

## Tobacco and estrogen metabolic polymorphisms and risk of non-small cell lung cancer in women

Michele L.Cote<sup>1,2,\*</sup>, Wonsuk Yoo<sup>2</sup>, Angela S.Wenzlaff<sup>1</sup>, Geoffrey M.Prysak<sup>1</sup>, Susan K.Santer<sup>1</sup>, Gina B.Claeys<sup>1</sup>, Alison L.Van Dyke<sup>1,3</sup>, Susan J.Land<sup>4,5</sup> and Ann G.Schwartz<sup>1,2</sup>

<sup>1</sup>Population Studies and Prevention Program, Karmanos Cancer Institute, 110 East Warren Avenue, Detroit, MI 48201, USA, <sup>2</sup>Department of Internal Medicine, <sup>3</sup>Cancer Biology Program and <sup>4</sup>Applied Genomics Technology Center, Department of Obstetrics and Gynecology, Wayne State University School of Medicine, Detroit, MI 48201, USA and <sup>5</sup>Molecular Biology and Genetics Program, Karmanos Cancer Institute, 110 East Warren Avenue, Detroit, MI 48201, USA

\*To whom correspondence should be addressed. Tel: +1 313 578 4204; Fax: +1 313 578 4306; Email: cotem@karmanos.org

To explore the potential role for estrogen in lung cancer susceptibility, candidate single-nucleotide polymorphism (SNPs) in tobacco and estrogen metabolism genes were evaluated. Population-based cases ( $n = 504$ ) included women aged 18–74, diagnosed with NSCLC in metropolitan Detroit between November 2001 and October 2005. Population-based controls ( $n = 527$ ) were identified through random digit dialing and matched on race and age. Eleven SNPs in 10 different genes were examined in relation to risk: *CYP1A1* MspI, *CYP1A1* Ile<sup>462</sup>Val, *CYP1B1* Leu<sup>432</sup>Val, *CYP17*, *CYP19A1*, *XRCC1* Gln<sup>399</sup>Arg, *COMT* Val<sup>158</sup>Met, *NQO1* Pro<sup>187</sup>Ser, *GSTM1*, *GSTT1* and *GSTP1* Ile<sup>105</sup>Val. Lung cancer risk associated with individual SNPs was seen for *GSTP1* [A allele; odds ratio (OR) = 1.85; 95% confidence interval (CI), 1.04–3.27] and *XRCC1* (A/A genotype; OR = 1.68; 95% CI, 1.01–2.79) in white women and *CYP1B1* (G allele; OR = 11.1; 95% CI, 1.18–104) in black women smokers. White women smokers carrying two risk genotypes at the following loci were at increased risk of lung cancer compared with individuals not carrying risk alleles at these loci: *CYP17* and *GSTM1*, *COMT* and *GSTM1*, *CYP17* and *GSTT1*, *XRCC1* and *GSTP1*, *CYP1B1* and *XRCC1* and *COMT* and *XRCC1*. The most parsimonious model of lung cancer risk in white smoking women included age, family history of lung cancer, history of chronic lung disease, pack-years, body mass index, *XRCC1* A/A genotype, *GSTM1* null and *COMT* A/G or G/G genotype. These findings support the need for continued study of estrogen in relation to lung cancer risk. Polymorphisms in the tobacco metabolism, estrogen metabolism and DNA repair pathways will be useful in developing more predictive models of individual risk.

### Introduction

Lung cancer is the second most commonly diagnosed cancer in men and women and is the leading cause of cancer-related deaths for both groups in the USA (1). It has been suggested that women may be at increased risk of lung cancer compared with their male counterparts, and several sources of evidence, though inconsistent, suggest that estrogens may play a role in lung cancer. Variation in risk factors and tumor characteristics between men and women has been reported in studies showing that women are more likely to have adenocarcinomas of the lung (1), a higher risk in never smokers (2), higher levels of polycyclic aromatic hydrocarbon–DNA adducts at any given level of

smoking (3), higher levels of expression of the gene encoding cytochrome P450 (CYP) 1A1 (3,4), more frequent G:C to T:A transversions in *p53* (5) and more frequent epidermal growth factor receptor mutations (6) than men. Evidence supporting an estrogen component to risk based on epidemiologic studies has been inconsistent, with reports of increased risk of adenocarcinomas with ever/never use of estrogen replacement therapy (7,8), as well as reduced risk with ever/never use of oral contraceptives and hormone replacement therapy (HRT) (9,10) or increased risk associated with long duration of HRT use (11). In our own work, we found an inverse relationship between HRT use and non-small cell lung cancer (NSCLC) risk in post-menopausal women (12). This association was strongest in women with tumors expressing estrogen receptors (ERs). In addition to exogenous exposure to hormones, some findings report that lung cancer risk may be associated with endogenous hormones. A prospective cohort study of Canadian women found increased risk of lung cancer [Hazard Ratio = 1.42; 95% confidence intervals (CI), 1.06–1.88] among women with five or more births compared with nulliparous women (11). Additionally, among parous women, those who had their first birth after age 30 were at reduced risk compared with those who had their first birth before age 23, after adjustment for various other risk factors, including smoking (11). A population-based cohort study of never-smoking Japanese women reported that the interaction between later age at menopause and earlier age at menarche was associated with increased risk of lung cancer (Relative Risk = 2.76; 95% CI, 1.32–5.77) after adjusting for age, region and passive smoking exposure, although age at menopause and menarche were not individually associated with risk (7). Thus, the roles that endogenous and exogenous hormones play in lung cancer development have not been elucidated through epidemiologic studies.

Laboratory-based evidence supporting a role for estrogens in lung cancer development and progression has also been published. Several studies suggest frequent expression of ER in lung tissue, lung tumors and lung cancer cell lines (13–18). Further epidemiologic evidence suggests that ER may be associated with NSCLC prognosis (19–22). Estrogen bound to ERs affects cell growth, and multiple pathways of ER action in the lung are currently being studied (13,14,18).

Numerous studies have taken a candidate gene approach to understanding susceptibility to lung cancer. The focus of this work has been on single-nucleotide polymorphisms (SNPs) in genes coding for enzymes involved in tobacco carcinogen metabolism (e.g. *CYP1A1* and *NQO1*) and genes involved in DNA repair (e.g. *XRCC1*) (23,24). A number of the same genes involved in tobacco carcinogen metabolism pathways are also active in the estrogen metabolism pathway, including *CYP1A1*, *CYP1B1* and the glutathione *S*-transferases (*GSTs*) (24). In addition, catechol-O-methyltransferase (*COMT*) enzymes appear to play a role in both estrogen metabolism and possibly nicotine dependence by breaking down catecholamines in the brain (25–27). *CYP17* and *CYP19A1* are both key enzymes in biosynthesis of estradiol, but do not have an established role in tobacco carcinogen metabolism (28–30). To further explore the role of estrogen in lung cancer development, we evaluated the role of the following SNPs in these tobacco and estrogen metabolism genes: *CYP1A1* MspI, *CYP1A1* Ile<sup>462</sup>Val, *CYP1B1* Leu<sup>432</sup>Val, *CYP17*, *CYP19A1*, *XRCC1* Gln<sup>399</sup>Arg, *COMT* Val<sup>158</sup>Met, *NQO1* Pro<sup>187</sup>Ser, *GSTM1*, *GSTT1* and *GSTP1* Ile<sup>105</sup>Val along with tobacco use and estrogen exposures in determining risk of NSCLC in women in a large population-based study. In addition, we further examine the role of these SNPs as part of a risk prediction model, including both traditional and genetic risk factors.

### Materials and methods

#### Study subjects

Subjects were enrolled through the population-based Metropolitan Detroit Cancer Surveillance System, a participant in the National Cancer Institute's

**Abbreviations:** BMI, body mass index; CI, confidence interval; *COMT*, catechol-O-methyltransferase; *CYP*, cytochrome P450; *ER*, estrogen receptor; *ETS*, environmental tobacco smoke; *GST*, glutathione *S*-transferase; *HRT*, hormone replacement therapy; *NSCLC*, non-small cell lung cancer; *OR*, odds ratio; *PCR*, polymerase chain reaction; *SNP*, single-nucleotide polymorphism.

Surveillance, Epidemiology and End Results program. This study has been described in detail elsewhere (12). Women aged 18–74, diagnosed with primary NSCLC in Wayne, Macomb and Oakland counties between 1 November 2001 and 31 October 2005 were eligible to participate. Ascertainment was originally limited to those with adenocarcinoma histology, but was broadened after 1 November 2004 to include all NSCLC histologic types since many histological diagnoses at the time of rapid case ascertainment were not more specific. The majority of cases (71.6%) were adenocarcinoma histology, 8.9% were squamous cell carcinoma, 3.0% large cell carcinoma and 16.5% were NSCLC, unspecified, reflecting this sampling method.

Of the 1092 women identified as eligible for the study, 582 women completed an in-person interview and provided a DNA sample (53.3%), 133 were too ill (12.2%), 273 refused (25.0%) and a working phone number could not be found for another 97 (8.9%). Seven women participated in the interview portion of the study but did not provide a DNA sample (0.6%). Women self-reporting race other than black or white ( $n = 16$ ) were excluded from these analyses. Participants reporting a previous history of breast cancer (64 cases and 27 controls) were excluded from this analysis due to the associations between reproductive factors, estrogen-related SNPs and breast cancer risk thus ensuring that associations detected were not driven by differences in breast cancer risk factors between cases and controls. In total, interview data and DNA samples were available for 504 women with lung cancer.

Population-based controls were identified through random digit dialing. Exclusion criteria included a previous diagnosis of a primary lung cancer; however, due to the low prevalence of this condition, no controls were excluded for this reason. Control women were frequency matched to cases on race and 5 years age group. Of the households willing to complete the brief eligibility screening questionnaire, 565 (68.5%) participated. Eleven controls reporting race other than black or white and 27 with a previous breast cancer diagnosis were excluded from this analysis, leaving 527 controls with interview data and DNA samples for analysis.

#### Data and biospecimen collection

All local institutional and review boards approved this study. Informed consent was obtained from each subject prior to study participation in an in-person interview. Data collected included demographic information, self-reported race, smoking history, health history, reproductive history and environmental tobacco smoke (ETS) exposure. Health history included self-report of physician diagnoses of asthma, emphysema, allergies, pneumonia, bronchitis, chronic obstructive pulmonary disease, tuberculosis and cancer. Emphysema, chronic obstructive pulmonary disease and chronic bronchitis were combined to create a broad 'chronic obstructive lung disease' variable. Diagnoses reported within 1 year of lung cancer diagnosis (for cases) or interview (for the controls) were excluded. Non-smokers were individuals who reported smoking <100 cigarettes in their lifetime. Ex-smokers were women who reported quitting smoking  $\geq 2$  years prior to diagnosis (cases) or interview (controls). Both ex-smokers and current smokers were further categorized based on number of pack-years of smoking, based on the median amount of smoking in the control population. Light smokers were individuals with  $\leq 25$  pack-years of smoking, and heavy smokers included women with  $\geq 26$  pack-years of use. Dates and dosages for oral contraceptive use and HRT were collected. Estrogen only, estrogen combined with progesterone, progesterone only and unknown HRT formulations were included as HRT use in analyses. Family history of lung cancer was coded as yes or no based on detailed first-degree family history information.

#### Genotyping

**DNA extraction.** DNA was isolated from whole blood with a Gentra Autopure system (Qiagen, Valencia, CA), cells from inside the mouth (Gentra Puregene kit) or paraffin-embedded tissue (QIAamp DNA Mini Kit, Qiagen) following the manufacturers' protocols. When multiple biospecimens were obtained from a study participant, DNA extracted from blood was used preferentially, followed by DNA extracted from mouthwash samples and DNA extracted from normal tissue within paraffin blocks. In cases, DNA was obtained from blood ( $n = 428$ , 84.9%), buccal cells ( $n = 72$ , 14.3%) and normal tissue blocks ( $n = 4$ , 0.8%). Controls provided blood ( $n = 484$ , 91.8%) or buccal cells ( $n = 43$ , 7.2%) for genotyping.

**Preamplification.** DNA isolated from buccal cells or paraffin-embedded tissue was preamplified using a nested polymerase chain reaction (PCR) strategy. Preamplification (outer amplification) was carried out in a 25  $\mu$ l reaction containing 2.5 mmol/l  $MgCl_2$ , 0.5  $\mu$ mol/l of the gene-specific outer forward and reverse primers, 1.25 U AmpliTaq Gold polymerase and 200  $\mu$ mol/l of deoxyadenosine triphosphate, deoxycytidine triphosphate, deoxyguanosine triphosphate and deoxythymidine triphosphate. The outer amplification mixture was denatured at 95°C for 10 min and amplification was achieved by 15 cycles of 94°C for 30 s, a primer-specific annealing temperature for 30 s and 72°C for

1 min, followed by a final extension at 72°C for 10 min. The outer amplification was performed on a Mastercycler® Gradient thermocycler (Eppendorf, Westbury, NY). PCR failure was responsible for the following missing genotype data: eight cases, eight controls (*CYP1A1 MspI*); two cases, four controls (*CYP1A1 Ile<sup>462</sup>Val*); five cases, one control (*CYP1B1*); two cases (*CYP17*); four cases, one control (*CYP19A1*); four cases, one control (*NQO1*); two cases, three controls (*GSTM1*); one case (*GSTT1*); four cases, one control (*GSTP1*); seven cases and four controls (*COMT*) and two cases (*XRCC1*). Additionally, we did not attempt to genotype five cases and 24 controls at *CYP19A1* due to time and financial constraints at the end of the study period. Overall, genotyping success rates were >98% for all genotypes.

**Restriction fragment length polymorphism.** Genotypes for the *CYP1A1 MspI* (rs4646903) polymorphism were ascertained using restriction fragment length polymorphism protocols as described by Drakoulis *et al.* (31). Briefly, products of the enzyme digests were run on an 8% polyacrylamide gel and visualized by ethidium bromide staining. Sequencing verified a random sample of 5% of the subjects with 100% concordance. Laboratory techniques and primers are described in greater detail in Wenzlaff *et al.* (32).

**TaqMan assays.** Genotypes for *CYP1A1 Ile<sup>462</sup>Val* (rs1048943), *CYP17* (rs743572), *CYP1B1 Leu<sup>432</sup>Val* (rs1056836), *XRCC1 Gln<sup>399</sup>Arg* (rs25487) and *COMT Val<sup>158</sup>Met* (rs4680) polymorphisms were ascertained through the creation of custom TaqMan assays (TaqMan, Applied Biosystems, Foster City, CA) (see Table II footnotes for specific primers). The *CYP19A1* tetranucleotide repeat polymorphism (TTTA) inner reaction contained 1 $\times$  True Allele PCR Mix (Applied Biosystems), 333 nM *CYP19 IF* (5'-NED-TGAATGT-GCCTTTTGTGAAATCATA-3') and *CYP19 R* primer and either 1  $\mu$ l of the preamplified outer reaction or 20 ng genomic DNA. The NED-labeled inner PCR products were mixed with ROX-labeled 500HD size standards (Applied Biosystems) and electrophoresed on a AB3100 (Applied Biosystems). CEPH 1347-02 DNA (Applied Biosystems) was used as a standard DNA and amplified and electrophoresed on each 96-well plate. *NQO1 Pro<sup>187</sup>Ser* polymorphism (rs1800566) genotyping was performed using a 5'-nuclease assay (TaqMan, Applied Biosystems) for all subjects and has been described previously (33). Likewise, the *Ile<sup>105</sup>Val* polymorphisms of *GSTP1* (rs1695) and null mutations of *GSTM1* and *GSTT1* were also determined by using the TaqMan 5'-nuclease assay and have been described previously (34). For all markers, 5% of the products were randomly sequenced and 10% of genotypes were carried out in duplicate, with 100% concordance between the genotype results.

#### Statistical analysis

Genotype frequencies for each polymorphism were calculated for cases and controls, stratifying by race. Hardy–Weinberg equilibrium tests of genotype distribution in controls, by race, were conducted for all genes except *GSTM1*, *GSTT1* and *CYP19A1* because of the nature of the mutations. Comparisons of dichotomous risk factors were made between cases and controls using Pearson's chi-square tests; comparisons of means were conducted using Student's *t*-tests. Heterozygotes were combined with homozygotes to assess risk of lung cancer associated with carrying certain genotypes. Specifically, individuals with at least one C allele at *CYP1A1 MspI* (TC or CC) were compared with those with the TT genotype. Other combinations were as follows: *CYP1A1 Ile<sup>462</sup>Val* individuals carrying at least one G allele (Ile/Val or Val/Val) were compared with those with the AA (Ile/Ile) genotype; at *CYP1B1 Ile<sup>432</sup>Val*, those with at least one G allele (Ile/Val or Val/Val) were compared with those with the CC (Ile/Ile) genotype; at *CYP17*, those with at least one G allele were compared with the AA genotype; at *NQO1*, those with at least one T allele (Pro/Ser or Ser/Ser) were compared with the CC (Pro/Pro) genotype; at *GSTP1*, those with at least one A allele (Ile/Ile or Ile/Val) were compared with the GG (Val/Val) genotype; at *COMT*, those with at least one A allele (Met/Met or Val/Met) were compared with the GG (Val/Val) genotype and at *XRCC1*, those with the AA genotype (Gln/Gln) were compared with those with any G allele (Gln/Arg or Arg/Arg). For the repeat polymorphism in *CYP19A1*, TTTA<sub>10</sub> was established as the cut point based on the frequency distribution in the control population. The effect of each variant on lung cancer risk was tested in unconditional logistic regression models, adjusting for age at diagnosis (cases)/interview (controls), self-reported race (white/black), pack-years of smoking, current body mass index (BMI), family history of lung cancer and history of a chronic obstructive lung disease. Odds ratios (OR) and 95% CIs for each genotype were calculated from coefficients in the adjusted models. Analyses were repeated after stratification by age (<60/60+), race, smoking exposure history (never/ever and light/heavy use among smokers) and HRT use (ever/never).

Selecting the group that represented the majority of our lung cancer cases (white ever smokers), two gene combinations were analyzed in unconditional logistic regression models to assess risk of lung cancer associated with carrying

no risk genotypes, one risk genotype or two risk genotypes. It was determined *a priori* that we would analyze single SNPs that were found to be statistically significant in the white population. Data are presented for combinations of the *GSTs* and *XRCC1* with the other SNPs analyzed in this study. In addition, using data from this subgroup of white ever-smoking women, we performed a model building procedure using forward selection which included both environmental risk factors (age, current BMI, pack-years, obstructive lung disease and family history) and the candidate genotypes. To assess which environmental risk factors were important and whether the addition of genotype information into this model would improve the fit, we used the likelihood ratio test to calculate the statistical significance of nested models as new terms were added. Differences in the 2 ln (likelihood of full model/likelihood of null model) between nested models (notated as 'G') were compared with the  $\chi^2$  distribution, with the degrees of freedom equal to the difference in the degrees of freedom between the two models. In addition to the main effects of these genes, we considered two- and three-way gene-gene interactions. We also considered the interaction of each gene with environmental risk factors: age (<60/60+ years

of age), BMI tertiles ( $\leq 24.9/25.0-29.9/30+$ ), personal smoking history (0-25/26+ pack-years), personal history of an obstructive lung disease (yes/no) and family history of lung cancer (yes/no). When evidence of gene-gene or gene-environment interaction was found, we examined cases and controls separately using a log-linear model. Adjusted ORs for variables included in the most parsimonious model are presented. The significance level was set at  $\alpha = 0.05$  and all tests were two sided. All statistical analyses were done using SAS version 9.1 (SAS Institute, Cary, NC).

## Results

Table I presents the characteristics of case and control women, overall and stratified by race. Among white women, cases and controls were frequency matched on age (at diagnosis or interview), with the mean for both groups of ~60 years of age ( $P = 0.19$ ), with the majority of the subjects reporting to be post-menopausal (95.1% of cases and

**Table I.** Characteristics of NSCLC cases and controls

Characteristic	Total sample			White <sup>a</sup>			Black <sup>a</sup>		
	Cases <i>N</i> = 504 (48.9%)	Controls <i>N</i> = 527 (51.1%)	<i>P</i> -value	Cases <i>N</i> = 389 (48.9%)	Controls <i>N</i> = 406 (51.7%)	<i>P</i> -value	Cases <i>N</i> = 115 (48.7%)	Controls <i>N</i> = 121 (51.3%)	<i>P</i> -value
Age (mean, SD)	59.7 (9.2)	59.0 (9.5)	0.27	60.3 (9.1)	59.5 (9.4)	0.19	57.4 (9.3)	57.5 (9.8)	0.94
Current BMI (mean, SD)	26.2 (5.8)	29.4 (7.3)	<0.0001	26.0 (5.6)	28.6 (6.6)	<0.0001	26.7 (6.6)	32.2 (8.6)	<0.0001
Education, <i>n</i> (%)			<0.0001			<0.0001			0.48
Less than high school	94 (18.7)	45 (8.5)		65 (16.8)	22 (5.4)		29 (25.2)	23 (19.0)	
High School graduate or equivalent	200 (39.8)	168 (31.9)		166 (42.8)	132 (32.5)		34 (29.6)	36 (29.8)	
Some college	209 (41.5)	314 (59.6)		157 (40.4)	252 (62.1)		52 (45.2)	62 (51.2)	
Smoking status, <i>n</i> (%)			<0.0001			<0.0001			<0.0001
Never	40 (7.9)	259 (49.5)		32 (8.2)	204 (50.6)		8 (7.0)	55 (45.9)	
Exsmoker <sup>b</sup>	159 (31.6)	164 (31.4)		166 (34.2)	130 (32.3)		26 (22.6)	34 (28.3)	
Current smoker	305 (60.5)	100 (19.1)		224 (57.6)	69 (17.1)		81 (70.4)	31 (25.8)	
Pack-years (mean, SD)	46.9 (29.2)	25.3 (22.8)	<0.0001	51.2 (29.6)	26.2 (22.8)	<0.0001	32.7 (22.9)	22.6 (22.7)	0.005
Exposure to ETS as a child, <i>n</i> (%)			0.04			0.11			0.18
No	108 (21.4)	142 (26.9)		80 (20.6)	103 (25.4)		28 (24.3)	39 (32.2)	
Yes	396 (78.6)	385 (73.1)		309 (79.4)	303 (74.6)		87 (75.7)	82 (67.8)	
Years (mean, SD among exposed)	16.5 (3.8)	15.8 (6.1)	0.08	16.8 (3.5)	16.0 (6.5)	0.07	16.5 (4.8)	16.2 (4.5)	0.76
Household exposure to ETS as an adult, <i>n</i> (%)			<0.0001			<0.0001			0.94
No	98 (19.5)	177 (33.7)		69 (17.8)	146 (36.0)		29 (25.2)	31 (25.6)	
Yes	405 (80.5)	349 (66.3)		319 (82.2)	259 (64.0)		86 (74.8)	90 (74.4)	
Years (mean, SD among exposed)	27.0 (15.2)	19.8 (13.3)	<0.0001	28.0 (14.6)	19.9 (13.2)	<0.0001	23.7 (16.7)	19.8 (13.8)	0.09
Work exposure to ETS, <i>n</i> (%)			<0.0001			<0.0001			0.05
No	108 (21.5)	184 (35.0)		80 (20.6)	140 (34.6)		28 (24.6)	44 (36.4)	
Yes	395 (78.5)	341 (65.0)		309 (79.4)	264 (65.4)		86 (75.4)	77 (63.6)	
Years (mean, SD among exposed)	16.4 (10.7)	13.2 (11.0)	<0.0001	16.7 (11.0)	12.9 (11.1)	<0.0001	15.4 (9.4)	14.3 (10.7)	0.47
History of chronic obstructive lung disease, <i>n</i> (%) <sup>c</sup>			<0.0001			<0.0001			0.38
No	346 (68.7)	453 (86.0)		252 (64.8)	349 (86.0)		94 (81.7)	104 (86.0)	
Yes	158 (31.3)	74 (14.0)		137 (35.2)	57 (14.0)		21 (18.3)	17 (14.0)	
First-degree family history of lung cancer, <i>n</i> (%)			<0.0001			<0.0001			0.01
No	374 (74.2)	463 (87.9)		287 (73.8)	356 (87.7)		87 (75.6)	107 (88.4)	
Yes	130 (25.8)	64 (12.1)		102 (26.2)	50 (12.3)		28 (24.4)	14 (11.5)	
HRT use, <i>n</i> (%)			0.13			0.25			0.24
Ever	233 (46.2)	269 (51.0)		203 (52.3)	177 (43.6)		30 (26.1)	40 (33.1)	
Never	270 (53.8)	258 (49.0)		185 (47.7)	229 (56.4)		85 (73.9)	81 (66.9)	

<sup>a</sup>White cases were missing one response for education, one for household exposure to ETS as an adult and one response for HRT use. White controls were missing three responses for smoking status and one response for household exposure to ETS as an adult. Black cases were missing one response for work exposure to ETS and black controls were missing one response for smoking status.

<sup>b</sup>Quit smoking at least 2 years prior to diagnosis (cases) or interview (controls).

<sup>c</sup>Chronic bronchitis, chronic obstructive pulmonary disease or emphysema.

85.5% of controls,  $P < 0.0001$ , data not shown). White controls had higher BMIs than case women ( $P < 0.0001$ ) and had attained higher levels of education ( $P < 0.0001$ ). White control women were less likely to be current smokers ( $P < 0.0001$ ). Among those who did smoke (either current or ex-smokers), white control women reported significantly fewer pack-years of smoking ( $P < 0.0001$ ). Exposure to ETS as a child was similar among all white women ( $P = 0.11$ ), but among those who were exposed (approximately three quarters of women in both groups), the mean number of years was similar for both groups ( $P = 0.07$ ). As adults, both household exposure to ETS and work exposure to ETS was more common among the white case women (both  $P < 0.0001$ ), and the number of years exposed at home and work was also significantly higher (both  $P < 0.0001$ ). White cases were also more likely than controls to report a personal history of chronic obstructive lung disease ( $P < 0.0001$ ) and to report a family history of lung cancer in a first-degree relative ( $P < 0.0001$ ). There were no differences in reported ever/never use of HRT between the white case and control women ( $P = 0.25$ ).

The mean age reported among black cases and controls was similar ( $P = 0.94$ ), with 88.7% of cases and 81.8% of controls reporting to be post-menopausal ( $P = 0.14$ , data not shown). Black controls had higher BMIs than case women ( $P < 0.0001$ ). Educational attainment was similar for both groups ( $P = 0.48$ ). Black control women were less likely to be current smokers ( $P < 0.0001$ ), and those that were ever smokers reported significantly fewer pack-years of smoking compared with black case women ( $P < 0.0001$ ). Exposure to ETS as a child and as an adult at home was similar among all black women ( $P = 0.18$  and  $P = 0.94$ , respectively). Work exposure to ETS was more common among the black case women ( $P = 0.05$ ), but the number of years exposed did not differ significantly ( $P = 0.47$ ). Black cases were not more likely than controls to report a personal history of chronic obstructive lung disease ( $P = 0.38$ ), but were more likely to report a family history of lung cancer in a first-degree relative ( $P < 0.01$ ). There were no differences in reported ever/never use of HRT between the black case and control women ( $P = 0.24$ ).

Table II reports the genotype frequency by case or control status, stratified by race. The following genotype distributions in controls differed between white and black women: *CYP1A1 MspI* ( $P = 0.03$ ), *CYP1B1* ( $P < 0.0001$ ), *CYP19A1* ( $P < 0.0001$ ), *GSTM1* ( $P < 0.0001$ ), *GSTP1* ( $P < 0.005$ ), *COMT* ( $P < 0.0001$ ) and *XRCC1* ( $P < 0.0001$ ). We examined Hardy–Weinberg equilibrium in the white controls and black controls separately, and no evidence of deviation from Hardy–Weinberg equilibrium in either group was found. The only significant difference between cases and controls was seen in black women for the *CYP17* genotypes ( $P = 0.01$ ). All other genotype frequencies were similar between cases and controls in unadjusted analyses.

Table III presents lung cancer risk associated with each genotype for all women combined, then stratified by race and by smoking status and race and HRT use, after adjusting for BMI, age, family history of lung cancer, pack-years of smoking (among smokers) and obstructive lung disease (in the smokers groups only). Among all women, those with *GSTM1* null genotypes were more likely to have lung cancer than those with a *GSTM1* present genotype (OR = 1.40; 95% CI, 1.02–1.91). Women carrying at least one A allele at *GSTP1* were also at greater risk of having lung cancer compared with those with the GG genotype (OR = 1.59; 95% CI, 1.02–2.47); a similar association was seen for white women only (OR = 1.85; 95% CI, 1.04–3.27). Women carrying the A/A genotype at *XRCC1* were 1.6-fold more likely to have lung cancer compared with women with the at least one G allele (OR = 1.62; 95% CI, 1.01–2.60), and this association was also seen for white women only (OR = 1.68; 95% CI, 1.01–2.79), particularly in ever-smoking white women (OR = 2.11; 95% CI, 1.13–3.94). Among black ever-smoking women, those carrying the C/G or G/G genotype were 11.09-fold more likely to have lung cancer compared with those carrying the C/C genotype at *CYP1B1* (OR = 11.09; 95% CI, 1.18–104.25), although it should be noted that these results are based on six individuals with the C/C genotype. There were no other associations between single

genotypes and lung cancer risk among the entire group, by race or by HRT use.

Table IV presents lung cancer risk associated with single genotypes and two gene combinations (the *GSTs* and *XRCC1*) for white ever-smoking women, with ORs and 95% CIs adjusted for age, BMI, pack-years of smoking, family history of lung cancer and history of obstructive lung disease. *GSTM1* and *GSTT1* were included because, although not statistically significant at the 0.05 level in our population, they are among the most commonly analyzed gene–gene combinations in the literature and have both been identified singularly in larger pooled and meta-analyses as having modest associations with lung cancer risk (35,36). When examining two genes and coding risk as 0 risk genotypes, 1 risk genotype or 2 risk genotypes, the combination of carrying *GSTM1* null genotype and the *CYP17* risk genotype (A/G and G/G) was associated with an increased risk of lung cancer compared with individuals who carried 0 risk genotypes at these loci (OR = 1.83; 95% CI, 1.02–3.30). Increased risk was also seen for individuals carrying one or two risk genotypes at *GSTM1* and *COMT* compared with those with no risk genotypes at these loci (OR = 2.28; 95% CI, 1.07–4.84 and OR = 2.57; 95% CI, 1.18–5.59, respectively). Risk was also increased for individuals carrying either risk genotype for *GSTM1* or *XRCC1* (OR = 1.73; 95% CI, 1.12–2.65) compared with those with no risk genotypes at these loci. Individuals with both *GSTM1* and *XRCC1* risk genotypes had non-statistically significant increased risk compared with those with no risk genotypes (OR = 2.12; 95% CI, 0.92–4.89). Increased risk of lung cancer was also seen among individuals who carry both the *GSTT1* and *CYP17* risk genotypes compared with those who do not (OR = 2.06; 95% CI, 1.02–4.19). Increased risk was also seen for those carrying either the *GSTT1* or *XRCC1* risk genotypes (OR = 1.76; 95% CI, 1.11–2.77) and for individuals carrying *GSTT1* and *GSTP1* risk genotypes compared with those with no risk alleles at these loci (OR = 2.56; 95% CI, 1.02–6.42). Individuals with *GSTP1* and *XRCC1* risk genotypes were >3-fold more likely to have lung cancer compared with those with no risk genotypes at these loci (OR = 3.57; 95% CI, 1.38–9.21). Individuals with both risk genotypes for *XRCC1* and *CYP1B1* were also >3-fold more likely to have lung cancer compared with those with neither risk genotype (OR = 3.39; 95% CI, 1.35–8.53), and those with *XRCC1* and *COMT* risk genotypes were >2-fold more likely to have lung cancer compared with those without risk genotypes at these loci (OR = 2.54; 95% CI, 1.10–5.83).

Estimates for each variable in the best fitting unconditional regression model are presented in Table V for white ever-smoking women. The most parsimonious model consisted of traditional risk factors for lung cancer, including age at diagnosis ( $P = 0.34$ ), family history of lung cancer in a first-degree relative ( $P = 0.0011$ ), personal history of a chronic obstructive lung disease ( $P = 0.008$ ) and BMI (with a higher BMI showing a protective effect) ( $P = 0.0003$ ). Among the traditional risk factors, pack-years of cigarette smoking was the most significant predictor of lung cancer risk ( $P < 0.0001$ ). Three of the SNPs examined in this study were also selected in the most parsimonious model, namely *GSTM1* ( $P = 0.11$ ), *COMT* A/G or G/G genotype ( $P = 0.59$ ), and *XRCC1* A/A was the strongest genetic predictor of risk that was examined in this study ( $P = 0.02$ ). The addition of *XRCC1* provided a better fit than the risk factor model alone ( $G = 5.91$ ), and the addition of *GSTM1* further improved the fit ( $G = 5.47$ ). Only the addition of *COMT* significantly improved the fit of the model once *XRCC1* and *GSTM1* were included with the traditional risk factors ( $G = 4.11$ ), resulting in the most parsimonious model. It should be noted that after the inclusion of the traditional risk factors associated with lung cancer, adding any combination of two SNPs was found to result in a significantly improved model fit, even when the main effects of each single SNP by itself were not significant. No statistically significant interactions were noted between the SNPs, although there was a significant interaction between *GSTM1* (null/present) and BMI (underweight or normal/overweight/obese) ( $P = 0.007$ ). After stratification by case or control status, this association was not seen among case women ( $P = 0.26$ ), only control women ( $P = 0.02$ ) and may represent a chance finding. No other

**Table II.** Genotype distribution in women with lung cancer and controls, by race

Genotype	Total sample			White			Black		
	Cases N = 504, n (%)	Controls N = 527, n (%)	P- value	Cases N = 389, n (%)	Controls N = 406, n (%)	P- value	Cases N = 115, n (%)	Controls N = 121, n (%)	P- value
<i>CYP1A1</i> <i>MspI</i>									
T/T	373 (75.2)	402 (77.4)	0.23	308 (80.0)	320 (80.0)	0.60	65 (58.6)	82 (68.9)	0.16
C/T	109 (22.0)	110 (21.2)		70 (18.2)	76 (19.0)		39 (35.1)	34 (28.6)	
C/C	14 (2.8)	7 (1.4)		7 (1.8)	4 (1.0)		7 (6.3)	3 (2.5)	
<i>CYP1A1</i> <i>Ile<sup>462</sup>Val<sup>a</sup></i>									
A/A	464 (92.4)	489 (93.5)	0.25	359 (92.8)	372 (92.5)	0.38	105 (91.3)	117 (96.7)	0.08
A/G	38 (7.6)	32 (6.1)		28 (7.2)	28 (7.0)		10 (8.7)	4 (3.3)	
G/G	0 (0)	2 (0.4)		0 (0)	2 (0.5)		0 (0)	0 (0)	
<i>CYP1B1</i> <sup>b</sup>									
C/C	120 (24.0)	144 (27.4)	0.45	119 (30.9)	136 (33.6)	0.70	1 (0.9)	8 (6.6)	0.07
C/G	234 (46.9)	240 (45.6)		190 (49.4)	195 (48.1)		44 (38.6)	45 (37.2)	
G/G	145 (29.1)	142 (27.0)		76 (19.7)	74 (18.3)		69 (60.5)	68 (56.2)	
<i>CYP17</i> <sup>c</sup>									
A1/A1	180 (35.9)	209 (39.7)	0.34	136 (35.1)	159 (39.2)	0.36	44 (38.6)	50 (41.3)	0.01
A1/A2	246 (49.0)	251 (47.6)		200 (51.5)	189 (46.5)		46 (40.4)	62 (51.2)	
A2/A2	76 (15.1)	67 (12.7)		52 (13.4)	58 (14.3)		24 (21.0)	9 (7.4)	
<i>CYP19A1</i> <sup>d</sup>									
<10	222 (44.8)	217 (43.2)	0.61	137 (35.8)	133 (34.6)	0.74	85 (75.9)	84 (71.2)	0.42
10+	273 (55.2)	285 (56.8)		246 (64.2)	251 (65.4)		27 (24.1)	34 (28.8)	
<i>NQO1</i>									
C/C	348 (69.6)	350 (66.5)	0.40	271 (70.0)	271 (66.9)	0.32	77 (67.5)	79 (65.3)	0.82
C/T	129 (25.8)	155 (29.5)		97 (25.1)	119 (29.4)		32 (28.0)	36 (29.8)	
T/T	23 (4.6)	21 (4.0)		19 (4.9)	15 (3.7)		4 (3.5)	6 (4.9)	
<i>GSTM1</i>									
0 (+)	256 (51.0)	299 (57.1)	0.05	178 (45.9)	206 (51.1)	0.14	78 (68.4)	93 (76.9)	0.15
1 (null)	246 (49.0)	225 (42.9)		210 (54.1)	197 (48.9)		36 (31.6)	28 (23.1)	
<i>GSTT1</i>									
0 (+)	398 (79.3)	419 (79.5)	0.88	311 (80.1)	324 (79.8)	0.90	87 (75.7)	95 (78.5)	0.60
1 (null)	105 (20.7)	108 (20.5)		77 (19.9)	82 (20.2)		28 (24.3)	26 (21.5)	
<i>GSTP1</i>									
G/G	63 (12.6)	84 (16.1)	0.30	38 (9.8)	56 (13.8)	0.21	25 (21.9)	28 (23.3)	0.86
A/G	238 (47.6)	244 (46.3)		177 (45.9)	184 (45.3)		61 (53.5)	60 (50.0)	
A/A	199 (39.8)	198 (37.6)		171 (44.3)	166 (40.9)		28 (24.6)	32 (26.7)	
<i>COMT</i> <sup>e</sup>									
G/G	134 (27.0)	151 (28.9)	0.47	78 (20.3)	92 (22.8)	0.46	56 (50.0)	59 (49.2)	0.79
A/G	251 (50.5)	244 (46.6)		205 (53.2)	197 (48.9)		46 (41.1)	47 (39.2)	
A/A	112 (22.5)	128 (24.5)		102 (26.5)	114 (28.3)		10 (8.9)	14 (11.7)	
<i>XRCC1</i> <sup>f</sup>									
G/G	258 (51.4)	248 (47.1)	0.05	172 (44.4)	160 (39.4)	0.06	86 (74.8)	88 (72.7)	0.80
A/G	182 (36.2)	228 (43.2)		159 (41.1)	200 (49.3)		23 (20.0)	28 (23.2)	
A/A	62 (12.4)	51 (9.7)		56 (14.5)	46 (11.3)		6 (5.2)	5 (4.1)	

<sup>a</sup>*CYP1A1* primers: 5'-GGCAAGCGGAAGTGTATCG-3' and 5'-CAGGATAGCCAGGAAGAGAAAGAC-3'. *CYP1A1* probe sequences: 5'-VIC-TGAGACCATGCCCCGC-MGB-3' and 5'-FAM-AGACCGTTGCCCGC-MGB-3'.

<sup>b</sup>*CYP1B1* primers: 5'-CACTGCCAACACCTCTGTCTTG-3' and 5'-AAGAATCGAGCTGGATCAAAGTTC-3'. *CYP1B1* probe sequences: 5'-6FAM-ATGACCCACTGAAGTG-MGBNFQ-3' and 5'-VIC-ATGACCCAGTGAAGTG-MGBNFQ-3'.

<sup>c</sup>*CYP17* primers: 5'-CACGAGCTCCACATGGT-3' and 5'-CCTCCTTGTGCCCTAGAGTTG-3'. *CYP17* probe sequences: 5'-VIC-ATAGACAGCAGTGGAGTAG-MGB-3' and 5'-FAM-ACAGCGTGGAGTAG-MGB-3'.

<sup>d</sup>Individuals are recorded as '10+' if they have at least one set of repeats that is ≥10 repeats in length and <10 if both sets of repeats are <10 repeats.

<sup>e</sup>*COMT* primers: 5'-CGACTGTGCCGCCATCA-3' and 5'-AACGGGTCAGGCATGCA-3'. *COMT* probe sequences: 5'-FAM-CTTGTCCTTCACGCCAGCGAAAT-TAMRA-3' and 5'-VIC-CCTTGTCTTCATGCCAGCGAA-TAMRA-3'.

<sup>f</sup>*XRCC1* primers: 5'-GTGGGTGCTGGACTGTCA-3' and 5'-AACGGGTCAGGCATGCA-3'. *XRCC1* probe sequences: 5'-VIC-CCCTCCCAGAGGTAA-MGB-3' and 5'-FAM-CCTCCCGAGGTAA-MGB-3'.

interactions between the traditional risk factors for lung cancer risk and the genetic risk factors were identified.

## Discussion

The associations between the SNPs involved in tobacco carcinogen metabolism and lung cancer have been examined in numerous studies with variable results, as we have recently reviewed (37). There is evidence from studies of human lung tissue suggesting that the pathways involved in carcinogen metabolism and sex steroid synthesis interact with *GSTT1*, *CYP1B1* and *NQO1* correlated with plasma estradiol levels or ER expression (24). *NQO1* expression was also

correlated with hormonal factors in human lung tumor tissue. Further work by this group suggests that ER-α regulates *CYP1A1* protein expression when exposed to cigarette smoke extract in normal human bronchial epithelial cells and increases both basal expression of *CYP1B1* at the messenger RNA level and protein level (38). The modest ORs and 95% CIs estimated for the association between lung cancer and polymorphisms in *GSTM1*, *GSTT1*, *GSTP1*, *NQO1*, *CYP1A1* and *XRCC1*, both individually and combined, in our study population of women with NSCLC are similar to the ranges reported in the literature. The genes involved in sex steroid synthesis (*CYP19A1*, *COMT*, *CYP17* and *CYP1B1*) have been less well studied in relation to lung cancer risk and are described in more detail in the

**Table III.** Risk of NSCLC associated with polymorphisms in candidate genes, by cigarette smoking, hormone use and race

	Total <sup>a</sup>		White women <sup>b</sup>		White ever smokers <sup>b</sup>		White non-smokers <sup>c</sup>		White ever HRT <sup>b</sup>		White never HRT <sup>b</sup>	
	OR <i>n</i> = 504 cases, <i>n</i> = 527 controls	95% CI	OR <i>n</i> = 389 cases, <i>n</i> = 406 controls	95% CI	OR <i>n</i> = 351 cases, <i>n</i> = 196 controls	95% CI	OR <i>n</i> = 31 cases, <i>n</i> = 199 controls	95% CI	OR <i>n</i> = 233 cases, <i>n</i> = 269 controls	95% CI	OR <i>n</i> = 270 cases, <i>n</i> = 258 controls	95% CI
<i>CYP1A1 MspI</i>												
T/T	1.00		1.00		1.00		1.00		1.00		1.00	
T/C and C/C	0.95	0.66–1.36	0.78	0.50–1.22	0.78	0.48–1.27	0.54	0.15–1.89	0.96	0.53–1.16	0.59	0.30–1.16
<i>CYP1A1 Ile<sup>462</sup>Val</i>												
A/A	1.00		1.00		1.00		1.00		1.00		1.00	
A/G and G/G	0.77	0.42–1.41	0.57	0.29–1.14	0.58	0.27–1.23	0.78	0.17–3.63	0.83	0.33–2.12	0.36	0.13–1.04
<i>CYP1B1</i>												
C/C	1.00		1.00		1.00		1.00		1.00		1.00	
C/G and G/G	1.26	0.88–1.81	1.15	0.79–1.68	1.19	0.77–1.82	1.06	0.47–2.39	1.22	0.74–2.04	1.05	0.59–1.86
<i>CYP17</i>												
A1/A1	1.00		1.00		1.00		1.00		1.00		1.00	
A1/A2 and A2/A2	1.21	0.88–1.65	1.16	0.81–1.67	1.11	0.73–1.68	1.52	0.67–3.46	1.41	0.86–2.30	0.83	0.47–1.46
<i>CYP19A1</i>												
<10	1.00		1.00		1.00		1.00		1.00		1.00	
10+	0.93	0.67–1.30	1.02	0.70–1.49	1.06	0.69–1.64	0.81	0.37–1.80	0.76	0.46–1.27	1.45	0.82–2.57
<i>NQO1</i>												
C/C	1.00		1.00		1.00		1.00		1.00		1.00	
C/T and T/T	0.93	0.67–1.28	0.94	0.65–1.38	0.93	0.61–1.43	0.97	0.42–2.25	0.72	0.43–1.20	1.25	0.71–2.20
<i>GSTM1</i>												
Present	1.00		1.00		1.00		1.00		1.00		1.00	
Null	1.40	1.02–1.91	1.28	0.90–1.83	1.38	0.92–2.07	0.97	0.46–2.08	1.20	0.74–1.94	1.45	0.85–2.48
<i>GSTT1</i>												
Present	1.00		1.00		1.00		1.00		1.00		1.00	
Null	1.23	0.85–1.78	1.20	0.77–1.85	1.51	0.91–2.53	0.47	0.15–1.46	1.45	0.82–2.54	0.92	0.45–1.87
<i>GSTP1</i>												
G/G	1.00		1.00		1.00		1.00		1.00		1.00	
A/G and A/A	1.59	1.02–2.47	1.85	1.04–3.27	1.69	0.90–3.16	4.75	0.62–36.57	1.50	0.69–3.28	2.21	0.95–5.16
<i>COMT</i>												
G/G	1.00		1.00		1.00		1.00		1.00		1.00	
A/G and A/A	1.10	0.78–1.56	1.13	0.73–1.74	1.22	0.74–2.01	0.79	0.34–1.86	1.51	0.85–2.68	0.80	0.41–1.58
<i>XRCC1</i>												
A/G and G/G	1.00		1.00		1.00		1.00		1.00		1.00	
A/A	1.62	1.01–2.60	1.68	1.01–2.79	2.11	1.13–3.94	0.92	0.29–2.90	1.70	0.82–3.55	1.57	0.76–3.26
	Black women <sup>b</sup>		Black ever smokers <sup>b</sup>		Black non-smokers <sup>c</sup>		Black ever HRT <sup>b</sup>		Black never HRT <sup>b</sup>			
	OR <i>n</i> = 115 cases, <i>n</i> = 121 controls	95% CI	OR <i>n</i> = 102 cases, <i>n</i> = 63 controls	95% CI	OR <i>n</i> = 8 cases, <i>n</i> = 53 controls	95% CI	OR <i>n</i> = 30 cases, <i>n</i> = 40 controls	95% CI	OR <i>n</i> = 85 cases, <i>n</i> = 81 controls	95% CI		
<i>CYP1A1 MspI</i>												
T/T			1.00		1.00		1.00		1.00		1.00	
T/C and C/C			1.48	0.78–2.80	1.42	0.68–2.93	1.01	0.18–5.85	0.62	0.18–2.18	1.99	0.88–4.48
<i>CYP1A1 Ile<sup>462</sup>Val</i>												
A/A			1.00		1.00		1.00 <sup>d</sup>		1.00		1.00	
A/G and G/G			2.95	0.63–13.77	2.80	0.55–14.27	<sup>d</sup>		2.88	0.27–30.64	1.64	0.21–12.58
<i>CYP1B1</i>												
C/C			1.00		1.00		1.00 <sup>d</sup>		1.00 <sup>d</sup>		1.00	
C/G and G/G			8.29	0.95–72.28	11.09	1.18–104.25	<sup>d</sup>		<sup>d</sup>		5.95	0.61–58.46
<i>CYP17</i>												
A1/A1			1.00		1.00		1.00		1.00		1.00	
A1/A2 and A2/A2			1.28	0.69–2.35	1.00	0.50–2.03	2.30	0.40–13.25	2.81	0.79–10.05	1.07	0.50–2.29
<i>CYP19A1</i>												
<10			1.00		1.00		1.00		1.00		1.00	
10+			0.73	0.36–1.46	0.62	0.28–1.41	1.40	0.29–6.83	0.87	0.24–3.12	0.82	0.34–1.99
<i>NQO1</i>												
C/C			1.00		1.00		1.00		1.00		1.00	
C/T and T/T			0.90	0.47–1.72	0.94	0.44–2.00	1.36	0.25–7.52	0.49	0.14–1.67	1.29	0.55–3.01
<i>GSTM1</i>												
Present			1.00		1.00		1.00		1.00		1.00	
Null			1.80	0.92–3.51	1.82	0.81–4.08	2.53	0.55–11.69	1.40	0.39–5.01	1.60	0.69–3.71
<i>GSTT1</i>												
Present			1.00		1.00		1.00		1.00		1.00	
Null			1.31	0.65–2.65	1.84	0.79–4.28	0.30	0.03–3.24	2.51	0.62–10.19	1.16	0.48–2.79
<i>GSTP1</i>												
G/G			1.00		1.00		1.00		1.00		1.00	
A/G and A/A			1.25	0.62–2.54	1.17	0.51–2.66	2.28	0.25–20.55	1.18	0.31–4.53	1.45	0.59–3.56
<i>COMT</i>												
G/G			1.00		1.00		1.00		1.00		1.00	
A/G and A/A			1.07	0.59–1.95	1.13	0.56–2.28	1.02	0.22–4.63	2.00	0.62–6.49	0.83	0.39–1.77
<i>XRCC1</i>												
A/G and G/G			1.00		1.00		1.00 <sup>d</sup>		1.00		1.00	
A/A			1.40	0.35–5.58	4.52	0.48–42.99	<sup>d</sup>		2.24	0.20–24.54	0.92	0.15–5.76

<sup>a</sup>Adjusted for race, BMI, age, family history of lung cancer, pack-years of smoking and obstructive lung disease.<sup>b</sup>Adjusted for BMI, age, family history of lung cancer, pack-years of smoking and obstructive lung disease.<sup>c</sup>Adjusted for BMI, age and family history of lung cancer.<sup>d</sup>Unable to calculate due to sample size.

Table IV. Risk of lung cancer and GST combinations in white ever-smoking women

Single gene	OR <sup>a</sup>	95% CI	Combinations	<i>GSTM1</i>		<i>GSTT1</i>		<i>GSTP1</i>		<i>XRCC1</i>	
				OR <sup>a</sup>	95% CI	OR <sup>a</sup>	95% CI	OR <sup>a</sup>	95% CI	OR <sup>a</sup>	95% CI
<i>CYP1A1 MspI</i>			<i>CYP1A1 MspI</i>								
T/T	1.00		0	1.00		1.00		1.00		1.00	
T/C and C/C	0.78	0.48–1.27	1	1.22	0.79–1.89	1.00	0.64–1.54	1.44	0.72–2.91	1.21	0.78–1.89
			2	1.00	0.51–1.96	1.25	0.47–3.32	1.22	0.53–2.79	1.18	0.31–4.51
<i>CYP1A1 Ile<sup>462</sup>Val</i>			<i>CYP1A1 Ile<sup>462</sup>Val</i>								
A/A	1.00		0	1.00		1.00		1.00		1.00	
A/G and G/G	0.58	0.27–1.23	1	1.31	0.86–1.99	1.18	0.73–1.91	1.77	0.90–3.48	1.44	0.84–2.47
			2	0.83	0.31–2.26	0.95	0.20–4.38	0.94	0.31–2.81	0.84	0.16–4.34
<i>CYP1B1</i>			<i>CYP1B1</i>								
C/C	1.00		0	1.00		1.00		1.00		1.00	
C/G and G/G	1.19	0.77–1.82	1	1.75	0.98–3.13	1.26	0.79–2.02	0.79	0.22–2.84	1.13	0.72–1.77
			2	1.74	0.96–3.15	1.72	0.86–3.43	1.03	0.28–3.87	3.39	1.35–8.53
<i>CYP17</i>			<i>CYP17</i>								
A1/A1	1.00		0	1.00		1.00		1.00		1.00	
A1/A2 and A2/A2	1.11	0.73–1.68	1	1.75	1.00–3.06	0.87	0.55–1.37	1.20	0.50–2.92	1.32	0.85–2.05
			2	1.83	1.02–3.30	2.06	1.02–4.19	1.46	0.62–3.44	1.97	0.88–4.42
<i>CYP19A1</i>			<i>CYP19A1</i>								
<10	1.00		0	1.00		1.00		1.00		1.00	
10+	1.06	0.69–1.64	1	0.97	0.53–1.78	0.95	0.59–1.53	1.85	0.69–4.92	1.14	0.72–1.80
			2	1.35	0.70–2.59	1.91	0.89–4.11	2.39	0.87–6.53	1.98	0.89–4.43
<i>NQO1</i>			<i>NQO1</i>								
C/C	1.00		0	1.00		1.00		1.00		1.00	
C/T and T/T	0.93	0.61–1.43	1	1.00	0.63–1.60	1.10	0.72–1.69	0.45	0.11–1.84	0.89	0.56–1.41
			2	1.33	0.71–2.48	1.35	0.58–3.13	1.26	0.55–2.91	2.36	0.89–6.30
<i>COMT</i>			<i>COMT</i>								
G/G	1.00		0	1.00		1.00		1.00		1.00	
A/G and A/A	1.22	0.74–2.01	1	2.28	1.07–4.84	1.04	0.59–1.82	1.51	0.43–5.33	1.17	0.69–1.97
			2	2.57	1.18–5.59	1.96	0.92–4.20	1.90	0.57–6.34	2.54	1.11–5.83
<i>XRCC1</i>			<i>XRCC1</i>								
A/G and G/G	1.00		0	1.00		1.00		1.00		1.00	
A/A	2.11	1.13–3.94	1	1.73	1.12–2.65	1.76	1.11–2.77	1.95	1.00–3.80		
			2	2.12	0.92–4.89	3.61	0.75–17.34	3.57	1.38–9.21		
<i>GSTP1</i>			<i>GSTP1</i>								
G/G	1.00		0	1.00		1.00					
A/G and A/A	1.69	0.90–3.16	1	1.79	0.70–4.57	1.57	0.75–3.26				
			2	2.17	0.85–5.54	2.56	1.02–6.42				
<i>GSTT1</i>			<i>GSTT1</i>								
Present	1.00		0	1.00							
Null	1.51	0.91–2.53	1	1.42	0.92–2.19						
			2	1.92	0.96–3.86						
<i>GSTM1</i>											
Present	1.00										
Null	1.38	0.92–2.07									

<sup>a</sup>Adjusted for BMI, age, family history of lung cancer, pack-years of smoking and obstructive lung disease.

following paragraphs. Our goal was to examine SNPs in a set of genes involved in one or both of these pathways to investigate whether they were related to development of NSCLC in women. In addition, the SNPs examined here were also considered in the risk model incorporating both traditional and genetic risk factors for lung cancer.

Genes involved in sex steroid synthesis

The *CYP19A1* gene codes for the aromatase enzyme, which converts testosterone to estradiol, and has a common tetranucleotide simple tandem repeat polymorphism (TTTA) in intron 4 (39). Longer repeats (which various studies define as between 7 and 10 repeats) have been associated with higher levels of circulating estrogen levels in older men (40) and women (41). Studies of hormone-related cancers report conflicting findings with regard to this polymorphism and cancer risk, with inconsistent findings reported for breast (42,43), prostate (44,45) and endometrial cancer (46) studies. It has been suggested that the associations identified may be due to linkage disequilibrium across this locus. We did not identify any association with carrying TTTA<sub>10+</sub> repeats and lung cancer risk, either alone or in combination with other SNPs.

*COMT* is involved in methylation of catechol estrogens and reduced activity of this enzyme may increase the accumulation of catecholestrogens, leading to oxidative DNA damage (47,48). In the *COMT* gene, a G→A transition at codon 158 results in the substitution of methionine for valine (47) and carriage of the Val/Met or Met/Met genotypes is thought to result in lower enzyme activity (48). An analysis of 105 SNPs in 31 genes was the first to report risk associated with *COMT* (Val<sup>158</sup>Met) and NSCLC in 365 Norwegian lung cancer cases and 413 smokers without lung cancer (49). Zienoldddy *et al.* (49) found that individuals who carried the Val/Met or Met/Met genotypes were at 1.69-fold increased risk of NSCLC compared with those with the Val/Val genotype, after adjusting for age, sex and pack-years of smoking (95% CI, 1.16–2.47). Our findings in the total study population (OR = 1.10) and in the white ever-smoking women (OR = 1.22) were similar in magnitude but did not reach statistical significance after adjustment for BMI, age, family history of lung cancer, pack-years of smoking and obstructive lung disease. When combined with carrying the null mutation in *GSTM1*, white women smokers with either one or both risk genotypes at *GSTM1* or *COMT* were at statistically significantly increased risk of lung cancer compared

**Table V.** Estimates of main effects for environmental and genetic risk factors associated with lung cancer risk, from the most parsimonious model, white ever-smoking women

	OR (95% CI)	P-value
Age at diagnosis/interview	0.99 (0.97–1.01)	0.34
Family history of lung cancer	2.56 (1.46–4.50)	0.0011
History of chronic obstructive lung disease	1.93 (1.19–3.12)	0.008
Pack-years of cigarette smoking	1.04 (1.03–1.05)	<0.0001
BMI	0.93 (0.90–0.97)	0.0003
<i>XRCC1</i> A/A genotype	2.10 (1.12–3.94)	0.02
<i>GSTM1</i> null	1.40 (0.93–2.12)	0.11
<i>COMT</i> A/G or G/G genotype	1.15 (0.69–1.92)	0.59

with individuals who carried *GSTM1* non-null and *COMT* GG genotypes (OR = 2.57; 95% CI, 1.18–5.59) after adjustment for BMI, age, family history of lung cancer, pack-years of smoking and obstructive lung disease. In combination with the *XRCC1* risk genotype, carrying the *COMT* A allele also resulted in 2.5-fold increased risk. These findings suggest that a combination of accumulated catechol estrogens and associated oxidative damage with impaired detoxification and/or DNA repair may lead to increased lung cancer risk.

*CYP17* is another potential candidate gene in hormone-associated cancers based on its role in steroid hormone biosynthesis. A single polymorphism (T→C, A1→A2) in the 5' untranslated region had been associated with higher levels of estradiol among carriers of the A2 allele (50), but more recent work suggests that circulating estrogens and androgens do not significantly differ by *CYP17* genotype at this locus (51). This polymorphism has not been shown to be associated with increased risk of prostate or endometrial cancers, either alone or in combination with other steroid hormone-related genes (44,52). An analysis of major single-gene effects of *CYP17* variants found no associations with breast or prostate cancer, although the combination of two SNPs in *CYP17* was mildly associated with high-grade prostate cancers (53). This SNP has not been previously studied in people with lung cancer. In our population of white women smokers, the *CYP17* A1/A2 or A2/A2 genotypes did not increase risk of lung cancer until combined with the null mutations of *GSTM1* or *GSTT1* compared with women with no risk genotypes at these loci. These findings provide additional support to the idea that increased estrogen biosynthesis in combination with reduced detoxification of metabolites might contribute to lung cancer risk in women.

*CYP1B1*, in addition to activating procarcinogens to reactive metabolites (i.e. in the metabolism of tobacco smoke), is also thought to mediate the hydroxylation of 17 $\beta$ -estradiol. The Val allele at *Leu432Val* involves a change from an aliphatic amino acid to a smaller aliphatic amino acid, enabling higher catalytic efficiency for the 4-hydroxylation of estradiol and may play a dual role in carcinogenesis depending on the substrate (54). We recently reported increased risk of lung cancer in non-smokers associated with this single SNP and also in combination with other phase II enzymes (32). We did replicate these findings in our current population, but only for black women smokers, and we did find a statistically significant increased risk of lung cancer among white smoking women who carried both risk genotypes for *CYP1B1* and *XRCC1* (OR = 3.39; 95% CI, 1.35–8.53). We are unable to disentangle the metabolic pathway (estrogen or tobacco metabolism) through which these genotypes are affecting lung cancer risk, but the findings suggest that the inability to repair DNA damage caused by the 4-hydroxestradiol may contribute to lung cancer risk.

#### Risk model of NSCLC in white ever-smoking women

We also presented a systematic model building procedure that incorporated traditional non-modifiable (i.e. age and family history of lung cancer), potentially modifiable (i.e. personal history of chronic obstructive lung disease, pack-years of smoking and BMI) and genetic risk factors. Given the differences in the allele frequencies reported here between black and white women and the potentially different

pathways by which never smokers and smokers develop lung cancer, our model building was limited to ever-smoking white women. This represented the largest subgroup in our study population. In addition, recent work by Etzel *et al.* (55) presents a risk prediction model in African-Americans and includes some of the participants from our study in the validation of their model. For these reasons, the model presented was limited to white ever-smoking women. The traditional risk factors, namely age, family history of lung cancer, personal history of lung disease, BMI and pack-years of smoking were more strongly associated with lung cancer risk than any of the SNPs tested; however, genotypes at *XRCC1*, *GSTM1* and *COMT* all contributed to the most parsimonious risk model. The model presented here is limited by both sample size and the number of SNPs examined, and should not be considered a predictive risk model. Instead, our findings highlight the importance of examining the associations identified between SNPs in genes in candidate pathways versus single genes and suggest that future lung cancer modeling efforts will benefit from the inclusion of genotyping results.

The need for an individual risk prediction model for lung cancer has been well described (56). Past work in this area has focused on biomarkers found in sputum samples from individuals suspected to be at increased risk due to personal smoking histories and airflow obstruction (57) or asbestos exposures (58). A more recently proposed model included smoking pack-year histories, past diagnoses of pneumonia or another cancer, family history of early/late onset lung cancer and occupational exposure to asbestos. It was estimated that approximately two-thirds of lung cancer cases in a 5 year period could be predicted through use of this model (59). The most detailed model to date provided separate risk estimates for never, former and current smokers and considered 10 different exposures, family history of smoking-related and other cancers and information about smoking initiation, cessation and pack-years (60). The discriminatory abilities of all of these models are similar and modest in predictive ability, but have the advantage of using questionnaire and/or clinical data that are relatively easy to obtain compared with genetic information. The addition of information from pathway-based candidate genes represents the next step in the development of better risk prediction models for lung cancer.

There are limitations to the findings presented here. First, the majority of the subjects in our study had lung adenocarcinomas, so our findings may not be representative of all NSCLCs. In addition, while our population included a substantial number of black women, we were still underpowered to examine gene–environment interactions in these women. Since many of the SNPs studied varied in frequency significantly by self-reported race, we chose to present our data stratified by race or limited to white women smokers (our largest subgroup) only. Thus, our ability to adequately assess risk in other subgroups, such as non-smokers, was limited. The ORs and CIs for single SNP and gene–gene combinations presented here were not adjusted for multiple comparisons and need to be validated in other populations. Lastly, only a limited number of candidate SNPs were genotyped and included in this analysis, which is not fully representative of the complexity in the tobacco or estrogen metabolism pathways.

The strengths of this study include the detailed, in-person data collected from each subject, including information on past medical and reproductive histories, over-the-counter and prescribed medication use, occupational exposures, BMI, family history of lung cancer and smoking history. These data are from a large, population-based sample of women and are probably representative of the women in their birth cohorts. The ability to integrate risk factors identified via questionnaires, clinical exam and through the use of data derived from genotyping is an additional strength of this study. Data from this study can be analyzed with other comparable data to better elucidate the gene–gene and gene–environment interactions which influence individual lung cancer risk and can provide further validation of existing models in under-studied populations (i.e. women, blacks and non-smokers). The continued study of the role of estrogen and estrogen metabolism as a contributor to lung cancer risk is also warranted.



## Funding

National Institutes of Health (R01-CA87895, N01-PC35145, P30CA22453).

## Acknowledgements

The authors thank Lynda Forbes, Yvonne Bush, Kelly Montgomery, Pat Campagna, and the staff of the Metropolitan Detroit Cancer Surveillance System for data collection and management.

*Conflict of Interest Statement:* None declared.

## References

- Ries, L.A.G. et al. (eds) (2008) *SEER Cancer Statistics Review, 1975–2005*. National Cancer Institute, Bethesda, MD.
- Zang, E.A. et al. (1996) Differences in lung cancer risk between men and women: examination of the evidence. *J. Natl Cancer Inst.*, **88**, 183–192.
- Ryberg, D. et al. (1994) Different susceptibility to smoking-induced DNA damage among male and female lung cancer patients. *Cancer Res.*, **54**, 5801–5803.
- Mollerup, S. et al. (1999) Sex differences in lung CYP1A1 expression and DNA adduct levels among lung cancer patients. *Cancer Res.*, **59**, 3317–3320.
- Kure, E.H. et al. (1996) p53 mutations in lung tumours: relationship to gender and lung DNA adduct levels. *Carcinogenesis*, **17**, 2201–2205.
- Tam, I.Y. et al. (2006) Distinct epidermal growth factor receptor and KRAS mutation patterns in non-small cell lung cancer patients with different tobacco exposure and clinicopathologic features. *Clin. Cancer Res.*, **12**, 1647–1653.
- Liu, Y. et al. (2005) Reproductive factors, hormone use and the risk of lung cancer among middle-aged never-smoking Japanese women: a large-scale population-based cohort study. *Int. J. Cancer*, **117**, 662–666.
- Taioli, E. et al. (1994) Re: endocrine factors and adenocarcinoma of the lung in women. *J. Natl Cancer Inst.*, **86**, 869–870.
- Kreuzer, M. et al. (2003) Hormonal factors and risk of lung cancer among women? *Int. J. Epidemiol.*, **32**, 263–271.
- Schabath, M.B. et al. (2004) Hormone replacement therapy and lung cancer risk: a case-control analysis. *Clin. Cancer Res.*, **10**, 113–123.
- Kabat, G.C. et al. (2007) Reproductive and hormonal factors and risk of lung cancer in women: a prospective cohort study. *Int. J. Cancer*, **120**, 2214–2220.
- Schwartz, A.G. et al. (2007) Reproductive factors, hormone use, estrogen receptor expression and risk of non small-cell lung cancer in women. *J. Clin. Oncol.*, **25**, 5785–5792.
- Beattie, C.W. et al. (1985) Steroid receptors in human lung cancer. *Cancer Res.*, **45**, 4206–4214.
- Cagle, P.T. et al. (1990) Estrogen and progesterone receptors in bronchogenic carcinoma. *Cancer Res.*, **50**, 6632–6635.
- Chaudhuri, P.K. et al. (1982) Steroid receptors in human lung cancer cytosols. *Cancer Lett.*, **16**, 327–332.
- Kaiser, U. et al. (1996) Steroid-hormone receptors in cell lines and tumor biopsies of human lung cancer. *Int. J. Cancer*, **67**, 357–364.
- Mollerup, S. et al. (2002) Expression of estrogen receptors alpha and beta in human lung tissue and cell lines. *Lung Cancer*, **37**, 153–159.
- Stabile, L.P. et al. (2002) Human non-small cell lung tumors and cells derived from normal lung express both estrogen receptor alpha and beta and show biological responses to estrogen. *Cancer Res.*, **62**, 2141–2150.
- Kawai, H. et al. (2005) Estrogen receptor alpha and beta are prognostic factors in non-small cell lung cancer. *Clin. Cancer Res.*, **11**, 5084–5089.
- Schwartz, A.G. et al. (2005) Nuclear estrogen receptor beta in lung cancer: expression and survival differences by sex. *Clin. Cancer Res.*, **11**, 7280–7287.
- Skov, B.G. et al. (2008) Oestrogen receptor beta over expression in males with non-small cell lung cancer is associated with better survival. *Lung Cancer*, **59**, 88–94.
- Wu, C.T. et al. (2005) The significance of estrogen receptor beta in 301 surgically treated non-small cell lung cancers. *J. Thorac. Cardiovasc. Surg.*, **130**, 979–986.
- Hung, R.J. et al. (2005) Genetic polymorphisms in the base excision repair pathway and cancer risk: a HuGE review. *Am. J. Epidemiol.*, **162**, 925–942.
- Spivack, S.D. et al. (2003) Phase I and II carcinogen metabolism gene expression in human lung tissue and tumors. *Clin. Cancer Res.*, **9**, 6002–6011.
- Ball, P. et al. (1980) Catecholestrogens (2- and 4-hydroxyestrogens): chemistry, biogenesis, metabolism, occurrence and physiological significance. *Acta Endocrinol. Suppl. (Copenh.)*, **232**, 1–127.
- Colilla, S. et al. (2005) Association of catechol-O-methyltransferase with smoking cessation in two independent studies of women. *Pharmacogenet. Genomics*, **15**, 393–398.
- Martucci, C.P. et al. (1993) P450 enzymes of estrogen metabolism. *Pharmacol. Ther.*, **57**, 237–257.
- Feigelson, H.S. et al. (1998) Cytochrome P450c17alpha gene (CYP17) polymorphism is associated with serum estrogen and progesterone concentrations. *Cancer Res.*, **58**, 585–587.
- Picado-Leonard, J. et al. (1987) Cloning and sequence of the human gene for P450c17 (steroid 17 alpha-hydroxylase/17,20 lyase): similarity with the gene for P450c21. *DNA*, **6**, 439–448.
- Means, G.D. et al. (1989) Structural analysis of the gene encoding human aromatase cytochrome P-450, the enzyme responsible for estrogen biosynthesis. *J. Biol. Chem.*, **264**, 19385–19391.
- Drakoulis, N. et al. (1994) Polymorphisms in the human CYP1A1 gene as susceptibility factors for lung cancer: exon-7 mutation (4889 A to G), and a T to C mutation in the 3'-flanking region. *Clin. Investig.*, **72**, 240–248.
- Wenzlaff, A.S. et al. (2005) CYP1A1 and CYP1B1 polymorphisms and risk of lung cancer among never smokers: a population-based study. *Carcinogenesis*, **26**, 2207–2212.
- Bock, C.H. et al. (2005) NQO1 T allele associated with decreased risk of later age at diagnosis lung cancer among never smokers: results from a population-based study. *Carcinogenesis*, **26**, 381–386.
- Wenzlaff, A.S. et al. (2005) GSTM1, GSTT1 and GSTP1 polymorphisms, environmental tobacco smoke exposure and risk of lung cancer among never smokers: a population-based study. *Carcinogenesis*, **26**, 395–401.
- Benhamou, S. et al. (2002) Meta- and pooled analyses of the effects of glutathione S-transferase M1 polymorphisms and smoking on lung cancer risk. *Carcinogenesis*, **23**, 1343–1350.
- Raimondi, S. et al. (2006) Meta- and pooled analysis of GSTT1 and lung cancer: a HuGE-GSEC review. *Am. J. Epidemiol.*, **164**, 1027–1042.
- Schwartz, A.G. et al. (2007) The molecular epidemiology of lung cancer. *Carcinogenesis*, **28**, 507–518.
- Han, W. et al. (2005) Estrogen receptor alpha increases basal and cigarette smoke extract-induced expression of CYP1A1 and CYP1B1, but not GSTP1, in normal human bronchial epithelial cells. *Mol. Carcinog.*, **44**, 202–211.
- Polymeropoulos, M.H. et al. (1991) Tetranucleotide repeat polymorphism at the human aromatase cytochrome P-450 gene (CYP19). *Nucleic Acids Res.*, **19**, 195.
- Gennari, L. et al. (2004) A polymorphic CYP19 TTTA repeat influences aromatase activity and estrogen levels in elderly men: effects on bone metabolism. *J. Clin. Endocrinol. Metab.*, **89**, 2803–2810.
- Dick, I.M. et al. (2005) Association of an aromatase TTTA repeat polymorphism with circulating estrogen, bone structure, and biochemistry in older women. *Am. J. Physiol. Endocrinol. Metab.*, **288**, E989–E995.
- Baxter, S.W. et al. (2001) Polymorphic variation in CYP19 and the risk of breast cancer. *Carcinogenesis*, **22**, 347–349.
- Healey, C.S. et al. (2000) Polymorphisms in the human aromatase cytochrome P450 gene (CYP19) and breast cancer risk. *Carcinogenesis*, **21**, 189–193.
- Cussenot, O. et al. (2007) Combination of polymorphisms from genes related to estrogen metabolism and risk of prostate cancers: the hidden face of estrogens. *J. Clin. Oncol.*, **25**, 3596–3602.
- Li, L. et al. (2004) No association between a tetranucleotide repeat polymorphism of CYP19 and prostate cancer. *Cancer Epidemiol. Biomarkers Prev.*, **13**, 2280–2281.
- Paynter, R.A. et al. (2005) CYP19 (aromatase) haplotypes and endometrial cancer risk. *Int. J. Cancer*, **116**, 267–274.
- Lachman, H.M. et al. (1996) Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. *Pharmacogenetics*, **6**, 243–250.
- Palmatier, M.A. et al. (1999) Global variation in the frequencies of functionally different catechol-O-methyltransferase alleles. *Biol. Psychiatry*, **46**, 557–567.
- Zienolddiny, S. et al. (2008) A comprehensive analysis of phase I and phase II metabolism gene polymorphisms and risk of non-small cell lung cancer in smokers. *Carcinogenesis*, **29**, 1164–1169.
- Carey, A.H. et al. (1994) Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. *Hum. Mol. Genet.*, **3**, 1873–1876.

51. Olson, S.H. *et al.* (2007) Variants in estrogen biosynthesis genes, sex steroid hormone levels, and endometrial cancer: a HuGE review. *Am. J. Epidemiol.*, **165**, 235–245.
52. Gaudet, M.M. *et al.* (2008) Genetic variation in CYP17 and endometrial cancer risk. *Hum. Genet.*, **123**, 155–162.
53. Setiawan, V.W. *et al.* (2007) CYP17 genetic variation and risk of breast and prostate cancer from the National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3). *Cancer Epidemiol. Biomarkers Prev.*, **16**, 2237–2246.
54. Hanna, I.H. *et al.* (2000) Cytochrome P450 1B1 (CYP1B1) pharmacogenetics: association of polymorphisms with functional differences in estrogen hydroxylation activity. *Cancer Res.*, **60**, 3440–3444.
55. Etzel, C.J. *et al.* (2008) Development and validation of a lung cancer risk prediction model for African-Americans. *Cancer Prev. Res.*, **1**, 255–265.
56. Cassidy, A. *et al.* (2007) Lung cancer risk prediction: a tool for early detection. *Int. J. Cancer*, **120**, 1–6.
57. Prindiville, S.A. *et al.* (2003) Sputum cytological atypia as a predictor of incident lung cancer in a cohort of heavy smokers with airflow obstruction. *Cancer Epidemiol. Biomarkers Prev.*, **12**, 987–993.
58. Bach, P.B. *et al.* (2003) Variations in lung cancer risk among smokers. *J. Natl Cancer Inst.*, **95**, 470–478.
59. Cassidy, A. *et al.* (2008) The LLP risk model: an individual risk prediction model for lung cancer. *Br. J. Cancer*, **98**, 270–276.
60. Spitz, M.R. *et al.* (2007) A risk model for prediction of lung cancer. *J. Natl Cancer Inst.*, **99**, 715–726.

*Received August 20, 2008; revised January 20, 2009;  
accepted January 20, 2009*