CYP1A1 and GSTM1 genetic polymorphisms and lung cancer risk in Caucasian non-smokers: a pooled analysis

Rayjean J.Hung1,2, Paolo Boffetta1,17, Jürgen Brockmöller3, Dorota Butkiewicz4, Ingolf Cascorbi5, Margie L.Clapper6, Seymour Garte7, Aage Haugen8, Ari Hirvonen9, Sisko Anttila9, Ivan Kalina10, Loïc Le Marchand11, Stephanie J.London12, Agneta Rannug13, Marjorie Romkes14, Ivan Kalina10, Bernadette Schoket15, Laura Gaspari16 and Emanuela Taioli16

1Unit of Environmental Cancer Epidemiology, International Agency for Research on Cancer, 150 Cours Albert-Thomas, F-69372 Lyon Cedex 08, France, 2Department of Epidemiology, School of Public Health, University of California at Los Angeles, Los Angeles, CA, USA, 3Institute of Clinical Pharmacology, Georg August University Goettingen, Germany, 4Department of Tumor Biology, Centre of Oncology, Gliwice, Poland, 5Institute of Pharmacology, Ernst Moritz Arndt University Greifswald, Germany, 6Divisions of Population Science and Medical Science, Fox Chase Cancer Center, Philadelphia, PA, USA, 7Genetics Research Institute ONLUS, Milan, Italy, 8Department of Toxicology, National Institute of Occupational Health, Oslo, Norway, 9Finnish Institute of Occupational Health, Helsinki, Finland, 10Department of Medical Biology, School of Medicine, P.J.Safárik University, Kosice, Slovakia, 11Cancer Research Center of Hawaii, University of Hawaii, Honolulu, HI, USA, 12Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA, 13Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, 14Environmental and Occupational Health, Center for Clinical Pharmacology, University of Pittsburgh, Pittsburgh, PA, USA, 15Department of Biochemistry, National Institute of Environmental Health, Budapest, Hungary and 16Department of Environmental Medicine, IRCCS University of Milano, Milan, Italy

17To whom correspondence should be addressed Email: boffetta@iarc.fr

Polymorphisms for genes encoding the metabolic enzymes cytochrome P450 1A1 (CYP1A1) and glutathione S-transferase M1 (GSTM1) might contribute to the variability in individual susceptibility to lung cancer. The role of CYP1A1 and GSTM1 in lung carcinogenesis might be more important at low levels of exposure to carcinogens. Non-smokers represent a population at low exposure, however, they are often overlooked because of the small number of cases. We therefore conducted a pooled analysis of 14 case-control studies on lung cancer in Caucasian non-smokers with comparable information on genetic polymorphisms included in the International Collaborative Study on Genetic Susceptibility to Environmental Carcinogens. We pooled the raw data from a total of 302 cases and 1631 controls with random effects models. We also evaluated the possibility of inclusion bias and conducted influence analyses. The odds ratio (OR) of lung cancer for the variant CYP1A1 Ile462Val polymorphism (Ile/Val, Val/Val) was 2.99 [95% confidence interval (95% CI) 1.51–5.91]; this effect was stronger on lung adenocarcinoma (OR 4.85, 95% CI 2.03–11.6). After excluding outlying or imprecise studies, we did not observe a significant effect of the CYP1A1 Mspl (T380I) polymorphism or GSTM1 null genotype (OR 1.20, 95% CI 0.89–1.63). Furthermore, our analyses suggested a combined effect of the CYP1A1 Ile462Val polymorphism and GSTM1 null genotype. The OR for the combination of the CYP1A1 Ile462Val variant and GSTM1 null genotype was 4.67 (95% CI 2.00–10.9) compared with the concurrent presence of the CYP1A1 wild-type and GSTM1 non-null genotype. We did not observe a modification of the effect of the GSTM1 null genotype according to exposure to environmental tobacco smoke and urban/rural residence. Our study therefore suggests that the CYP1A1 Ile462Val variant allele might play a role in lung carcinogenesis among non-smokers, possibly in combination with the GSTM1 null genotype.

Introduction

Lung cancer is the most common malignancy and the leading cause of cancer death in men world wide, with an estimated 900 000 new cases and 810 000 deaths per year. It is also the second most lethal cancer in women, after breast cancer (1,2). Tobacco smoking has been established as the most important etiological factor of lung cancer for both men and women (3–5). However, only a fraction of smokers will eventually develop lung cancer, depending on the extent of smoking, exposure to other environmental carcinogens and individual susceptibility. Furthermore, the etiology of lung cancer arising in lifetime non-smokers is still not fully established. Besides involuntary smoking, lung cancer in non-smokers has also been associated with dietary factors, occupational exposures, prior lung disease and indoor air pollutants. However, the excess risks of the above factors are relatively low (6–9). To identify the high risk individuals and therefore contribute to the prevention of cancer, it is important to study host susceptibility and its interaction with environmental factors in non-smoking-related lung cancer.

Polymorphisms of the genes encoding phase I and phase II xenobiotic metabolizing enzymes have been shown to be associated with susceptibility to lung cancer in a number of epidemiological studies (10). However, most of these studies are limited by lack of adequate statistical power (11,12). To overcome this limitation, the International Collaborative Study on Genetic Susceptibility to Environmental Carcinogens (GSEC) was begun and is on-going to pool raw data of studies on metabolic genetic polymorphisms and cancer risk. The detailed data collecting and pooling procedures were described in a previous publication (13). To evaluate host susceptibility and possible interactions between genes and environmental factors in lung cancer and overcome the problem of the small sample size of individual studies, we conducted a pooled analysis among lifetime non-smokers based on the GSEC...
dataset, focusing on the genetic polymorphisms of cytochrome p450 subfamily I polypeptide 1 (CYP1A1) and glutathione S-transferase M1 (GSTM1). Although a number of epidemiological studies have addressed the association between genetic host factors and tobacco-related lung cancer, only a very few of them have examined the effect of host genetic susceptibility in non-smokers. It has been suggested that genetic host factors may play a more important role in cancer development upon low dose exposure to carcinogens, and non-smokers might represent such a population (14,15).

CYP1A1 plays an important role in the metabolism of polycyclic aromatic hydrocarbons (PAHs), an important group of lung carcinogens (16). Two genetically linked polymorphisms of CYP1A1, MspI (T<sup>2455</sup>C) and He<sup>626</sup>Val, have been associated with increased risk of lung cancer. A base substitution of thymine by cytosine in a non-coding region of the gene at position 3801 creates a MspI recognition site (CYP1A1*2A), which does not exist in the predominant genotype (T/T). A one base substitution A:T → G:C at position 2455 in the hemebinding region of exon 7 induces an amino acid change of isoleucine to valine at codon 462 (He<sup>626</sup>Val polymorphism) (16). This polymorphism is usually linked with the MspI polymorphism (CYP1A1*2B). The presence of the He<sup>626</sup>Val polymorphism alone (CYP1A1*2C) is very rare in Caucasians. Finally, another polymorphism in exon 7, i.e. a base substitution of cytosine by adenine at position 2453, leading to the Thr<sup>661</sup>Asn polymorphism (CYP1A1*4), has been described (17). Unfortunately, many older studies could not distinguish between the base changes at positions 2455 and 2453 and what has been described as ‘exon 7 polymorphism’ may refer to either. Contributions to clarify the situation, including a recommended nomenclature, have been published (18,19).

The above listed CYP1A1 polymorphisms have been shown to increase microsomal catalytic activity for converting procarcinogens, including PAH and aromatic amines (20). Positive associations of CYP1A1 genetic polymorphisms and lung cancer risk were pointed out in early Japanese studies (21,22). Studies had to contain information on the genetic polymorphism of never-smoking cases and never-smoking controls to be eligible for inclusion in our pooled analysis. Data from 14 studies were available, 13 of which have been previously published (see references in Table I). The GSEC data set includes data on genetic polymorphisms of the CYP1A1, GSTT1, GSTM1, NAT2, CYP2E1 and CYP2D6 genes. Among non-smokers, there were sufficient numbers of subjects to conduct a pooled analysis on GSTM1 and CYP1A1. Only three of the study results on the relationship of CYP1A1 and GSTM1 polymorphisms and lung cancer risk in non-smokers have been previously reported (28–30).

### Materials and methods

**Identification of the studies**

We obtained the original data of 14 studies on genetic polymorphism and risk of lung cancer included in the GSEC International Collaborative Study. Part of these data were included in previous analyses of lung cancer risk and CYP1A1 and GSTM1 polymorphism (27, L. Le Marchand et al., in preparation). To avoid the potential confounding of ethnicity, our analyses were restricted to the Caucasian component of the studies. The data collecting procedure of GSEC was described in a previous publication (13). The participation in GSEC was voluntary and therefore some relevant studies were not included in our analysis (e.g. studies on CYP1A1 from Japan; 21,22). Studies had to contain information on the genetic polymorphism of never-smoking cases and never-smoking controls to be eligible for inclusion in our pooled analysis. Data from 14 studies were available, 13 of which have been previously published (see references in Table I). The GSEC data set includes data on genetic polymorphisms of the CYP1A1, GSTT1, GSTM1, NAT2, CYP2E1 and CYP2D6 genes. Among non-smokers, there were sufficient numbers of subjects to conduct a pooled analysis on GSTM1 and CYP1A1. Only three of the study results on the relationship of CYP1A1 and GSTM1 polymorphisms and lung cancer risk in non-smokers have been previously reported (28–30).

<table>
<thead>
<tr>
<th>Study no.</th>
<th>Year of publication</th>
<th>Reference</th>
<th>Country</th>
<th>Cases (n)</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1991</td>
<td>(48,49)</td>
<td>Norway</td>
<td>28</td>
<td>182</td>
</tr>
<tr>
<td>2</td>
<td>1992</td>
<td>(50,51)</td>
<td>Finland</td>
<td>19</td>
<td>225</td>
</tr>
<tr>
<td>3</td>
<td>1993</td>
<td>(29,52)</td>
<td>Germany</td>
<td>8</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>1994</td>
<td>(53)</td>
<td>Sweden</td>
<td>9</td>
<td>148</td>
</tr>
<tr>
<td>5</td>
<td>1996</td>
<td>(17)</td>
<td>Germany</td>
<td>12</td>
<td>201</td>
</tr>
<tr>
<td>6</td>
<td>1996</td>
<td>(54,55)</td>
<td>US</td>
<td>7</td>
<td>164</td>
</tr>
<tr>
<td>7</td>
<td>1998</td>
<td>(45)</td>
<td>Hungary</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>1998</td>
<td>(50)</td>
<td>Slovakia</td>
<td>26</td>
<td>243</td>
</tr>
<tr>
<td>9</td>
<td>1999</td>
<td>(57)</td>
<td>US (Hawaii)</td>
<td>8</td>
<td>54</td>
</tr>
<tr>
<td>10</td>
<td>1998</td>
<td>(57,58)</td>
<td>US</td>
<td>6</td>
<td>60</td>
</tr>
<tr>
<td>11</td>
<td>1999</td>
<td>(59)</td>
<td>Poland</td>
<td>23</td>
<td>143</td>
</tr>
<tr>
<td>12</td>
<td>2000</td>
<td>(28)</td>
<td>Europe, Brazil</td>
<td>123</td>
<td>123</td>
</tr>
<tr>
<td>13</td>
<td>2000</td>
<td>(60)</td>
<td>US</td>
<td>16</td>
<td>30</td>
</tr>
<tr>
<td>14</td>
<td>Unpublished*</td>
<td></td>
<td>UK</td>
<td>4</td>
<td>27</td>
</tr>
</tbody>
</table>

*M.L. Clapper, personal communication.

Table I. Summary of the characteristics of the case–control studies included in the pooled analysis

---

GSTM1 detoxifies hydrophobic electrophiles derived from the metabolism of xenobiotics, including PAH-derived epoxides, by catalyzing their conjugation to glutathione (24). Although the expression of GSTM1 in lung is relatively low (25), loss of GSTM1 enzyme activity (GSTM1 null genotype) has been associated with host susceptibility in smoking-related lung cancer (26), however, its role in non-smoking-related lung cancer is still not established. Although the relative risk of lung cancer associated with the GSTM1 null genotype is unlikely to be greater than 1.5, it might be responsible for a substantial proportion of cancer cases because of the high frequency of deficiency which ranges from 40 to 60% in most ethnic groups.

Pooled analysis is a particularly useful approach to analyze subgroups of the study subjects that are scientifically relevant but were not evaluated in the individual studies because of small numbers. We therefore pooled data from 14 case–control studies on lung cancer with comparable information on genetic polymorphisms in non-smokers.

---

876
Among the 14 studies, comprising a total of 302 cases of lung cancer, four studies were from the USA, including one from Hawaii, nine studies were from Europe, including two from Germany and one each from Finland, Sweden, the UK, Norway, Hungary, Slovakia and Poland, and the last one was a multicentric analysis of patients from various European countries and Brazil. The time period of subject recruitment varied among studies, but was mainly between 1988 and 1997. All the subjects were Caucasians and had never smoked in their lives, although the precise definition of never-smoking status varied slightly among the studies. The cases were histologically confirmed lung cancer patients. Nine studies recruited controls from healthy populations, two from hospital patients and three from a combination of hospital patients and healthy subjects. The characteristics of the individual studies are summarized in Table I. The unpublished study recruited incidence cases from the tissue bank facility of University of Pittsburgh Cancer Institute and subjects were genotyped by a PCR-based method. Further information on the unpublished study can be requested from the authors.

Information on GSTM1 polymorphism was available from 13 studies, including 284 cases; results on CYP1A1 MspI and Ile62Val polymorphisms were available from eight (70 cases) and seven studies (82 cases), respectively. Only one study provided information on the CYP1A1 Thr461Asn polymorphism (17).

### Statistical analysis

GSTM1 polymorphism was dichotomized into the null genotype and the non-null type, whereas the CYP1A1 MspI and CYP1A1 Ile62Val polymorphisms were categorized into homozygous wild-type and variant allele-containing genotypes. We estimated the study-specific odds ratios (OR) of lung cancer for each polymorphism using unconditional logistic regression modeling. Such results might vary slightly from those reported for some of the studies because of differences in the inclusion criteria of cases and controls and in the statistical analyses. Although the tests of heterogeneity among study-specific ORs were not significant, we were still concerned that differences in study design among studies might contradict the assumption of homogeneity in the fixed-effects model. We therefore decided to use random-effects models (31). Random-effects models account for heterogeneity among studies by adding a term which represents unknown variance between study results.

Studies in which the OR could not be estimated (because one or more cells in the four-fold table had no subjects) were excluded from the pooled analysis. We used the STATA package for pooling study-specific ORs. We also stratified on gender to evaluate the potential modification by gender of the effect of genetic polymorphisms.

The inclusion of the studies in our analysis might suffer from participation bias. We therefore performed Egger’s test and a funnel plot to assess the possible inclusion bias of our pooled data set (32). The funnel plot includes the study-specific log effect estimates against their standard errors, based on the fact that the smaller studies (with larger standard error) should scatter more widely around the true effect due to less precision, whereas larger studies spread narrowly around the true effect, which results in a plot of inverted funnel shape if no bias exists. The asymmetry or skewness of the plot indicates the potential existence of bias in the pooled data. Egger’s test is a regression-based test for detecting skewness in the funnel plot.

We also conducted influence analyses by taking out outlying studies or studies with high variances. By comparing the full to the restricted data set, we evaluated the influence of a single study on the pooled estimates and inclusion bias. We also conducted the same analysis restricted to lung adenocarcinoma, which accounts for ~50% of the cases.

Since the CYP1A1 MspI and Ile62Val polymorphisms are genetically linked, we repeated our analysis after restricting the data set to the five studies (nos 4, 5, 9, 10 and 13) that provided genotyping information on both polymorphisms, to explore the effects of CYP1A1 MspI alone (CYP1A1-2A) and the concurrent presence of the Ile62Val and MspI polymorphism variants. There was an insufficient number of subjects for an analysis of the effect of the Ile62Val variant (Ile/Val, Val/Val) alone.

Besides the main effect of the CYP1A1 and GSTM1 polymorphisms on lung cancer, we were also interested in the possible combined effect of the CYP1A1 variants and GSTM1 null genotype. Wild-type CYP1A1 and the non-null genotype of GSTM1 was used as reference group to assess the combined effects of the two genes. We further evaluated the possible interaction between genetic polymorphisms and exposure to environmental tobacco smoke (ETS) and urban/rural residence, used as a proxy for exposure to outdoor air pollution. The information on ETS was available in three studies (nos 7, 9 and 12) and information on residence was available in six studies (nos 5, 7–9, 11 and 12). Interaction analyses focused on GSTM1, which has a higher statistical power for estimating effect modification. For ETS, the group of subjects with non-null genotype and no current ETS exposure was used as the reference group; for residence, the group of subjects with non-null genotype and living in a rural area was used as the reference group. All analyses concerning ETS and residence were adjusted for age, gender and study center.

### Results

Among the controls enrolled in the studies, the frequencies of CYP1A1 MspI polymorphism variants (genotype T/C or C/C) ranged from 15 to 33% and the frequencies of CYP1A1 Ile62Val variants (Ile/Val or Val/Val) ranged from 4 to 18% among all studies. For GSTM1, the frequencies of the null genotype ranged from 40 to 73% (Table I).

Table II displays the study-specific OR estimates computed directly from the raw counts underlying Table I. The OR for CYP1A1 MspI polymorphism variants varied from 0.3 to 5.15. The P value of the test for heterogeneity among them was 0.6. The result of study no. 5 (17) was different from the others, and after excluding this study the P value of the heterogeneity test increased to 0.9. The study-specific estimates of CYP1A1 Ile62Val variants ranged from 0.90 to 10.9. The P value of the test for heterogeneity was 0.3; after excluding study no. 5 the P value of the heterogeneity test increased to 0.7. For the GSTM1 null genotype, the 95% confidence interval (95% CI) of the study-specific estimates were distributed around the null, with about half of the studies showing risk

<table>
<thead>
<tr>
<th>Study no.</th>
<th>CYP1A1 MspI (T/T versus T/C, C/C)</th>
<th>CYP1A1 Ile62Val (Ile/Ile versus Ile/Val, Val/Val)</th>
<th>GSTM1 (Present versus null)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95%CI)</td>
<td>OR (95%CI)</td>
<td>OR (95%CI)</td>
</tr>
<tr>
<td>1</td>
<td>0.30 (0.01, 6.38)</td>
<td>2.04 (0.36, 11.6)</td>
<td>1.01 (0.40, 2.56)</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>0.68 (0.56, 1.79)</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>0.75 (0.11, 5.11)</td>
</tr>
<tr>
<td>4</td>
<td>2.05 (0.48, 8.70)</td>
<td>1.73 (0.20, 15.2)</td>
<td>1.01 (0.26, 3.90)</td>
</tr>
<tr>
<td>5</td>
<td>5.15 (1.54, 17.2)</td>
<td>10.9 (2.64, 45.2)</td>
<td>1.28 (0.39, 4.18)</td>
</tr>
<tr>
<td>6</td>
<td>1.50 (0.19, 11.9)</td>
<td>2.01 (0.38, 10.6)</td>
<td>0.84 (0.13, 5.26)</td>
</tr>
<tr>
<td>7</td>
<td>0.96 (0.11, 9.06)</td>
<td>2.08 (0.20, 21.8)</td>
<td>1.25 (0.55, 2.82)</td>
</tr>
<tr>
<td>8</td>
<td>3.53 (0.54, 17.2)</td>
<td>0.90 (0.10, 8.57)</td>
<td>0.44 (0.09, 2.18)</td>
</tr>
<tr>
<td>9</td>
<td>1.71 (0.57, 5.12)</td>
<td>1.77 (0.72, 4.34)</td>
<td>1.51 (0.91, 2.51)</td>
</tr>
<tr>
<td>10</td>
<td>0.60 (0.04, 9.16)</td>
<td>7.29 (1.52, 35.0)</td>
<td>0.47 (0.13, 1.68)</td>
</tr>
<tr>
<td>11</td>
<td>1.30 (0.10, 17.7)</td>
<td>0.80 (0.10, 6.55)</td>
<td>0.47 (0.13, 1.68)</td>
</tr>
</tbody>
</table>

Downloaded from https://academic.oup.com/carcin/article-abstract/24/5/875/2390467 by guest on 29 April 2018
effects and half showing protective effects. There was little evidence of heterogeneity since the $P$ value of heterogeneity was 0.8.

Figure 1 shows the funnel plot of the results of the pooled analysis of the CYP1A1 MspI polymorphism. There was evidence of participation bias ($P$ value of Egger’s test 0.01), which was reduced after exclusion of the results of study no. 5 ($P$ value of Egger’s test 0.09). Results for the CYP1A1 Ile<sup>462</sup>Val polymorphism are plotted in Figure 2. There was no evidence of inclusion bias ($P$ value of Egger’s test 0.7).

The plot of the results on GSTM1 null genotype against their standard error is shown in Figure 3. There was evidence of inclusion bias ($P$ value of Egger’s test 0.02), with more imprecise studies having results mainly around or below the null value. The exclusion of the five studies with largest variance (studies nos 3, 6, 7, 9 and 14) increased the $P$ value of Egger’s test to 0.06.

Table III shows the pooled OR estimates and 95%CI for CYP1A1 MspI, the CYP1A1 Ile<sup>462</sup>Val variant and the GSTM1 null genotype based on random-effects models. Both the results based on full and restricted data sets are displayed. The OR of the CYP1A1 MspI polymorphism variant was >2, but it decreased and became insignificant after we excluded the study with outlying results. Stratifying on gender, the OR

![Fig. 1. Funnel plot of CYP1A1 MspI polymorphism (log ORs of lung cancer and their standard errors).](image1)

![Fig. 2. Funnel plot of CYP1A1 Ile<sup>462</sup>Val polymorphism (log ORs of lung cancer and their standard errors).](image2)

![Fig. 3. Funnel plot of GSTM1 null genotype (log ORs of lung cancer and their standard errors).](image3)

<p>| Table III. Pooled odds ratios and confidence intervals for CYP1A1 and GSTM1 polymorphisms |
|-------------------------------------------------|-------------------------------------------------|</p>
<table>
<thead>
<tr>
<th>Case ($n$)</th>
<th>Control ($n$)</th>
<th>Full dataset</th>
<th>Restricted dataset&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>All lung cancer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP1A1 MspI</td>
<td>Wild-type (T/T)</td>
<td>49</td>
<td>364</td>
</tr>
<tr>
<td></td>
<td>T/C, C/C</td>
<td>21</td>
<td>89</td>
</tr>
<tr>
<td>CYP1A1 Ile&lt;sup&gt;462&lt;/sup&gt;Val</td>
<td>Wild-type (Ile/Ile)</td>
<td>61</td>
<td>591</td>
</tr>
<tr>
<td></td>
<td>Ile/Val, Val/Val</td>
<td>21</td>
<td>61</td>
</tr>
<tr>
<td>GSTM1</td>
<td>Present</td>
<td>132</td>
<td>706</td>
</tr>
<tr>
<td></td>
<td>Null</td>
<td>152</td>
<td>727</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP1A1 MspI</td>
<td>Wild-type (T/T)</td>
<td>24</td>
<td>305</td>
</tr>
<tr>
<td></td>
<td>T/C, C/C</td>
<td>11</td>
<td>76</td>
</tr>
<tr>
<td>CYP1A1 Ile&lt;sup&gt;462&lt;/sup&gt;Val</td>
<td>Wild-type (Ile/Ile)</td>
<td>20</td>
<td>346</td>
</tr>
<tr>
<td></td>
<td>Ile/Val, Val/Val</td>
<td>11</td>
<td>40</td>
</tr>
<tr>
<td>GSTM1</td>
<td>Present</td>
<td>70</td>
<td>590</td>
</tr>
<tr>
<td></td>
<td>Null</td>
<td>68</td>
<td>600</td>
</tr>
</tbody>
</table>

The number of cases and controls might not correspond to those shown in Table I because of missing data on polymorphisms for some individuals.

<sup>a</sup>CYP1A1 MspI, excludes study no. 5; CYP1A1 Ile<sup>462</sup>Val, excludes study no. 5; GSTM1, excludes studies nos 3, 6, 7, 9 and 14.

<sup>b</sup>Adjusted for study.
was 2.48 (95% CI 0.73–8.40) for men and 2.64 (95% CI 0.70–10.0) for women (P value of heterogeneity 0.9).

The OR for the CYP1A1 Ille<sup>462</sup>Val variants (Ille/Val or Val/Val) versus the Ille/Ille genotype increased three times in the full data set (OR 2.99, 95% CI 1.51–5.91). After excluding the outlying study, the CYP1A1 Ille<sup>462</sup>Val variant still showed a very strong association with lung cancer risk (OR 2.21, 95% CI 1.12–4.37). The gender-specific OR was 3.52 (95% CI 1.35–9.17) for men and was 2.57 (95% CI 0.58–11.5) for women (P value of heterogeneity 0.7). The OR of the GSTM1 null genotype was only marginally increased, and was not substantially modified when we excluded imprecise results which might have contributed to inclusion bias. After stratifying on gender, the OR for men was 0.94 (95% CI 0.53–1.67) and for women 1.36 (95% CI 0.93–2.00) (P value of heterogeneity 0.3). For lung adenocarcinoma, comprising about half of the total cases, we did not observe a major difference in the effects of the CYP1A1 MspI polymorphism and GSTM1 null genotype compared with the results based on all lung cancer cases. However, an even stronger effect was suggested for the CYP1A1 Ille<sup>462</sup>Val variants.

The results of the analyses restricted to studies providing data on both the CYP1A1 MspI and Ille<sup>462</sup>Val polymorphisms are displayed in Table IV. Compared with subjects having neither the MspI polymorphism nor a Ille<sup>462</sup>Val variant, having both MspI and Ille<sup>462</sup>Val variants is associated with an increased risk of lung cancer (OR 5.30, 95% CI 1.94–14.50).

Table V shows the results of the analysis of the combined effect of both the GSTM1 null genotype and CYP1A1 polymorphisms. Compared with the subjects with both wild-type genes, individuals with both a CYP1A1 Ille<sup>462</sup>Val variant and the GSTM1 null genotype were at an increased risk of lung cancer (P value of interaction 0.01). A similar effect, although less strong, was suggested for subjects with a CYP1A1 MspI polymorphism and the GSTM1 null genotype (P value of interaction 0.1).

The analyses of effect modification by GSTM1 status on ETS and urban residence did not provide evidence of an interaction between the polymorphism and either exposure. When unexposed individuals with the non-null genotype were chosen as the reference group, the OR in the group of exposed individuals with the null genotype was 1.23 (95% CI 0.51–2.98) for ETS and 1.16 (95% CI 0.61–2.21) for urban residence. Similar results were obtained when the analysis was restricted to adenocarcinoma cases.

Discussion

In this analysis we evaluated the effect of genetic polymorphisms on lung cancer risk among lifetime non-smokers, an interesting population for studies on genetic susceptibility, since they are not exposed to a strong carcinogen such as tobacco smoke. We found that the CYP1A1 Ille<sup>462</sup>Val variant was associated with an increased risk of lung cancer. This effect was particularly strong for adenocarcinoma and for individuals with the GSTM1 null genotype. The strength of the association was not reduced after resolving the potential inclusion of heterogeneous studies. We also observed a slight and not statistically significant increase in risk with either the CYP1A1 MspI polymorphism, the GSTM1 null genotype or their combination. The analyses restricted to the studies with both information on CYP1A1 MspI and Ille<sup>462</sup>Val polymorphisms suggested that the MspI polymorphism alone does not seem to have an effect on the risk of lung cancer, however, a strong effect was observed when the MspI and Ille<sup>462</sup>Val variants are present concurrently.

The preference for a pooled analysis over a meta-analysis is due to the advantage of using individual-based rather than group-based data. Furthermore, although published studies did include non-smokers, only a few of them reported separate results for this subgroup: a meta-analysis of results on non-smokers would therefore not be very informative. Another major strength of a pooled analysis is the ability to conduct analyses on potential effect modification, which requires large numbers and might be difficult to achieve in individual studies.

Some methodological limitations of our pooled analysis need to be considered. First, the result can be biased if the
studies included represent a biased sample of studies in general. Our results revealed evidence of inclusion bias in the case of the CYP1A1 MspI polymorphism and GSTM1 null genotype. However, we also demonstrated the possibility of reducing this bias by removing studies in influence analyses. The validity of the estimates was therefore increased, however, we faced losing precision.

The validity of the pooled estimate can also be threatened by the heterogeneity of the studies. This heterogeneity can result from differences in study populations, study design and often methodological aspects. Since the studies included in the pooled analysis were conducted in different populations and did not follow a standard protocol, heterogeneity may result from effect modification by other genes/exposures with different distributions across studies and variable misclassifications of genotypes. The most important difference among studies is the control recruitment. Studies with hospital controls might result in the OR of the CYP1A1 Thr461Asn polymorphism from hospital and healthy controls.

A further limitation might be population mixing, i.e. inclusion of individuals from populations with different genetic backgrounds. It has been shown that population mixing is a form of confounding by ethnicity, which can be controlled by adjusting for ethnic background (33). Although we have restricted our analyses to Caucasian populations and adjusted for study, the adjustment might still be incomplete since some studies were conducted in ethnically heterogeneous populations (e.g. the USA) or in multiple centers. However, even in the case of incomplete adjustment, the bias resulting from population mixing is likely to be of small magnitude when data are collected within a broad ethnic group (33).

Misclassification of exposure can also be a source of bias. Misclassification of non-smoking status, if differential between cases and controls, might bias the point estimates either towards or away from the null. The extent of such misclassification has been estimated as 1–2% in a previous validation study (34): if this result is applicable to the studies included in our pooled analysis, the validity of our result would only be marginally affected. Another possible misclassification involves genotyping. The genotyping in the 14 studies was mainly done by PCR, which may be subject to error. In the case of the GSTM1 null genotype, false null genotyping can occur when the PCR conditions are not optimized and false positive results can result from contamination. However, it is likely that such misclassification is non-differential with respect to case–control status, resulting in the bias towards the null.

Another more severe methodological limitation is the potential misclassification in genotyping of CYP1A1 Ile62Val polymorphisms. As we mentioned earlier, some of the older studies could not distinguish CYP1A1 Ile62Val polymorphisms from CYP1A1 Thr61Asn polymorphisms, which was discovered later on and is only 2 bp away. However, this methodological limitation of genotyping may result in misclassification, which is not differential between cases and controls and is not likely to generate a positive association like the one we observe. This problem, however, limits our ability to attribute an increased susceptibility to lung cancer to either polymorphism.

The polymorphisms of CYP1A1 have been shown to be associated with increased aryl hydrocarbon hydroxylase enzyme activity and gene induction (21,35–38), although some discrepancy remains in the literature (39–41). Our results show that the CYP1A1 Ile62Val polymorphism variant (Ile/Val or Val/Val) may be associated with increased risk of lung cancer in non-smokers. A similar result was reported in a previous pooled analysis of part of these data, which however also included smokers and non-Caucasians (R.Le Marchand et al., in preparation). The strong positive association of CYP1A1 Ile62Val polymorphisms and lung cancer risk as we observed disagrees with the recent meta-analysis by Houlston (23), which, however, did not separate smoking lung cancer cases from non-smoking lung cancer cases. Any analysis of lung cancer without separation of smokers and non-smokers would be strongly influenced by the results among smokers, who account for at least 90% of the cases. Therefore, the difference between our analyses and relevant meta-analyses might reflect a true difference in the effect of genetic polymorphisms in non-smokers and smokers.

We found an association between CYP1A1 polymorphism and lung cancer, specifically an effect of the MspI polymorphism (I/C or C/C), which, however, seemed to be due to the results of a single study (17). Furthermore, since the MspI recognition site locates in the 3' non-coding region, an association of the CYP1A1 MspI polymorphism might be due in part to linkage disequilibrium with the CYP1A1 Ile62Val polymorphism. The results of analyses on the MspI polymorphism alone (CYP1A1*2A allele) and the concurrent presence of the MspI and Ile62Val polymorphisms are consistent with this hypothesis. We should, however, stress that the concurrent presence of the MspI and Ile62Val polymorphisms is not equal to the linkage of the MspI and Ile62Val polymorphisms (CYP1A1*2B allele). The effect of the 2B allele per se cannot be analyzed epidemiologically, because the allele combination is unknown for subjects with both heterozygous MspI and Ile62Val variants. We can rule out a major influence of the MspI polymorphism in lung cancer among non-smokers.

The GSTM1 null genotype expresses no enzyme activity in detoxifying tobacco-related procarcinogens (24). We found only a very small and not significant increase in risk of lung cancer among subjects with the GSTM1 null genotype. A similar conclusion arose from the results of a previous pooled analysis of some of these data, which also included smokers and non-Caucasians (27). One of the plausible reasons may be that other GSTs or other detoxifying enzymes, which we did not take into account, might play an important role when GSTM1 activity is absent.

A possible combined effect between the genetic polymorphisms of GSTM1 and CYP1A1 was suggested before (11). However, there is no consistent evidence supporting such an interaction so far. Epidemiological studies showed an increased risk of lung cancer with the combination of the GSTM1 null genotype and CYP1A1 variants in Asian populations (42,43). A further study suggested an interaction between CYP1A1 hyper-inducibility and the GSTM1 null genotype in a French population (44). However, a study of DNA adducts and the CYP1A1 MspI polymorphism and GSTM1 genotype in Hungarian lung cancer patients did not show a combined effect of these two genetic polymorphisms (45). Our analysis shows a strong gene–gene interaction between the CYP1A1 Ile62Val variant and GSTM1 null genotype. CYP1A1 takes part in the activation of PAHs into diol epoxide metabolites in the lung.
and GSTM1 plays an important role in the detoxification of diol epoxide metabolites. The balance between activation by CYP1A1 and detoxification by GSTM1 and other enzymes therefore may be a step in the initiation of lung cancer (46). The strong interaction can be reconciled with the very small effect suggested for the GSTM1 null genotype alone, because of the rarity of CYP1A1 variants.

Our results did not suggest a major effect modification by the GSTM1 null genotype on the carcinogenic effect of either ETS or urban residence among non-smokers. They agreed with a study that showed no effect of the GSTM1 null genotype on DNA or protein adduct levels in non-smoking healthy residents of rural and urban areas (47). However, the statistical power of our gene–environmental interaction analysis was relatively low.

Age and gender were not adjusted for in the estimation of the main effect of polymorphisms since genetic polymorphisms should not be affected by age and gender, and there is no difference in results after adjusting for age and gender (e.g. GSTM1 pooled OR 1.04, 95% CI 0.75–1.45). The stratified analysis on gender showed slightly different effects of genetic polymorphisms between males and females, however, the difference was not significant.

In conclusion, our results suggest a possible role of CYP1A1 Ile62Val polymorphisms in lung carcinogenesis among non-smokers, particularly in combination with the GSTM1 null genotype. The results for the CYP1A1MspI polymorphism variants and GSTM1 null genotype excluded an important effect of each of these polymorphisms alone on lung cancer risk. As we know that carcinogenesis is a multi-factorial process, it is unlikely that any single genetic polymorphism accounts completely for the individual variability in host susceptibility of lung cancer. There was a combined effect of CYP1A1 Ile62Val polymorphisms and the GSTM1 null genotype observed in our analyses, in agreement with a biological hypothesis of a possible gene–gene interaction. Our results stress the need for larger studies addressing the interaction between polymorphisms in different genes.

Acknowledgement
RayJean Hung worked on this study under the tenure of a Special Training Award from the International Agency for Research on Cancer.

References
Control Clin. Trials, 7, 177–188.

Received October 7, 2002; revised February 5, 2003; accepted February 11, 2003