Genetic Polymorphisms in N-acetyltransferase-2 and Microsomal Epoxide Hydrolase, Cumulative Cigarette Smoking, and Lung Cancer

Wei Zhou, Geoffrey Liu, Sally W. Thurston, Li Lian Xu, David P. Miller, John C. Wain, Thomas J. Lynch, Li Su, and David C. Christiani

Abstract

N-acetyltransferase-2 (NAT2) and microsomal epoxide hydrolase (mEH) are polymorphic genes that metabolize different tobacco carcinogens. Smaller studies found inconsistent relationships between NAT2 or mEH polymorphisms and lung cancer risk. To determine whether there is gene-environment interaction between NAT2 polymorphisms, alone or in combination with mEH polymorphisms, and cumulative smoking exposure in the development of lung cancer, we conducted a case control study of 1115 Caucasian lung cancer patients and 1250 spouse and friend controls. The results were analyzed using generalized additive models and logistic regression, adjusting for relevant covariates. There was no overall relationship between NAT2 genotype and lung cancer risk; the adjusted odds ratio (OR) of the rapid versus slow acetylator genotypes was 0.96 [95% confidence interval (CI), 0.79–1.16]. However, gene-environment interaction analyses revealed that the adjusted ORs increased significantly as pack-years increased. For nonsmokers, the fitted OR was 0.66 (95% CI, 0.44–0.99), whereas for heavy smokers (80 pack-years), the OR increased to 1.22 (95% CI, 0.89–1.67). When comparing the extreme genotype combinations of the NAT2 rapid acetylator, higher mEH activity genotype to the NAT2 slow acetylator, and very low mEH activity genotype, the corresponding ORs at 0 and 80 pack-years were 0.30 (95% CI, 0.14–0.62) and 2.19 (95% CI, 1.26–3.81), respectively. Results were similar with ORs derived from stratified models. In conclusion, NAT2 rapid acetylator genotypes are protective against lung cancer in nonsmokers but are risk factors in heavy smokers. The joint effects of NAT2 and mEH polymorphisms are consistent with an independent, additive effect of these two genes, modified by smoking history.

Introduction

NAT2 and mEH are polymorphic genes encoding enzymes that metabolize environmental and tobacco smoke carcinogens. NAT2 has a complex action involved in both the activation and detoxification of potential carcinogens, NAT2 inactivates aromatic amines through N-acetylation, which competes with the cytochrome P450 pathway; however, the enzyme also activates certain arylamine metabolites originating in cigarette smoke through N- and O-acetylation (1, 2). mEH has similar dual functions; mEH catalyzes the hydrolysis of arene, alkene, and aliphatic epoxides to less reactive and more water soluble dihydriodols but activates some polycyclic aromatic hydrocarbons found in cigarette smoke into a more carcinogenic form (3). Hence, the roles of NAT2 and mEH in carcinogenesis may depend on individual exposures to different environmental substrates.

There are seven widely studied genetic polymorphisms associated with decreased enzyme activity and/or variable stability for the human NAT2 gene: 191G→A, 237C→T, 341T→C, 481C→T, 590G→A, 803A→G, and 857G→A (2). For mEH, there are two polymorphisms that affect enzyme activity: 113T→C (exon 3) and 130A→G (exon 4), where the variant genotypes confer decreased and increased activity, respectively (4). Because tobacco smoking is the major environmental risk factor for lung cancer, these genetic polymorphisms may modify host susceptibility to lung cancer through their roles in metabolizing tobacco procarcinogens. Most genotype or phenotype studies found no overall lung cancer risk related to NAT2 rapid or slow acetylators (5–9), whereas the slow acetylator genotype (10–12) and the homozygous rapid acetylation genotype (13) were associated with increased lung cancer risk in specific studies or populations. One study suggested that the rapid acetylator genotype was a protective factor in never-smokers but a risk factor in smokers, but the results were not conclusive, and the study population was too small for formal interaction analyses (n = 185, P > 0.05; Ref. 14). In the case of mEH polymorphisms, inconsistent relationships have also been found with respect to lung cancer risk (15–20).

The abbreviations used are: NAT2, N-acetyltransferase-2; mEH, microsomal epoxide hydrolase; GAM, generalized additive model; SR-PY, square root of pack-years; OR, odds ratio; CI, confidence interval.
We recently reported a gene-environment interaction between mEH polymorphisms and cumulative smoking exposure in lung cancer development, where the higher activity genotypes (compared with the very low activity genotype) were protective against lung cancer in nonsmokers but were a significant risk factor in heavy smokers (21). Because NAT2 and mEH have similar putative dual functions and genetically polymorphic variants that affect enzyme function, a gene-smoking interaction may also modify the relationship between NAT2 polymorphisms and lung cancer risk.

The primary aim of this study was to determine whether cumulative smoking exposure altered the relationship between NAT2 polymorphisms and lung cancer risk in a gene-environment interaction analysis. If NAT2 genotypes were found to interact with cigarette smoking history, a secondary aim of this study was to determine whether the combined effects of NAT2 and mEH polymorphisms on lung cancer risk were additive, because their corresponding enzymes metabolized different tobacco-related substrates (1–3).

Materials and Methods

Study Population. The study was approved by the Human Subjects Committees of Massachusetts General Hospital and the Harvard School of Public Health, Boston. Eligible cases (patients with histologically confirmed incident lung cancers) at Massachusetts General Hospital were recruited between December 1992 and December 2000. Controls were recruited first among the friends and nonblood related family members of the lung cancer cases with no specific matching characteristics. In a small minority of cases (<10%), such individuals were not available, and controls were recruited among friends and nonblood related family members of nonlung cancer patients admitted to the cardiothoracic wards. This selection of controls was chosen in an attempt to balance cases and controls with respect to race, age, and smoking variables while avoiding overmatching. Interviewer-administered questionnaires collected information on demographic and detailed smoking histories from each subject.

Cumulative Smoking Exposure. The questionnaire asked detailed smoking questions, including the age of onset of regular smoking, the average number of cigarettes smoked during different time periods, duration of smoking, and details on periods of smoking cessation. Pack-years were calculated as the product of the number of years an individual smoked and the average number of cigarettes smoked per day (converted into a standard pack of 20 cigarettes). Total periods of smoking cessation for periods >6 months were excluded from the total number of years of smoking in all calculations. Nonsmokers were defined as individuals who smoked fewer than 100 cigarettes in their lifetime.

NAT2 and mEH Genotyping. Peripheral blood samples were obtained from each case or control, and DNA was extracted using the Puregene DNA Isolation Kit (Gentra Systems, Minneapolis, MN). NAT2 polymorphisms were detected using previously reported PCR-RFLP methods (22, 23). In brief, the four most common NAT2 polymorphisms, 481C→T, 590G→A, 803A→G, and 857G→A, were identified by different restriction fragment analysis on digestion of the 547-bp PCR product, using restriction enzymes KpnI (481), TaqI (590), DdeI (803), and BamHI (857; New England BioLabs, Beverly, MA). DNA-digested PCR products were separated by size on a 2.5% agarose gel (Sigma Chemical Co., St. Louis, MO).

The genotyping methods for mEH polymorphisms have been described in detail (21). In brief, two separate PCR assays were used to detect the polymorphisms in exons 3 and 4 of mEH, and EcoRV and RsaI enzyme digestion (New England BioLabs) were used respectively for restriction fragment analyses.

For quality control, a random 5% of the samples were repeated to assess the reproducibility of results, and a second investigator rechecked all data entry and analyses.

Statistical Analysis. NAT2 and mEH genotypes were categorized according to their corresponding genetic polymorphisms. For NAT2, the wild type of each genetic polymorphism corresponds with rapid acetylation, and the variant genotypes correspond to slow acetylation. Thus, individuals carrying wild-type alleles (i.e., two “rapid” alleles, NAT2*4/*4) were designated as carrying the homozygous rapid acetylator genotype; individuals with only one copy of a rapid allele (heterozygous variant genotype, NAT2*4) were identified as heterozygous rapid acetylators, and those with two “slow” alleles (homozygous variant genotypes; NAT2*5A/*5A, *5B/*5B, *5C/*5C, *6, or *7) were labeled slow acetylators. A similar system was used to categorize the mEH genotype into very low, low, intermediate, and high activity groups, according to different mEH genotypes in exons 3 and 4 (21).

Gene-environment interaction and gene–gene joint effects analyses involving multiple strata require enormous sample sizes. Combining different genotype strata on the basis of biological or epidemiological evidence has been a reasonable and practical approach. We combined NAT2 homozygous and heterozygous rapid acetylators into one group, comparing it with the slow acetylators, similar to approaches taken by others (9, 11, 12, 14, 23). On the basis of our previous epidemiological research on mEH genotypes and lung cancer risk (21), mEH was dichotomized into the genotype representing very low enzyme activity (homozygous for histidine in both exons and 4) versus all other genotypes combined.

Although individuals of all races were recruited for this study, we restricted our analyses to Caucasians to minimize the possibility of a different gene effect, i.e., effect modification by race or unmeasured confounders that vary by race. We included all Caucasians with complete information on age, gender, smoking status (non, ex-, and current smokers), pack-years of smoking, and years since smoking cessation (for ex-smokers). Logistic regression and GAM (24) were used to examine the relationship between the log odds of lung cancer and each covariate, after adjusting for possible confounding factors, such as age (years), gender, pack-years of smoking, smoking status, and years since smoking cessation (defined for ex-smokers; this variable was defined as zero for nonsmokers and current smokers). Some variables were transformed to approach normality. GAM allows the relationship between lung cancer risk and each covariate to be an unspecified smooth function. GAM plots showed a linear association between lung cancer risk and SR-PY but not for the untransformed variable. Subsequent analyses were based on logistic regression models, using SR-PY, age, and years since smoking cessation as continuous variables. In the NAT2 genotype-smoking primary interaction analysis, we included in our initial model all possible interactions among NAT2 genotype and two smoking variables (smoking status and SR-PY). To examine the combined effects of the NAT2 and mEH genotypes with smoking, we constructed a model that included all of the possible interactions among NAT2 and mEH genotypes and SR-PY, as well as any interactions between smoking status and either NAT2, mEH, or SR-PY that were included in the final models of NAT2 (or mEH) genotype-smoking interaction analysis. When the highest order
interaction term was not significant, we considered models with only lower order interactions. Interaction terms were removed from the model when the likelihood ratio tests (comparing a candidate smaller model with all larger models in which it was nested) were not significant. Where appropriate, OR and 95% CI for the risk of lung cancer were calculated from these models. A lack of fit test was performed to summarize the goodness of fit of each logistic regression model (25). Statistical analyses were all undertaken using the S-plus (MathSoft, Inc, Cambridge, MA) and SAS statistical packages (SAS Institute, Cary, NC).

Results

Baseline Characteristics. Over 85% of eligible cases and 90% of eligible controls participated in this study with no significant demographic differences (age and gender) between enrolled and unenrolled eligible participants. A total of 2592 (98.9%) of enrolled subjects were successfully genotyped for all NAT2 polymorphisms. The distributions of race, gender, age, and smoking characteristics for those with genotype data were similar to the corresponding distributions observed for the entire study population. Complete information on age, gender, and smoking variables was available for 2453 subjects (94.6%). We restricted our analysis to the 2365 Caucasians with complete data. Of these, there were 1115 lung cancer cases and 1250 controls. There was 100% concordance for the 5% randomly repeated samples and 100% concordance between the laboratory data and data entry. Histological and clinical stage data were described in detail in other reports (21, 26).

The mean age (and SD) of all cases was 65 ± 10.8 years (range 26–91) and of controls, 58.5 ± 12.3 years (range 19–100). Females represented 46.5% of cases and 54.2% of controls. Among cases, the mean pack-years of ex-smokers was 55.2 ± 35.9 and of current smokers, 64.3 ± 35.7; among controls, the numbers were 29.3 ± 27.7 and 38.1 ± 25.2, respectively. The years since smoking cessation for ex-smokers were 14.1 ± 11 and 19.3 ± 12 for cases and controls, respectively.

Of the seven widely studied genetic NAT2 polymorphisms, we identified the most frequently occurring four: 481C→T, 590G→A, 803A→G, and 857G→A (11, 22, 23). We also determined two other polymorphisms (341T→C and 256C→T) in 312 random DNA samples, using standard methods (27), and found that 99.7% of the 341T→C polymorphism was linked with the 481C→T or 803A→G polymorphisms, and 100% of the 282C→T polymorphism results were concordant with the 590G→A or 857G→A polymorphisms (2, 7, 13, 14). The seventh polymorphism (191G→A), generally restricted to African populations, was not evaluated in our Caucasian-restricted samples (7, 13).

Distribution of NAT2 and mEH Polymorphisms Among Cases and Controls. All four polymorphic alleles of NAT2 were consistent with the Hardy-Weinberg equilibrium in our control population (P > 0.05, χ2 goodness of fit). The frequencies of genotypes representing slow, heterozygous rapid, and homozygous rapid acetylators were 60.2, 35.8, and 4% in cases; these frequencies were very similar in controls (59.9, 35.8, and 4.2%, respectively; P > 0.05). As pack-years increased, the frequencies of slow acetylator genotypes decreased in cases (χ2 trend test, P < 0.05) but not in controls (P > 0.05), e.g., the frequencies were 67% in cases and 56% in controls for nonsmokers and 61% in cases and 55% in controls with pack-years between 60 and 79, respectively.

A total of 896 cases and 1131 controls were genotyped for both NAT2 and mEH polymorphisms. The frequencies of the genotype combinations of NAT2 and mEH by pack-years and smoking status are shown in Table 1. Baseline characteristics of mEH genotypes have been reported previously (21).

Association Between NAT2 Genotypes and Lung Cancer Risk. No overall relationship between NAT2 genotypes and lung cancer risk was found. Compared with the slow acetylator genotypes, the crude ORs of lung cancer were 0.99 (95% CI, 0.84–1.18) for the heterozygous rapid acetylator genotype and 0.95 (95% CI, 0.63–1.43) for the homozygous rapid acetylator genotype. When adjusted for age, gender, SR-PY, smoking status, and years since smoking cessation, the ORs were 0.98 (95% CI, 0.80–1.20) and 0.76 (95% CI, 0.46–1.23), respectively. When the heterozygous and homozygous rapid acetylator genotypes were combined, the adjusted OR of the rapid versus slow acetylator genotypes was 0.96 (95% CI, 0.79–1.16).

Interaction Between NAT2 Genotype and Cumulative Cigarette Smoking in Lung Cancer Risk. NAT2 genotypes representing homozygous and heterozygous rapid acetylators were combined into one group for this analysis. In the gene-tobacco smoking interaction analysis, the three- and two-way interactions between smoking status and genotype were not significant, and these were removed from the model. The final model

### Table 1

<table>
<thead>
<tr>
<th>NAT2 Genotypes</th>
<th>Cases (%)</th>
<th>Controls (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>S + VL</td>
</tr>
<tr>
<td>Total</td>
<td>896</td>
<td>8.4</td>
</tr>
<tr>
<td>Pack-years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>51</td>
<td>13.7</td>
</tr>
<tr>
<td>1–19</td>
<td>88</td>
<td>11.4</td>
</tr>
<tr>
<td>20–39</td>
<td>185</td>
<td>6.0</td>
</tr>
<tr>
<td>40–59</td>
<td>228</td>
<td>8.3</td>
</tr>
<tr>
<td>60–79</td>
<td>135</td>
<td>8.2</td>
</tr>
<tr>
<td>≥80</td>
<td>209</td>
<td>8.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smoking status</th>
<th>Ex-</th>
<th>Current</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>479</td>
<td>7.9</td>
</tr>
<tr>
<td></td>
<td>51.2</td>
<td>51.4</td>
</tr>
<tr>
<td></td>
<td>6.7</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>34.0</td>
<td>36.1</td>
</tr>
<tr>
<td></td>
<td>524</td>
<td>209</td>
</tr>
<tr>
<td></td>
<td>8.8</td>
<td>11.5</td>
</tr>
<tr>
<td></td>
<td>51.9</td>
<td>55.0</td>
</tr>
<tr>
<td></td>
<td>5.2</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>34.2</td>
<td>28.7</td>
</tr>
</tbody>
</table>

*Individuals with combinations of slow (S) or rapid (R) acetylator genotypes and very low (VL) or combined higher mEH activity genotypes (H): Rapid acetylators were defined as individuals with either the heterozygote (*4/*4) and homozygote (*4/*4) rapid allele. Combined higher activity mEH genotypes (H) consisted of low, intermediate, and high activity genotypes. VL required the His/His genotype in both exons 3 and 4 polymorphisms.

*Nonsmokers were individuals with pack-years = 0.
NAT2 and mEH Polymorphisms, Smoking, and Lung Cancer

included interactions between NAT2 genotype and SR-PY \((P = 0.04)\) and between smoking status and SR-PY \((P = 0.03)\) and was adjusted for age, gender, NAT2 genotype, SR-PY, smoking status, and years since smoking cessation. In this model, the ORs of the rapid versus slow acetylator genotypes increased as SR-PY increased (Table 2). At 0 pack-years (nonsmokers), the fitted OR was 0.66 (95% CI, 0.44–0.99), indicating that the rapid acetylator genotype was protective against lung cancer. In contrast, at 80 pack-years, the fitted OR increased to 1.22 (95% CI, 0.85–1.67). No statistically significant results were found when cases of different histological types or disease stages were compared separately to all controls \((P > 0.05)\) for all comparisons.

**Table 2** Fitted ORs (95% CIs) of NAT2 and combined NAT2-mEH genotypes for different pack-years

<table>
<thead>
<tr>
<th>Pack-years of smoking</th>
<th>0</th>
<th>40</th>
<th>80</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAT2 alone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R vs. S (^b)</td>
<td>0.66 (0.44–0.99)</td>
<td>1.02 (0.83–1.26)</td>
<td>1.22 (0.89–1.67)</td>
<td>1.40 (0.92–2.14)</td>
</tr>
<tr>
<td>NAT2 and mEH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R + VL vs. S + VL (^b)</td>
<td>0.64 (0.40–1.01)</td>
<td>1.07 (0.85–1.34)</td>
<td>1.32 (0.94–1.86)</td>
<td>1.56 (0.98–2.48)</td>
</tr>
<tr>
<td>S + H vs. S + VL (^a)</td>
<td>0.47 (0.27–0.83)</td>
<td>1.14 (0.85–1.54)</td>
<td>1.65 (1.07–2.56)</td>
<td>2.19 (1.23–3.95)</td>
</tr>
<tr>
<td>R + H vs. S + VL (^b)</td>
<td>0.30 (0.14–0.62)</td>
<td>1.22 (0.84–1.78)</td>
<td>2.19 (1.26–3.81)</td>
<td>3.42 (1.61–7.26)</td>
</tr>
<tr>
<td>R + H vs. S + VL (^c)</td>
<td>0.35 (0.16–0.78)</td>
<td>1.27 (0.85–1.91)</td>
<td>2.16 (1.19–3.92)</td>
<td>3.25 (1.45–7.30)</td>
</tr>
</tbody>
</table>

\(^a\) R, NAT2 rapid acetylator genotypes; S, NAT2 slow acetylator genotypes; VL, very low mEH activity genotype; H, combined low, intermediate, and high mEH activity genotypes.  
\(^b\) Logistic regression model (final model) included variables for age, gender, SR-PY, smoking status (current, ex-, and nonsmokers), years since smoking cessation, genotypes (NAT2 alone or NAT2 and mEH), interaction terms between each genotype and SR-PY, and interactions between smoking status and SR-PY.  
\(^c\) The initial model (gene-gene interaction) included variables of age, gender, SR-PY, smoking status, years since smoking cessation, mEH genotype, NAT2 genotype, interaction between smoking status and SR-PY, all two-way interactions between NAT2, mEH, and SR-PY, and the three-way interaction between NAT2, mEH, and SR-PY. This model provides algebraically equivalent ORs to the gene-gene joint-effect model.

We investigated the interaction between NAT2 and mEH (gene-gene interaction) in lung cancer risk, because NAT2 and mEH have similar dual pathways in metabolizing tobacco smoking chemicals, and similar interactions between these genetic polymorphisms and SR-PY had been observed (21). In the initial model, we included the three-way interaction between NAT2, mEH, and SR-PY, all nested two-way interactions, and the main effects. In addition, the interaction between smoking status and SR-PY was included, whereas all other interactions with smoking status were not included in this model, based on previous analyses of our data (21). The three-way interaction term was not statistically significant \((P > 0.25)\), and it was then removed from the model. The final NAT2 and mEH tobacco smoking interaction model included the interaction terms between NAT2 genotype and SR-PY \((P = 0.03)\), mEH genotype and SR-PY \((P < 0.01)\), and smoking status and SR-PY \((P = 0.07)\), in addition to age, gender, NAT2 genotype, mEH genotype, SR-PY, smoking status, and years since smoking cessation. Compared with the combination of the NAT2 slow acetylator and very low mEH activity genotype, the ORs of each of the other three genotype combinations increased as pack-years increased. The ORs of the extreme combination of the NAT2 rapid acetylator and higher mEH activity genotypes increased most significantly (Table 2; Fig. 1). For nonsmokers (0 pack-years), the fitted OR for this comparison was 0.30 (95% CI, 0.14–0.62); at 80 pack-years, the fitted OR was 2.19 (95% CI, 1.26–3.81).

Although no significant interaction between NAT2, mEH, and SR-PY was observed, we wanted to examine the sensitivity of our results to the presence or absence of the three-way interaction between NAT2, mEH, and SR-PY. We compared the fitted ORs from our final model with the ORs from the model that included this three-way interaction and all nested two-way interactions and main effects (initial model). When comparing the final (non-gene-gene interaction) with the initial model (gene-gene interaction, equivalent to the joint-effects model), the fitted ORs were essentially unchanged (Fig. 1), whereas the CIs were slightly wider under the initial model (Table 2).

**Discussion**

The relationship between NAT2 genotypes and lung cancer risk has been examined extensively. Although results have been contradictory, most studies did not account for notable interactions with smoking exposure (5–13). The one study that examined this role was underpowered for formal interaction testing (14). Our study is the first to report a statistically significant interaction between NAT2 genotype and pack-years of smoking in lung cancer risk; NAT2 rapid acetylator genotypes (when compared with the slow acetylator genotypes) decreased lung cancer risk in nonsmokers and increased risk in heavy smokers. This NAT2 genotype-cumulative cigarette smoking interaction may explain partly the inconsistent results reported previously, based on variation in baseline smoking characteristics across different study populations.

Our study has several important strengths, including a large sample size that permitted formal analysis of interactions between NAT2 genetic polymorphisms and smoking. This large sample size permitted the use of lifelong nonsmokers as a stable baseline comparison group in gene-environment interaction analyses. In addition, complete smoking data were collected systematically, diagnoses of lung cancer were histologically confirmed, and females were well represented in both cases and controls. In an attempt to avoid overmatching, we did not individually match our controls to cases on sex, race, or smoking variables (28, 29). Our method resulted in a reasonable balance of these characteristics between cases and controls, because the distributions of smoking variables in our controls were similar to the general Massachusetts population overall and in age- and gender-specific strata (26),\(^{2,5}\)

The use of fitted ORs can occasionally lead to the identification of artifactual relationships among variables. Before we

\(^{2}\) Internet address: http://www.state.ma.us/dph/mtcp/report/mats.htm.  
\(^{5}\) W. Zhou et al., unpublished data.
carried out the gene-smoking interaction analysis, we used a two-step test to identify possible misclassification of subjects with respect to cigarette smoking exposure (30). An association between NAT2 genotype and SR-PY was found in the case-only analysis (P = 0.04) but not in the control-only analysis (P > 0.05). Because the association between NAT2 genotype and SR-PY was not statistically significant in controls, the true smoking exposure is independent of the genotype among controls; if there is misclassification of smoking exposure, it is likely that the possible misclassification is nondifferential with respect to the genotype. According to Garcia-Closas et al. (30), the pattern of association observed in our study may actually underestimate the true gene-smoking interaction but would not lead to an artifactual statistical interaction. In addition to fitted ORs, we obtained ORs from separate analyses that stratified at different pack-years of cigarette smoking (heavy versus non or light smokers and at different quartiles of pack-years). The crude and adjusted ORs for different strata were similar in magnitudes and directions to the fitted ORs derived from the interaction models (data not shown). Additionally, we decomposed pack-years into its component parts of smoking intensity (mean number of cigarettes per day) and duration (in years) and found similar gene-smoking interaction associations when either cigarettes per day or smoking years (both as continuous variables) were substituted for pack-years into its component parts of smoking intensity (as defined by genotype) in our study were very similar.

There are biological reasons to consider an interaction between NAT2 polymorphisms and smoking variables, because the purported mechanism for modifying lung cancer risk by these polymorphisms is dependent on their role in metabolizing tobacco-related substrates. This interaction is particularly important in light of evidence for a dual role of activation and detoxification of tobacco procarcinogens and carcinogens by the NAT2 enzyme. For nonsmokers, NAT2 appears to provide a competitive alternative pathway to the cytochrome P4501A2 and P4501A1-catalyzed N-oxidation pathway for reactive hydroxylamines without an alternative mechanism. These intermediates are then further metabolized into DNA-binding electrophiles (34). NAT2 N-acetylates and detoxifies these aromatic amine compounds (35, 36) and is a possible explanation for why the rapid acetylator genotypes showed a possible protective effect in nonsmokers.

By contrast in smokers, the amount of inhaled cigarette smoke appears to differentially induce cytochrome P4501A2 and 1A1-catalyzed N-oxidation, whereby heavy smoking significantly induces enzyme activity and increases the number of these reactive intermediates (37). NAT2 can then O-acetylate these hydroxylamines to acetoxy esters, which can spontaneously disintegrate into highly reactive aryl nitrenium (38). These reactive products have been associated with other genetic derangements, such as p53 mutations in tobacco tissues (39, 40). Thus, higher NAT2 activity may result in an increased lung cancer risk in heavy smokers.

The varying relationship between NAT2 genotypes and lung cancer risk at different pack-years is supported by indirect evidence from several adduct studies. Adducts (4-aminobiphenyl-hemoglobin) were higher in slow acetylators at low or null nicotine-cotinine levels, whereas this difference was less evident at increasing nicotine-cotinine levels, and adduct levels were higher in rapid acetylators at the highest nicotine-cotinine levels (41, 42). In a Caucasian study of subjects >60 years old,
rapid acetylator smokers had significantly higher frequencies of chromosome aberration than slow acetylator smokers (43).

Specific evaluations of the smoking habits of the underlying lung cancer study populations in other studies provide some corroborative evidence. The possible protective effect of the rapid acetylator genotypes for lung cancer observed among nonsmokers in this population has also been reported in two previous studies with nonsmoker predominant populations (11, 14). In contrast, the homozygous rapid acetylator genotype was reported to be a risk factor for lung cancer in another study of Caucasians; 77% of the 155 cases in this particular study were smokers with >20 pack-years of smoking (13). However, the lack of uniformity in the design, analysis, smoking data, and reporting may explain discrepancies across previous reports of NAT2 polymorphisms and lung cancer risk (Table 3).

This is the first report showing that NAT2 and mEH genotypes interact independently with smoking exposure in modifying lung cancer risk, and our results are consistent with the concept that NAT2 and mEH metabolize different tobacco-associated substrates. We found that the extreme joint genotype combination of these two genes had a very strong interaction with cumulative smoking exposure, stronger than with either gene alone. In addition, the intermediate genotype combinations also showed consistent, significant trends with increasing smoking exposure (Table 2; Fig. 1). The results of the comparison of the extreme joint genotype combinations were very similar in models with and without a gene–gene interaction, consistent with an additive effect of the two polymorphic genes at each level of smoking exposure (Fig. 1).

The results of this study provide empirical evidence for two important interrelated paradoxes. Firstly, genetic polymorphism-environmental relationships can influence cancer risk. Our results illustrate how negative overall results can hide important relationships that are only uncovered when an environmental factor integral to the proposed pathogenetic mechanism (such as smoking) is incorporated into the analysis. Secondly, combinations of different genetic polymorphisms can have additive effects on cancer risk. Cigarette smoke contains >4000 constituents, including ~50 carcinogens (44); the carcinogenic effects of these compounds are mediated by multiple biotransformation enzymes in complex interrelated pathways. Our results highlight the potential for individual risk profiling through the evaluation of multiple genetic polymorphisms.

In conclusion, cumulative cigarette smoking exposure appears to play a critical role in altering the direction and magnitude of the association between NAT2 polymorphisms and the risk of lung cancer. Lung cancer risk by combined genotypes of both NAT2 and mEH is dependent on cumulative smoking exposures. The increase in risk conferred by carrying the higher risk mEH genotypes (low, intermediate, and high activity genotypes) and the higher risk rapid acetylator NAT2 genotypes appears to be independent and additive.

Acknowledgments

We thank the following staff members of the Lung Cancer Susceptibility Group: Barbara Bean, Andrea Solomon, Linda Lineback, Lucy Ann Principe, Salvatore Mucci, Richard Rivera-Massa, Lisa I. Wang, Rong Fan, Eike Schaffner, Yang Sai, Stephanie Shih, Maria Fragoso, and Jessica Shim and the generous support of Dr. Panos Fidias and the physicians and surgeons of the Massachusetts General Hospital Cancer Center.

References

Genetic Polymorphisms in $N$-acetyltransferase-2 and Microsomal Epoxide Hydrolase, Cumulative Cigarette Smoking, and Lung Cancer

Wei Zhou, Geoffrey Liu, Sally W. Thurston, et al.