Original Contribution

Genetic Variants on 15q25.1, Smoking, and Lung Cancer: An Assessment of Mediation and Interaction

Tyler J. VanderWeele*, Kofi Asomaning, Eric J. Tchetgen Tchetgen, Younghun Han, Margaret R. Spitz, Sanjay Shete, Xifeng Wu, Valerie Gaborieau, Ying Wang, John McLaughlin, Rayjean J. Hung, Paul Brennan, Christopher I. Amos, David C. Christiani, and Xihong Lin

* Correspondence to Dr. Tyler J. VanderWeele, Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115 (e-mail: tvanderw@hsph.harvard.edu).

Genome-wide association studies have identified variants on chromosome 15q25.1 that increase the risks of both lung cancer and nicotine dependence and associated smoking behavior. However, there remains debate as to whether the association with lung cancer is direct or is mediated by pathways related to smoking behavior. Here, the authors apply a novel method for mediation analysis, allowing for gene-environment interaction, to a lung cancer case-control study (1992–2004) conducted at Massachusetts General Hospital using 2 single nucleotide polymorphisms, rs8034191 and rs1051730, on 15q25.1. The results are validated using data from 3 other lung cancer studies. Tests for additive interaction ($P = 2 \times 10^{-10}$ and $P = 1 \times 10^{-9}$) and multiplicative interaction ($P = 0.01$ and $P = 0.01$) were significant. Pooled analyses yielded a direct-effect odds ratio of 1.26 (95% confidence interval (CI): 1.19, 1.33; $P = 2 \times 10^{-15}$) for rs8034191 and an indirect-effect odds ratio of 1.01 (95% CI: 1.00, 1.01; $P = 0.09$); the proportion of increased risk mediated by smoking was 3.2%. For rs1051730, direct- and indirect-effect odds ratios were 1.26 (95% CI: 1.19, 1.33; $P = 1 \times 10^{-15}$) and 1.00 (95% CI: 0.99, 1.01; $P = 0.22$), respectively, with a proportion mediated of 2.3%. Adjustment for measurement error in smoking behavior allowing up to 75% measurement error increased the proportions mediated to 12.5% and 9.2%, respectively. These analyses indicate that the association of the variants with lung cancer operates primarily through other pathways.

gene-environment interaction; lung neoplasms; mediation; pathway analysis; smoking

Abbreviations: CI, confidence interval; MGH, Massachusetts General Hospital; SNP, single nucleotide polymorphism.

Three genome-wide association studies (1–3) have found associations between genetic variants on chromosome 15q25.1 and lung cancer. These variants are known to be associated with smoking behavior as well (3–9), raising the question of whether the association of the variants with lung cancer operates primarily through smoking or through other pathways (1–5, 10–14). In addition to possible effects of genetic variants on 15q25.1 on lung cancer risk either through smoking or independent of smoking, Thorgeirsson et al. (3, 13) noted a third possible explanation for the associations: that the variant may increase individuals' vulnerability to the harmful effect of tobacco smoke, a form of gene-environment interaction. Prior studies attempting to discriminate between these possibilities have been limited by lack of adequate methods

to accommodate interaction (5, 12, 14) in assessing direct and indirect effects and inadequate handling of case-control data (14). Traditional mediation methods do not allow for interaction between the effects of the exposure (the genetic variant) and the effects of the mediator (smoking), though such interaction would be present if the variant increased vulnerability to the effect of smoking (3, 13).

The method (15) we demonstrate here overcomes these limitations and is applied to a case-control lung cancer study of 1,836 cases and 1,452 controls (16) conducted at Massachusetts General Hospital (MGH). Allowing for such gene-environment interaction in estimating direct and indirect effects may be important, since prior literature has noted the possibility that the variant increases vulnerability to the

effects of smoking on lung cancer (13), and there is evidence that carriers of the variant allele extract more nicotine and toxins from each cigarette (17). Analyses using the same method were also applied to 3 other genome-wide case-control studies of lung cancer (1, 2) to replicate results.

MATERIALS AND METHODS

We drew 1,836 cases and 1,452 controls from a case-control study assessing the molecular epidemiology of lung cancer, which was conducted at MGH from 1992 to 2004 and is described in detail elsewhere (16). Briefly, eligible cases included any person over the age of 18 years with a diagnosis of primary lung cancer that was further confirmed by an MGH lung pathologist. The controls were recruited from the friends or spouses of cancer patients or the friends or spouses of other surgery patients in the same hospital. Potential controls that carried a previous diagnosis of any cancer (other than nonmelanoma skin cancer) were excluded from participation. Interviewer-administered questionnaires collected information on sociodemographic variables, including age (years; continuous), sex, educational history (college degree or more; yes/no), smoking intensity (cigarettes/day), and duration of smoking (years), from each subject. The study was reviewed and approved by the institutional review boards of MGH and the Harvard School of Public Health.

To confirm the findings, we replicated the analyses in 3 additional case-control studies of lung cancer: a study conducted at the University of Texas M. D. Anderson Cancer Center (2) with 2,827 cases and 2,345 controls (1995–2006); a central European study conducted by the International Agency for Research on Cancer (1) with 1,871 cases and 2,472 controls (1998–2002); and a study conducted in Toronto, Ontario, Canada (1), with 333 cases and 501 controls (1997–2002). All of the analyses were limited to Caucasians.

We selected the 2 single nucleotide polymorphisms (SNPs) rs8034191 and rs1051730, based on published reports that have shown them to have the most consistent statistically significant associations with smoking behavior. The association region of 15q25.1 is one of high linkage disequilibrium, and the 2 SNPs are highly correlated with other significant SNPs in the area and fairly representative of the region. In the MGH study, peripheral blood samples were obtained from all study participants at the time of enrollment. DNA was extracted from peripheral blood samples using the Puregene DNA Isolation Kit (Gentra Systems, Minneapolis, Minnesota). The polymorphisms in the MGH study were genotyped using a 5'-nuclease assay (TaqMan) and the ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Foster City, California). Genotyping was performed by laboratory personnel blinded to clinical variables and case-control status, and analysis of a randomly selected 5% of the samples was repeated to validate genotyping procedures. Blinded genotyping results were independently reviewed by 2 of the authors. To check for genotyping error, departures from Hardy-Weinberg equilibrium in controls were examined.

Genotype data from the M. D. Anderson study were obtained with Illumina HumanHap 300 BeadChips for 1,154 cases and 1,137 controls at the Johns Hopkins Center for Inherited Disease Research (Baltimore, Maryland), and the

remaining cases and controls were genotyped using TaqMan. The International Agency for Research on Cancer's central Europe study (1) was also genotyped with the HumanHap300 BeadChip using the Illumina Infinium platform, and genotyping was conducted at the Centre National Genotypage (Paris, France). Genotype data for the Toronto study were obtained by genotyping cases and controls with the HumanHap300 BeadChip with the Illumina Infinium platform at McGill University and the Genome Quebec Innovation Centre (Montreal, Canada). Further details on genotyping in these studies can be found elsewhere (1, 2).

Number of cigarettes smoked per day was used as a measure of smoking intensity and has been shown to be a good marker for nicotine dependence (18–20). Linear regression was used for models of smoking intensity, measured as the square root of cigarettes per day so as to better approximate a linear fit. Analyses using total cigarettes per day gave similar results. Logistic regression was used to model lung cancer status both with and without a smoking × variant interaction term. A multiplicative model was used for the number of risk alleles throughout. Covariates included in the models were genotype, age, sex, college education, and smoking duration; models for lung cancer also included smoking intensity (square root of cigarettes per day). Analyses which omitted smoking duration as a covariate gave qualitatively similar conclusions.

The linear regression for smoking intensity was weighted; cases were weighted by the prevalence of lung cancer divided by the proportion of cases, and controls were weighted by 1 minus the prevalence divided by the proportion of controls in the study (15, 21). Weights were further adjusted by sampling fractions in studies (M. D. Anderson and Toronto) in which sampling fractions varied by smoking status (22). The weighting takes into account the fact that in the case-control study design, cases were selected by lung cancer status, not by smoking (15, 21). The weighted regression corresponds to the associations that would be observed in a cohort study of the same population. Robust standard errors were used to account for weighting and possible nonnormality. When sampling fractions of cases and controls varied by smoking status (M. D. Anderson and Toronto)—for example, oversampling of controls to match the cases according to smoking behavior—an offset term was used in the logistic regression to account for sampling design (22).

The regression for smoking intensity and the regression for lung cancer risk were combined to obtain direct and indirect effects using odds ratios for mediation analysis for a dichotomous outcome (15). The direct effect can be interpreted as the odds ratio for lung cancer among persons with the genetic variant versus those without the variant if smoking behavior were what it would have been without the variant. The indirect effect can be interpreted as the odds ratio for lung cancer for those with the genetic variant present comparing the risk if smoking behavior were what it would have been with versus without the genetic variant. Direct and indirect effects are averaged over all individuals (smokers and nonsmokers) and are also evaluated at the mean population level of the covariates. The proportion mediated is reported on the risk difference scale (15) and is obtained by $OR^d \times (OR^i - 1)/(OR^d \times OR^i - 1)$, where OR^d is the direct-effect odds ratio and OR^i is the indirect-effect odds ratio.

Analyses assume that conditional on the covariates, there is no confounding of 1) the exposure-outcome relation, 2) the mediator-outcome relation, or 3) the exposure-mediator relation and that 4) there is no effect of the exposure that itself confounds the mediator-outcome relation (15). No confounding of the effect of the exposure on the mediator and on the outcome (assumptions 1 and 3), when the exposure is a genetic variant with analysis restricted to a single ethnic group, is likely to hold approximately and is generally assumed in genetic studies. The robustness of results to the confounding assumptions can be examined through sensitivity analysis techniques (23).

P values for interaction on the additive scale are reported using the relative excess risk due to interaction (24–26); P values for multiplicative interaction were obtained by means of a Wald test of the interaction coefficient in the logistic regression. Measures correspond to a 1-allele change in the genetic variants and to a 1-unit change in the cigarette measure. The measure of additive interaction assesses the extent to which the odds ratio for a 1-unit increase in both exposures exceeds the sum of the odds ratios for a 1-unit increase in each exposure considered separately. The measure of multiplicative interaction assesses the log of the ratio of the odds ratios for a 1-unit increase in both exposures relative to the product of the odds ratios for a 1-unit increase in each exposure considered separately. Analyses for measurement error are conducted using bias analysis (27). Results from the 4 studies were combined on the log-odds scale using sample-size-based meta-analysis. Analyses were implemented with SAS 9.2 (SAS Institute, Inc., Cary, North Carolina) and R 2.4 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Table 1 summarizes the demographic characteristics of the 4 studies used in the analysis. Models for lung cancer and for smoking intensity (cigarettes per day) can be combined to calculate indirect effects mediated by smoking and direct effects through other pathways (15). Table 2 shows direct and indirect effects for rs8034191, along with tests for gene-by-smoking interaction from the 4 studies and a pooled meta-analysis (full details are available in Web Table 1 (http://aje.oxfordjournals.org/)). Analyses from the MGH study indicated strong evidence for a direct effect and suggested that the indirect effect is small. Ignoring possible gene-environment interaction gave a direct-effect odds ratio of 1.35 (95% confidence interval (CI): 1.21, 1.52; $P = 3 \times 10^{-7}$) and an indirect-effect odds ratio of 1.01 (95% CI: 0.99, 1.02; $P = 0.15$) per rs8034191 C allele, with 3.6% of the increased risk being mediated by smoking. Tests for interaction were significant on the additive risk scale ($P = 1 \times 10^{-3}$), indicating that the effect of smoking varied by genotype, with weaker evidence on the multiplicative scale ($P = 0.17$). Allowing for smoking-by-gene interaction gives, for changes from 0 to 1 C allele and 0 to 2 C alleles, respectively, direct-effect odds ratios of 1.31 (95% CI: 1.15, 1.49; $P = 3 \times 10^{-5}$) and 1.72 (95% CI: 1.34, 2.21; $P = 2 \times 10^{-5}$) and indirect-effect odds ratios of 1.01 (95% CI: 0.99, 1.02; $P = 0.15$) and 1.03

(95% CI: 0.99, 1.07; $P = 0.16$), with 6.3% of the increased risk being mediated by smoking intensity in the latter scenario. The confidence interval for the indirect effect is relatively narrow; the mediated effect is a relatively small portion.

Table 3 shows results of similar analyses for rs1051730 (full details are available in Web Table 2). In the MGH study, ignoring possible interaction gives a direct-effect odds ratio of 1.32 (95% CI: 1.18, 1.48; $P = 2 \times 10^{-6}$) and an indirect-effect odds ratio of 1.01 (95% CI: 0.99, 1.02; $P = 0.19$) per A allele, with a proportion mediated of 3.5%. Tests for interaction between the effects of smoking and rs1051730 alleles indicated additive interaction ($P = 0.004$), with weaker evidence for multiplicative interaction ($P = 0.26$). Allowing for smoking-by-gene interaction gives direct-effect odds ratios of 1.29 (95% CI: 1.14, 1.46; $P = 8 \times 10^{-5}$) and 1.66 (95% CI: 1.30, 2.12; $P = 1 \times 10^{-8}$) and indirect-effect odds ratios of 1.01 (95% CI: 0.99, 1.02; $P = 0.22$) and 1.02 (95% CI: 0.99, 1.06; $P = 0.20$) for changes from 0 to 1 A allele and 0 to 2 A alleles, respectively, with 5.7% of the increased risk being mediated by smoking in the latter scenario.

As shown in Tables 2 and 3, all 3 replication studies exhibited patterns similar to those of the MGH study, and overall replication P values for the 3 studies again indicated significant direct effects and small indirect effects. For both rs8034191 and rs1051730, there was moderately strong evidence in the replication for interaction on both additive ($P = 6 \times 10^{-8}$; $P = 1 \times 10^{-7}$) and multiplicative ($P = 0.03$; $P = 0.03$) scales. Likewise, pooled estimates from all 4 studies, reported in the Abstract, indicated large, highly significant direct effects ($P = 2 \times 10^{-15}$; $P = 1 \times 10^{-15}$), indirect effect odds ratios close to 1 ($P = 0.09$; $P = 0.22$), and interaction on both additive ($P = 2 \times 10^{-10}$; $P = 1 \times 10^{-9}$) and multiplicative ($P = 0.01$; $P = 0.01$) scales.

We conducted further analyses using all 4 studies to allow for the possibility that the cigarettes/day measure was recorded with error and that the measure does not capture all of the relevant smoking behavior. Assuming that the cigarettes/day measure explains only 50% of the variability in the biologically relevant measure gives corrected direct- and indirect-effects odds ratios of 1.24 and 1.01, respectively, for rs8034191, with a proportion mediated of 6.3% and odds ratios of 1.25 and 1.01, respectively, for rs1051730, with a proportion mediated of 4.7%. Assuming that cigarettes/day explains only 25% of the variability in the biologically relevant measure (i.e., 75% measurement error) gives odds ratios of 1.23 and 1.03 for rs8034191 with a proportion mediated of 12.5% and odds ratios of 1.24 and 1.02 for rs1051730 with a proportion mediated of 9.2%. Measurement error may have attenuated estimates of the proportion mediated, but even in the more extreme scenario the majority of the effect appears to be direct.

DISCUSSION

The analyses here indicate that the associations of these genetic variants with lung cancer operate principally through pathways other than changing smoking intensity.

Table 1. Characteristics of 4 Study Populations Included in an Analysis of Direct and Indirect Effects of Genetic Variants on the Risk of Lung Cancer[a]

| | Massachusetts General Hospital (1992–2004)[b] | | | | | | M. D. Anderson Cancer Center (1995–2006)[c] | | | | | | International Agency for Research on Cancer (1998–2002)[d] | | | | | | Toronto, Ontario, Canada (1997–2002)[e] | | | | | |
| | Cases (n = 1,836) | | | Controls (n = 1,452) | | | Cases (n = 2,827) | | | Controls (n = 2,345) | | | Cases (n = 1,867) | | | Controls (n = 2,463) | | | Cases (n = 333) | | | Controls (n = 501) | | |
	No.	%	Mean	No.	%	Mean	No.	%	Mean	No.	%	Mean	No.	%	Mean	No.	%	Mean	No.	%	Mean	No.	%	Mean
Smoking status																								
Ever smoker	1,671			943			2,365			1,836			1,728			1,584			238			281		
Never smoker	165			509			462			509			139			879			95			220		
Smoking intensity																								
No. of cigarettes/day			25.42			13.97			23.51			8.46			18.06			10.83			20.65			9.75
Duration, years			38.5			18.93			30.91			10.50			35.74			20.38			34.44			12.01
Age, years			64.86			58.58			62.93			57.65			60.27			59.63			64.71			52.10
College education		31.3			33.5			18.9			31.1			14.5			23.2			23.9			55.4	
Sex																								
Male		50.1			56.1			52.5			59.4			78.3			73.8			52.6			37.6	
Female		49.9			43.9			47.5			40.6			21.7			27.2			47.4			62.4	
No. of rs8034191 C alleles																								
0		33.8			43.3			38.6			42.3			35.6			43.2			38.9			42.5	
1		48.5			43.7			46.4			46.8			47.8			45.4			44.7			46.5	
2		17.7			13.0			15.0			10.9			16.7			11.4			16.4			11.0	
No. of rs1051730 A alleles																								
0		34.9			43.4			38.6			42.2			36.1			43.4			36.9			42.5	
1		46.2			43.6			46.4			47.0			47.3			45.0			46.7			46.3	
2		18.9			13.0			15.0			10.8			16.6			11.6			16.4			11.1	

[a] All analyses were limited to Caucasians.
[b] Data were obtained from the study by Miller et al. (16).
[c] Data were obtained from the study by Amos et al. (2).
[d] Data were obtained from the study by Hung et al. (1).
[e] Data were obtained from the study by Hung et al. (1).

Table 2. Direct and Indirect Effects on Lung Cancer and Assessment of Variant-Smoking Interaction for rs8034191

| Analysis | Original Study (MGH) | | | Replication Studies | | | | | | | | | Meta-Analysis | | | |
| | | | | M. D. Anderson Cancer Center | | | IARC | | | Toronto, Ontario, Canada | | | | | | |
	OR	95% CI	PM, %	OR	95% CI	PM, %	OR	95% CI	PM, %	OR	95% CI	PM, %	Pooled OR	95% CI	PM, %	*P* Value
Without SNP × smoking interaction			3.6			6.8			0.2			0.4			3.2	
Direct effect	1.35	1.21, 1.52		1.18	1.08, 1.29		1.26	1.14, 1.40		1.33	1.02, 1.73		1.26	1.19, 1.33		1.8×10^{-15}
Indirect effect	1.01	0.99, 1.02		1.01	0.99, 1.03		1.00	0.99, 1.02		1.00	0.99, 1.01		1.00	1.00, 1.01		0.09
Total effect	1.36	1.22, 1.54		1.19	1.09, 1.30		1.26	1.14, 1.40		1.33	1.02, 1.74		1.26	1.19, 1.34		6.7×10^{-16}
With SNP × smoking interaction																
T/C vs. TT			4.2			13			0.3			0.7			4.2	
Direct effect	1.31	1.15, 1.49		1.08	0.97, 1.21		1.24	1.11, 1.38		1.30	0.99, 1.71		1.20	1.12, 1.28		5.6×10^{-8}
Indirect effect	1.01	0.99, 1.02		1.01	0.99, 1.03		1.00	0.99, 1.02		1.00	0.99, 1.01		1.01	1.00, 1.01		0.09
Total effect	1.32	1.16, 1.50		1.10	0.98, 1.23		1.24	1.11, 1.39		1.30	0.99, 1.71		1.21	1.13, 1.29		1.9×10^{-8}
C/C vs. TT			6.3			15.2			0.3			2.3			5.6	
Direct effect	1.72	1.34, 2.21		1.18	0.96, 1.47		1.53	1.23, 1.92		1.70	0.99, 2.93		1.44	1.27, 1.63		2.3×10^{-8}
Indirect effect	1.03	0.99, 1.07		1.03	0.99, 1.07		1.00	0.97, 1.04		1.00	0.97, 1.05		1.02	1.00, 1.04		0.07
Total effect	1.77	1.38, 2.26		1.22	0.98, 1.51		1.54	1.23, 1.92		1.72	1.00, 2.94		1.46	1.29, 1.66		4.4×10^{-9}
Interaction analyses																
Additive interaction	0.06	0.02, 0.10		0.05	0.03, 0.08		0.07	0.03, 0.10		0.04	−0.06, 0.15		0.06	0.04, 0.08		2.3×10^{-10}
Multiplicative interaction	1.04	0.98, 1.09		1.05	1.01, 1.09		1.03	0.96, 1.09		1.04	0.93, 1.15		1.04	1.01, 1.07		0.01

Abbreviations: CI, confidence interval; IARC, International Agency for Research on Cancer; MGH, Massachusetts General Hospital; OR, odds ratio; PM, proportion mediated; SNP, single nucleotide polymorphism.

Table 3. Direct and Indirect Effects on Lung Cancer and Assessment of Variant-Smoking Interaction for rs1051730

Analysis	Original Study (MGH)			Replication Studies									Meta-Analysis			
				M. D. Anderson Cancer Center			IARC			Toronto, Ontario, Canada						
	OR	95% CI	PM, %	OR	95% CI	PM, %	OR	95% CI	PM, %	OR	95% CI	PM, %	Pooled OR	95% CI	PM, %	*P* Value
Without SNP × smoking interaction			3.5			4.9			0			0.4			2.3	
Direct effect	1.32	1.18, 1.48		1.19	1.09, 1.30		1.27	1.14, 1.40		1.42	1.09, 1.86		1.26	1.19, 1.33		1.1×10^{-15}
Indirect effect	1.01	0.99, 1.02		1.01	0.99, 1.02		1.00	0.98, 1.01		1.00	0.99, 1.01		1.00	0.99, 1.01		0.22
Total effect	1.33	1.19, 1.50		1.20	1.10, 1.31		1.27	1.14, 1.40		1.42	1.09, 1.86		1.26	1.19, 1.34		4.4×10^{-16}
With SNP × smoking interaction																
A/G vs. GG			4.1			9.3			0			0.5			3	
Direct effect	1.29	1.14, 1.46		1.09	0.98, 1.22		1.24	1.10, 1.38		1.41	1.07, 1.87		1.20	1.12, 1.28		3.7×10^{-8}
Indirect effect	1.01	0.99, 1.02		1.01	0.99, 1.02		1.00	0.98, 1.01		1.00	0.99, 1.01		1.01	0.99, 1.01		0.22
Total effect	1.30	1.15, 1.47		1.10	0.98, 1.23		1.24	1.10, 1.38		1.42	1.07, 1.87		1.21	1.13, 1.29		1.2×10^{-8}
A/A vs. GG			5.7			11.1			0			0.7			4	
Direct effect	1.66	1.30, 2.12		1.20	0.97, 1.48		1.53	1.22, 1.91		2.00	1.15, 3.48		1.44	1.27, 1.64		1.4×10^{-8}
Indirect effect	1.02	0.99, 1.06		1.02	0.98, 1.06		1.00	0.96, 1.03		1.00	0.97, 1.04		1.01	0.99, 1.03		0.2
Total effect	1.70	1.33, 2.17		1.22	0.99, 1.51		1.53	1.22, 1.91		2.01	1.16, 3.47		1.46	1.29, 1.66		5.1×10^{-9}
Interaction analyses																
Additive interaction	0.05	0.02, 0.09		0.05	0.03, 0.08		0.07	0.04, 0.11		0.01	−0.10, 0.13		0.06	0.04, 0.08		1.4×10^{-9}
Multiplicative interaction	1.03	0.98, 1.08		1.05	1.01, 1.09		1.03	0.97, 1.10		1.01	0.90, 1.12		1.04	1.00, 1.07		0.01

Abbreviations: CI, confidence interval; IARC, International Agency for Research on Cancer; MGH, Massachusetts General Hospital; OR, odds ratio; PM, proportion mediated; SNP, single nucleotide polymorphism.

Although this may initially appear surprising, further evidence supports these results. In the studies conducted here and as verified in larger studies (6–9), the effect size of the variants on smoking is only approximately 1 cigarette per day, which may be of limited biologic relevance for lung cancer.

Our conclusion is further supported by the fact that recent studies have found no association between the variants and lung cancer among nonsmokers (28, 29). These studies, in conjunction with the results presented here, suggest that although the association is not principally mediated by changing smoking behavior, it occurs in the presence of smoking. The strong empirical evidence here of variant-by-smoking interaction, on both additive and multiplicative scales, provides further support for this. The direct association of the variants with lung cancer may operate only for smokers, even though the variants do not substantially increase smoking intensity itself.

The interpretation of direct-effect estimates is complicated in the presence of a smoking-by-variant interaction: The direct effect may vary by smoking status. There appears to be a direct association for smokers, but perhaps not for nonsmokers. The natural direct effects estimated in our analyses essentially average over the direct effects for smokers and nonsmokers. Importantly, however, the indirect effects were small in all of our analyses. The associations of the variants with lung cancer do not operate primarily through changing the number of cigarettes smoked per day.

Certain biologic hypotheses are consistent with the statistical evidence. As noted above, it may be that the variant serves to increase the amount of nicotine and toxins extracted from each cigarette (17); such an effect would only be observed for smokers. Such an effect would also be observed even if the variant did not operate primarily by changing the number of cigarettes smoked per day. Smoking (or nicotine) has an effect on the regulation or is involved in some downstream action (e.g., expression) of the genes for which these SNPs are markers (30). In addition, nicotine and the tobacco derivatives N'-nitrosonornicotine and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone have strong affinity for the nicotinic acetylcholine receptors that are also known to be present in lung tissue (31). Nicotine has been shown to be involved in lung carcinogenesis through activation of the nictonic acetylcholine receptors in nonneuronal cells (32). If the variant or any functional SNPs for which the variants are markers served to activate nicotinic receptors (33), such an effect would again only be observed for smokers.

The result that most of the association operates through pathways other than smoking seems reasonably robust to measurement error. It is possible that our estimates are inaccurate due to unmeasured confounders that affect both smoking and lung cancer. However, such factors (e.g., low socioeconomic status) would probably affect smoking and lung cancer in the same direction, and sensitivity analysis (23) indicates that this would likely lead to underestimation of the direct effect and overestimation of indirect effects. This would yet further support our conclusion that the vast majority of the association is direct, and so we have not explored such sensitivity analyses further. The method we employed for mediation (15) allows for potential variant-by-smoking interaction, as had been previously suggested

(3, 13), and correctly handles case-control data. These are advantages not shared by other mediation analysis approaches. Our conclusions appear to be on fairly solid ground.

Our results here are also of historical interest. Over 50 years ago, Fisher (34) suggested that there might be a genetic variant responsible for both smoking behavior and lung cancer. He proposed that this common genetic cause might explain the association between smoking and lung cancer and thus that smoking may not itself in fact have a causal effect on lung cancer. Our results here show that, in some respects, Fisher was at least slightly correct. In previous studies, the variants on chromosome 15q25.1 have been shown to affect smoking behavior (3–9); here we have provided fairly conclusive evidence that these variants also affect lung cancer through pathways other than by increasing smoking behavior. Thus, there is indeed a common genetic cause of smoking and lung cancer. Fisher was partially correct, but only partially. As Cornfield et al. (35) clearly demonstrated, in response to Fisher, using sensitivity analysis for a hypothetical genetic variant, the effect sizes of the variants on smoking and on lung cancer here are much too small to try to explain away the causal effect of smoking itself on lung cancer.

REFERENCES

1. Hung RJ, McKay JD, Gaborieau V, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*. 2008;452(7187):633–637.

2. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet.* 2008;40(5):616–622.

3. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature.* 2008;452(7187):638–642.

4. Saccone SF, Hinrichs AL, Saccone NL, et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet.* 2007;16(1):36–49.

5. Spitz MR, Amos CI, Dong Q, et al. The CHRNA5-A3 region on chromosome 15q24-25.1 is a risk factor both for nicotine dependence and for lung cancer. *J Natl Cancer Inst.* 2008; 100(21):1552–1556.

6. Liu JZ, Tozzi F, Waterworth DM, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Wellcome Trust Case Control Consortium. *Nat Genet.* 2010;42(5):436–440.

7. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet.* 2010;42(5):441–447.

8. Thorgeirsson TE, Gudbjartsson DF, Surakka I, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. ENGAGE Consortium. *Nat Genet.* 2010;42(5): 448–453.

9. Saccone NL, Culverhouse RC, Schwantes-An TH, et al. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet.* 2010;6(8):e1001053. (doi:10.1371/journal.pgen.1001053).

10. Chanock SJ, Hunter DJ. Genomics: when the smoke clears. *Nature.* 2008;452(7187):537–538.

11. Wacholder S, Chatterjee N, Caporaso N. Intermediacy and gene-environment interaction: the example of CHRNA5-A3 region, smoking, nicotine dependence, and lung cancer. *J Natl Cancer Inst.* 2008;100(21):1488–1491.

12. Lips EH, Gaborieau V, McKay JD, et al. Association between a 15q25 gene variant, smoking quantity and tobacco-related cancers among 17 000 individuals. *Int J Epidemiol.* 2010; 39(2):563–577.

13. Thorgeirsson TE, Stefansson K. Commentary: gene-environment interactions and smoking-related cancers. *Int J Epidemiol.* 2010;39(2):577–579.

14. Wang J, Spitz MR, Amos CI, et al. Mediating effects of smoking and chronic obstructive pulmonary disease on the relation between the CHRNA5-A3 genetic locus and lung cancer risk. *Cancer.* 2010;116(14):3458–3462.

15. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol.* 2010; 172(12):1339–1348.

16. Miller DP, Liu G, De Vivo I, et al. Combinations of the variant genotypes of GSTP1, GSTM1, and p53 are associated with an increased lung cancer risk. *Cancer Res.* 2002;62(10): 2819–2823.

17. Le Marchand L, Derby KS, Murphy SE, et al. Smokers with the CHRNA lung cancer-associated variants are exposed to higher levels of nicotine equivalents and a carcinogenic tobacco-specific nitrosamine. *Cancer Res.* 2008;68(22): 9137–9140.

18. Heatherton TF, Kozlowski LT, Frecker RC, et al. Measuring the heaviness of smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. *Br J Addict.* 1989;84(7):791–799.

19. Berrettini W, Yuan X, Tozzi F, et al. Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry.* 2008;13(4):368–373.

20. Stevens VL, Bierut LJ, Talbot JT, et al. Nicotinic receptor gene variants influence susceptibility to heavy smoking. *Cancer Epidemiol Biomarkers Prev.* 2008;17(12):3517–3525.

21. van der Laan MJ. Estimation based on case-control designs with known prevalence probability. *Int J Biostat.* 2008; 4(Article 17):1–57.

22. Breslow NE, Cain KC. Logistic regression for two-stage case-control data. *Biometrika.* 1988;75(1):11–20.

23. VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology.* 2010;21(4): 540–551.

24. Rothman KJ. *Modern Epidemiology.* 1st ed. Boston, MA: Little, Brown and Company; 1986.

25. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology.* 1992;3(5):452–456.

26. Knol MJ, van der Tweel I, Grobbee DE, et al. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. *Int J Epidemiol.* 2007; 36(5):1111–1118.

27. Wang N, Lin X, Gutierrez R, et al. Bias analysis and SIMEX inference in generalized linear mixed measurement error models. *J Am Stat Assoc.* 1998;93(441):249–261.

28. Li Y, Sheu CC, Ye Y, et al. Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. *Lancet Oncol.* 2010;11(4):321–330.

29. Truong T, Hung RJ, Amos CI, et al. Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. *J Natl Cancer Inst.* 2010;102(13):959–971.

30. Catassi A, Servent D, Paleari L, et al. Multiple roles of nicotine on cell proliferation and inhibition of apoptosis: implications on lung carcinogenesis. *Mutat Res.* 2008; 659(3):221–231.

31. Egleton RD, Brown KC, Dasgupta P. Nicotinic acetylcholine receptors in cancer: multiple roles in proliferation and inhibition of apoptosis. *Trends Pharmacol Sci.* 2008;29(3): 151–158.

32. Minna JD. Nicotine exposure and bronchial epithelial cell nicotinic acetylcholine receptor expression in the pathogenesis of lung cancer. *J Clin Invest.* 2003;111(1):31–33.

33. Bierut LJ, Stitzel JA, Wang JC, et al. Variants in nicotinic receptors and risk for nicotine dependence. *Am J Psychiatry.* 2008;165(9):1163–1171.

34. Fisher RA. Lung cancer and cigarettes. *Nature.* 1958; 182(4628):108.

35. Cornfield J, Haenszel W, Hammond EC, et al. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst.* 1959;22(1):173–203.