Taixing City, Jiangsu, China; [5]Department of Statistics, UCLA, Los Angeles, CA, USA; [6]Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, CA USA; [7]Fudan University School of Public Health, Shanghai, China; [8]Departments of Pathology, Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; [9]Division of Pulmonary and Critical Care Medicine, UCLA David Geffen School of Medicine, Los Angeles, CA, USA; [10]Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA; [11]Department of Social and Preventive Medicine, State University of New York (SUNY) at Buffalo, NY, USA; [12]Department of Urology, UCLA David Geffen School of Medicine, Los Angeles, CA, USA; [13]Departments of Pathology and Laboratory Medicine, UCLA David Geffen

School of Medicine, Los Angeles, CA, USA; [14]School of Public Health, Peking University, Beijing, China, [15]Gates Foundation Beijing Office, China

Correspondence to:

Request for reprints:

Zuo-Feng Zhang, M.D., Ph.D.

Department of Epidemiology

School of Public Health

University of California, Los Angeles (UCLA)

71-225 CHS, Box 951772

650 Charles E Young Drive, South

Los Angeles, CA 90095-1772

U.S.A.

Phone:310-825-8418

Fax: 310-206-6039

Email: zfzhang@ucla.edu

Running Title: 8q24 variants and smoking-related cancers

Keywords: 8q24 polymorphisms, smoking-related cancers, neoplasms, lung, oropharynx, nasopharynx, larynx, esophagus, stomach, liver, bladder, kidney, SNP

Abstract

Recent genome-wide association (GWA) studies identified key single nucleotide polymorphisms (SNPs) in the 8q24 region to be associated with prostate cancer. 8q24 SNPs have also been associated with colorectal cancer, suggesting this region may not be specifically associated to just prostate cancer. To date, the association between these polymorphisms and tobacco smoking-related cancer sites remains unknown. Using epidemiological data and biological samples previously collected in three case-control studies from U.S. and Chinese populations, we selected and genotyped one SNP from each of the three previously determined "regions" within the 8q24 loci: rs1447295 (region 1), rs16901979 (region 2), and rs6983267 (region 3), and examined their association with cancers of the lung, oropharynx, nasopharynx, larynx, esophagus, stomach, liver, bladder, and kidney. We observed noteworthy associations between rs6983267 and upper aero-digestive tract (UADT) cancers (OR_{adj}=1.69, 95% CI=1.28, 2.24), particularly in oropharynx (OR_{adj}=1.80, 95% CI=1.30, 2.49) and larynx (OR_{adj}=2.04, 95% CI=1.12, 3.72). We also observed a suggestive association between rs6983267 and liver cancer (OR_{adj}=1.51, 95% CI=0.99, 2.31). When we stratified our analysis by smoking status, rs6983267 was positively associated with lung cancer among ever-smokers (OR_{adj}=1.45, 95% CI=1.05, 2.00) and inversely associated with bladder cancer among ever-smokers (OR_{adj}=0.35, 95% CI=0.14, 0.83). Associations were observed between rs16901979 and UADT cancer among never-smokers, and between rs1447295 and liver cancer among ever-smokers. Our results suggest variants of the 8q24 chromosome may play an important role in smoking-related cancer development.

Functional and large epidemiological studies should be conducted to further investigate the association of 8q24 SNPs with smoking-related cancers.

Introduction

Tobacco smoking is responsible for over 20% of all cancer deaths worldwide (1) and is a known cause of lung, aero-digestive, urinary tract, and cervical cancers. Recent IARC reviews have found growing epidemiologic evidence supporting associations between tobacco smoking with liver and stomach cancer as well (2, 3). Among developed countries, smoking cessation can decrease cancer risk (4, 5); however, in developing countries, particularly China, smoking prevalence continues to increase (6). It was estimated from twin cohort studies that heritable factor may attribute to 26% in lung cancer and 31% in bladder cancer development(7). Thus, understanding the role of genetics within smoking-related cancers continues to be of importance.

The 8q24 chromosome region has been of increasing research interest in cancer development and epidemiology. Amplification within the 8q24 loci has been observed within a diverse group of cancers (8-15). Recent genome-wide association (GWA) studies identified associations between genetic variants or single nucleotide polymorphisms (SNPs): DG8S737, rs1447295, rs16901979, and rs6983267, along the 8q24 region and prostate cancer among multiple study populations: Icelandic, Swedish, European-American, African American, and the Multiethnic Cohort (16-19). Haiman and colleagues, using fine mapping markers, designated 8q24 into 3 "regions" and identified SNPs that showed the strongest single association in 2 "regions" ("region 2": rs16901979 and "region 3": rs6983267) (20). DG8S737 (rs1447295) variants from "region 1" were previously observed to have the strong associations with prostate cancer (16). These SNPs and additional 8q24 variants have been subsequently confirmed by genetic association studies (21-27). Furthermore, studies have investigated the

associations between variants of 8q24 region and cancers of the breast (23), colon (25, 28-30), endometrium(31) and testes (32). It is still unknown whether SNPs at 8q24 region are associated with tobacco smoking-related cancer sites.

Some epidemiologic studies have suggested that tobacco smoking may be associated with colorectal cancers (33). Additionally, a number of studies observed associations between colorectal cancer and SNPs rs6983267 (25, 28-30); therefore, we hypothesize that 8q24 SNPs may be associated with smoking-related cancers. To test this hypothesis, we selected one SNP from each "region" to investigate their potential associations with nine smoking-related cancer sites (lung, oropharynx, larynx, esophagus, stomach, liver, bladder, and pilot studies on nasopharynx and kidney), using data from three case-control studies: Los Angeles County (the LA study), Memorial Sloan Kettering Cancer Center (the MSKCC study), and China Taixing study,

Material and Methods

Los Angeles (LA) study

Details of this population-based case-control study have been described previously (34, 35) Study participation criteria included the following: (i) all subjects were residents of Los Angeles County at the time of recruitment (for controls) or diagnosis (for cases), (ii) during the study period were 18-65 years of age, and (iii) were able to speak either English or Spanish. Newly diagnosed pathologically confirmed cases were identified using the rapid ascertainment system of the Cancer Surveillance Program for Los Angeles County (34). Lung cancer cases (N=611) and the UADT cancer cases (N=601, oropharynx, larynx, nasopharynx, esophagus, and others) were interviewed from

1999 to 2004. Population-based controls (n=1040) who were lung and upper aero-digestive tract (UADT) cancer free were identified through a formal algorithm providing a list of households within the neighborhood of each individual case. Recruitment rates were 39% for eligible lung cancer cases, 46% for eligible UADT cancer cases, and 79% for contacted eligible controls. Cases and controls were matched by age (10 year categories) and gender. Informed consent approved by the Institutional Review Boards of University of California, Los Angeles, and University of Southern California, were obtained from all study participants.

China Taixing Study

Specific details regarding this study population were previously reported (36-38). In brief, this was a population-based case-control study conducted in Taixing City, Jiangsu Province, China. Eligible cases were residents of Taixing City (living in Taixing for 10 years or more), 20 years of age or older, and newly diagnosed with esophagus, stomach, or liver cancer from June 1, 2000 to December 30, 2000. All cases were pathologically or clinically confirmed and reported to the Taixing Tumor Registry at the Taixing CDC. A total of 206 stomach cancer cases, 204 liver cancer cases, 218 esophageal cancer cases, and 464 population-based healthy controls were interviewed using an epidemiological questionnaire. Control groups were randomly selected from a generated list of residents, frequency-matched with cases on gender, age group (5-years), and residential village (or residential block in the city). In the six-month study period the recruitment rates were 89.4% for controls, 65% for stomach, 57% for liver, and 67% for esophageal cancer cases.

Memorial Sloan Kettering Cancer Center (MSKCC) Study

Detailed information of this study population was previously reported (39, 40). Briefly, this was a hospital-based case-control study conducted at MSKCC. Eligible cases of bladder and kidney cancer were seen at MSKCC from August 1, 1993 to June 30 1997. Cases were recruited according to the following criteria: had a pathologically confirmed diagnosis, lived in the U.S. for one year or more, and were in stable medical condition. All cases were either newly diagnosed or undergoing surgical procedure for their relevant cancer. A total of 233 cases with bladder cancer and 34 cases with kidney cancer were interviewed. Controls were recruited based on the following criteria: consented in writing to participate in the study; resided in the United States for at least one year; and were in stable medical condition. During the four-year study period, 178 controls were recruited from the MSKCC blood bank or were patients with a negative diagnosis for cancers at MSKCC. This study was approved by the Institutional Review Board on Human Subjects of MSKCC, and all study participants signed informed consents.

Epidemiological Data collection

Epidemiologic data were collected by trained interviewers, using study specific standardized questionnaires. The detailed standard questionnaires of all three studies included the following information: (1) demographic factors; (2) personal habits: cigarette smoking, passive smoking, alcohol consumption, coffee and tea consumption, etc; (3) history of occupational and environmental exposures; (4) family history of cancer; (5) dietary factors (food frequency questionnaire); (6) medical history; and (7)

questions regarding environmental exposures that were specific to each of these nine cancer sites. The personal interview process took approximately 40 minutes to one hour.

Biological Specimen Collection

For the Los Angeles study, buccal cells were collected from both cases and controls, using the brushing of buccal mucosa and rinsing with mouthwash method (41). Response rates for interviewed participants providing buccal cells were 89% for controls and 89%, 68%, 88%, and 90% for lung, oropharyngeal/nasopharyngeal, laryngeal, and esophageal cancer cases, respectively. In the Taixing study, peripheral blood samples were collected from interviewed participants with response rates of 97.5% for controls, 95% for stomach and liver cancer cases and 94% for esophageal cancer cases, respectively. Lastly, for the MSKCC study, peripheral blood samples were collected from both cases and controls, and normal and tumor tissue samples from cases who had undergone radical cystectomy. Biological specimens were available for 166 healthy controls, 174 bladder and 20 kidney cancer cases. Biological specimens were transported and stored in freezers of -70 degree Celsius of the Molecular Epidemiology Laboratory, UCLA School of Public Health.

Genotyping by TaqMan Assays

DNA samples were isolated from biological specimens using a modified phenol-chloroform method and assayed for purity and concentration by spectrometry. (41) We selected from each "region" the strongest single association SNPs, "region 1": rs1447295, "region 2": rs16901979, and "region 3": rs6983267. SNP genotyping was

performed using the TaqMan allelic discrimination method with the ABI 7900HT Real Time PCR System (TaqMan; Applied Biosystems, Foster City, CA). Aliquots of DNA from cases and controls were randomized onto PCR plates, into which a reaction mix containing Applied Biosystems Taqman universal master mix, and a probe for either SNP (Applied Biosystems, Foster City, CA) was added. Specific primers and probes were custom-designed by the ABI Taqman system. Modified from the protocols of ABI Taqman manual, after holding the plates at 92 °C for 10 minutes, they underwent 60 thermocycles of denaturing at 92 °C for 15 seconds and annealing at 62 °C for 80 seconds. Following PCR amplification, end-point fluorescence was read using the ABI Primer 7900HT instrument and genotypes were scored using SDS 2.3 Allelic Discrimination Software from Applied Biosystems. For quality control (QC), we genotyped 5% duplicated samples randomly selected to evaluate reproducibility and concordance rate was >99%. The automatic call rates were $\geq$96% for all three SNPs. Furthermore, as QC process, all laboratory researchers were blinded to the case or control statuses and to the identity of quality control samples.

Statistical Analysis

Analysis was performed using SAS v9.2 software (Cary, NC). Tests for Hardy-Weinberg equilibrium (HWE) and differences in minor allele frequencies (MAFs) were evaluated for all three SNPs using the chi-squared test. Unconditional logistic regression models were employed to determine crude and adjusted odds ratios (OR) and 95% confidence intervals (CI) for the association between SNPs and each cancer site. For the Los Angeles County study we adjusted for the following variables: age, gender, ethnicity,

educational level, and tobacco smoking. Education level and tobacco smoking were treated as continuous variables. Age was adjusted for in fine categories (under 34, 35-36, 37-38, 39-40, 41-42, 43-44, 45-46, 47-48, 49-50, 51-52, 53-54, 55-56, 57-58, 59-62), and controls who were more than 3 years younger than the youngest case or 3 years older than the oldest case were excluded from the analysis. This resulted in 11 excluded controls for lung cancer and 1 excluded control for UADT cancers. For all UADT cancers, the variable alcohol drinking was also applied to the model. All models for the Taixing study were adjusted for age, gender, smoking pack-year and alcohol drinking. For stomach cancer, we also adjusted for *H. pylori* infection status, and for liver cancer we adjusted for HBsAg status. Age and smoking variables were continuous while the remaining variables were categorical. Alcohol drinking was defined in categories of never, occasionally, often, or everyday in the Taixing study. Lastly, for the MSKCC study of bladder cancer, we adjusted for the following factors: gender, age (< 55, 55-<60, 60-<65, $\geq$65), race (white vs nonwhite), and smoking (never vs. ever). We first analyzed SNP genotypes (CC, CA, AA or TT, TG, GG) as a continuous variable (additive models) and as dummy variables for each cancer site. These results were used to identify the appropriateness of the use of a dominant or recessive model. For each site, changes in the odds ratios for these three SNPs across levels of tobacco smoking were evaluated using unconditional logistic regression adjusting for previous mentioned confounding factors and ORs for interaction were estimated by including smoking (never or ever), SNP genotypes (0 and 1) according to the dominant or recessive model, and product terms of smoking by each SNP.

To account for false positive findings due to multiple testing we calculated the false positive report probability (FPRP), using an approach presented in Wacholder (42). We set the FPRP threshold at 0.5, since this is an initial study investigating the associations between three 8q24 SNPs and smoking-related tumor sites. Due to the overwhelming evidence of associations between 8q24 variants with prostate and colon cancer, we assigned a prior probability range of 0.01-0.1 to detect an OR of 1.5 or 0.67.

Results

The baseline characteristics of each study's population were previously reported (34-39) and a short summary for study populations can be found on Supplementary Table 1. The LA study consisted of a multiethnic population, non-Hispanic Whites (59%), African American (12%), Hispanic 17%, and others (12%, predominantly Asian). In the MSKCC study, the majority of participants were Whites (92.1% of cases and 96.8% of controls). Table 1 presents the distribution of genotypes of controls stratified by study sites and ethnicities. The distributions of 8q24 SNPs were consistent with the HWE ($p \geq 0.05$) among Whites (the LA and MSKCC studies), Mexican (the LA study), and Asian American (the LA study), African American (the LA study, 2 SNPs) and Chinese (the Taixing study, 2 SNPs). However, the distribution of rs16901979 in African Americans in the LA study, and of rs6983267 in the Chinese population did not meet HWE ($p=0.021$ and $p=0.028$ respectively). There were notable variations in the distribution of MAFs in both African-American (the LA study) and Chinese (the Taixing study) when compared to Whites (the LA study) for all SNPs (rs1447295: $p<0.0001$ and $=0.014$; rs16901979: $p<0.0001$ and <0.0001; rs6983267: $p<0.0001$ and $=0.0016$, respectively).

Table 2 presents ORs and 95% CIs for rs1447295, rs16901979, and rs6983267. After initial analyses by genotyping of each SNP, we determined the dominant model was appropriate for rs1447295 and rs16901979 in all cancer sites, whereas for rs6983267, the recessive model was appropriate for all sites except stomach and liver cancers, where the dominant model was employed. Using a recessive model and adjusting for potential confounding factors, rs6983267 (region 3) was positively associated with UADT cancers (OR_{adj}=1.69, 95% CI=1.28, 2.24). When stratified by tumor site, rs6983267 was associated with cancers of the oropharynx (OR_{adj}=1.80 95% CI=1.30, 2.49) and larynx (OR_{adj}=2.04 95% CI=1.12, 3.72). Using the dominant model, there was a suggestive positive association between rs6983267 and liver cancer (OR_{adj} =1.51, 95% CI=0.99, 2.31). Lastly, in a pilot study, using the dominant model, we observed an inverse association of rs16901979 (region 2) with kidney cancer (OR=0.48, 95% CI=0.23, 1.00, data not shown). No obvious associations were observed between rs1447295 (region 1) and each smoking-related cancer.

Table 3 shows the adjusted ORs for all three SNPs and cancer sites with at least 75 cases stratified by smoking status (never vs. ever). In assessing the relationship between rs6983267 (region3) and lung cancer stratified by smoking, we observed adjusted ORs of 1.45 (95% CI=1.05-2.00) for ever-smokers and 1.00 (95% CI=0.58-1.70) for never-smokers; suggesting possible interaction between smoking and the SNP rs6983267 on lung cancer. Associations between rs6983267 and UADT cancers were observed in both ever-smokers (OR_{adj}=1.56, 95% CI=1.01, 2.39) and never-smokers (OR_{adj}=1.79, 95% CI=1.23, 2.61), suggesting the SNP rs6983267 may be independent of tobacco smoking for UADT cancers. Among smokers, the SNP rs6983267 was observed

to be positively associated with oropharyngeal cancer (OR_{adj}=2.01, 95% CI=1.29, 3.85) and laryngeal cancer (OR_{adj}=2.05, 95% CI=1.09, 3.85) and inversely associated with bladder cancer (OR_{adj}=0.35, 95% CI=0.14, 0.83).

The SNP rs16901979 (region 2) was positively associated with UADT among never smokers (OR_{adj}=1.86, 95% CI=1.06, 3.28). When stratified by tumor site, rs16901979 was associated the cancer of the oropharynx (OR_{adj}=2.28, 95% CI=1.19, 4.39). Among ever-smokers, no obvious association was observed between rs16901979 and all tumor sites listed in Table 3. For rs1447295 (region 1), when stratified by smoking, the only noteworthy change in odds ratio was found in liver cancer (p=0.025), OR_{adj}=1.96 (95% CI=1.07-3.59) among smokers and 0.90 (95% CI=0.49-1.65) among never-smokers with an adjusted OR for interaction of 1.95 (95% CI: 1.09, 3.51).

Table 4 shows the FPRP for the observed associations presented in Tables 2 and 3. Assuming a prior probability of 0.01, we find two of our observed associations below FPRP threshold of 50%: rs6983267 and UADT cancers has an 11% probability of being a false positive and when stratified by tumor site, cancer of the oropharynx has a 22% probability of being a false positive. If we increase our prior probability to 0.1, the observed associations among ever-smokers and cancers of the lung and UADT, as well as oropharynx also have a less than 50% probability of false positivity.

Discussion

Positive associations were observed between rs6983267 and UADT cancer in this study. When analyzed by genotypes, the GG genotype was strongly associated with the UADT cancer after adjusting for potential confounders. There was a clear dose-response

relationship between rs6983267 and the UADT cancer (p for trend=0.0071). When stratified by tobacco smoking, the adjusted ORs were 1.56 (95% CI=1.01, 2.39) for never-smokers and 1.79 (95% CI=1.23, 2.61) for smokers. There is no clear indication that the rs6983267 modifies the association between tobacco smoking and the UADT cancer, although the point estimate of the adjusted OR was slightly higher among smokers. Among UADT cancer when stratified by tumor site, both cancers of oropharynx and larynx were positively associated with the rs6983267. Similarly, no clear difference of the associations was found between smokers and non-smokers for both tumor sites. Small sample sizes of esophageal and nasopharyngeal cancers did not allow us to evaluate the associations precisely with rs6983267. Although no overall association was observed between rs6983267 and lung cancer, a positive association was found for smokers (OR_{adj} =1.45, 95% CI=1.05-2.00) and a null association for never-smokers, indicating possible effect modification of the rs6983267 on smoking and lung cancer. The SNP rs6983267 was inversely associated with bladder cancer (OR_{adj}=0.52, 95%CI=0.25, 1.07). When stratified by tobacco smoking, the adjusted ORs were 0.35 (95% CI=0.14-0.83) among smokers and 1.16 (95% CI=0.28-4.77) among never smokers, suggesting the possibility of effect modification.

Our observations that rs6983267 was positively associated with UADT cancers, independent of tobacco smoking, positive associated with lung cancer only among smokers, and inversely associated with bladder cancer dependent of tobacco smoking status, implicates this SNP an important candidate marker for smoking related cancers with etiological heterogeneity (43, 44). Our observations suggest that the SNP rs6983267 may play an important role in tobacco-related carcinogenesis involving target specific

carcinogens, including metabolic, DNA repair and other related pathways. Among non-smokers, we observed positive associations between rs16901979 (region 2) and UADT as well as oropharyngeal cancers, and between rs6983267 and UADT cancer. A higher proportion of UADT cancer cases are diagnosed among non-smokers, which may be associated with HPV infection, alcohol drinking, and other factors such as genetic predisposition. Our results indicate that both rs16901979 and rs6983267 may play a role in non-smoking related pathways of UADT cancers.

SNPs of the 8q24 chromosome are notable for their associations in prostate cancer(16-20, 25-27) and increasing evidence with colorectal cancer(25, 29, 30, 45); however, this region is one with few recognized genes and known functionality. 8q24 chromosome is located upstream of c-Myc proto-oncogene and located close to the pseudogene POU5F1P1. To our current knowledge 8q24 SNPs have not been investigated in any of the mentioned nine smoking-related cancer sites. However, the 8q24 chromosome has often been observed to be amplified in liver(9), lung(11), kidney(46), bladder(13, 47), and oral cancers(10, 48), suggesting that our results may not be due to chance. The clear association of 8q24 with prostate cancer suggests a potential hormone-related or other carcinogenic pathways which may be associated with expression of microRNAs in the 8q24 region(49). Our results and those of previous studies shows that SNPs of "region 3" are more often observed to be associated in cancer sites other than prostate, indicating that this specific "region" may be involved in other carcinogenic pathways, such as a tobacco-related carcinogenic pathway, or a combination of different pathways. Recent studies have observed SNPs between 128.47 to 129.54 Mb, i.e. "region 3." to be associated colorectal and ovarian cancers (25, 28-30, 50).

Ghoussaini and colleagues reported this cancer associated "region" may be narrower than previously believed, spanning only 128.47 to 128.50 Mb(50). Further research will be required to determine whether 8q24 loci, specifically "region 3," are associated with smoking-related carcinogenesis. Studies of SNPs in LD with rs6983267, SNPs within "region 3," and those between 128.47 to 128.50 Mb in relation to smoking-related cancers may also be useful to detect new markers and reveal possible underlying biological mechanisms. Lastly, we cannot exclude the possibility that SNPs beyond "region 3" may also be associated with tobacco-related carcinogenesis and that our results for rs6982267 were due to its high MAF providing us with more precision to detect the observed associations. Thus, functional studies and studies with larger sample sizes should be conducted to further investigate the association of these SNPs with smoking-related cancers.

Two minor deviations in HWE were observed (rs16901979 genotype distributions in African-Americans and rs6983267 genotype distribution in the Chinese population); however, the allelic proportions remained consistent with the previously published literature(20, 24). Chance finding, selection bias, or laboratory genotyping error may potentially lead to the HWE deviations. Since we observed a high QC concordance rate for all 3 SNPs (>99%) in our lab, the possibility of genotyping error is unlikely. Controls in both the Los Angeles and Taixing City studies were randomly selected from the population at risk using algorithms to capture an accurate representation of their respective cities (34, 36). After removing the African-American population in our analysis of rs16901979, we observed similar associations. The association between rs6983267 and liver cancer needs investigation by other studies.

Multiple comparison issue may be of concern from multiple testing of tumor sites and SNPs involved in this study. We have performed the false positive reporting probability analyses. Using our FPRP cutoff of 50% and a prior probability of 0.01, it is likely that the observed association between rs6983267 and UADT cancers (FPRP=11%), specifically cancer of the oropharynx (FPRP=22%), is not due to chance from multiple hypothesis testing.

The potential of selection bias may exist due to the poor survival of many of these cancers—liver, esophagus, stomach, and lung (globally, these sites have survival rates <30% (51)). The relatively low case participation rate was due to death before they were interviewed. For instance, among eligible lung cancer cases, 25% died before we could contact them. If 8q24 SNPs played a role in the prognosis of smoking-related cancer sites, selective-survival bias would have affected our observed associations. Because of the lack of studies investigating such effects on the prognosis of these cancers, we were unable to estimate whether such bias was present in this study. The sample sizes in the Taixing study and in the stratified analyses of the LA study may affect the precision of our measurements. As a result, the interval estimates from both the Chinese study (over 200 cases for each site and over 400 controls) and the MSKCC study (172 cases/157 controls) are imprecise. Among esophageal cancer analyses, we combined Los Angeles and Taixing study sites to increase precision and observed no obvious associations after adjusting for potential confounding variables. Strengths of our study include a relatively large sample size in our lung and UADT cancer sites, diverse populations allowing us to investigate ethnic-specific genotype distributions along multiple cancers, and the potential to adjust for a variety of confounders.

In conclusion, our results support the hypothesis that the 8q24 variants, particularly rs6983267, play a role in smoking-related cancer sites, particularly in upper aero-digestive tract cancers and lung cancer among smokers. Laboratory-based functional studies and large epidemiological studies in multiple populations should be conducted to further investigate the association of 8q24 SNPs with smoking-related cancers.

Funding

This work was supported in part by NIH National Institute of Environmental Health Sciences, National Cancer Institute, Department of Health and Human Services, Grants ES06718, ES01167, DA11386, CA90833, CA77954, CA09142, CA96134, the Alper Research Center for Environmental Genomics of the UCLA Jonsson Comprehensive Cancer Center and UCLA Center for Occupational and Environmental Health.

Acknowledgements

We would like the thank Dr. Thomas M. Mack and Dr. Wendy Cozen for their contributions for the Los Angeles population-based lung and UADT case-control study. Also, we are indebted to the studies' participants for their time and dedication.

Reference

1. Danaei, G., Vander Hoorn, S., Lopez, A. D., Murray, C. J. L., and Ezzati, M. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. The Lancet, *366:* 1784-93.

2. IARC. Monographs on the evaluation of carcinogenic risks to humans. Tobacco smoke and involuntary smoking. Lyon: IARC, 2004.

3. Ezzati, M., Henley, S. J., Lopez, A. D., and Thun, M. J. Role of smoking in global and regional cancer epidemiology: current patterns and data needs. Int J Cancer, *116:* 963-71, 2005.

4. Vineis, P., Alavanja, M., Buffler, P., Fontham, E., Franceschi, S., Gao, Y. T., et al. Tobacco and cancer: recent epidemiological evidence. J Natl Cancer Inst, *96:* 99-106, 2004.

5. Peto, R., Darby, S., Deo, H., Silcocks, P., Whitley, E., and Doll, R. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. Bmj, *321:* 323-9, 2000.

6. Zhang, H., and Cai, B. The impact of tobacco on lung health in China. Respirology, *8:* 17-21, 2003.

7. Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med, *343:* 78-85, 2000.

8. Alitalo, K., Schwab, M., Lin, C. C., Varmus, H. E., and Bishop, J. M. Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-myc) in malignant neuroendocrine cells from a human colon carcinoma. Proc Natl Acad Sci U S A, *80:* 1707-11, 1983.

9. Chan, K.-L., Guan, X.-Y., and Ng, I. O.-L. High-throughput tissue microarray analysis of c-myc activation in chronic liver diseases and hepatocellular carcinoma. Human Pathology, *35:* 1324-31, 2004.

10. Garnis, C., Coe, B. P., Ishkanian, A., Zhang, L., Rosin, M. P., and Lam, W. L. Novel regions of amplification on 8q distinct from the MYC locus and frequently altered in oral dysplasia and cancer. Genes Chromosomes Cancer, *39:* 93-8, 2004.

11. Kang, J. U., Koo, S. H., Kwon, K. C., Park, J. W., Shin, S. Y., Kim, J. M., et al. High frequency of genetic alterations in non-small cell lung cancer detected by multi-target fluorescence in situ hybridization. J Korean Med Sci, *22 Suppl:* S47-51, 2007.

12. Sonoda, G., du Manoir, S., Godwin, A. K., Bell, D. W., Liu, Z., Hogan, M., et al. Detection of DNA gains and losses in primary endometrial carcinomas by comparative genomic hybridization. Genes Chromosomes Cancer, *18:* 115-25, 1997.

13. Zaharieva, B., Simon, R., Ruiz, C., Oeggerli, M., Mihatsch, M. J., Gasser, T., et al. High-throughput tissue microarray analysis of CMYC amplificationin urinary bladder cancer. Int J Cancer, *117:* 952-6, 2005.

14. Nessling, M., Richter, K., Schwaenen, C., Roerig, P., Wrobel, G., Wessendorf, S., et al. Candidate genes in breast cancer revealed by microarray-based comparative genomic hybridization of archived tissue. Cancer Res, *65:* 439-47, 2005.

15. Visakorpi, T., Kallioniemi, A. H., Syvanen, A. C., Hyytinen, E. R., Karhu, R., Tammela, T., et al. Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. Cancer Res, *55:* 342-7, 1995.

16. Amundadottir, L. T., Sulem, P., Gudmundsson, J., Helgason, A., Baker, A., Agnarsson, B. A., et al. A common variant associated with prostate cancer in European and African populations. Nat Genet, *38:* 652-8, 2006.

17. Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L. T., Gudbjartsson, D., Helgason, A., et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet, *39:* 631-37, 2007.

18. Yeager, M., Orr, N., Hayes, R. B., Jacobs, K. B., Kraft, P., Wacholder, S., et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet, *39:* 645-49, 2007.

19. Freedman, M. L., Haiman, C. A., Patterson, N., McDonald, G. J., Tandon, A., Waliszewska, A., et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc Natl Acad Sci U S A, *103:* 14068-73, 2006.

20. Haiman, C. A., Patterson, N., Freedman, M. L., Myers, S. R., Pike, M. C., Waliszewska, A., et al. Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet, *39:* 638-44, 2007.

21. Zheng, S. L., Sun, J., Wiklund, F., Smith, S., Stattin, P., Li, G., et al. Cumulative association of five genetic variants with prostate cancer. N Engl J Med, *358:* 910-9, 2008.

22. Sun, J., Lange, E. M., Isaacs, S. D., Liu, W., Wiley, K. E., Lange, L., et al. Chromosome 8q24 risk variants in hereditary and non-hereditary prostate cancer patients. Prostate, 2008.

23. Schumacher, F. R., Feigelson, H. S., Cox, D. G., Haiman, C. A., Albanes, D., Buring, J., et al. A common 8q24 variant in prostate and breast cancer from a large nested case-control study. Cancer Res, *67:* 2951-6, 2007.

24. Robbins, C., Torres, J. B., Hooker, S., Bonilla, C., Hernandez, W., Candreva, A., et al. Confirmation study of prostate cancer risk variants at 8q24 in African Americans identifies a novel risk locus. Genome Res, *17:* 1717-22, 2007.

25. Haiman, C. A., Le Marchand, L., Yamamato, J., Stram, D. O., Sheng, X., Kolonel, L. N., et al. A common genetic risk factor for colorectal and prostate cancer. Nat Genet, *39:* 954-56, 2007.

26. Salinas, C. A., Kwon, E., Carlson, C. S., Koopmeiners, J. S., Feng, Z., Karyadi, D. M., et al. Multiple independent genetic variants in the 8q24 region are associated with prostate cancer risk. Cancer Epidemiol Biomarkers Prev, *17:* 1203-13, 2008.

27. Thomas, G., Jacobs, K. B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet, *40:* 310-5, 2008.

28. Berndt, S. I., Potter, J. D., Hazra, A., Yeager, M., Thomas, G., Makar, K. W., et al. Pooled analysis of genetic variation at chromosome 8q24 and colorectal neoplasia risk. Hum Mol Genet, 2008.

29. Gruber, S. B., Moreno, V., Rozek, L. S., Rennert, H. S., Lejbkowicz, F., Bonner, J. D., et al. Genetic Variation in 8q24 Associated with Risk of Colorectal Cancer. Cancer Biol Ther, *6*, 2007.

30. Li, L., Plummer, S. J., Thompson, C. L., Merkulova, A., Acheson, L. S., Tucker, T. C., et al. A common 8q24 variant and the risk of colon cancer: a population-based case-control study. Cancer Epidemiol Biomarkers Prev, *17:* 339-42, 2008.

31. Setiawan, V. W., Ursin, G., Horn-Ross, P. L., Van Den Berg, D., Le Marchand, L., Henderson, B. E., et al. Germ line variation at 8q24 and endometrial cancer risk. Cancer Epidemiol Biomarkers Prev, *16:* 2166-8, 2007.

32. Cook, M. B., Graubard, B. I., Quraishi, S. M., Yeager, M., Chanock, S. J., Crenshaw, A., et al. Genetic variants in the 8q24 locus and risk of testicular germ cell tumors. Hum Genet, *123:* 409-18, 2008.

33. Giovannucci, E., and Martinez, M. E. Tobacco, colorectal cancer, and adenomas: a review of the evidence. J Natl Cancer Inst, *88:* 1717-30, 1996.

34. Cui, Y., Morgenstern, H., Greenland, S., Tashkin, D. P., Mao, J., Cao, W., et al. Polymorphism of Xeroderma Pigmentosum group G and the risk of lung cancer and squamous cell carcinomas of the oropharynx, larynx and esophagus. Int J Cancer, *118:* 714-20, 2006.

35. Hashibe, M., Morgenstern, H., Cui, Y., Tashkin, D. P., Zhang, Z. F., Cozen, W., et al. Marijuana use and the risk of lung and upper aerodigestive tract cancers: results of a population-based case-control study. Cancer Epidemiol Biomarkers Prev, *15:* 1829-34, 2006.

36. Mu, L. N., Lu, Q. Y., Yu, S. Z., Jiang, Q. W., Cao, W., You, N. C., et al. Green tea drinking and multigenetic index on the risk of stomach cancer in a Chinese population. Int J Cancer, *116:* 972-83, 2005.

37. Mu, L. N., Cao, W., Zhang, Z. F., Cai, L., Jiang, Q. W., You, N. C., et al. Methylenetetrahydrofolate reductase (MTHFR) C677T and A1298C polymorphisms and the risk of primary hepatocellular carcinoma (HCC) in a Chinese population. Cancer Causes Control, *18:* 665-75, 2007.

38. Lu, H., Cai, L., Mu, L. N., Lu, Q. Y., Zhao, J., Cui, Y., et al. Dietary mineral and trace element intake and squamous cell carcinoma of the esophagus in a Chinese population. Nutr Cancer, *55:* 63-70, 2006.

39. Cao, W., Cai, L., Rao, J. Y., Pantuck, A., Lu, M. L., Dalbagni, G., et al. Tobacco smoking, GSTP1 polymorphism, and bladder carcinoma. Cancer, *104:* 2400-8, 2005.

40. Hung, R. J., Zhang, Z. F., Rao, J. Y., Pantuck, A., Reuter, V. E., Heber, D., et al. Protective effects of plasma carotenoids on the risk of bladder cancer. J Urol, *176:* 1192-7, 2006.

41. Cao, W., Hashibe, M., Rao, J. Y., Morgenstern, H., and Zhang, Z. F. Comparison of methods for DNA extraction from paraffin-embedded tissues and buccal cells. Cancer Detect Prev, *27:* 397-404, 2003.

42. Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L., and Rothman, N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst, *96:* 434-42, 2004.

43. Dobrossy, L. Epidemiology of head and neck cancer: magnitude of the problem. Cancer Metastasis Rev, *24:* 9-17, 2005.

44. Gallo, A., and Cha, C. Updates on esophageal and gastric cancers. World J Gastroenterol, *12:* 3237-42, 2006.

45. Poynter, J. N., Figueiredo, J. C., Conti, D. V., Kennedy, K., Gallinger, S., Siegmund, K. D., et al. Variants on 9p24 and 8q24 are associated with risk of colorectal cancer: results from the Colon Cancer Family Registry. Cancer Res, *67:* 11128-32, 2007.

46. Furge, K. A., Chen, J., Koeman, J., Swiatek, P., Dykema, K., Lucin, K., et al. Detection of DNA copy number changes and oncogenic signaling abnormalities from

gene expression data reveals MYC activation in high-grade papillary renal cell carcinoma. Cancer Res, *67:* 3171-6, 2007.

47. Visapaa, H., Seligson, D., Eeva, M., Gaber, F., Rao, J., Belldegrun, A., et al. 8q24 amplification in transitional cell carcinoma of bladder. Appl Immunohistochem Mol Morphol, *11:* 33-6, 2003.

48. Wolff, E., Girod, S., Liehr, T., Vorderwulbecke, U., Ries, J., Steininger, H., et al. Oral squamous cell carcinomas are characterized by a rather uniform pattern of genomic imbalances detected by comparative genomic hybridisation. Oral Oncol, *34:* 186-90, 1998.

49. Huppi, K., Volfovsky, N., Runfola, T., Jones, T. L., Mackiewicz, M., Martin, S. E., et al. The Identification of MicroRNAs in a Genomically Unstable Region of Human Chromosome 8q24. Molecular Cancer Research, *6:* 212-21, 2008.

50. Ghoussaini, M., Song, H., Koessler, T., Al Olama, A. A., Kote-Jarai, Z., Driver, K. E., et al. Multiple Loci with different cancer specificities within the 8q24 gene desert. J Natl Cancer Inst, *100:* 962-6, 2008.

51. Parkin, D. M., Bray, F., Ferlay, J., and Pisani, P. Global cancer statistics, 2002. CA Cancer J Clin, *55:* 74-108, 2005.

Table 1. Genotype and Allele frequencies of 8q24 variants, stratified by Ethnicity and Study.

Study	White (LA study)	White-only (MSKCC Study)	African American (LA study)	Mexican (LA study)	Asian American (LA study)	Chinese (Taixing City Study)
Variable	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
rs1447295						
CC	457 (79.2)	140 (89.2)	42 (53.2)	113 (81.3)	30 (58.8)	276 (71.1)
CA	113 (19.6)	17 (10.8)	30 (38.0)	25 (18.0)	19 (37.3)	101 (26.0)
AA	7 (1.2)	0	7 (8.9)	1 (0.7)	2 (3.9)	11 (2.8)
C	513.5 (89.0)	140.5 (94.6)	57 (72.2)	125.5 (90.3)	39.5 (77.5)	326.5 (84.2)
A	63.5 (11.0)	8.5 (5.4)	22 (27.8)	13.5 (9.7)	11.5 (22.5)	21.5 (15.9)
P-value*	0.996	0.473	0.625	0.763	0.634	0.634
rs16901979						
CC	531 (92.0)	149 (94.3)	23 (29.1)	128 (92.8)	33 (64.7)	207 (54.5)
CA	44 (7.6)	9 (5.7)	48 (60.8)	9 (6.5)	13 (25.5)	143 (37.6)
AA	2 (0.4)	0	8 (10.1)	1 (7)	5 (9.8)	30 (7.9)
C	553 (95.8)	153.5 (97.2)	47 (59.5)	132.5 (96.0)	39.5 (77.5)	278.5 (73.3)
A	24 (4.2)	4.5 (2.8)	32 (40.5)	5.5 (4.0)	11.5 (22.5	101.5 (26.7)
P-value*	0.30	0.713	0.021	0.082	0.053	0.449
rs6983267						
TT	138 (24.0)	34 (21.7)	1 (1.3)	27 (19.6)	17 (32.7)	146 (37.9)
TG	287 (49.9)	73 (46.5)	23 (29.1)	61 (44.2)	28 (53.9)	165 (42.9)
GG	150 (26.1)	50 (31.9)	55 (69.6)	50 (36.2)	7 (13.5)	74 (19.2)
T	281.5 (49.0)	70.5 (44.9)	12 (15.8)	57.5 (41.7)	31 (59.5)	228.5 (59.4)
G	293.5 (51.0)	86.5 (55.1)	66.5 (84.2)	80.5 (58.3)	21 (40.4)	156.5 (40.7)
P-value*	0.975	0.450	0.409	0.287	0.394	0.028

* Hardy-Weinberg Test

Table 2. Association between 8q24 SNPs and 9 smoking-related cancer sites (UADT = upper aero-digestive tract cancers)

8q24 SNPs	rs1447295 C>A			rs16901979 C>A				rs6983267 G>T			
Cancer Site genotype	Case/ control	OR_{crude} (95% CI)	OR_{adj} (95% CI)*	genotype	Case/ control	OR_{crude} (95% CI)	OR_{adj} (95% CI)*	genotype	Case/ control	OR_{crude} (95% CI)	OR_{adj} (95% CI)*
LA study											
Lung											
CC	403/714	1.00	1.00	CC	429/786	1.00	1.00	TT	123/194	1.00	1.00
CA	124/197	1.12 (0.86, 1.44)	0.94 (0.69, 1.28)	CA	85/127	1.23 (0.91, 1.65)	0.79 (0.52, 1.18)	TG	225/441	0.81 (0.61, 1.06)	0.78 (0.56, 1.10)
AA	13/20	1.15 (0.57, 2.34)	0.73 (0.31, 1.72)	AA	16/19	2.18 (1.11, 4.27)	1.25 (0.55, 2.85)	GG	192/291	1.04 (0.78, 1.39)	1.02 (0.70, 1.58)
Ptrend		0.38	0.49			0.016	0.68			0.56	0.80
CA&AA	137/217	1.12 (0.87, 1.43)	0.92 (0.68, 1.24)	CA&AA	104/143	1.33 (1.01, 1.76)	0.84 (0.57, 1.23)	GG**	192/291	1.20 (0.96, 1.51)	1.21 (0.90, 1.61)
UADT (squamous)											
CC	301/714	1.00	1.00	CC	314/785	1.00	1.00	TT	78/194	1.00	1.00
CA	82/196	1.11 (0.84, 1.48)	1.00 (0.73, 1.36)	CA	73/127	1.44 (1.05, 1.97)	1.31 (0.89, 1.93)	TG	154/441	0.87 (0.63, 1.20)	0.87 (0.61, 1.23)
AA	8/20	0.95 (0.41, 2.18)	0.67 (0.26, 1.70)	AA	14/16	2.19 (1.06, 4.54)	1.65 (0.71, 3.82)	GG	168/290	**1.44 (1.04, 1.99)**	**1.53 (1.06, 2.21)**
Ptrend		0.59	0.66			0.0030	0.11			0.0056	0.0071
CA&AA	90/216	0.93 (0.41, 2.12)	0.67 (0.27, 1.69)	CA&AA	87/143	1.52 (1.13, 2.05)	1.34 (0.92, 1.95)	GG**	168/290	**1.59 (1.24, 2.02)**	**1.69 (1.28, 2.24)**
UADT stratified											
Oropharynx											
CC	183/714	1.00	1.00	CC	193/785	1.00	1.00	TT	51/194	1.00	1.00
CA	51/196	1.02 (0.72, 1.44)	0.96 (0.66, 1.40)	CA	41/127	1.31 (0.89, 1.93)	1.45 (0.91, 2.31)	TG	86/441	0.74 (0.51, 1.09)	0.75 (0.50, 1.13)
AA	4/20	0.78 (0.26, 2.31)	0.68 (0.22, 2.13)	AA	5/16	1.27 (0.46, 3.51)	1.18 (0.39, 3.58)	GG	100/290	1.31 (0.89, 1.92)	1.48 (0.97, 2.26)
Ptrend		0.87	0.61			0.18	0.19			0.052	0.024
CA&AA	55/216	0.99 (0.71, 1.39)	0.94 (0.65, 1.35)	CA&AA	46/143	1.31 (0.91, 1.89)	1.42 (0.91, 2.23)	GG**	100/290	**1.60 (1.19, 2.14)**	**1.80 (1.30, 2.49)**
Larynx											
CC	55/714	1.00	1.00	CC	59/785	1.00	1.00	TT	14/194	1.00	1.00
CA	21/196	1.39 (0.82, 2.36)	1.65 (0.92, 2.97)	CA	13/127	1.36 (0.73, 2.56)	0.78 (0.32, 1.92)	TG	25/441	0.79 (0.40, 1.54)	0.68 (0.32, 1.47)
AA	2/20	1.30 (0.30, 5.70)	0.69 (0.08, 5.60)	AA	5/16	4.16 (1.47, 11.75)	2.79 (0.68, 11.54)	GG	37/290	1.77 (0.93, 3.36)	1.56 (0.71, 3.46)
Ptrend		0.24	0.90			0.016	0.55			0.022	0.14
CA&AA	23/216	1.38 (0.83, 2.30)	1.55 (0.88, 2.75)	CA&AA	18/143	1.68 (0.96, 2.92)	0.95 (0.41, 2.19)	GG**	37/290	**2.08 (1.30, 3.33)**	**2.04 (1.12, 3.72)**
Nasopharynx											
CC	29/714	1.00	1.00	CC	24/785	1.00	1.00	TT	6/194	1.00	1.00
CA	8/196	1.01 (0.45, 2.23)	0.76 (0.32, 1.81)	CA	12/127	3.09 (1.51, 6.34)	1.88 (0.80, 4.43)	TG	23/441	1.69 (0.68, 4.21)	1.75 (0.66, 4.65)
AA	2/20	2.46 (0.55, 11.0)	1.64 (0.32, 8.51)	AA	3/16	6.13 (1.68, 22.5)	2.48 (0.53, 11.5)	GG	11/290	1.23 (0.45, 3.37)	1.65 (0.54, 5.03)
Ptrend		0.50	0.93			0.00012	0.11			0.85	0.42
CA&AA	10/216	1.14 (0.55, 2.38)	0.85 (0.38, 1.91)	CA&AA	15/143	3.43 (1.76, 6.70)	1.96 (0.87, 4.45)	TG&GG	34/731	1.50 (0.62, 3.63)	1.72 (0.67, 4.45)
Esophagus (squamous)											
CC	22/714	1.00	1.00	CC	23/785	1.00	1.00	TT	5/194	1.00	1.00

CA	8/196	1.33 (0.58, 3.02)	1.43 (0.58, 3.47)	CA	6/127	1.61 (0.64, 4.04)	1.28 (0.42, 3.86)	TG	13/441	1.14 (0.40, 3.25)	1.21 (0.40, 3.68)
AA	0/20	--	--	AA	1/16	2.13 (0.27, 16.8)	1.48 (0.16, 14.1)	GG	13/290	1.74 (0.61, 3.25)	2.49 (0.75, 8.24)
Ptrend		0.89	0.90			0.23	0.62			0.24	0.10
CA&AA	8/216	1.20 (0.53, 2.74)	1.26 (0.52, 3.09)	CA&AA	7/143	1.57 (0.70, 3.97)	1.30 (0.45, 3.78)	GG**	13/290	1.58 (0.77, 3.27)	2.16 (0.93, 5.06)
Other (squamous)[†]											
CC	12/714	1.00	1.00	CC	15/785	1.00	1.00	TT	2/194	1.00	1.00
CA	4/196	1.21 (0.38, 3.81)	1.45 (0.44, 4.78)	CA	1/127	0.42 (0.05, 3.15)	0.43 (0.04, 4.40)	TG	7/441	1.54 (0.32, 7.48)	1.26 (0.25, 6.38)
AA	0/20	--	--	AA	0/16	--	--	GG	7/290	2.34 (0.48, 11.4)	2.26 (0.44, 11.5)
Ptrend		0.98	0.75			0.32	0.42			0.25	0.25
CA&AA	4/216	1.10 (0.35, 3.45)	1.35 (0.41, 4.47)	CA&AA	1/143	0.37 (0.05, 2.79)	0.40 (0.04, 4.13)	TG&GG	14/731	1.86 (0.42, 8.24)	1.62 (0.35, 7.40)
Esophagus (adenocarcinoma)											
CC	50/714	1.00	1.00	CC	60/785	1.00	1.00	TT	14/194	1.00	1.00
CA	16/196	1.17 (0.65, 2.09)	1.23 (0.66, 2.29)	CA	7/127	0.72 (0.32, 1.61)	1.03 (0.43, 2.50)	TG	37/441	1.16 (0.61, 2.2)	1.22 (0.63, 2.37)
AA	1/20	0.72 (0.09, 5.43)	0.65 (0.08, 5.50)	AA	0/16	--	--	GG	15/290	0.72 (0.34, 1.52)	0.95 (0.43, 2.10)
Ptrend		0.81	0.76			0.21	0.74			0.33	0.89
CA&AA	17/216	1.12 (0.64, 1.99)	1.17 (0.64, 2.15)	CA&AA	7/143	0.64 (0.28, 1.43)	0.95 (0.39, 2.29)	GG**	15/290	0.64 (0.36, 1.16)	0.83 (0.44, 1.55)
Taixing Study											
Esophagus											
CC	137/276	1.00	1.00	CC	112/207	1.00	1.00	TT	66/146	1.00	1.00
CA	59/101	1.18 (0.80, 1.72)	1.18 (0.79, 1.74)	CA	74/143	0.96 (0.67, 1.38)	0.96 (0.66, 1.40)	TG	95/165	1.27 (0.87, 1.87)	1.18 (0.79, 1.75)
AA	2/11	0.37 (0.08, 1.68)	0.41 (0.09, 1.91)	AA	14/30	0.86 (0.44, 1.69)	1.00 (0.50, 1.99)	GG	40/74	1.20 (0.74, 1.94)	1.06 (0.64, 1.76)
Ptrend		0.98	0.92			0.66	0.89			0.36	0.70
CA&AA	61/112	1.10 (0.76, 1.59)	1.10 (0.75, 1.63)	CA&AA	88/112	0.94 (0.67, 1.33)	0.97 (0.68, 1.38)	GG**	40/74	1.04 (0.68, 1.60)	0.97 (0.62, 1.52)
Stomach											
CC	140/276	1.00	1.00	CC	107/207	1.00	1.00	TT	61/146	1.00	1.00
CA	39/101	0.76 (0.50, 1.16)	0.80 (0.51, 1.24)	CA	69/143	0.93 (0.65, 1.35)	0.96 (0.65, 1.42)	TG	94/165	1.36 (0.92, 2.02)	1.21 (0.79, 1.83)
AA	8/11	1.43 (0.56, 3.65)	1.57 (0.56, 4.40)	AA	16/30	1.03 (0.54, 1.98)	1.26 (0.62, 2.55)	GG	32/74	1.04 (0.62, 1.73)	0.88 (0.51, 1.51)
Ptrend		0.63	0.79			0.89	0.76			0.62	0.82
CA&AA	47/112	0.83 (0.56, 1.23)	0.86 (0.57, 1.32)	CA&AA	85/173	0.95 (0.67, 1.35)	1.00 (0.69, 1.45)	TG&GG	126/239	1.26 (0.87, 1.82)	1.10 (0.74, 1.63)
Liver											
CC	128/276	1.00	1.00	CC	99/207	1.00	1.00	TT	54/146	1.00	1.00
CA	52/101	1.11 (0.75, 1.65)	1.30 (0.84, 2.02)	CA	74/143	1.08 (0.75, 1.56)	1.20 (0.80, 1.81)	TG	88/165	1.44 (0.96, 2.16)	1.51 (0.96, 2.38)
AA	7/11	1.37 (0.52, 3.62)	1.36 (0.47 3.93)	AA	14/30	0.98 (0.50, 1.92)	1.13 (0.53, 2.43)	GG	45/74	1.64 (1.01, 2.67)	1.54 (0.89, 2.65)
Ptrend		0.45	0.22			0.84	0.46			0.034	0.091
CA&AA	59/112	1.14 (0.78, 1.66)	1.31 (0.86, 1.99)	CA&AA	88/173	1.06 (0.75, 1.51)	1.19 (0.80, 1.76)	TG&GG	134/239	1.51 (1.03, 2.19)	**1.51 (0.99, 2.31)**
MSKCC											
Bladder											
CC	152/140	1.00	1.00	CC	154/149	1.00	1.00	TT	37/34	1.00	1.00
CA	18/17	0.98 (0.48, 1.97)	0.59 (0.21, 1.64)	CA	17/9	1.83 (0.79, 4.23)	2.31 (0.74, 7.23)	TG	90/73	1.13 (0.65, 1.98)	0.92 (0.41, 2.09)
AA	2/0	--		AA	0/0	--	--	GG	44/50	0.81 (0.44, 1.50)	0.49 (0.19, 1.25)
Ptrend		0.60	0.47			0.16	0.15			0.43	0.11

CA&AA	20/17	1.08 (0.55, 2.15)	0.63 (0.23, 1.72)	CA&AA	--	--	GG**	44/50	0.74 (0.46, 1.20)	0.52 (0.25, 1.07)

* lung cancer adjusted for gender, smoking, education, race, and age; UADTs adjusted for gender, smoking, education, race, age, and alcohol drinking; Taixing city esophagus adjusted for: gender, smoking, education, age, alcohol drinking; stomach cancer adjusted for age, gender, smoking pack-year, alcohol drinking and *H.Pylori*; infection; liver cancer adjusted for age, gender, smoking pack-year, alcohol drinking, and HBsAg status; bladder cancer adjusted for gender, smoking, race, and age.

** For rs6983267 TT & TG served as the referent in the recessive model for the following sites: lung, UADT (combined), oropharynx, larynx, esophagus (squamous and adenocarcinoma), bladder, and kidney.

†Other (squamous) are sites 30.0, 31.1, and 31.1 as defined by ICD-O-2.

Table 3. Association between 8q24 SNPs and 7 smoking-related cancer sites, stratified by smoking status (UADT = upper aero-digestive tract cancers)

8q24 SNPs	rs1447295 C>A			rs16901979 C>A			rs6983267 G>T		
Cancer Site	genotype	Never Smoker OR_{adj} (95% CI)	Ever-Smoker OR_{adj} (95% CI)	genotype	Never Smoker OR_{adj} (95% CI)	Ever-Smoker OR_{adj} (95% CI)	genotype	Never Smoker OR_{adj} (95% CI)	Ever-Smoker OR_{adj} (95% CI)
Lung	CC	1.00	1.00	CC	1.00	1.00	TT & GT	1.00	1.00
	CA & AA	0.72 (0.39, 1.34)	0.93 (0.67, 1.28)	CA & AA	1.01 (0.49, 2.06)	0.79 (0.52, 1.21)	GG	1.00 (0.58, 1.70)	**1.45 (1.05, 2.00)**
	p-value	0.30	0.64	p-value	0.99	0.28	p-value	0.99	0.024
UADT	CC	1.00	1.00	CC	1.00	1.00	TT & GT	1.00	1.00
(squamous)	CA & AA	1.10 (0.67, 1.78)	0.89 (0.60, 1.31)	CC & CA	**1.86 (1.06, 3.28)**	1.07 (0.64, 1.79)	GG	**1.56 (1.01, 2.39)**	**1.79 (1.23, 2.61)**
	p-value	0.71	0.54	p-value	0.032	0.81	p-value	0.045	0.0025
Oropharynx	CC	1.00	1.00	CC	1.00	1.00	TT & GT	1.00	1.00
	CA & AA	1.06 (0.60, 1.89)	0.88 (0.55, 1.41)	CA & AA	**2.28 (1.19, 4.39)**	1.03 (0.55, 1.92)	GG	1.56 (0.94, 2.57)	**2.01 (1.29, 3.11)**
	p-value	0.833	0.590	p-value	0.014	0.934	p-value	0.084	0.0017
Larynx	CC	1.00	1.00	CC	1.00	1.00	TT & GT	1.00	1.00
	CA & AA	1.61 (0.35, 7.51)	0.98 (0.54, 1.85)	CA & AA	1.88 (0.27, 12.93)	0.87 (0.35, 2.17)	GG	2.81(0.57, 13.77)	**2.05 (1.09, 3.85)**
	p-value	0.55	0.94	p-value	0.52	0.77	p-value	0.20	0.025
Esophagus-	CC	1.00	1.00	CC	1.00	1.00	TT & GT	1.00	1.00
Taixing	CA & AA	0.91 (0.52, 1.60)	1.41 (0.82, 2.41)	CA & AA	1.03 (0.61, 1.76)	0.91 (0.56, 1.49)	GG	0.98 (0.51, 1.89)	0.88 (0.47, 1.63)
	p-value	0.75	0.22	p-value	0.91	0.72	p-value	0.95	0.68
Stomach	CC	1.00	1.00	CC	1.00	1.00	TT	1.00	1.00
	CA & AA	0.54 (0.29, 1.01)	1.38 (0.77, 2.48)	CA & AA	1.00 (0.58, 1.72)	1.04 (0.62, 1.73)	GT & GG	0.89 (0.50, 1.57)	1.28 (0.73, 2.23)
	p-value	0.053	0.28	p-value	0.99	0.89	p-value	0.68	0.39
Liver	CC	1.00	1.00	CC	1.00	1.00	TT	1.00	1.00
	CA & AA	0.90 (0.49, 1.65)	**1.96 (1.07, 3.59)**	CA & AA	1.27 (0.72, 2.26)	1.11 (0.64, 1.91)	GT & GG	1.68 (0.88, 3.21)	1.34 (0.76, 2.36)
	p-value	0.73	0.030	p-value	0.41	0.74	p-value	0.12	0.32
Bladder	CC	1.00	1.00	CC	1.00	1.00	TT & GT	1.00	1.00
	CA & AA	1.95 (0.37, 10.1)	0.34 (0.10, 1.14)	CA & AA	5.14 (0.69, 38.3)	1.46 (0.38, 5.62)	GG	1.16 (0.28, 4.77)	**0.35 (0.14, 0.83)**
	p-value	0.43	0.080	p-value	0.11	0.58	p-value	0.83	0.017

* lung cancer adjusted for gender, education, race, and age; UADTs adjusted for gender, smoking, education, race, age, and alcohol drinking; Taixing city esophagus adjusted for: gender, education, age, and alcohol drinking; stomach cancer adjusted for age, gender, smoking pack-year, alcohol drinking and *H.Pylori*; infection; liver cancer adjusted for age, gender, alcohol drinking, and HBsAg status; bladder cancer adjusted for gender, race, and age; kidney cancer adjusted for gender, and age.

Table 4. False positive report probability (FPRP) values for associations between 8q24 variants and smoking-related cancer sites

					Prior probability				
SNP	Stratum	OR_{adj} (95% CI)*	Power†	Reported *p-values**	0.5	0.25	0.1	0.01	0.001
rs6983267	UADT	1.69 (1.28, 2.24)	0.90	0.00021	**0.0013**	**0.0038**	**0.011**	**0.11**	0.56
rs6983267	Oral pharynx	1.80 (1.30, 2.49)	0.76	0.00042	**0.0028**	**0.0085**	**0.025**	**0.22**	1.0
rs6983267	Larynx	2.04 (1.12, 3.72)	0.43	0.020	**0.11**	**0.28**	0.53	0.93	0.99
rs6983267	Liver	1.51 (0.99, 2.31)	0.54	0.034	**0.11**	**0.26**	0.52	0.92	0.99
rs16901979	UADT never-smokers	1.86 (1.06, 3.28)	0.33	0.032	**0.12**	**0.30**	0.56	0.93	0.99
rs16901979	Oral-pharynx never-smokers	2.28 (1.19, 4.39)	0.27	0.014	**0.12**	**0.28**	0.54	0.93	0.99
rs6983267	Lung ever-smokers	1.45 (1.05, 2.00)	0.83	0.024	**0.039**	**0.11**	**0.27**	0.80	0.99
rs6983267	UADT never-smokers	1.56 (1.01, 2.39)	0.50	0.045	**0.087**	**0.22**	**0.46**	0.91	0.99
rs6983267	UADT ever-smokers	1.79 (1.23, 2.61)	0.74	0.002	**0.014**	**0.040**	**0.11**	0.58	0.93
rs6983267	Oral-pharynx ever-smokers	2.01 (1.29, 3.11)	0.55	0.002	**0.018**	**0.052**	**0.14**	0.64	0.95
rs6983267	Larynx ever-smokers	2.05 (1.09, 3.85)	0.27	0.25	**0.13**	**0.32**	0.58	0.94	0.99
rs1447295	Liver, ever-smokers	1.96 (1.07, 3.59)	0.30**	0.032	**0.13**	**0.32**	0.58	0.94	0.99
rs6983267	Bladder, ever-smokers	0.35 (0.14, 0.83)	0.12	0.017	**0.19**	**0.42**	0.68	0.96	1.0

*Odds ratios from tables 2 and 3.

†Statistical power calculated using the recessive model, except where noted, is the power to detect an odds ratio of 1.5 or 0.67, at α level of 0.05.

** Statistical power calculated using the dominant model.

Supplementary Table 1. Baseline characteristics of cases and controls from the LA study, Taixing City study, and MSKCC study

	LA Study			Taixing City Study				MSKCC study	
	Lung Cancer Cases (%)	UADT cancer Cases (%)	Controls (%)	Stomach Cancer Cases (%)	Esophageal Cancer Cases (%)	Liver Cancer Cases (%)	Controls (%)	Bladder Cancer Cases (%)	Controls (%)
Total	611	601	1040	206	218		415	233	204
Age range	32-59	20-59	17-65	30-82	30 – 84	22-83	21-84	32-84	17-80
Age, mean	52.2	50.3	49.9	61.5	60.6	53.8	57.7	64.8	42.0
Gender									
Males	303 (49.6)	391 (74.2)	623 (59.9)	138 (67.0)	141 (64.7)	159 (77.9)	287 (69.2)	206 (83.4)	156 (77.2)
Females	308 (50.4)	136 (25.8)	417 (40.1)	68 (33.0)	77 (35.3)	45 (22.1)	128 (30.8)	41 (16.6)	46 (22.8)
Education									
≤ High school	265 (43.4)	240 (45.5)	300 (28.9)	204 (99.5)	215 (100.0)	204 (100.0)	405 (97.6)	95 (40.8)	34 (16.7)
>High School	346 (56.6)	287 (54.5)	739 (71.1)	1 (0.5)	0 (0.0)	0 (0.0)	10 (2.4)	138 (59.2)	170 (83.3)
Smoking									
Never	110 (18.0)	164 (31.1)	491 (47.3)	92 (45.8)	94 (43.1)	85 (44.3)	217 (52.4)	42 (17.3)	92 (46.0)
Ever	501 (82.0)	363 (68.9)	548 (52.7)	109 (54.2)	117 (53.7)	107 (55.7)	197 (47.9)	201 (82.7)	108 (54)