Combinations of the Variant Genotypes of GSTP1, GSTM1, and p53 Are Associated with an Increased Lung Cancer Risk

David P. Miller, Geoffrey Liu, Immaculata De Vivo, Thomas J. Lynch, John C. Wain, Li Su, and David C. Christiani

ABSTRACT

GSTP1 and GSTM1 are genes involved in Phase II metabolism, whereas p53 is a tumor suppressor gene. Individually, functional polymorphisms of these genes have been studied as risk factors for lung cancer. Small sample sizes have hindered the detection of possible increases in risk associated with having two or more “at risk” polymorphisms of these three genes. In a large Caucasian population, we examined the association of combined variant genotypes [or double-variants (DVs)] of these three genes and lung cancer risk, compared with their corresponding “double-wild-type” genotypes. Because these DVs may promote lung carcinogenesis at an earlier age, a subgroup of individuals aged 55 years or younger was examined separately.

Using a case-control design, individuals were genotyped for GSTM1, GSTP1, and p53 codon 72 using PCR-RFLP techniques. All of the analyses used multiple logistic regression. Indicator variables were created to evaluate the risk for individuals with the following DVs: GSTP1 GG + GSTM1-null and GSTP1 GG + p53 Arg/Pro or Pro/Pro.

A total of 1694 cases and controls were evaluated. In the whole population, those with the double variants have a higher risk of lung cancer compared with those with the double-wild-type genotypes, supporting our original hypothesis. Individuals with the GSTP1 and GSTM1, DV (P1-M1 DV) had a marginally significant higher risk of lung cancer compared with their double-wild-type counterparts [adjusted odds ratio (AOR), 1.60; 95% confidence interval (CI), 0.95–2.70]. A significantly higher risk was found for the GSTP1, p53 DV (P1-p53 DV; AOR, 1.99; 95% CI, 1.12–3.53). Among individuals aged 55 or younger, these risks were even higher: for the P1-M1 DV the AOR was 4.03 (95% CI, 1.47–11.1); for the P1-p53 DV the AOR was 5.10 (95% CI, 1.42–18.30).

Specific DVs of GSTM1, GSTP1, and p53 codon 72 are associated with a higher lung cancer risk. This susceptibility is highest among younger individuals.

INTRODUCTION

Genetic susceptibility is one of the primary hypotheses used to explain why a minority of smokers develops lung cancer. Genetic polymorphisms involved in metabolism of carcinogens have been studied as possible risk factors for lung cancer (1–10). GSTs,1 a major group of enzymes whose main classes are α, μ, π, and θ, are directly involved in the detoxification step of this metabolic process (Phase II; Ref. 11). The expression of these enzymes differs among organs. GST π has the highest expression in the lung and is one of the main detoxifiers of the activated form of benzopyrene (12–14). GST π is encoded by a polymorphic gene, GSTP1. Polymorphisms of GSTP1 are a consequence of a single bp substitution, where (A) adenine is replaced by (G) guanine, leading to an amino acid substitution in which Isoleucine (I105) is replaced by Valine (V105). This substitution results in a lower enzymatic activity (15, 16) and is associated with higher hydrophobic adduct levels in lung tissue (7) and higher levels of polycyclic aromatic hydrocarbon-DNA adducts in human lymphocytes (17). Recent studies examining the association between GSTP1 polymorphisms and lung cancer have found no statistically significant associations (2, 3, 8, 10, 18, 19). The polymorphism in the GSTM1 gene (which encodes for the GST μ enzyme) is a germ line deletion (known as GSTM1-null), which results in a total loss of enzymatic function. The GSTM1-null genotype has been studied extensively with contradictory results as to whether there is any association with lung cancer risk (2, 4, 5, 7–9, 20–23).

More recently, polymorphic genes involved in cell cycle regulation, apoptosis, and tumor suppression have been studied as possible risk factors for lung cancer (9, 24–26). p53 is one of these genes. Somatic point mutations in this gene have been associated with increased risk in various cancers. The gene is located on chromosome 17p13 and is one of the most commonly mutated genes in all of the human cancers (27, 28). The codon 72 p53 polymorphism is a result of a single bp substitution: guanine is replaced by cytosine leading to an arginine (Arg) replaced by proline (Pro). The wild-type p53 gene operates by suppressing cellular transformation by activated oncogenes, thus inhibiting the growth of malignant cells (29, 30). Point mutations in p53 alter the DNA binding properties and transcription factor function, resulting in cellular proliferation (31). Several studies have reported that p53 Pro/Pro is associated with a higher risk of lung cancer (9, 24–26). Mutations in the p53 gene have also been associated with altered repair and enhanced cytotoxicity because of DNA damage by benzo(a)pyrene diol epoxide adducts (32, 33).

Individuals carrying more than one of the at risk polymorphisms may have a greater risk of developing lung cancer. Two studies found that the combination of GSTP1 GG and GSTM1-null genotypes was associated with a higher risk of lung cancer (2, 3). Another study found that lung cancer risk associated with the GSTM1-null polymorphism increased in the presence of the p53 Pro allele (9). We confirmed recently this association in our study population (34). Small sample size was a limiting factor in the previous studies. We reported recently a significant association between GSTP1 GG and early age onset lung cancer (diagnosed at 55 years or younger),4 the presence of DVs may be associated with early age onset lung carcinoma through an accelerated accumulation of DNA damage leading to carcinogenesis.

In this study, we tested the following a priori hypotheses on a large sample of 1694 participants (767 cases and 927 controls): (a) the GSTP1 GG, GSTM1-null DV combination (P1-M1 DV) is associated with increased lung cancer risk; (b) the GSTP1 GG, p53 variant (Pro/Pro + Arg/Pro) DV combination (P1-p53 DV) is associated with

1 Supported by NIH Grants CA 74386, ES/CA 06409, and ES00002.
2 To whom requests for reprints should be addressed, at Occupational Health Program, Harvard School of Public Health, 665 Huntington Avenue, Building 1 Room 1407, Boston, MA 02115. Phone: (617) 432-3323; Fax: (617) 432-3441; E-mail: dchris@hohp.harvard.edu.
3 The abbreviations used are: GST, glutathione S-transferase; DV, double-variant; MGH, Massachusetts General Hospital; AOR, adjusted odds ratio.
lung cancer risk. We repeated the analyses for those participants <55 years of age to assess whether the risks were higher among younger individuals.

**MATERIALS AND METHODS**

Subjects for this study were recruited as part of an ongoing case-control study initiated in 1992 at the MGH in Boston, Massachusetts. The study was approved by the Institutional Review Board at both MGH and Harvard School of Public Health. Eligible cases included any person over 18 years of age with a diagnosis of primary lung cancer that was evaluated by the pulmonary, thoracic surgery, or hematology-oncology units at MGH for either surgery (from 1992), chemotherapy and radiation treatment (from 1996), or any combination of treatment modalities. An MGH lung pathologist confirmed all of the cases histologically. Controls were recruited first among friends and nonblood-related family members of the cases (usually spouses). If friends of lung cancer patients were not available, controls were recruited from friends and family of patients receiving thoracic surgery, chemotherapy, or radiation treatment for a condition other than lung cancer. To determine whether our set of controls were similar to the Massachusetts general population, we compared important covariate data obtained from our controls with information provided by the Massachusetts Tobacco Survey, 1993–1997. A research nurse administered the health questionnaires. Some participants opted to complete the questionnaires at home and returned them by mail in a self-addressed stamped envelope. Participants were contacted by telephone when there was missing data. The participation rate was >85% and did not differ between cases and controls. The analysis was restricted to the 1694 participants for whom, at time of analysis, all of the data (genotypic as well as questionnaire data) was available. To reduce potential variation in allele frequency by ethnicity, only Caucasians were considered in the analysis.

Blood samples were collected from all of the participants at the time of recruitment. Two to three 10-mI EDTA tubes and one 5-mI silicon-coated tube were used for sample collection. Samples were processed in the molecular epidemiology laboratory at the Harvard School of Public Health. DNA was extracted from whole blood for the purpose of genotyping.

All of the genotyping was performed using PCR-RFLP techniques and blinded to case or control status. Genotyping for GSTP1, GSTM1, and p53 have been described previously (18, 24, 35). For quality control, a random 5% of the samples were repeated with 100% concordance. Two authors reviewed independently 100% of the agarose gels and genotype data entry.

Other variables were obtained through the health questionnaire (20, 24, 35). Cumulative exposure to cigarette smoking was estimated in pack-years by multiplying the mean number of packs smoked per day by the number of years of smoking, taking into account periods of smoking cessation.

Population characteristics were tabulated, and significant differences in the distribution of the principal covariates were tested using the χ², Fisher exact, and Student t tests, where appropriate. The distributions of the genes were checked among controls to verify if they were in Hardy-Weinberg equilibrium. The number of people in each gene-gene category was evaluated to ensure that none of the categories had fewer than five people. Multiple logistic regression was used to assess the association between the DV combinations and lung cancer. In each model GSTP1 GG and AG genotypes were considered separately, thus avoiding assumptions required when combining the heterozygote and the homozygote variant. The p53 Pro/Pro genotype because of the low frequency of the homozygote variant and the similar risk associated with the heterozygote and homozygote variant (24). Two models were used to evaluate the association between the two pairs of DV and lung cancer. For each model, we obtained both crude odds ratios, and odds ratios adjusted for age, gender, smoking status (indicator variables for nonsmoker, ex-smoker, and current smoker), and a continuous variable for cumulative smoking exposure (pack-years). The first model (Model 1) assessed the independent effect of PI-M1 DV and PI-p53 DV by including indicator variables for both sets of genes (Fig. 1). In Model 1, the combined effect of each pair of polymorphic variants was represented by the product of the two AORs for the variant polymorphisms, assuming an independent effect of each polymorphism on lung cancer risk. Model 2 evaluated the joint effects of each pair of polymorphic variants and made no assumption of independence between each polymorphism pair. In the second model, indicator variables were created to represent the six possible polymorphic combinations: GSTP1 has three polymorphisms (AA, AG, and GG), GSTM1 has two (null and wild-type), and p53 was dichotomized (AA and AP+PP). The polymorphism combinations investigated were GSTP1 + GSTM1 (3 × 2 = 6 combinations), and GSTP1 + p53 (3 × 2 = 6 combinations). If the two genes had a simple additive effect, the result of the product of AORs from Model 1 should be similar to the results of the corresponding AORs from Model 2. Differences in the AORs from the two models would imply a deviation from an additive, independent effect of the two polymorphisms. The statistical significance of the odd ratios was tested using the Wald test (36). The log odds of the DV were compared, and percentage differences were calculated to estimate the magnitude of difference between Models 1 and 2.

To determine whether the DVs have a role in early age onset of lung cancer, the adjusted analysis for Models 1 and 2 were repeated in the subgroup ≤55 years of age. The average age of lung cancer patients in our study was 65 years; thus, restricting the analysis to those ≤55 years of age identifies a subgroup much younger than average lung cancer patients. Also, using age 55 or younger as a cutoff allowed an appropriate sample (n = 470; 27%) for the analysis. The SAS statistical package was used to perform all of the analyses (37).

**RESULTS**

Table 1 shows the distribution of demographic, histological, and genotypic characteristics among cases and controls. Cases were older than controls (67 versus 61 years; P < 0.01). Women were well represented among cases (45%); they were over-represented among controls (55%; P < 0.01). The frequency of current smokers was higher among cases than controls. The frequency of ex-smokers was similar in both groups, whereas the frequency of nonsmokers was significantly higher among controls (P < 0.01). Among ever-smokers, cases had accumulated much more pack-years of smoking than controls (52 versus 27 pack-years; P < 0.01, respectively). The distribution of smoking variables in our controls was similar to the general Massachusetts population over age 45. The proportion of non-, ex-, and current smokers were 36%, 46%, and 18% in our controls and 36%, 47%, and 17% in the general Massachusetts population over age 45, respectively. For current smokers, mean cigarettes per day (controls: 21.2 cigarettes; Massachusetts: 21.4 cigarettes) and earliest age of smoking (controls: 17.9 years; Massachusetts: 17.9) were similar. For ex-smokers, the proportions of those who had quit smoking for more than 5 years were 87.4% (controls) and 85.5% (Massachusetts).

Education level was not available for 45 study participants (32 cases and 13 controls). The number of college graduates was higher among controls than cases (29% versus 23%; P = 0.02).

The distribution of the GSTP1 and GSTM1 polymorphisms did not
differ significantly between cases and controls (P = 0.30 and P = 0.75, respectively). The distribution of the p53 Pro/Pro and Arg/Pro genotypes was slightly higher in cases compared with controls (13% and 39% versus 10% and 37%; P = 0.03). The genotype frequencies were in Hardy-Weinberg equilibrium among the controls. Almost half the lung cancer cases had adenocarcinomas (44%); one quarter had squamous cell carcinomas, whereas the remaining 32% comprised all of the other histological cell types. Staging information was available for 88% of the cases: stage I and II, 55%; stage III and IV, 33%. The relatively high percentage of early stage (potentially curable) cancers in our sample reflected a referral bias because of the surgical expertise at MGH, as well as the selection criteria for cases between 1992 and 1996, which favored cases eligible for surgical treatment (i.e., early stage I and II).

Table 2 presents the number of participants stratified by case status and genotype combination. The smallest cell size was 36 (P1-p53 DV among controls), reflecting sufficient numbers to conduct our planned analysis.

Crude and AORs from Models 1 and 2 are shown in Table 3. We observed a marginally statistically significant association for P1-p53 DV compared with the double-wild-type (OR, 1.59; 95% CI, 1.00–2.57; P = 0.06). The differences in the log odds from Model 1 and Model 2 were 87% for the P1-M1 DV and 47% for the P1-p53 DV. The AOR for the P1-M1 DV was suggestive of a higher lung cancer risk (AOR, 1.60; 95% CI, 0.95–2.70; P = 0.07). The AOR for the P1-p53 DV was highly significant (AOR, 1.99; 95% CI, 1.12–3.53; P < 0.01). The percentage differences of the log odds between Models 1 and 2 were 68% for the P1-M1 DV and 44% for the P1-p53 DV.

Finally, Table 4 shows the AOR for participants ≤55 years of age. The AORs associated with both DVs were much greater than for the whole population: for the P1-M1 DV the AOR was 4.03 (95% CI, 1.47–11.08; P < 0.01); for the P1-p53 DV the AOR was 5.10 (95% CI, 1.42–18.30; P < 0.01). The respective differences of the DV risk between Models 1 and 2 were 7% and 5%. The association between both sets of DVs and lung cancer risk among participants >55 years was not statistically significant: AOR for P1-M1 DV was 1.16 (95% CI, 0.62–2.19; P = 0.62), and AOR for P1-p53 DV was 1.66 (95% CI, 0.86–3.22; P = 0.09; data not shown).

DISCUSSION

Small studies have reported the association of combined gene polymorphisms with lung cancer risk. Jourenkova-Mironova et al. (Ref. 2; n = 322) reported a higher risk of lung cancer associated with the combination of the GSTM1 Null, GSTP1 AG+GG, and GSTM3 AA. The interaction between GSTP1 and GSTM3 was evaluated in each strata of the GSTM1 polymorphism, and was restricted to Caucasian smokers only. The small numbers in each comparison category was a major limiting factor in this study where the smallest and largest categories had 4 and 39 individuals, respectively. Kihara and Noda (Ref. 3; n = 542) reported an indication of a potential interaction between the GSTP1 and GSTM1 genes in a population of Japanese male smokers ages 50–69 years in which a higher risk of lung cancer was associated with the combination of the variant allele for GSTP1 and the GSTM1-null genotype. Their study lacked data on females. To-Figuera et al. (Ref. 9; n = 286) reported a higher risk of lung cancer among northwestern Mediterranean smokers (93% men) associated with the GSTM1-null and p53 Pro/Pro + Arg/Pro genotypes. However, the small sample size and the inability to adjust for important confounders, such as smoking, limited interpretation of this study. We reported previously (n = 862) an increased risk of lung cancer associated with the combination of the CYP1A1Msp1 and GSTM1 null genotypes (20) but were unable to significantly confirm a gene–gene interaction. When the effect of the combined genotypes (joint effect) is different from the sum of the independent effects of each genotype considered, a greater than additive effect may be present.

Our study supports that the P1-M1 DV in both genders is associated with increased lung cancer risk in a large Caucasian population. Furthermore, this study is the first, to our knowledge, to describe an increased risk of lung cancer associated with the combination of GSTP1 and GSTM1 genes in a population of Japanese male smokers ages 50–69 years in which a higher risk of lung cancer was associated with the combination of the variant allele for GSTP1 and the GSTM1-null genotype. Their study lacked data on females. To-Figuera et al. (Ref. 9; n = 286) reported a higher risk of lung cancer among northwestern Mediterranean smokers (93% men) associated with the GSTM1-null and p53 Pro/Pro + Arg/Pro genotypes. However, the small sample size and the inability to adjust for important confounders, such as smoking, limited interpretation of this study. We reported previously (n = 862) an increased risk of lung cancer associated with the combination of the CYP1A1Msp1 and GSTM1 null genotypes (20) but were unable to significantly confirm a gene–gene interaction. When the effect of the combined genotypes (joint effect) is different from the sum of the independent effects of each genotype considered, a greater than additive effect may be present.

Our study supports that the P1-M1 DV in both genders is associated with increased lung cancer risk in a large Caucasian population. Furthermore, this study is the first, to our knowledge, to describe an increased risk of lung cancer associated with the P1-p53 DV. In general, the lung cancer risk associated with the DVs from the joint effects models (Model 2) was greater than the simple sum of the independent risks associated with the variant genes as independent risk factors (Model 1). The mechanism behind these effects for the DVs is unclear. Tobacco carcinogens are metabolized typically in a two-phase activation-deactivating sequence. Phase I results in the activation of carcinogens, usually mediated through enzymes encoded by the CYP family of genes. The detoxification phase (Phase II) occurs when the activated carcinogen is processed and rendered more hydrophilic, thus excretable. GSTP1 and GSTM1 are directly involved in Phase II metabolism (11, 13). Furthermore, the GSTP1 polymorphism codes for an enzyme, GST class π, which is a major metabolizer of the activated products of benzo(a)pyrene, a principal carcinogen in cigarette smoke.
smoke (12–14). The presence of either the GSTP1 GG or GSTM1-null leads to a reduced Phase II function, which leads to higher levels of activated carcinogens. In turn, DNA adduct formation occurs, if not repaired results in accumulated DNA damage. GSTP1 GG and GSTM1-null are independently associated with higher levels of DNA adducts compared with their wild-type counterparts in humans (7, 17). p53 mutations have a reduced repair capability in cells damaged by benzo(a)pyrene diol epoxide-DNA adducts (32, 33). The reduction in protein function alters the “gatekeeper” ability of the p53 protein to stop cell cycle progression to allow repair (32, 33), leading to cell damage and carcinogenesis. The p53 Pro/Pro polymorphism is associated with reduced apoptotic kinetics (38). DNA damage mediated through decreased Phase II function (GSTP1 GG and/or GSTM1-null) combined with altered function of a tumor suppressor gene (p53 Pro/Pro) may lead to carcinogenesis.

It is important to note that the presence of a joint effect greater than the sum of the independent effects does not necessarily imply that the two genes being studied are in the same biological pathways. Kinzler and Vogelstein (39) provided an example of this phenomena in somatic mutations: a gene involved in cell proliferation and another involved in cell adhesion function; two different pathways. Even although alteration of either gene may result in slight progression toward neoplasia, it is the combined alteration of these two genes that results in substantial progression toward neoplasia (39). Unique interactions of multiple variants can affect biological phenotypes for which individual variants may have poor prediction power.

The risk of lung cancer associated with the P1-M1 DV and P1-p53 DV were greatest among participants ≤55 years of age. A younger age at diagnosis may reflect an underlying genetic predisposition. We reported previously a strong association between GSTP1 GG and early age onset lung cancer (2); the current study strengthens these associations by reporting consistent results in DVs that involve the GSTP1 GG genotype. An interesting finding is that in these younger individuals there appears to be only an additive effect of each polymorphic variant in these DVs. However, the wide confidence intervals in this subgroup may mask substantial deviations from an additive effect. Alternatively, time to diagnosis may be an important factor in producing greater than additive effects, where the subtle permissive effects result in substantially higher risk over a lifetime of accumulated DNA damage.

Our study benefited from a large sample size and relatively homogeneous ethnic group (Caucasians), but this may affect generalizability to other populations. Although cases smoked more than controls, as expected, and the number of men among cases was higher than among the controls, these factors are accounted and adjusted for in the multivariate analysis. In addition, we evaluated both a joint effects model and a model that assumed independence of these genetic polymorphisms to show that the joint effects were at least additive if not more than additive. Although our sample size was larger than case control studies reported previously, our study was still not large enough to formally test for the significance of an interaction term (40, 41).

In conclusion, we have shown that the joint effects of the two sets of DVs (GSTP1 GG, GSTM1-null and GSTP1 GG, p53 variant) were associated with lung cancer risk. Furthermore, the risk is greatest among participants ≤55 years of age, additionally supporting the role that several genes may play in early age onset lung cancer. Additional studies will help elucidate the possible mechanisms involved in the steps leading to carcinogenesis (i.e., correlation between intermediate biomarkers associated with at risk and lung cancer risk). Although the combined DV risks generally had a greater than additive effect, larger sample sizes are needed to consider differences because of gender, histology, and to also clarify the association with early age onset lung cancer.

**ACKNOWLEDGMENTS**

We thank Linda Lineback, Barbara Bean, Jeanne Jackson, and Andrea Solomon for patient recruitment and data collection; Lucy Ann Principe, Salvatore V. Mucci, and Richard Rivera-Massa for data entry, Stephanie Shih for sample preparation and genotyping; and Shannon Magari for editing advice and comments.

**REFERENCES**


Combinations of the Variant Genotypes of GSTP1, GSTM1, and p53 Are Associated with an Increased Lung Cancer Risk

David P. Miller, Geoffrey Liu, Immaculata De Vivo, et al.