Associations between Polymorphisms and Haplotypes in the UDP-Glucuronosyl Transferase 1A Gene Family with Lung Cancer Risk

Kathleen M Schieffer1,*, Andrea Y Angstadt2,*, Junjia Zhu1, Philip Lazarus3 and Carla J Gallagher4

1Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA 17033, USA
2Department of Pharmaceutical Sciences, Washington State University College of Pharmacy, Spokane, WA 99201, USA
3Department of Pharmacology, Penn State College of Medicine, Hershey, PA 17033, USA
4Next Generation Sequencing, Eurofins MWG Operon LLC, Louisville, KY 40204, USA

Abstract

Title: Associations between polymorphisms and haplotypes in the UDP-Glucuronosyl transferase 1A gene family with lung cancer risk.

Background: Over 5,000 compounds have been identified in cigarette smoke, of which 73 are considered carcinogenic to either laboratory animals or humans by the International Agency for Research on Cancer. The UDP-glucuronosyl transferases (UGTs) are Phase II drug-metabolizing enzymes that catalyze the metabolism of several cigarette smoke carcinogens for excretion. There are hundreds of genetic variants that span the nine genes in the UGT1A gene cluster, but the effect of UGT1A polymorphisms on lung cancer risk has not been studied in American Caucasians. This study hypothesized that UGT1A variants, either individually or as a haplotype, would be associated with lung cancer risk in American Caucasians.

Methods and Findings: To examine the effect of genetic variation in the nine UGT1A genes on lung cancer risk, a comprehensive association and haplotype study was conducted using American Caucasian lung cancer cases and matched healthy controls on ninety-six tagSNPs. Known lung cancer risk factors including age, sex, smoking status, and pack-years of smoking were controlled for in all analyses. Multiple SNPs in the UGT1A gene cluster were associated with lung cancer risk in various stratified analyses before a multiple testing correction was applied. A significant association was found between two haplotypes in the UGT1A haplotype block 2 (rs7569014, rs7421795, rs1817154) and small cell carcinoma risk in American Caucasians. Multiple SNPs were found to affect lung cancer risk that did not remain significant following multiple testing correction. The UGT1A haplotype block 2 may be associated with small cell lung carcinoma. However, single SNP association studies did not yield significant results after multiple testing correction.

Conclusions: UGT1A variants may play only a minor role in other lung cancer risk in American Caucasians; however, this needs to be confirmed in larger studies.

Keywords: Lung; Cancer; UGT; Carcinogen metabolism

Introduction

Lung cancer is the leading cause of cancer mortality in the United States, accounting for more deaths than the next three most common cancers combined [1]. Mortality due to lung cancer has increased in the United States about 4.3% between 1999 and 2010 [1]. It is estimated that tobacco smoke causes about 80% of all lung cancers, however, only about 15% of lifetime smokers develop this malignancy [2-4]. Heritable traits in the population may explain why only a percentage of smokers develop lung cancer [5]. Therefore, identifying genetic markers of high-risk may allow individuals to be targeted for lung cancer screening or prevention programs. However, the lack of genetic markers for lung cancer risk is evident. This study sought to identify genetic variants involved in carcinogen detoxification. The International Agency for Research on Cancer has identified 73 of the ~5,000 compounds in cigarette smoke to be carcinogenic to laboratory animals or humans and many of these carcinogens are substrates for the phase II drug-metabolizing enzymes UDP-glucuronosyl transferases (UGTs) [6-9].

The UGT gene family (comprised of UGT1 and UGT2) catalyzes the conversion of tobacco smoke carcinogens to water-soluble compounds that can be readily excreted from the body [10]. The UGT1A gene locus is located on chromosome 2q37 and nine functional proteins are coded from the UGT1A locus: UGT1A1, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9, and UGT1A10 [10]. There are hundreds of UGT1A genetic variants but association with lung cancer risk is not well-studied. Many of the UGT1A gene variants exhibit allele frequencies up to 40-50% in the general population and are found to be in high linkage disequilibrium (LD) [11]. Mediated by the UGT family of enzymes, glucuronidation is the major detoxification pathway for several tobacco smoke carcinogens, [12-19]. All tested UGTs except UGT1A4, UGT1A6, and UGT2B4 exhibit glucuronidation activity against a variety of benzo(a)pyrene (BaP) metabolites [15,18,19]. In particular, UGT1A10 exhibits high glucuronidation activity against many BaP metabolites [18]. UGT1A7 and UGT1A10 are well expressed in aerodigestive tract and UGT1A7, UGT1A8, and UGT1A10 exhibited glucuronidation activity against 7-OH-BaP [20]. Almost all UGT1A enzymes (except UGT1A6) exhibit glucuronidation activity against 2-amino-1-methyl-6-phenylimidazo (4,5-b) pyridine (PhIP) and its bioactivated metabolite (N-OH-PhIP), with UGT1A10 exhibiting the highest activity [19]. UGTs are the primary mechanism of detoxification
for tobacco-specific nitrosoamines (TSNAs), with five UGTs displaying glucuronidation activity against 4-(methyl nitrosoamino)-1(3-pyridyl)-1-butanol (NNAL) [13]. Therefore, UGT1A genetic variants are hypothesized to increase lung cancer risk due to their involvement in detoxification of multiple tobacco smoke carcinogens [21-23].

Despite the prominent role in detoxification of many tobacco smoke carcinogens, few studies have examined genetic variation in UGTs for association with lung cancer risk. GWAS studies are not well-powered to detect variations in UGT genes since high homology leads to difficulty designing assays. Genetic association studies should be performed to address the potential association between these genes and disease. A common deletion polymorphism in the UGT2B17 gene was significantly associated with a reduced rate of glucuronidation of NNAL in human liver microsomes and was associated with increased lung cancer risk in Caucasian women [17,9]. A case-control study of a Japanese population found an association between UGT1A6*3 and increased lung cancer risk [24]. Additionally, various UGT1A6 polymorphisms were associated with increased or decreased risk of lung cancer in the Chinese population [25].

The goal of the present study was to understand the effect of polymorphisms in the UGT1A gene cluster on lung cancer risk. A comprehensive study, including individual SNP and haplotype analysis, of ninety-six tag and coding SNPs within and surrounding the nine UGT1A genes was conducted.

Methods

Subjects

Caucasian subjects were recruited at the H. Lee Moffitt Cancer Center (Tampa, FL) from 2000-2003 for this case-control study and details have been described elsewhere [9]. During recruitment, subjects were race-, sex-, and age-matched to healthy controls. Subjects with remaining DNA available were utilized in this study. Cases (n=407) were newly diagnosed subjects with histologically confirmed lung cancer and no previous history of cancer. Controls (n=582) were selected from community residents attending the Lifetime Cancer Screening Facility of the Moffitt Cancer Center and were randomly selected from thousands of community residents who underwent prostate-specific antigen testing, skin examinations, endoscopy, or mammography. Spiral computed tomography for lung cancer was not completed at the clinic. Control IDs were matched against the hospital patient database to identify any subjects who may have developed cancer. Controls with newly diagnosed cancer were excluded from this study. Of the patients who were asked to participate in this study, 99% of hospital patients and 97% of clinic patients signed a consent form approved by the institutional review board. A trained interviewer administered a structured lifestyle and smoking history questionnaire 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). Medical charts from patients were reviewed to obtain diagnostic and pathology records.

SNP selection and genotyping

Oral buccal cell swabs or blood samples were collected for genomic DNA isolation. DNA was isolated using standard phenol-chloroform isolation [26]. Picogreen analysis was used to quantify the amount of double-stranded DNA for each sample (Life Technologies, Grand Island, NY). Genotypes for SNPs within and surrounding the UGT1A genes representing people with European ancestry (CEPH) were downloaded from the International HapMap Project [27]. LD in the UGT1A gene cluster was determined using Haploview software [28].

LD was estimated between all pairs of SNPs using the D’ statistic and haplotype block structure was determined using the solid spine of LD option, with the block extended if pairwise D’ between SNPs was >0.80. TagSNPs, representing the genetic variants in the UGT1A gene cluster and 5 kb flanking each side of the gene, were identified using the Tagger program implemented in Haploview software, with the minimum r-squared value=0.8. The minor allele frequency cutoff was 1%. From this analysis, tagSNPs were complemented with coding SNPs to total ninety-six SNPs used in the design of an Illumina GoldenGate genotyping assay and run on the Illumina Bead XPress Instrument (Illumina, San Diego, CA). Illumina GenomeStudio software (Illumina, San Diego, CA) was used for automated calling of genotypes.

Statistical analysis

Hardy-Weinberg equilibrium (HWE), allele frequencies, and identification of haplotype blocks in the study dataset were conducted using the control sample set in the Haploview software, defining blocks by the solid spine of LD. SNPs were excluded if the call rate was <90% and/or a HWE p<1 x 10^-3 [29]. Power analysis was performed using PS: Power and Sample Size Calculation [30]. Power is reduced for stratified analyses.

Unconditional logistic regression was used to calculate odds ratios (ORs) and 95% confidence intervals (CIs) for associations between individual SNPs using R SNPassoc package and lung cancer risk [31]. Three statistical models were tested for individual SNP logistic regression analysis: log-additive, dominant, and recessive. Demographic characteristics between cases and controls were compared using the chi-squared test for categorical variables and non-parametric Wilcoxon rank sum test for continuous variables. If variables appeared non-normally distributed, log transformation was performed, t-test performed and confirmed by Wilcoxon rank sum test. Likelihood ratio tests were used to evaluate the fit of each model.

Sas Proc Haplotype

This procedure was used to conduct the haplotype analysis using the haplotype block definitions from the Haploview software [32]. The procedure utilizes the expectation maximization (EM) algorithm to generate maximum likelihood estimates of haplotype frequencies given a multi-locus sample of genetic marker genotypes under the assumption of HWE. The initializing method was INIT=RANDOM, which initializes haplotype frequencies with random values from a uniform (0,1) distribution. The haplotype frequency threshold was set to 0.05, and haplotypes with a lower frequency were excluded from subsequent logistic regression analysis. The standard errors and the confidence intervals for each haplotype were estimated by default, under a binomial assumption. The total probability of an individual having a particular haplotype compared to all other haplotype possibilities was created. These values were used in the following statistical analysis assuming an additive statistical model (comparing the probability of one haplotype to all other haplotype possibilities).

Multivariate models were adjusted for potential confounding variables that were selected a priori: age, sex, smoking status, pack-years of smoking, and education. Associations stratified by sex, pack-years of smoking, smoking status, and histology (adenocarcinoma, squamous cell carcinoma, non-small cell carcinoma, and small cell carcinoma) were also investigated. Corrected p-values, with significance measured at <0.05, were identified after adjusting for the effects of multiple testing within each genomic region by calculating False Discovery Rate (FDR) [33].
Haplotypic analysis was performed with SAS version 9.3 (SAS Institute, Inc., Cary, NC) and individual SNP analysis was performed with R version 3.1.0 (R Foundation for Statistical Computing, Vienna, Austria). PhEWS-View software was used to visually integrate study results, to discover novel relationships between SNPs and phenotypes, and to produce forest plots [34].

Results

Study population

A summary of the population demographics is described in Table 1. A significant difference between cases and controls was found for age, sex, smoking status, pack-years of smoking, and education. Overall, cases were older than controls, more likely to be former or current smokers, have a higher number of pack-years of smoking, and less likely to complete a college degree. Males had a higher representation in the case population (56%) while women were more represented in the control population (52%). The majority of cases were adenocarcinoma (38%) and squamous cell carcinoma (25%). Histologies representing greater than 10% of the case population were analyzed in stratification analysis.

UGT1A and lung cancer risk

Of the ninety-six SNPs analyzed, sixteen failed the genotyping assay (rs10176426, rs11692021, rs12327959, rs12463641, rs12479240, rs12615708, rs1300947, rs13579741, rs17682878, rs17864678, rs17868323, rs2011404, rs28898605, rs3892170, rs6719561, and rs7571337). Six SNPs were genotyped at a rate <90% (rs2741028: 88.7%, rs17868323, rs2011404, rs28898605, rs3892170, rs6719561, and rs7571337). Seventy-two SNPs were analyzed for association with lung cancer risk (Table 2). Unconditional logistic regression analysis was conducted on individual SNPs and haplotypes within the UGT1A gene family for the effect on lung cancer risk adjusting for age, education, sex, and pack-years of smoking (Figure 1). Twenty-six SNPs were trending toward significance based on an uncorrected p-value <0.05. One SNP (rs17868322) was associated in multiple categories (Table 3). Following multiple testing correction, none of the individual SNPs, were significant before or after stratification by sex, pack-years of smoking, smoking status, age, or histology (Table 4).

UGT1A haplotypes and small cell lung carcinoma

Haploview software was used on the controls to divide the

| SNP Name | Gene Name | Accession | Location |
|----------|-----------|-----------|----------|
| rs10176119 | UGT1A7 | NM_019077.2 | flanking_5UTR |
| rs17864678 | UGT1A10 | NM_019075.2 | flanking_5UTR |
| rs10189426 | UGT1A9 | NM_021027.2 | flanking_3UTR |
| rs12615708 | UGT1A9 | NM_019077.2 | flanking_5UTR |
| rs10203853 | LOC339766 | XM_291007.6 | intron |
| rs10210058 | UGT1A6 | NM_205862.1 | flanking_5UTR |
| rs11692021 | UGT1A10 | NM_019075.2 | flanking_3UTR |
| rs11695770 | UGT1A3 | NM_019093.2 | flanking_3UTR |
| rs11888492 | UGT1A8 | NM_019076.4 | flanking_5UTR |
| rs11891311 | UGT1A3 | NM_019093.2 | flanking_3UTR |
| rs11893247 | UGT1A9 | NM_021027.2 | flanking_5UTR |
| rs12327959 | UGT1A10 | NM_019075.2 | flanking_5UTR |
| rs12463641 | UGT1A4 | NM_007120.2 | flanking_5UTR |
| rs12774680 | UGT1A6 | NM_205862.1 | flanking_5UTR |
| rs12475068 | UGT1A5 | NM_019078.1 | flanking_5UTR |
| rs12479240 | UGT1A9 | NM_021027.2 | flanking_5UTR |
| rs12615708 | UGT1A9 | NM_021027.2 | flanking_5UTR |
| rs12988520 | UGT1A6 | NM_019093.2 | flanking_3UTR |
| rs13009407 | UGT1A3 | NM_019093.2 | flanking_3UTR |
| rs13418420 | UGT1A9 | NM_021027.2 | flanking_5UTR |
| rs1455322 | UGT1A9 | NM_021027.2 | flanking_5UTR |
| rs1500482 | UGT1A8 | NM_019076.4 | flanking_5UTR |
| rs1597941 | UGT1A8 | NM_019076.4 | flanking_5UTR |
| rs1597942 | UGT1A4 | NM_007120.2 | flanking_5UTR |
| rs1500482 | UGT1A8 | NM_019076.4 | flanking_5UTR |
| rs1604144 | UGT1A6 | NM_205862.1 | flanking_5UTR |
| rs17862859 | UGT1A6 | NM_205862.1 | flanking_5UTR |
| rs17862878 | UGT1A1 | NM_000463.2 | flanking_5UTR |
| rs17863762 | UGT1A8 | NM_019076.4 | flanking_5UTR |
| rs17863773 | UGT1A9 | NM_021027.2 | flanking_5UTR |
| rs17863775 | UGT1A9 | NM_021027.2 | flanking_5UTR |
| rs17864678 | UGT1A10 | NM_019075.2 | flanking_5UTR |
| rs17864683 | UGT1A9 | NM_021027.2 | flanking_5UTR |
| rs17868322 | UGT1A9 | NM_019076.4 | flanking_5UTR |
| rs17868323 | UGT1A7 | NM_019077.2 | flanking_5UTR |
| rs17868336 | UGT1A3 | NM_019093.2 | flanking_3UTR |
| rs17868337 | UGT1A3 | NM_019093.2 | flanking_3UTR |
| rs1817154 | UGT1A10 | NM_019075.2 | flanking_5UTR |
| rs1823803 | UGT1A10 | NM_019075.2 | flanking_5UTR |
| rs1875263 | UGT1A4 | NM_007120.2 | flanking_5UTR |
| rs2008595 | UGT1A3 | NM_019093.2 | flanking_5UTR |
| rs2011404 | UGT1A4 | NM_007120.2 | flanking_5UTR |
| rs2070959 | UGT1A6 | NM_001072.2 | flanking_5UTR |
| rs2221198 | UGT1A1 | NM_000463.2 | flanking_5UTR |
| rs224873 | UGT1A9 | NM_021027.2 | flanking_5UTR |
| rs3202538 | UGT1A8 | NM_019076.4 | flanking_5UTR |
| rs6202373 | UGT1A10 | NM_019075.2 | flanking_3UTR |
| rs6202379 | UGT1A9 | NM_021027.2 | flanking_3UTR |
| rs6202381 | UGT1A9 | NM_021027.2 | flanking_3UTR |

Table 1: Summary of Population Demographics.
Table 2: List of SNPs analyzed.

| SNP       | Category          | Statistical Model | Odds Ratio (95% CI) | Uncorrected p-value | p-value | FDR    |
|-----------|-------------------|-------------------|---------------------|---------------------|---------|--------|
| rs2741028 | UGT1A8            | flanking_5UTR     | 2.00 (1.03-3.86)    | 0.039               | 0.98    |        |
| rs2741029 | UGT1A8            | flanking_3UTR     | 2.14 (1.06-4.32)    | 0.046               | 0.98    |        |
| rs2741034 | UGT1A10           | flanking_5UTR     | 2.31 (1.08-5.05)    | 0.023               | 0.98    |        |
| rs2741048 | UGT1A9            | flanking_5UTR     | 2.38 (1.12-5.03)    | 0.035               | 0.98    |        |
| rs28989590| UGT1A5            | flanking_5UTR     | 2.44 (1.12-5.31)    | 0.020               | 0.98    |        |
| rs28989605| UGT1A4            | flanking_3UTR     | 2.89 (1.21-6.08)    | 0.028               | 0.98    |        |
| rs28989648| UGT1A3            | flanking_3UTR     | 2.97 (1.21-7.23)    | 0.043               | 0.98    |        |
| rs28984388| UGT1A7            | flanking_5UTR     | 3.06 (1.23-7.45)    | 0.023               | 0.98    |        |
| rs28969676| UGT1A8            | flanking_5UTR     | 3.14 (1.21-7.93)    | 0.028               | 0.98    |        |
| rs28969678| UGT1A10           | flanking_5UTR     | 3.16 (1.21-7.93)    | 0.028               | 0.98    |        |

Table 3: Association between UGT1A9 promoter variant rs1788332 and lung adenocarcinoma in former smokers.

Table 4: Associations between UGT1A missense polymorphisms and lung cancer risk at the uncorrected p<0.05.

UGT1A gene region into eleven haplotype blocks and each haplotype (Figure 2A) was analyzed for impact on lung cancer risk. SAS PROC HAPLOTYPE was used to assign the probability that each individual possesses a particular haplotype compared to all other haplotype possibilities. Probabilities were analyzed in an unconditional logistic regression model controlling for the same covariates in the individual SNP analysis. Assuming an additive statistical model, the analysis reported the risk associated with a specific haplotype when compared to all other haplotypes in the population at frequencies >5%. A significant association between lung cancer risk and the UGT1A haplotypes in block 2 (rs7569014, rs7421795, rs1817154) was seen.
after multiple testing correction in a stratification of thirty-eight small cell carcinoma patients (Figure 2B). No associations were found when studying all cases and controls. The (G-G-G) haplotype decreased small cell carcinoma risk (OR=0.092, 95% CI=0.013-0.65, p-value=0.017) while the (G-G-T) haplotype increased risk (OR=6.64, 95% CI=1.44-30.6, p-value=0.017; Figure 2B). The SNPs in this haplotype block span exon 1 of UGT1A10 (Figure 2B).

Discussion

A major mode of detoxification of tobacco smoke carcinogens is glucuronidation by UGT enzymes. Since only 15% of lifetime smokers develop lung cancer, the hypothesis for this study was that genetic variation within the UGT1A gene family was associated with lung cancer risk [4]. Despite the role of the UGTs on tobacco carcinogen metabolism, few studies have analyzed the impact of UGT1A polymorphisms on lung cancer risk and none have evaluated American Caucasians. In the present study, twenty-six UGT1A SNPs were found to be potentially associated with lung cancer risk; however, none of the SNPs remained significant following multiple testing correction. Haplotype analysis was conducted to determine whether this lack of significance after multiple testing correction was due to high LD of SNPs spanning the UGT1A region and therefore difficult to differentiate statistically. Few associations with UGT1A haplotypes were detected, suggesting that UGT1A SNPs may not be associated with overall lung cancer risk in American Caucasians. A significant association was found between haplotypes in block 2 (rs7569014, rs7421795, rs1817154) that span exon 1 of UGT1A10. Due to the small population of small cell carcinoma cases, this finding should be confirmed in larger populations.

Associations between UGT1A genes and lung cancer risk have been identified in other populations, supporting the hypothesis that UGT1As play a role in disease risk for some races. The UGT1A7*3 polymorphism increased lung cancer risk in a Japanese population [24]. In addition, UGT1A6 19T>G, 541A>G and 552A>C significantly increased lung cancer risk and UGT1A6 105C>T and IVS1+130G>T polymorphisms significantly reduced lung cancer risk in a Chinese population [25]. Asian populations and Caucasian populations differ in SNP frequencies, extent of LD, and haplotypes, in addition to lung cancer and smoking patterns, so it is not surprising that our study in Caucasians identified different variants than the previous studies in Asian populations. Two SNPs in this study located in the intronic region of UGT1A6, 1A7, 1A8, 1A9, and 1A10 were also approaching significance for an increased risk in current smokers (rs2889590) and decreased risk for individuals with high pack-years of smoking (rs12474980) [35]. These data are consistent with the role for UGT1A7, 1A8, 1A9, and 1A10 in the metabolism of polycyclic aromatic hydrocarbons (PAHs) and N-OH-PhIP, both of which are present in tobacco smoke [18,19,35,36]. For reference, the p-values described for the remainder of this manuscript are not corrected for multiple testing. In the present study, rs12747980 was shown to decrease lung cancer risk in individuals with high-pack years of smoking (dominant model: OR=0.52, 95% CI=0.32-0.85, p-value=0.009; log-additive model: OR=0.6, 95% CI=0.6, 95% CI=0.39-0.90, p-value=0.014), correlating with a previous study (Table 5). Larger studies in American Caucasians are needed to better assess this result [25].

Several polymorphisms were significant prior to multiple testing correction in the present study, but the functional effect of has not been previously established. A missense polymorphism in exon 1 of UGT1A4 (rs6755571, UGT1A4*2) increases risk in current smokers in dominant model (OR=4.36, 95% CI=1.54-12.37, p-value=0.003) and log-additive model (OR=4.30, 95% CI=1.52-12.14, p-value=0.003) models (Table 3) before multiple testing correction. Of note, >80% power was only achieved in the current smokers stratification. After multiple testing correction, this SNP was not significant; however, it is of interest because UGT1A4 plays a role in the catalysis of tobacco-specific nitrosamines including NNAL, the PAH dibenzo (a,l) pyrene-trans-11,12-diol.

Figure 2: Haplovew software was used on the controls to divide the UGT1A gene region into eleven haplotype blocks and multiple testing correction in a stratification of thirty-eight small cell carcinoma patients.
and the heterocyclic amine 2-amino-9H-pyrido (2,3-b) indole (AαC) [16,37-39]. In addition, the UGT1A4*2 variant was linked to altered glucuronidation activities against NNAL in human liver microsomes [37]. UGT1A4 plays a role in the metabolism of nicotine, the addictive agent in tobacco [40]. The rate of nicotine metabolism has been shown to influence tobacco smoking patterns, with adult slow nicotine metabolizers smoking fewer cigarettes daily and weekly [41,42]. The UGT1A4*2 allele could be important in nicotine dependence because the association was only found in current smokers.

A UGT1A9 promoter variant (rs17868322) was associated with lung cancer risk, before multiple testing corrections, in all lung cancer cases, former smokers, and in subjects with adenocarcinoma (Table 2). Of note, >80% power was only achieved in the former smokers.

| SNP                  | Stratification                      | Model       | OR (95% CI)    | Uncorrected p value | FDR p value |
|----------------------|-------------------------------------|-------------|----------------|---------------------|-------------|
| rs10203853           | Males                               | Dominant    | 2.00 (1.03-3.86) | 0.039               | 0.98        |
| rs10210058           | Small Cell Carcinoma                | Dominant    | 2.00 (1.05-3.83) | 0.034               | 0.99        |
| rs11695770           | Young (<61 years)                   | Dominant    | 2.29 (1.04-5.05) | 0.044               | 0.98        |
| rs11888492           | Current Smokers                     | Dominant    | 2.10 (0.97-4.56) | 0.030               | 0.96        |
| rs12474980           | High Pack Years                     | Dominant    | 2.10 (0.97-4.56) | 0.030               | 0.96        |
| rs12474980           | High Pack Years                     | Log-Additive| 0.50 (0.25-1.01) | 0.038               | 0.98        |
| rs12475068           | Older (>61 years)                   | Dominant    | 2.00 (1.03-3.86) | 0.039               | 0.98        |
| rs12475068           | Older (>61 years)                   | Log-Additive| 0.50 (0.25-1.01) | 0.038               | 0.98        |
| rs1604144            | Squamous Cell Carcinoma             | Recessive   | 2.10 (1.28-3.67) | 0.044               | 0.98        |
| rs17868322           | All Lung Cancer                     | Dominant    | 2.10 (1.28-3.67) | 0.044               | 0.98        |
| rs17868322           | All Lung Cancer                     | Log-Additive| 0.26 (0.07-0.91) | 0.015               | 0.85        |
| rs17868322           | Adenocarcinoma                      | Dominant    | 0.57 (0.33-0.98) | 0.039               | 0.97        |
| rs17868322           | Adenocarcinoma                      | Log-Additive| 0.57 (0.33-0.98) | 0.039               | 0.97        |
| rs17868322           | Former Smokers                      | Dominant    | 1.56 (0.80-3.05) | 0.023               | 0.66        |
| rs17868322           | Former Smokers                      | Log-Additive| 0.25 (0.09-0.70) | 0.004               | 0.20        |
| rs2248733            | All Lung Cancer                     | Dominant    | 1.39 (1.02-1.9)  | 0.039               | 0.98        |
| rs2248733            | Squamous Cell Carcinoma             | Recessive   | 2.06 (1.03-4.10) | 0.042               | 0.97        |
| rs2248733            | Squamous Cell Carcinoma             | Log-Additive| 0.54 (0.33-0.90) | 0.014               | 0.47        |
| rs2302538            | Older (>61 years)                   | Recessive   | 2.00 (1.03-3.96) | 0.036               | 0.83        |
| rs2302538            | Current Smokers                     | Dominant    | 2.44 (1.21-4.90) | 0.009               | 0.13        |
| rs2302538            | Current Smokers                     | Log-Additive| 0.52 (0.23-1.23) | 0.023               | 0.66        |
| rs2602373            | Squamous Cell Carcinoma             | Dominant    | 2.00 (1.03-3.86) | 0.039               | 0.98        |
| rs2602381            | Never Smokers                       | Recessive   | 2.00 (1.03-3.86) | 0.039               | 0.98        |
| rs2741034            | Squamous Cell Carcinoma             | Dominant    | 2.00 (1.03-3.86) | 0.039               | 0.98        |
| rs2741034            | Squamous Cell Carcinoma             | Log-Additive| 0.57 (0.33-0.98) | 0.039               | 0.97        |
| rs3755321            | Young (<61 years)                   | Dominant    | 2.00 (1.03-3.86) | 0.039               | 0.98        |
| rs3755321            | Young (<61 years)                   | Log-Additive| 0.54 (0.33-0.90) | 0.014               | 0.47        |
| rs3771342            | Current Smokers                     | Dominant    | 2.00 (1.03-3.86) | 0.039               | 0.98        |
| rs3771342            | Current Smokers                     | Log-Additive| 0.57 (0.33-0.98) | 0.039               | 0.97        |
| rs4556969            | Young (<61 years)                   | Dominant    | 2.00 (1.03-3.86) | 0.039               | 0.98        |
| rs4639695            | Older (>61 years)                   | Dominant    | 2.00 (1.03-3.86) | 0.039               | 0.98        |
| rs6717546            | High Pack Years                     | Dominant    | 1.39 (1.02-1.9)  | 0.039               | 0.98        |
| rs6717546            | Current Smokers                     | Dominant    | 2.00 (1.03-3.86) | 0.039               | 0.98        |
| rs6717546            | Current Smokers                     | Log-Additive| 0.57 (0.33-0.98) | 0.039               | 0.97        |
| rs6725478            | Squamous Cell Carcinoma             | Dominant    | 1.50 (1.02-2.20) | 0.038               | 0.30        |
| rs6728940            | High Pack-Years                     | Recessive   | 0.38 (0.18-0.79) | 0.008               | 0.58        |
| rs6755571            | Current Smokers                     | Recessive   | 0.39 (0.18-0.80) | 0.008               | 0.58        |
| rs6755571            | Current Smokers                     | Dominant    | 0.39 (0.18-0.80) | 0.008               | 0.58        |
| rs7572563            | Current Smokers                     | Recessive   | 1.50 (1.00-2.97) | 0.045               | 0.31        |
| rs7586006            | Current Smokers                     | Log-Additive| 1.92 (1.04-3.56) | 0.034               | 0.30        |
| rs7586006            | Never Smokers                       | Dominant    | 2.00 (1.03-3.86) | 0.039               | 0.98        |
| rs7586006            | Never Smokers                       | Log-Additive| 0.31 (0.10-0.93) | 0.017               | 0.94        |
| rs7608175            | Current Smokers                     | Recessive   | 1.38 (0.99-2.04) | 0.030               | 0.34        |

Table 5: All SNPs displaying significant association before multiple testing correction.
stratification. Given that UGT1A9 metabolizes NNAL and PAHs, functional studies are warranted [13,38].

Haplotype analysis found a significant increase (G-G-T) (OR=6.64, 95% CI=1.44-30.6, p-value=0.017) and decrease (G-G-G) (OR=0.092, 95% CI=0.031-0.065, p-value=0.007) in small cell carcinoma risk for haplotypes containing the SNPs rs7569014, rs7421795, and rs1817154 (Figure 2B). Although this association remained significant after multiple testing corrections, it needs to be confirmed in a larger study due to the small population of small cell carcinoma patients. In addition, one of the SNPs contributing to the above haplotypes (rs1817154) was significant by itself prior to multiple testing corrections for small cell carcinoma risk (Table 5). The single SNP and haplotype analysis both support that there is a variant in this block associated with small cell carcinoma. These three SNPs span exon 1 of UGT1A10 and although these SNPs have not been identified with functional changes in UGT1A10 activity, they may be LD with other UGT1A10 variants, for example the 1664 bp deletion polymorphism located -190 to -1856 relative to the UGT1A10 translation start site that increases UGT1A10 gene expression [43]. Given that UGT1A10 exhibits high activity against many BaP metabolites and exhibits O-glucuronidation activity against NNAL, further studies exploring the potential role of these haplotypes in risk for small cell carcinoma are warranted [14,18,43].

Conclusions

In summary, while many UGT1A polymorphisms were observed to be potentially associated with lung cancer risk in this study, these results was not significant after multiple testing corrections was applied. A significant association between two UGT1A haplotypes and risk of small cell carcinoma was found after multiple testing corrections, but this needs to be confirmed in a larger study. These data are consistent with the generally low expression of UGT1A enzymes in lung and other respiratory tract tissues [20,44]. This contrasts with UGT2B17, which was shown to be relatively well expressed in lung and was associated with risk for lung adenocarcinoma [9,45-47]. It is interesting to note the UGT1A10 was shown to be expressed in lung and is spanned by the two haplotypes associated with small cell carcinoma risk [18]. Therefore, with the potential exception of small cell carcinoma, these data suggest that UGT1A variants may not be playing a major role in lung cancer risk.

References

1. U.S. Cancer Statistics Working Group (2015) United States Cancer Statistics: 1999-2012 Incidence and Mortality Web-based Report. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute.
2. U.S. Department of Health and Human Services (2014) 50 Years of Progress: A Report of the Surgeon General.
3. CDC’s Office on Smoking and Health (2014) Smoking and Tobacco Use; Fact Sheet: Health Effects of Cigarette Smoking.
4. Spitz MR, Wei Q, Dong Q, Amos CI, Wu X (2003) Genetic susceptibility to lung cancer: the role of DNA damage and repair. Cancer Epidemiol. Biomarkers Prev 12: 689-698.
5. Guo X, Peng S, Jiang G (1996) A genetic epidemiological study on lung cancer. Zhonghua Yu Fang Yi Xue Za Zhi. Chinese Journal of Preventive Medicine 30: 154-156.
6. Rodgman A, Periﬁeri T (2009) The Chemical Components of Tobacco and Tobacco Smoke, Second Edition. Boca Raton: CRC Press: 1483-1784p.
7. Hecht SS (2012) Research opportunities related to establishing standards for tobacco products under the Family Smoking Prevention and Tobacco Control Act. Nicotine Tob Res 14: 18-28.
8. Nagar S, Remmel RP (2006) Uridine diphosphoglucuronosyltransferase pharmacogenetics and cancer. Oncogene 25: 1659-1672.
9. Gallagher CJ, Muscat JE, Hicks AN, Zheng Y, Dyer AM, et al. (2007) The UDP-glucuronosyltransferase 2B17 gene deletion polymorphism: sex-specific association with urinary 4-(methyltrinitrosamino)-1-(3-pyridyl)-1-butanone glucuronidation phenotype and risk for lung cancer. Cancer Epidemiol. Biomarkers Prev 16: 823-828.
10. Mackenzie PJ, Gregory PA, Gardner-Stephen DA, Lewinsky RH, Jorgensen BR, et al. (2003) Regulation of UDP glucuronosyltransferase genes. Curr. Drug Metab 4: 249-257.
11. Mailtand ML, Grimisley C, Kultab-Bouls H, Witonksy D, Kasza KE, et al. (2006) Comparative genomics analysis of human sequence variation in the UGT1A1 gene cluster. Pharmacogenomics J 6: 52-62.
12. Jin CJ, Minors JD, Burchell B, Mackenzie PI (1993) The glucuronidation of hydroxylated metabolites of benzo[a]pyrene and 2-acetylaminofluorene by cDNA-expressed human UDP-glucuronosyltransferases. Carcinogenesis 14: 2637-2639.
13. Ren Q, Murphy SE, Zheng Z, Lazarus P (2000) O-Glucuronidation of the Lung Carcinogen 4-(methyltrinitrosamino)-1-(3-Pyridyl)-1-Butanol (NNAL) by Human UDP-Glucuronosyltransferases 2B7 and 1A9. Drug Metab. Dispos 28: 1352-1360.
14. Fang J, Beland FA, Doerge DR, Wiener D, Guillemette C, et al. (2002) Characterization of benzo(a)pyrene-trans,7,8-dihydrōdiol glucuronidation by human tissue microsomes and overexpressed UDP-glucuronosyltransferase enzymes. Cancer Res 62: 1978-1986.
15. Fang JL, Lazarus P (2004) Correlation between the UDP-glucuronosyltransferase (UGT1A1) TATAA box polymorphism and carcinogen detoxification phenotype: significantly decreased glucuronidating activity against benzo[a]pyrene-trans,7,8-dihydrodiol (-) in liver microsomes from subjects with the UGT1A1*2B variant. Cancer Epidemiol. Biomarkers Prev 13: 102-109.
16. Wiener D, Doerge DR, Fang JL, Upadhyaaya P Lazarus P (2004) Characterization of N-glucuronidation of the lung carcinogen 4-(methyltrinitrosamino)-1-(3-pyridyl)-1-butanol (NNAL) in human liver: importance of UDP-glucuronosyltransferase 1A4. Drug Metab Dispos 32: 72-79.
17. Lazarus P, Zheng Y, Runkle EA, Muscat JE, Wiener D (2005) Genotype-phenotype correlation between the polymorphic UGT2B17 gene deletion and NNAL glucuronidation activities in human liver microsomes. Pharmacogenet Genomics 15: 769-778.
18. Dellingwer RW, Fang J, Chen G, Weinberg R, Lazarus P (2006) Importance of UDP-Glucuronosyltransferase 1A10 (UGT1A10) in the Detoxification of Polycyclic Aromatic Hydrocarbons: Decreased Glucuronidative Activity of the UGT1A10 139Lys Isoform. Drug Metab Dispos 34: 943-949.
19. Dellingwer RW, Chen G, Blevins-Primeau AS, Krzeminski J, Amin S, et al. (2007) Characterization of Glucuronidation of PhIP and N-OH-PhIP by UDP-glucuronosyltransferase 1A10. Carcinogenesis 28: 2412-2418.
20. Zheng Z, Fang J, Lazarus P (2002) Glucuronidation: An Important Mechanism for Detoxification of Benzo[a] Pyrene Metabolites in Aerodigestive Tract Tissues. Drug Metab Dispos 30: 397-403.
21. Nobuo N, Shozo T, Minako N, Kazuo U (1976) Modification of the mutagenicity of benzo[a]pyrene by bacteria under enzymes producing water soluble conjugates Toxicol Lett 2: 205-211.
22. Chen G, Dellingwer RW, Sun D, Spratt TE, Lazarus P (2008) Glucuronidation of tobacco-specific nitrosamines by UGT2B10. Drug Metab Dispos 36: 824-830.
23. Chen G, Giambrone NE, Lazarus P (2012) Glucuronidation of trans-3-hydroxycycrodine by UGT1B17 and UGT2B10. Pharmacogenet Genomics 22: 183-190.
24. Araki J, Kobayashi Y, Iwasa M, Urawa N, Gabazza EC, et al. (2005) Polymorphism of UDP-glucuronosyltransferase 1A7 gene: a possible new risk factor for lung cancer. Eur J Cancer 41: 2360-2365.
25. Ren Q, Murphy SE, Zheng Z, Lazarus P (2000) O-Glucuronidation of the Lung Carcinogen 4-(methyltrinitrosamino)-1-(3-Pyridyl)-1-Butanol (NNAL) by Human UDP-Glucuronosyltransferases 2B7 and 1A9. Drug Metab. Dispos 28: 1352-1360.
26. Park JY, Muscat JE, Harwick RD, et al. (1997) CYP1A1 and GSTM1 polymorphisms and oral cancer risk. Cancer Epidemiol. Biomarkers Prev 6: 791-797.
27. The International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437: 1299-1302.
28. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263-265.

29. Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 76: 887-893.

30. Dupont WD, Plummer WD (1990) Power and sample size calculations. A review and computer progra. Control Clin Trials 11: 116-128.

31. González JR, Armengol L, Solé X, Guind E, Mercader JM, et al. (2007) SNPassoc: an R package to perform whole genome association studies. Bioinformatics 23: 644-645.

32. Czika W, Yu X (2004) Genetic Analysis of Complex Traits using SAS. In: Genetic Analysis of Complex Traits Using SAS, Saxton A, Ed. Cary NC: SAS Publishing pp. 179-200.

33. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) 57: 289-300.

34. Pendergrass SA, Dudek SM, Crawford DC, Ritchie MD (2012) Visually integrating and exploring high throughput Phenome-Wide Association Study (PheWAS) results using PheWAS-View. BioData Min 5: 5.

35. Hecht SS (1999) Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst 91: 1194-1210.

36. Wiener D, Fang JL, Dossett N, Lazarus P (2004) Correlation between UDP-glucuronosyltransferase genotypes and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone glucuronidation phenotype in human liver microsomes. Cancer Res. 64:1190-1196.

37. Olson KC, Sun D, Chen G, Sharma AK, Amin S, et al. (2011) Characterization of dibenzo[a,j]pyrene-trans-11,12-diol (dibenzo[def,p]chrysene) glucuronidation by UDP-glucuronosyltransferases. Chem Res Toxicol 24: 1549-1559.

39. Tang Y, LeMaster DM, Nauwelaers G, Gu D, Langouët S, et al. (2012) UDP-glucuronosyltransferase-mediated metabolic activation of the tobacco carcinogen 2-amino-3H-pyrido[3,4-b]indole. J Biol Chem 287: 14960-14972.

40. Kuehl GE, Murphy SE (2003) N-glucuronidation of nicotine and cotinine by human liver microsomes and heterologously expressed UDP-glucuronosyltransferases. Drug Metab Dispos 31: 1361-1368.

41. O’Loughlin J, Paradis G, Kim W, DiFranza J, Meshefiedjian G, et al. (2004) Genetically decreased CYP2A6 and the risk of tobacco dependence: a prospective study of novice smokers. Tob. Control 13: 422-426.

42. Schoedel KA, Hoffmann EB, Rao Y, Seiters EM, Tyndale RF (2004) Ethnic variation in CYP2A6 and association of genetically slow nicotine metabolism and smoking in adult Caucasians. Pharmacogenetics 14: 615-626.

43. Balliet RM, Chen G, Dellinger RW, Lazarus P (2010) UDP-glucuronosyltransferase 1A10: activity against the tobacco-specific nitrosamine, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol, and a potential role for a novel UGT1A10 promoter deletion polymorphism in cancer susceptibility. Drug Metab Dispos 38: 484-490.

44. Nakamura A, Nakajima M, Yamanaka H, Fujiwara R, Yokoi T (2008) Expression of UGT1A1 and UGT2B2 mRNA in human normal tissues and various cell lines. Drug Metab Dispos 36: 1461-1464.

45. Beaulieu M, Lévesque E, Hum DW, Bélanger A (1996) Isolation and characterization of a novel cDNA encoding a human UDP-glucuronosyltransferase active on C19 steroids. J Biol Chem 271: 22255-22262.

46. Nishimura M, Naito S (2006) Tissue-specific mRNA expression profiles of human phase I metabolizing enzymes except for cytochrome P450 and phase II metabolizing enzymes. Drug Metab Pharmacokinet 21: 357-374.

47. Jones NR, Lazarus P (2014) UGT2B gene expression analysis in multiple tobacco carcinogen-targeted tissues. Drug Metab Dispos 42: 529-536.