GSTM1 and *CYP1A1* Polymorphisms, Tobacco, Air Pollution, and Lung Cancer: A Study in Rural Thailand

Paola Pisani,[1] Petcharin Srivatanakul,[2] Juliette Randerson-Moor,[3] Sutee Vipasrinimit,[4] Somkiat Lalitwongsa,[4] Piyapong Unpunyo,[4] Saghir Bashir,[1] and D. Timothy Bishop[3]

[1]Descriptive Epidemiology Group, IARC, Lyons, France; [2]National Cancer Institute, Bangkok, Thailand; [3]Division of Genetic Epidemiology, Cancer Research UK Clinical Centre, St. James's Hospital, Leeds, United Kingdom; and [4]Lampang Provincial Hospital, Lampang, Thailand

Abstract

Objectives: **The Lampang Province is situated in the northern region of Thailand. Incidence rates of lung cancer are high for Asian standards, particularly in women. This study was conducted to quantify the risk of lung cancer associated with exposures prevalent in the area and to investigate possible interactions with genetic susceptibility. The presence of several large open-cast coal mines from 1955 close to electricity-generating plants was a particular focus of concern.**

Methods: **Two-hundred and eleven cases of primary lung cancers diagnosed in 1993 to 1995 and residents in the province were recruited at the Lampang Provincial Hospital (main referral center for treatment of the disease). Two sets of controls, frequency-matched to the cases by sex and age, were recruited (*a*) from the resident population (202 interviewed) and (*b*) from patients admitted to the hospital for diseases predominantly unrelated to tobacco smoking (211 interviewed). Sociodemographic information, complete residential history, and characteristics of the household (place of cooking, cooking fuel, and heating fuels), occupational history, and history of tobacco smoking were obtained by interview. Cases and controls (~50% of the population-based series) provided a blood sample. A point source air pollution exposure index was calculated for each village/township reported in residential histories based on the linear distance from the Mae Moh Center (the area of the electricity-generating plants), the year-specific gaseous (SO_2 and NO_2) or total suspended particulate emissions from the Mae Moh Power Plant, and the percentage of wind from the center. Odds ratios (OR) for the disease associated with categorical variables were estimated within unconditional logistic regression. Extraction of genomic DNA and genotyping of variants in *CYP1A1* and *GSTM1* were conducted to assess the extent of modification of risk by these genes that are involved in the metabolism of polycyclic aromatic hydrocarbons, a common component of the exposures.**

Results: **Overall, there was no evidence of relevant differences in the socioeconomic level of the three groups. The two control sets were similar with respect to lifelong tobacco habit and were subsequently pooled in analyses. Never-smokers were 7% of men and 33% of women. Smoking of local traditional products unfiltered and high in tar content is a common habit in the rural female population. ORs associated with smoking increased with duration of the habit and average daily amount, being 4.9 [95% confidence interval (95% CI), 2.5-9.7] for smokers of ≥7 cigarettes/d and 3.3 (95% CI, 1.7-6.2) for duration of 41 years or longer compared with nonsmokers. Smoking of local products was associated with an independent OR of 3.1 (95% CI, 1.7-5.6) adjusted for lifelong cumulative amount of tobacco smoked. Although most smokers had the habit for at least 16 years, the daily consumption was low compared with Western standards. Other potential sources of exposure to lung carcinogens (emission from the power-generating plants and domestic burning of coal and wood for cooking and heating) were not associated with increased risk of lung cancer. None of the three polymorphisms examined increased the risk of lung cancer or modified the risk associated with tobacco smoking.**

Conclusion: **In this rural population, 96% of male and 64% of female lung cancer incidence were explained by tobacco smoking. None of the potential sources of air pollution deriving from the combustion of coal and wood, or polymorphisms in the *CYP1A1* gene or deletion of the *GSTM1* had an effect on the risk of lung cancer, either together or separately.** (Cancer Epidemiol Biomarkers Prev 2006;15(4):667–74)

Introduction

Overwhelming evidence of the role of tobacco smoking in the occurrence of lung cancer has been accumulated since the 1960s (1). Quantitatively, it accounts for 37% of all cancer death in men and 9% in women in high-risk countries (2). Although declines in the prevalence of smoking are observed among men living in Western countries, rapid increases are reported in many developing countries where it remains an uncommon habit only in women (1). Further, many carcinogens of the respiratory tract have been identified as well as environmental sources of exposure to these agents.[5] These account for a relatively small proportion of the burden relative to tobacco in regions with widespread smoking but may have a significant effect in populations where tobacco smoking is a recent or rare habit. A high proportion of lung cancer is attributable to known environmental causes, raising the question of putative modulation of risk by genetic variants in metabolic genes (3, 4). Several polymorphisms in genes implicated in the metabolism of lung carcinogens have been identified and variation in their prevalence has been shown across ethnic groups (5).

The Lampang Province is situated in the northern region of Thailand. It contains several large open-cast coal mines opened from 1955 onward in proximity to functioning electricity-generating plants. The first started production in 1978 and

Received 8/30/05; revised 1/14/06; accepted 1/31/06.

Grant support: Electricity Generating Authority of Thailand and Cancer Research UK.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: S. Bashir is currently in SBTC International, Brussels, Belgium.

Requests for reprints: Paola Pisani, Descriptive Epidemiology Group, IARC, 150 Cours Albert-Thomas, Lyons, France. Phone: 33-47-2738522; Fax: 33-47-2738650. E-mail: pisani@iarc.fr

doi:10.1158/1055-9965.EPI-05-0667

[5] www-cie.iarc.fr.

there are currently 11 operating. The coal produced is soft coal/lignite, which is transported directly to the furnaces of the generating stations. The smoke emission from these plants is filtered to reduce particulate (this is reported as 99.5% efficient) and 30 tons/d are extracted. Despite the fact that pollution levels are within acceptable limits, there has been concern about the possible health effects to the population living in the vicinity of the plants. When the studies were initiated, a survey was conducted to measure cancer incidence in the years 1988 to 1992. Lung cancer was the most common cancer in men and second, after cancer of the cervix, in women, corresponding to age-standardized annual incidence rates per 100,000 population of 41.8 (men) and 20.1 (women). Sixty percent of the cases were histologically verified. The rates were comparable with those in the subsequent period of 1993 to 1997 (56.5 in men and 29.1 in women). Incidence rates in Thai provinces served by cancer registries around 1995 were as follows: 34.9 (men) and 25.9 (women) in Northwest Chang Mai, 18.5 (men) and 6.6 (women) in the rural province of Khon Kaen Northeast of Lampang, and 15.1 (men) and 4.7 (women) in the southern province of Songkhla (6). For comparison, the age-standardized rates in England around the same period were 51.1 (men) and 21.9 (women; ref. 6).

This study was conducted to quantify the risk of lung cancer attributable to identified causes of the disease prevalent in the area and to investigate putative interactions with genetic susceptibility. We investigated the role of two polymorphisms in the *CYP1A1* gene plus one in the *GSTM1* gene. Both genes are involved in the metabolism of polycyclic aromatic hydrocarbons, the main component of the pollutants and exposures suspected in this study.

Materials and Methods

Cases were primary lung cancers recruited at the Lampang Provincial Hospital (the main referral center for the treatment of the disease), diagnosed between 1993 and 1995, and residents in the province. Two sets of controls, frequency-matched to the cases by sex and age, were recruited from (*a*) the resident population (District Vital Statistics Office) and (*b*) patients admitted to the hospital for diseases predominantly unrelated to tobacco smoking. The second set was matched to the cases' residence (Lampang City or the countryside). Ethical clearance was granted by the Thai National Cancer Institute and the Provincial Health Authority. Under the approval, consent was to be obtained from each participant but it was not required to be written (one third of the population being illiterate; see Results). A trained nurse explained to each potential participant the aim and conduct of the study.

Suspected cases of lung cancer were interviewed before diagnosis. Two hundred and sixty-one patients were enrolled, of whom 211 (81%) were later confirmed and included in the study. A special effort was made for the duration of the study to confirm by histology or cytology the diagnosis so that the majority of the cases were microscopically verified. Forty-five percent of the tumors were squamous cell carcinomas, 21% adenocarcinomas, 11% large-cell carcinomas, and 8% small-cell carcinomas; 11% were mixed types and 3% remained unknown. We observed a similar distribution across histologies by gender. Similarly to the cases, hospital controls were approached at admission. Two hundred and seventy-four patients were invited to participate. Fifteen (5.5%) refused to participate and 48 (17.5%) were excluded because of the presence of disorders linked to tobacco smoking; 211 were interviewed. The causes of hospital admission of controls included diseases of the urinary system (27%), followed by cardiovascular diseases and diseases of the digestive system excluding the oral cavity (15% each), disorders of the eye and adnexa (8.5%), infectious diseases (6.6%), and other miscellaneous causes. One patient with a brain tumor and two with lymphoma were the only neoplastic cases included. For population controls, 250 individuals were identified by stratified random sampling (by sex and age) from the lists of the local residents and were contacted at home. Forty-eight refused to participate and 202 (81%) were interviewed.

Cases and controls were interviewed by trained nurses. The questionnaire included the following sections: sociodemographic variables (education, marital status, language, and religion); complete residential history by period, name of village/township, and characteristics of the household (place of cooking, cooking fuel, heating fuels, water source, and toilet facility); occupational history, job titles from a list of 48 major groups consistent with the International Labour Office classification 1990 (7); history of tobacco smoking, for the duration of defined time period, type of cigarettes, and amount of cigarettes; alcohol drinking (several periods characterized by type of beverage and amount); and chewing of betel nut and other products.

Cases and controls were asked to provide 10 mL venous blood. This was separated in four aliquots of 1 mL serum and two aliquots of buffy coats and stored at −20°C.

Compliance with blood donation was 168 cases (79.6%), 87 population controls (44.2%), and 200 hospital controls (94.8%).

DNA Extraction. DNA extraction from the buffy coats was carried out using the Nucleon BACC2 kit (Tepnel Life Sciences, Manchester, United Kingdom) according to the supplied protocol.

***GSTM1* Deletion.** A 230 bp fragment of the *GSTM1* gene was coamplified with a 157 bp fragment of the *GSTM4* gene using a three-primer system in which the forward primer was common to both *GSTM1* and *GSTM4*, whereas two reverse primers were specific to *GSTM1* and *GSTM4*, respectively. This assay could not distinguish the genotype heterozygous for the gene deletion from the wild-type genotype. One microliter of extracted DNA (~25-50 ng) was amplified using 0.625 unit AmpliTaq DNA polymerase in the supplied buffer containing 1.5 mmol/L $MgCl_2$, 200 μmol/L deoxynucleotide triphosphates, 100 ng common primer P1 (5′-CGCCATCTTGTGCTACATGTCCCG), and 50 ng of each GSTM-specific primer (*GSTM1*, P2, 5′-ATCTTCTCCTCTTCTGTCTC; *GSTM4*, P3, 5′-TTCTGGATTG-TAGCAGATCA). PCRs were done on an ABI9600 Thermal cycler at an annealing temperature of 52°C for 30 cycles and PCR products were visualized on a 1% agarose gel.

***CYP1A1* Genotyping.** Two allelic variants of the *CYP1A1* gene have been genotyped in this study.

The first, *CYP1A1**2A (which is also known as *CYP1A1 Msp*I or the *m1* allele), involves a T to C substitution (rs4646903) of the 3′ untranslated region creating an *Msp*I restriction site. A 335 bp fragment of the 3′ untranslated region was amplified and digested with *Msp*I. Wild-type DNA is not cut by the enzyme and retains the 335 bp fragment, whereas DNA carrying the variant is cut into two fragments of 206 and 129 bp. One microliter of extracted DNA (~25-50 ng) was amplified using 0.625 unit AmpliTaq DNA polymerase in the supplied buffer containing 1.5 mmol/L $MgCl_2$, 200 μmol/L deoxynucleotide triphosphates, and 50 ng forward and reverse primers (*CYP1A1 Msp*I-F, 5′-AAGAGGTGTAGCCGCTGCACT; *CYP1A1 Msp*I-R, 5′-TAGGAGTCTTGTCTCAGTCCT) in a 50 μL reaction. PCRs were done on an ABI9600 Thermal cycler at an annealing temperature of 59°C for 34 cycles and the PCR products digested with 5 units of *Msp*I in the supplied restriction buffer (New England Biolabs, Hitchin, United Kingdom). DNA fragments were then visualized on an 8% Tris-borate EDTA acrylamide gel.

The second allele, *CYP1A1**2C (also known as *CYP1A1*-E7 or the *m2* allele), is an A to G substitution (rs1048943) resulting in a missense mutation, $Ile^{462}Val$, in exon 7, which results in the loss of an *Nco*I restriction site. A 195 bp fragment of exon 7 was amplified and digested with *Nco*I. Wild-type DNA is cut

by the enzyme resulting in fragments of 163 and 32 bp. DNA carrying the variant is not cut, resulting in the presence of the 195 bp amplified fragment. One microliter of extracted DNA (~25-50 ng) is amplified using 0.625 unit AmpliTaq DNA polymerase in the supplied buffer containing 1.5 mmol/L $MgCl_2$, 200 μmol/L deoxynucleotide triphosphates, and 50 ng of forward and reverse primers (*CYP1A1*E7-F, 5′-GAACTGC-CACTTCAGCTGTCT; *CYP1A1*E7-R, 5′-CCAGGAAGAGAAA-GACCTCCCAGCGGGCCA) in a 50 μL reaction. PCRs were done on an ABI9600 Thermal cycler at an annealing temperature of 59°C for 34 cycles and the PCR products were digested with 5 units *Nco*I in the supplied restriction buffer (New England Biolabs). DNA fragments were then visualized on an 8% Tris-borate EDTA acrylamide gel.

Index of Exposure to Emissions from the Electricity-Generating Plants. The Ecology and Environment Division, Survey and Ecology Department, Electricity Generating Authority of Thailand provided historical information on wind speed and direction and measurements of air pollution. A point source air pollution exposure index was calculated for each village/township reported in residential histories based on three variables: linear (D) distance from the Mae Moh Lignite Center (the area of coal exploitation and electricity production); the year-specific gaseous (SO_2 and NO_2) or total suspended particulate emissions from the Mae Moh Power Plant (E_{year}) for years 1978 (first year of activity of the plant) to 1994; and the percentage of wind coming from the township direction (T_w) based on its angular distance from the Mae Moh Center. The point source air pollution index (P_e) was computed according to the following formula, which was derived from Brown et al. (8) and Buffler et al. (9):

$$P_e = \Sigma_{year} E_{year} \times 1/D^{1/2} \times T_w$$

Places of residence in surrounding provinces were estimated to be at the highest concentration of towns within the boundaries of the map that covered the whole of Lampang Province but only part of the neighboring provinces. Emission data were only available for years 1985 to 1994. Data for SO_2, NO_2, and total suspended particulate emission were extrapolated to years 1978 to 1984. Average percentage of wind in 16 directions was calculated using 1991, 1993, and 1994 wind rose diagrams. The percentages were then multiplied by each of the four wind speed measurements in each of the four main directions for each of the 3 years and then averaged. A cumulative pollution exposure index was calculated for each subject, summing point exposures across all places of residence reported in the residential history from 1978 using the above formula. For periods during which subjects lived outside Lampang province and outside neighboring districts within the relevant time frame of exposure, emissions were set to zero.

Index of Exposure to Domestic Fumes. Exposure to domestic fumes was estimated based on the type of fuel used to heat the residence and to cook (defined as coal or wood versus no heating or gas/electrical supplies), taking into account whether cooking was inside or outside, and the number of years living in this condition. These variables were combined in an exposure index calculated as the total time (years) spent using polluting fuel (coal or wood). Years of exposure due to cooking by coal or wood were halved if cooking was done outdoors.

Smoking History. The smoking histories covered different periods if changes in the number of cigarettes smoked per day or type of tobacco products occurred. Cigarette types were classified as filtered or unfiltered commercial products and local traditional hand-made khii yo and yamuan, both unfiltered. Traditional cigarettes are composed of dry uncured tobacco rolled in a dry banana leaf and seasoned with tamarind bark.

The lifelong cumulative amount of tobacco smoked was calculated as follows:

$$\Sigma_j \Sigma_i (\text{daily number})_{if} \times (\text{days})_{ij}$$

where i is the type of product and j is the period.

To allow for their bigger size, daily amounts of khii yoo were multiplied by 1.5.

Statistical Analyses. Odds ratios (OR) for the disease associated with categorical variables were estimated with unconditional logistic regression. Continuous variables were categorized by percentiles of the distribution of population controls. The overall population-attributable risk was calculated as the weighted sum of age- and sex-specific attributable risks (10) weighted for the proportion of cases in the sex-age stratum according to the incidence recorded by the local cancer registry (6). Weights were the sex-specific proportions of cases in each age group. The observed genotype distributions for the *CYP1A1* variants were tested for consistency with Hardy-Weinberg equilibrium among controls (only two phenotypes were available for *GSTM1*, making such a test uninformative).

Stata version 8 (Stata Corporation, College Station, TX) was used for all statistical analyses using standard methodology. Haplotype frequencies were estimated using the hapipf instruction.

Results

Table 1 compares demographic and exposure characteristics of the three study groups. Three hundred and ninety (63%) of the study population were male. Cases and controls were similar with respect to level of education. Over 95% of all three sets reported speaking the native language, being of Buddhist religion, and married or widowed (data not shown). Over 70% of cases and controls were farm laborers and 17% worked in service activities (drivers, street vendors, maids, and cooks). Less than 2% of either cases or controls were used in clerical or professional jobs. Overall, there was no evidence of relevant differences in the socioeconomic level of the three groups.

Four percent and 13% of the male and female cases, respectively, had never smoked. The prevalence of never-smokers in women was greater than in men in both control sets, 33% and 37% in hospital and population controls, respectively, versus 10% and 6%. Eleven percent of the smokers consumed only manufactured commercial cigarettes, whereas all others reported mixed consumption of manufactured and traditional products (88%). The consumption of commercial cigarettes was strongly associated with sex (being 1% in women versus 13% in men) and age in the two control sets (55% of controls up to age 55 years and 20% in age group ≥72 years). Fifty-one percent of hospital controls reported being ex-smoker compared with 38% in the population-based set. The mean duration of regular smoking among smokers was 32 years in population controls and 28 years in hospital controls. The average number of cigarettes consumed per day was low compared with Western levels with less than one third of the controls consuming ≥7 cigarettes/d. The median consumption in this upper category was 10 cigarettes/d, cases and controls pooled. The median daily consumption among long-term smokers (41+ years) was 5 cigarettes/d. The two control sets did not differ by cumulative amount of cigarettes smoked (P = 0.58). All controls were therefore pooled in the subsequent analyses except in those involving air pollution in which case only population controls were used.

Chewing of Betel nut was rare in men (overall 4%) and thrice more popular in women (overall 15%).

Table 2 shows ORs associated with tobacco smoking adjusted for sex and age. The risk of lung cancer increased with increasing duration of the habit and reached an OR of 3.3

[95% confidence interval (95% CI), 1.7-6.2] for >40 years of regular smoking compared with nonsmokers. Relatively low amounts of daily consumption were reported with only 14% of all smokers estimating consumption greater than the equivalent of 10 cigarettes/d. The ORs associated with ≥8 cigarettes/d was 4.9 (95% CI, 2.5-9.7). Seventy-eight percent of the smokers consumed the unfiltered traditional products, which were associated with an OR of 2.9 (95% CI, 1.7-5.2). Lifelong cumulative number of cigarettes smoked satisfactorily described the risk associated with tobacco smoking for various patterns of duration and amount. The OR increased regularly up to 4.9 (95% CI, 2.5-9.4) with increasing lifelong cumulative consumption. This variable was used to control confounding by tobacco in all the following analyses. The exclusion of 33 hospital controls admitted for cardiovascular or respiratory diseases did not materially modify the risks described (Table 2). When risk was analyzed separately by sex, ORs associated with cumulative exposure appeared greater in women compared with men although this difference was not statistically significant ($P = 0.18$).

Squamous cell carcinoma were more strongly associated with tobacco smoking than adenocarcinomas (Table 3). For squamous cell carcinoma, the trend of increasing OR was significant and reached an OR of 6.8 (95% CI, 2.8-16.8) in the higher category of exposure. There was a nonsignificant trend ($P = 0.13$) in the ORs among adenocarcinoma cases with increasing levels of exposure but none of the smoking categories was statistically significant. The risks in the group containing the remaining small cell, large cell, and mixed types, were of the same magnitude as those of squamous cell carcinoma and had similar significance.

None of the modestly elevated risks associated with the two indices of exposure to air pollutants from the Mae Moh area reached statistical significance (Table 4). ORs were slightly reduced when adjusted for tobacco smoking, indicating some confounding. Among never-smokers, the OR associated with nondomestic air pollution was more elevated but not statistically significantly (OR, 4.0; 95% CI, 0.7-23.0; Table 4). Cumulative exposure to domestic fumes from burning coal or wood was not associated with excess risk overall nor among

Table 1. Distribution of cases and controls by selected characteristics

	Cases	Hospital controls	Population controls	P of log-likelihood ratio test*
No. subjects	211	211	197	
Men (%)	66.4	59.7	62.9	NS
Education (%)	100	100	100	
Illiterate	33.2	38.9	35.1	
Primary school	64	59.7	61.9	NS
Vocational and high school	2.8	1.4	3.0	
Occupation (%)	100	100	100	
Agricultural workers	71.6	70.1	73.3	
Service workers†	16.6	20.9	13.3	
Professional and clerical	1.9	0.9	0.6	NS
Industry crafts and trade	4.3	3.8	9.4	
Other	5.7	4.3	3.3	
Smoking status (%)	100	100	100	
Never	7.1	19.0	16.8	
Occasional	0.5	0.9	2.0	
Ex-smoker	48.3	50.7	37.6	<0.01
Current	44.1	29.4	43.6	
%Nonsmokers				
Men	4.3	10.3	7.1	NS‡
Women	14.1	34.1	39.7	<0.01‡
Years of regular smoking§				
1-15	2.1	5.3	6.1	
16-40	48.2	58.0	43.9	0.02
41+	49.7	36.7	50.0	
Average cigarettes/d§				
1-3	30.1	32.1	45.0	
4-7	24.4	32.1	26.9	<0.01
7+	45.6	35.7	28.1	
%Smokers of manufactured cigarettes only				
Men	7.9	15.9	21.0	<0.01‡
Women	5.6	3.5	4.1	NS‡
%Smokers consuming commercial cigarettes by age				
Age (y)				
<55	16.1	44.6	41.8	
56-64	32.3	29.2	29.9	$P < 0.01$‖
65-71	25.8	15.4	20.9	
72+	25.8	10.8	7.5	
%Regular chewers of miang¶				
Men	71.4	72.2	72.6	NS
Women	81.8	62.3	72.5	0.03
%Regular chewers of betel nut**				
Men	2.1	5.6	4.8	NS
Women	12.7	14.1	17.8	NS

Abbreviation: NS, not significant.
*Test for independence of case/control status and variable on the row unless otherwise specified.
†Includes street vendors.
‡P for the difference of the three proportions within group.
§Smokers excluded.
‖P for trend by age.
¶Mixture of fermented tea leaves, coconut, peanut, and sugar.
**Betel nut, tobacco, red lime, and fresh leaf of piper betel.

Table 2. ORs associated with selected variables describing tobacco habit

	No. cases	No. controls	All controls, OR* (95% CI)	CVD and respiratory diseases excluded (33 hospital controls), OR* (95% CI)
Smoking status				
Never	16	80	1	
Ex-regular smokers	102	183	2.6 (1.4-4.8)	2.7 (1.5-5.0)
Current regular smokers	93	145	3.2 (1.7-6.0)	3.0 (1.6-5.6)
Duration, y				
1-15	4	19	0.8 (0.2-3.0)	1.1 (0.3-3.8)
16-40	94	167	2.9 (1.6-5.3)	2.9 (1.6-5.4)
41+	97	142	3.3 (1.7-6.2)	3.0 (1.6-5.8)
Average no. cigarettes/d				
1-3	58	126	2.2 (1.2-4.2)	2.2 (1.1-4.1)
4-7	47	97	2.5 (1.2-5.0)	2.6 (1.3-5.4)
7+	88	105	4.9 (2.5-9.7)	4.7 (2.4-9.3)
Types of cigarettes				
Manufactured filtered only	9	26	1.6 (0.6-4.2)	1.6 (0.6-4.1)
Only traditional and manufactured unfiltered	158	245	2.9 (1.7-5.2)	2.9 (1.7-5.1)
Smokers of all types	26	51	2.4 (1.2-4.9)	2.4 (1.1-5.0)
Lifelong cumulative tobacco smoked (000s)				
<60	37	100	1.9 (1.0-3.8)	1.9 (1.0-3.6)
61-126	56	110	2.8 (1.4-5.3)	2.8 (1.4-5.4)
127+	100	118	4.9 (2.5-9.4)	4.9 (2.5-9.5)
Ever traditional products				
Smokers of manufactured products only	15	52	1.4 (0.6-3.1)	1.3 (0.6-3.0)
Smokers of traditional products	180	276	3.1 (1.7-5.6)	3.1 (1.7-5.5)

Abbreviation: CVD, cardiovascular disease.
*Reference category for all variables listed are never-smokers. ORs are adjusted for sex and age.

nonsmokers (Table 4). None of the individual variables contributing to the cumulative index was significantly associated with the disease (details not shown).

The risk of lung cancer associated with smoking was analyzed in combination with the three genetic variants examined in this study. These variants had similar frequency distributions in the two control sets and their pooled distributions were in Hardy-Weinberg equilibrium (Table 5). The *GSTM1−/−* genotype occurred in about two thirds of the population, one third were homozygous carriers of the *CYP1A1**2A variant, whereas homozygous carriers of the *CYP1A1**2C variant were 6% and 8% in cases and controls, respectively. None of the variants was associated with the disease or with tobacco smoking. The possible effect modification of the genotypes on the risk due to tobacco smoking is reported in Table 6. This shows that the magnitude of the tobacco-associated risks is constant across the three genotypes examined, which, therefore, do not show any evidence of acting as modifiers. The same results were found when cases were restricted to squamous cell carcinoma and to adenocarcinoma (details not shown).

Based on the figures in Table 2, the proportion of the lung cancer incidence attributable to tobacco smoking was 98% and 64% of the cases were attributable to tobacco in men and women, respectively.

The frequency of *CYP1A1* haplotypes was estimated from both control groups; there was clear evidence of strong linkage disequilibrium between the *CYP1A1**2A variant and the *CYP1A1**2C variant ($P < 0.0001$) but no evidence for any difference in haplotype frequencies between the two control groups ($P = 0.62$). Finally, there was no evidence of any haplotype frequency differences between cases and controls ($P = 0.32$). For the combined control group, the estimated haplotype frequencies were 0.42 for the wild-type haplotype (Ile for *CYP1A1**2C, T for *CYP1A1**2A), 0.26 for the haplotype carrying the *CYP1A1**2A variant only (Ile, C), 0.02 for the *CYP1A1**2C variant only (Val, T), and 0.30 for the haplotype carrying both variants (Val, C), which is called the *CYP1A1**2B allele.

Discussion

The study presented was motivated and designed to evaluate the causes of the relatively high incidence of lung cancer in a Thai province where several installations of electricity-generating plants coal powered were a source of public concern. The area is largely rural as shown by the high prevalence of agriculture workers and the very low number of people used in industrial activities (Table 1). Pollution from industrial sources other than the one investigated as well as occupational exposures were judged therefore negligible. Road traffic is extensive only along the main provincial roads but certainly sparse in the countryside where most of the population lives.

Table 3. ORs and 95% CIs associated with tobacco smoking by histologic type (population and hospital controls pooled)

	Squamous cell carcinoma	Adenocarcinoma	Other and unknown
Never or occasional*	1	1	1
No. cases/controls			
Lifelong cigarettes smoked (000s)			
<60	2.7 (1.1-6.6)	0.80 (0.2-2.6)	3.1 (0.6-15.3)
61-127	3.1 (1.3-7.8)	1.3 (0.3-4.0)	5.7 (1.2-26.8)
128+	6.8 (2.8-16.8)	2.5 (0.8-7.3)	7.4 (1.5-36.4)
Test for trend, *P*	0.00	0.13	0.01

NOTE: ORs were adjusted for sex, age, and both.
*Reference category.

Table 4. ORs and 95% CIs associated with estimated exposure to air pollution from the Mae Mo Center and with domestic sources of fumes

	No. cases	No. controls	OR* (95% CI)	OR† (95% CI)
Cumulative index of exposure to SO_2 or NO_2, tons/$\sqrt{}$km/y				
<1,808	54	62	1‡	1‡
1,808-3,507	86	72	1.4 (0.8-2.2)	1.3 (0.8-2.1)
>3,507	71	66	1.2 (0.8-2.1)	1.2 (0.7-2.0)
Cumulative index of exposure to suspended particulate, tons/$\sqrt{}$km/y				
<214	56	61	1‡	1‡
214-392	84	70	1.3 (0.8-2.1)	1.2 (0.7-2.0)
>392	71	67	1.1 (0.7-1.9)	1.1 (0.7-1.8)
Cumulative index of exposure to domestic fumes, y				
<9	51	50	1‡	1‡
9-14	67	50	1.3 (0.8-2.2)	1.3 (0.7-2.2)
15-20	43	50	0.8 (0.5-1.5)	0.8 (0.4-1.4)
21+	50	49	1 (0.6-1.7)	0.8 (0.5-1.5)
Nonsmokers only				
Cumulative index of exposure to SO_2 or NO_2/suspended particulate§				
<1,808/<214	3	18	1‡	
≥1,809/≥215	13	26	3.4 (0.7-23.3)	
Cumulative index of exposure to domestic fumes, y				
<15	11	30	1‡	
≥15	5	14	0.4 (0.1-2.0)	

NOTE: ORs were adjusted for sex, age, and cumulative amount of cigarettes smoked. Only population controls were included.
*Adjusted for sex and age.
†Adjusted for sex, age, and cumulative amount of cigarettes smoked.
‡Reference category.
§The two indexes coincide in this subgroup.

The main sources of potential air pollutants identified were the domestic combustion of coal or wood and the power-generating plants. Information on place of cooking and type of fuel used for cooking and heating was obtained by questionnaire. These variables were considered as proxy estimates of exposure to domestic fumes. Nonexposed were subjects who did not use these types of fuel.

We also tested three polymorphisms in two genes involved in the metabolism of lung carcinogens, the *GSTM1* null/null genotype and the two *CYP1A1**2A and *CYP1A1**2C variants. We also examined the *CYP2D6**3, *CYP2D6**4, and *CYP2D6**5 alleles, the common inactive alleles of *CYP2D6* among Caucasians, but found them to be rare in this population and so these are not reported here. *GSTM1* and *CYP1A1* are involved in the metabolism of polycyclic aromatic hydrocarbons. These compounds are environmental contaminants formed during combustion of organic material and are, therefore, particularly relevant in our study where the major sources of pollution derived from tobacco smoking and the burning of coal and wood. Besides tobacco smoking, which alone explained 96% of male and 64% of female incidence, none of the other environmental factors analyzed was associated with the risk of lung cancer. No excess risk was associated with the cumulative index of exposure to exhausts from the Mae Moh plants, or duration of residence by distance from the production area. The relative risks of tobacco smoking that we found are comparable with those measured in many other populations and reflect the cumulative smoking experience of this community. Similarly, low risks for comparable levels of exposure were reported in studies conducted in China and Japan (11-15). The consumption of Western-style manufactured cigarettes is low in this rural region of Thailand where consumers prefer local traditional products consisting of chopped dry tobacco uncured, wrapped in a dry banana leaf. They are unfiltered and larger in size than standard cigarettes. The temperature of mainstream smoke is

Table 5. ORs and 95% CIs associated with polymorphisms in selected metabolic genes

	No. cases	No. controls	OR* (95% CI)	OR† (95% CI)
GSTM1				
−/+ or +/+‡	67	105	1	1
−/−	99	184	0.9 (0.6-1.3)	0.8 (0.5-1.2)
All	166	283		
*CYP1A1**2A SNP (rs4646903)				
T/T‡	26	53	1	1
T/C	87	155	1.2 (0.7-2.1)	1.2 (0.7-2.0)
C/C	55	78	1.5 (0.8-2.8)	1.5 (0.8-2.7)
All	168	286		
Hardy Weinberg equilibrium test§		$P = 0.15$		
*CYP1A1**2C SNP (rs1048943)				
Ile/Ile‡	78	135	1	1
Ile/Val	79	129	1.1 (0.7-1.6)	1.0 (0.7-1.5)
Val/Val	10	23	0.7 (0.3-1.6)	0.8 (0.3-1.7)
All	167	287		
Hardy-Weinberg equilibrium test§		$P = 0.33$		

*Adjusted for sex and age. Two control sets pooled.
†Adjusted for sex, age, and lifelong total tobacco smoked. Two control sets pooled.
‡Reference category.
§Exact probability.

Table 6. Effect of polymorphism in selected genes on the effect of lifelong amount of tobacco smoked

	GSTM1	
	+/+ or −/+	−/−
Never-smokers	1*	0.8
Medium	3.1†	1.8
High	2.9†	3.2†
Controls (%)	105 (37%)	184 (64%)

	*CYP1A1**2A		
	T/T	T/C	C/C
Never-smokers	1*	1.7	0.5
Medium	2.0	3.8	3.7
High	4.8	3.6	6.0§
Controls (%)	53 (19%)	155 (54%)	78 (27%)

	*CYP1A1**2C		
	Ile/Ile	Ile/Val	Val/Val
Never-smokers	1*	0.9	NA‡
Medium	1.6	3.0†	1.8
High	3.8†	2.5	2.5
Controls (%)	213 (74%)	208 (73%)	33 (12%)

Abbreviation: NA, not applicable.
*Reference category.
†$P < 0.05$.
‡Six controls but no cases occurred in this stratum.

high so that inhalation is likely to affect predominantly the bronchus and upper sections of the lungs. The high proportion of smokers in women is remarkable particularly by comparison with other Asian countries. Smoking in women is, however, a traditional habit rather than a recently acquired sign of Westernization; >92% of smoking women in fact had smoked only traditional products and had done so for over 15 years.

The *CYP1A1* gene encodes phase I enzymes that convert many lung carcinogens into DNA-binding metabolites with carcinogenic potential (16). The *GST* family produces phase II enzymes that detoxify reactive electrophils that act as mutagens (17). Variants in these genes may, therefore, modulate their action and result in varying susceptibility to develop cancer. Studies of lung cancer and polymorphisms in the *CYP1A1* gene were reviewed by D'Errico et al. (5), Houlston (18), and Bartsch et al. (19). Studies in Japan have consistently shown increased risks associated with the two variants we analyzed, particularly for squamous cell carcinoma. Excess risk confined to squamous cell carcinoma was found in three ethnic groups of Hawaii (20) and one of three case-control studies in China (21) the other two being negative for an association (22, 23). In a Swedish study (24), neither variant was associated with any subtype of lung cancer. Hong et al. (25) reported a significant deficit of the *CYP1A1**2C polymorphism in Korean lung cancer cases compared with controls, although notably genotypes were not in Hardy-Weinberg equilibrium, even in controls, making this result difficult to interpret. Studies of non-Asians conducted in America and Europe have been generally negative with few exceptions (26-29). However, the best summary of the accumulated evidence comes from a recent pooled analysis of 22 case-control studies that included over 2,400 cases and 3,300 controls and examined genotype in combination with the effect of tobacco smoking (30). This analysis concluded that the homozygous variant *CYP1A1**2A polymorphism increases the risk of lung cancer in Caucasians, whereas no significant increase was observed in Asians among whom the risk associated with tobacco smoking appeared constant by genotype. Our results are therefore in full agreement with this overview of the literature.

Published results on lung cancer risk and the *GSTM1*-null genotype are even more heterogeneous. Two meta-analyses estimated a weak but significant increase in risk associated with *GSTM1* deficiency (5, 31). Several case-control studies conducted in countries of diverse ethnicities were published later. Four in 10 found an overall positive association with lung cancer (29, 32-34), whereas the others were negative for an association (23, 35-40). When exposure to environmental carcinogens was taken into account, effect modification was found in some but not all studies. *GSTM1*-deficient subjects have been found at increased risk of lung cancer if moderately exposed to tobacco smoking (32), or never-smokers carrying the *CYP1A1**2C variant (38), or had a stronger protective effect from vegetable-derived isothiocyanates compared with non-deficient subjects (41). But interaction with tobacco smoking or exposure to asbestos or other environmental pollutant was not detected in other studies (20, 37, 42). In our data, *GSTM1* deficiency and the two *CYP1A1* variants examined were not associated with an increased risk of lung cancer even when polymorphisms in the two genes were combined. Moreover, they did not act as modifiers of the effect of tobacco smoking either overall or in the main histologic subtypes. The distributions of the genotypes that we observed were in Hardy-Weinberg equilibrium in the controls and in agreement with the frequencies estimated for Asians in the Genetic Susceptibility to Environmental Carcinogens (43), an international project that aims at describing the frequency of various alleles in metabolic genes, in different geographic and ethnic groups. We conclude that tobacco smoking remains the only identified risk factor for the disease in this Thai population.

Acknowledgments

We thank the cases and the controls for their participation in the study, the staff of the Lampang Provincial Health Office and Lampang Hospital for carrying out the interviews and collecting the specimens, Miranda Y. Ku for computing pollution exposure estimates, and Dr. Max Parkin for his advice in designing and conducting the study.

References

1. IARC. Tobacco smoke and involuntary smoking. IARC monographs, vol. 83. Lyons (France): IARC Press; 2004.
2. Peto R, Lopez AD, Boreham J, Thun M, Heath C, Jr., Doll R. Mortality from smoking worldwide. Br Med Bull 1996;52:12–21.
3. Vineis P. Molecular epidemiology: low-dose carcinogens and genetic susceptibility. Int J Cancer 1997;71:1–3.
4. Vineis P, Alavanja M, Garte S. Dose-response relationship in tobacco-related cancers of bladder and lung: a biochemical interpretation. Int J Cancer 2004;108:2–7.
5. d'Errico A, Malats N, Vineis P, Boffetta P. Review of studies of selected metabolic polymorphisms and cancer. In: Vineis P, Malats N, Lang M, et al., editors. Metabolic polymorphisms and susceptibility to cancer. IARC Scientific Publication no. 148. Lyons (France): IARC Press; 1999. p. 323–93.
6. Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB, editors. Cancer incidence in five continents. vol. VIII. IARC Scientific Publication no. 155. Lyons (France): IARC Press; 2004.
7. International Labour Office. International standard classification of occupations: ISCO. Geneva (Switzerland): International Labour Office; 1990.
8. Brown LM, Pottern LM, Blot WJ. Lung cancer in relation to environmental pollutants emitted from industrial sources. Environ Res 1984;34:250–61.
9. Buffler PA, Cooper SP, Stinnett S, et al. Air pollution and lung cancer mortality in Harris County, Texas, 1979-1981. Am J Epidemiol 1988;128:683–99.
10. Walter SD. Calculation of attributable risks from epidemiological data. Int J Epidemiol 1978;7:175–82.
11. Gao YT, Blot WJ, Zheng W, Fraumeni JF, Hsu CW. Lung cancer and smoking in Shanghai. Int J Epidemiol 1988;17:277–80.
12. Xu ZY, Blot WJ, Xiao HP, et al. Smoking, air pollution, and the high rates of lung cancer in Shenyang, China. J Natl Cancer Inst 1989;81:1800–6.
13. Liu Z. Smoking and lung cancer in China: combined analysis of eight case-control studies. Int J Epidemiol 1992;21:197–201.
14. Yuan JM, Ross RK, Wng XL, Gao YT, Henderson BE, Yu MC. Morbidity and mortality in relation to cigarette smoking in Shanghai, China. A prospective male cohort study. J Am Med Assoc 1996;275:1646–50.

15. Akiba S. Analysis of cancer risk related to longitudinal information on smoking habits. Environ Health Perspect 1994;102:15–20.
16. Kawajiri K. *CYP1A1*. In: Vineis P, Malats N, Lang M, et al., editors. Metabolic polymorphisms and susceptibility to cancer. In: IARC scientific publication no. 148. Lyons (France): IARC Press; 1999. p. 159–72.
17. Strange RC, Fryer AA. The glutathione *S*-transferases: influence of polymorphism on cancer susceptibility. In: Vineis P, Malats N, Lang M, et al., editors. Metabolic polymorphisms and susceptibility to cancer. IARC scientific publication no. 148. Lyons (France): IARC Press; 1999. p. 159–72.
18. Houlston RS. *CYP1A1* polymorphisms and lung cancer risk: a meta-analysis. Pharmacogenetics 2000;10:105–14.
19. Bartsch H, Nair U, Risch A, Rojas M, Wikman H, Alexandrov K. Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. Cancer Epidemiol Biomarkers Prev 2000;9:3–28.
20. Le Marchand L, Sivaraman L, Pierce L, et al. Associations of *CYP1A1*, *GSTM1*, and CYP2E1 polymorphisms with lung cancer suggest cell type specificities to tobacco carcinogens. Cancer Res 1998;58:4858–63.
21. Song N, Tan W, Xing D, Lin D. *CYP1A1* polymorphism and risk of lung cancer in relation to tobacco smoking: a case-control study in China. Carcinogenesis 2001;22:11–6.
22. Persson I, Johansson I, Lou YC, et al. Genetic polymorphism of xenobiotic metabolizing enzymes among Chinese lung cancer patients. Int J Cancer 1999;81:325–9.
23. Wang J, Deng Y, Li L, et al. Association of *GSTM1*, *CYP1A1* and CYP2E1 genetic polymorphisms with susceptibility to lung adenocarcinoma: a case-control study in Chinese population. Cancer Sci 2003;94:448–52.
24. Alexandrie AK, Nyberg F, Warholm M, Rannug A. Influence of *CYP1A1*, *GSTM1*, GSTT1, and NQO1 genotypes and cumulative smoking dose on lung cancer risk in a Swedish population. Cancer Epidemiol Biomarkers Prev 2004;13:908–14.
25. Hong YS, Chang JH, Kwon OJ, Ham YA, Choi JH. Polymorphism of the *CYP1A1* and glutathione-*S*-transferase gene in Korean lung cancer patients. Exp Mol Med 1998;30:192–8.
26. Cascorbi I, Henning S, Brockmoller J, et al. Substantially reduced risk of cancer of the aerodigestive tract in subjects with variant-463A of the myeloperoxidase gene. Cancer Res 2000;60:644–9.
27. Hamada GS, Sugimura H, Suzuki I, et al. The heme-binding region polymorphism of cytochrome P450IA1 (CypIA1), rather than the *Rsa*I polymorphism of IIE1 (CypIIE1), is associated with lung cancer in Rio de Janeiro. Cancer Epidemiol Biomarkers Prev 1995;4:63–7.
28. Xu X, Kelsey KT, Wiencke JK, Wain JC, Christiani DC. Cytochrome P450 *CYP1A1 Msp*I polymorphism and lung cancer susceptibility. Cancer Epidemiol Biomarkers Prev 1996;5:687–92.
29. Qiñones L, Lucas D, Godoy J, et al. *CYP1A1*, CYP2E1 and *GSTM1* genetic polymorphisms. The effect of single and combined genotypes on lung cancer susceptibility in Chilean people. Cancer Lett 2001;174:35–44.
30. Vineis P, Veglia F, Benhamou S, et al. *CYP1A1* T3801 C polymorphism and lung cancer: a pooled analysis of 2451 cases and 3358 controls. Int J Cancer 2003;104:650–7.
31. Houlston RS. Glutathione *S*-transferase M1 status and lung cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev 1999;8:675–82.
32. Chen S, Xue K, Xu L, Ma G, Wu J. Polymorphisms of the *CYP1A1* and *GSTM1* genes in relation to individual susceptibility to lung carcinoma in Chinese population. Mutat Res 2001;458:41–7.
33. Pinarbasi H, Silig Y, Cetinkaya O, Seyfikli Z, Pinarbasi E. Strong association between the *GSTM1*-null genotype and lung cancer in a Turkish population. Cancer Genet Cytogenet 2003;146:125–9.
34. Ruano-Ravina A, Figueiras A, Loidi L, Barros-Dios JM. *GSTM1* and GSTT1 polymorphisms, tobacco and risk of lung cancer: a case-control study from Galicia, Spain. Anticancer Res 2003;23:4333–7.
35. Stucker I, de Waziers I, Cenée S, et al. *GSTM1*, smoking and lung cancer: a case-control study. Int J Epidemiol 1999;28:829–35.
36. Lewis SJ, Cherry NM, Niven RM, Barber PV, Povey AC. *GSTM1*, GSTT1 and GSTP1 polymorphisms and lung cancer risk. Cancer Lett 2002;180: 165–71.
37. Malats N, Camus-Radon AM, Nyberg F, et al. Lung cancer risk in nonsmokers and *GSTM1* and GSTT1 genetic polymorphism. Cancer Epidemiol Biomarkers Prev 2000;9:827–33.
38. Hung RJ, Boffetta P, Brockmoller J, et al. E. *CYP1A1* and *GSTM1* genetic polymorphisms and lung cancer risk in Caucasian non-smokers: a pooled analysis. Carcinogenesis 2003;24:875–82.
39. Gsur A, Haidinger G, Hollaus P, et al. Genetic polymorphisms of *CYP1A1* and *GSTM1* and lung cancer risk. Anticancer Res 2001;21:2237–42.
40. Lu W, Xing D, Qi J, Tan W, Miao X, Lin D. Genetic polymorphism in lyeloperoxidase but not *GSTM1* is associated with risk of lung squamous cell carcinoma in a Chinese population. Int J Cancer 2002;192:275–9.
41. London SJ, Yuan JM, Chung FL, et al. Isothiocyanates, glutathione *S*-transferase M1 and T1 polymorphisms, and lung-cancer risk: a prospective study of men in Shanghai, China. Lancet 2000;356:724–9. Erratum in: Lancet 2000;356:2104.
42. Stucker I, Boffetta P, Antilla S, et al. Lack of interaction between asbestos exposure and glutathione *S*-transferase M1 and T1 genotypes in lung carcinogenesis. Cancer Epidemiol Biomarkers Prev 2001;10:1253–8.
43. Garte S, Gaspari L, Alexandrie AK, et al. Metabolic gene polymorphism frequencies in control populations. Cancer Epidemiol Biomarkers Prev 2001;10:1239–48.