Genetic variants on 15q25.1, smoking and lung cancer: an assessment of mediation and interaction

Tyler J. VanderWeele[1,2], Kofi Asomaning[3], Eric J. Tchetgen Tchetgen[1,2], Younghun Han[4], Margaret R Spitz[4], Sanjay Shete[4], Xifeng Wu[4], Valerie Gaborieau[5], Ying Wang[6], John McLaughlin[7], Rayjean J. Hung[6], Paul Brennan[5], Christopher I. Amos[4], David C. Christiani[1,3,8] and Xihong Lin[2]

Affiliations:

1 Department of Epidemiology Harvard School of Public Health.

2 Department of Biostatistics, Harvard School of Public Health.

3 Department of Environmental Health, Harvard School of Public Health

4 Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center

5 International Agency for Research on Cancer (IARC)

6 Samuel Lunenfeld Research Institute of Mount Sinai Hospital

7 Cancer Care Ontario

8 Massachusetts General Hospital/Harvard Medical School.

Correspondence to: Tyler J. VanderWeele, Harvard School of Public Health, Departments of Epidemiology and Biostatistics, 677 Huntington Avenue, Boston, MA, 02115; Phone: 617-432-7855; Fax: 617-432-1884; E-mail: tvanderw@hsph.harvard.edu

ABSTRACT

Genome-wide association studies have identified variants on chromosome 15q25.1 that increase the risk of both lung cancer as well as nicotine dependence and associated smoking behavior. However, there remains debate as to whether the association with lung cancer is direct or mediated by pathways related to smoking behavior. Here, we apply novel methodology for mediation analysis, allowing for gene-environment interaction, to a lung cancer case-control study at Massachusetts General Hospital using two SNPs, rs8034191 and rs1051730, on 15q25.1. The results are validated using three other lung cancer studies. Tests for additive interaction ($P=2\times10^{-10}$ and $P=1\times10^{-9}$) and multiplicative interaction (P=0.01 and P=0.01) were significant. Pooled analyses yielded a direct effect odds ratio of 1.26 (95% CI:1.19-1.33; $P=2\times10^{-15}$) 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 ratio were 1.26 (95% CI:1.19-1.33; $P=1\times10^{-15}$) and 1.00 (95% CI:0.99-1.01; P=0.22) respectively with proportion mediated 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 is primarily through other pathways.

Key words: Gene-Environment Interaction, Lung Cancer, Mediation, Pathway Analysis, Smoking

Introduction.

Three GWAS studies (1-3) found associations between genetic variants on 15q25.1 and lung cancer. These variants were known also to be associated with smoking behavior (3-9), raising the question of whether the association of the variants with lung cancer is primarily through smoking or through other pathways (1-5,10-14). In addition to possible effects of genetics variants on 15q25.1 on lung cancer either through or independent of smoking, Thorgeirsson et al. (3,13) note also a third possible explanation of 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 methodology 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 mediator (smoking) which would be present if the variant increased vulnerability to the effect of smoking (3,13). The methodology (15) we employ here overcomes these limitations and is applied to a case-control lung cancer study of 1836 cases and 1452 controls (16) conducted at the 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 methodology were applied also to three other GWAS lung cancer case-control studies (1,2) to replicate results.

Methods

1836 cases and 1452 controls were drawn from a case-control study assessing the molecular epidemiology of lung cancer, which began in 1992 at the Massachusetts General Hospital (MGH) 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 among 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 non-melanoma skin cancer) were excluded from participation. Interviewer-administered questionnaires collected information on variables including age (continuous), gender, educational history (college degree or more, yes/no), smoking intensity (cigarettes per day) and duration (years) from each subject. The study was reviewed and approved by the Institutional Review Boards of the Massachusetts General Hospital and the Harvard School of Public Health.

To confirm findings, the analyses were replicated in three additional case-control studies of lung cancer: MD Anderson with 2827 cases and 2345 controls (2); International Agency for Research on Cancer (IARC) with 1871 cases and 2472 control) (1); and Toronto with 333 cases and 501 controls. All the analyses were limited to Caucasians.

We selected the two 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 two 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 into the study. DNA was extracted from peripheral blood samples using the Puregene DNA Isolation Kit (Gentra Systems, Minneapolis, MN). The polymorphisms in the MGH study were genotyped by the 5'-nuclease assay (TaqMan) and using the ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA). Genotyping was performed by laboratory personnel blinded to clinical variables and case control status, and a random 5% of the samples were repeated to validate genotyping procedures. Blinded genotyping results were independently reviewed by two authors. To check for genotyping error, departures from Hardy Weinberg equilibrium in controls were examined.

Genotype data from the MD Anderson Study was obtained with Illumina HumanHap 300 BeadChips for 1154 cases and 1137 controls at the Center for Inherited Disease Research and the remaining cases and controls were genotyped using Taqman. The IARC central Europe study was also genotyped with HumanHap300 Beadchip using the Illumina Infinium platform and was conducted in Centre National Genotypage (CNG Paris France). Genotype data for the Toronto study was obtained by genotyping cases and controls with the HumanHap300 Beadchip with the Illumina Infinium platform at the

McGill University and Genome Quebec Innovation Centre (Montreal Canada). Further details on genotyping in these studies can be found elsewhere (1,2).

Cigarettes per day were 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 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; controls by the 1 minus the prevalence divided by the proportion of controls in the study (15,21). Weights were further adjusted by sampling fractions in studies (MD Andersen and Toronto) in which sampling fractions varied by smoking status (22). The weighting takes into account 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 non-normality. When sampling fractions of cases and controls varied by smoking status (MD Andersen and Toronto), e.g. 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 comparing the risk of lung cancer with the genetic variant present versus absent if smoking behavior were what it would have been without the genetic 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 non-smokers) 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 the (i) exposure-outcome relationship, (ii) mediator-outcome relationship, (iii) exposure-mediator relationship and that (iv) there is no effect of the exposure that itself confounds the mediator-outcome relationship (15). No confounding of the effect of the exposure on the mediator and on the outcome (assumptions i and iii), 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 a Wald test of the interaction coefficient in the logistic regression. Measures correspond to a one-allele change in the genetic variants and to a one-unit change in the cigarette measure. The measure of additive interaction assesses the extent to which the odds ratio for a one-unit increase in both exposures exceeds the sum of the odds ratios for a one-unit increase in each exposure considered separately. The measure of multiplicative interaction assesses the log of the ratio of the odds ratio for a one-unit increase in both exposures relative to the product of the odds ratios for a one-unit increase in each exposure considered separately. Analyses for measurement error are conducted using regression calibration (27). Results from the four studies were combined on the log-odds scale using sample size based meta-analysis. Analyses were implemented with SAS 9.2 and R 2.4.

Results

Table 1 summarizes the demographic characteristics of the four 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 reports direct and indirect effects for rs8034191, along with tests for gene-by-smoking interaction from the four studies and a pooled meta-analysis (full details are available in Web Table 1). 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 in meta-analyses gave a direct effect odds ratio of 1.35 (95% CI:1.21-1.52; $P=3\times10^{-7}$) 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 mediated by smoking. Tests for interaction were significant on the additive risk scale ($P=1\times10^{-3}$), indicating that the effect of smoking varies by genotype, with weaker evidence on the multiplicative scale (P=0.17). Allowing for smoking-by-gene interaction gives, for changes from 0-1 and from 0-2 C alleles respectively, direct effect odds ratios of 1.31 (95% CI:1.15-1.49; $P=3\times10^{-5}$) and 1.72 (95% CI:1.34-2.21; $P=2\times10^{-5}$) 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 increased risk 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 reports similar analyses for rs1051730 (full details are available in Web Table 2). Ignoring possible interaction gives a direct effect odds ratio of 1.32 (95% CI:1.18-1.48; $P=2\times10^{-6}$) and an indirect effect odds ratio of 1.01 (95% CI:0.99-1.02; P=0.19) per A allele, with proportion mediated 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; $P=8\times10^{-5}$) and 1.66 (95% CI:1.27-1.64; $P=1\times10^{-8}$) and indirect effect odds ratios of 1.01 (95% CI:0.99-1.01; P=0.22) and 1.01 (95% CI:0.99-1.03; P=0.20) for changes from 0-1 and from 0-2 A alleles respectively, with 4.0% of increased risk mediated by smoking in the latter scenario.

As shown in Tables 2 and 3 all three replication studies exhibited similar patterns and overall replication P-values for the three 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 ($P= 6\times10^{-8}$;$P= 1\times10^{-7}$) and multiplicative (P=0.03;P=0.03) scales. Likewise, pooled estimates from all four studies, reported in the summary, indicated large highly significant direct effects ($P=2\times10^{-15}$; $P=1\times10^{-15}$), small indirect effects (P=0.09;P=0.22) and interaction on both additive ($P=2\times10^{-10}$;$P=1\times10^{-9}$) and multiplicative scales (P=0.01;P=0.01).

Further analyses using all four studies were conducted to allow for the possibility that the measure of the cigarettes per day was recorded with error and that the measure does not capture all of the relevant smoking behavior. Assuming that the daily cigarette measure explains only 50% of the variability in the biologically relevant measure gives adjusted direct and indirect effects odds ratios of 1.24 and 1.01 with proportion mediated 6.3% for rs8034191 and odds ratios of 1.25 and 1.01 with proportion mediated 4.7% for rs1051730. Assuming that the daily cigarette measure explains only 25% of the variability in (i.e. 75% measurement error) gives odds ratios of 1.23 and 1.03 with proportion mediated 12.5% for rs8034191 and odds ratios of 1.24 and 1.02 with proportion mediated 9.2% for rs1051730. Measurement error may attenuate 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 the variants with lung cancer are principally through pathways other than by changing smoking intensity. Although this may initially appear surprising, further evidence supports these results. In the studies 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 biological relevance for lung cancer. The result that most of the association is 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 socio-economic status, would likely affect smoking and lung cancer in the same direction and sensitivity analysis (23) indicates that this would likely lead to underestimation of 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 methodology for mediation that we have employed (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 conclusion is further supported by the fact that recent studies have found no association between the variants and lung cancer amongst non-smokers (28,29). These studies, in conjunction with the results 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 only operate for smokers, even though the variants do not substantially increase smoking behavior 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 non-smokers. The natural direct effects estimated in our analyses essentially average over the direct effects for smokers and non-smokers. 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 per day.

Certain biological hypotheses are consistent with the statistical evidence. As noted above, it may be that the variant serves to increase the 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 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). Also nicotine and the tobacco derivatives N'-nitrosonornicotine (NNN) and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) have high affinity for the nicotinic acetylcholine receptors that are also known to be present in lung tissue (30). Nicotine has been shown to be involved in lung carcinogenesis through the activation of the nictonic acetylcholine receptors in non-neuronal cells (32). If the variant

or any functional SNPs for which the variants are markers served to activate nicotinic receptors (33) such effect would again only be observed for smokers.

Our results here are also of historic interest. Over fifty years ago, Fisher (34) suggested that there might be a genetic variant responsible for both smoking behavior and lung cancer. He proposed this common genetic cause might explain the association between smoking and lung cancer and thus that smoking may not in fact itself have a causal effect on lung cancer. Our results here show that, in some respects, Fisher was partially correct. In previous studies, the variants on 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. There is thus indeed a common genetic cause of smoking and lung cancer. Fisher was partially correct, but only partially. As clearly demonstrated by Cornfield (35) 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.

Acknowledgements: The authors thank the editor and two referees for helpful comments. We acknowledge funding from NIH grants. ES017876, HD060696, CA076404, CA134294, CA074386, CA092824, P50CA70907-05, CA121197, CPRIT RP10043, U19 CA148127, R01CA133996 and Canadian Cancer Society Research Institute (CCSRI no. 020214). The authors declare no conflicts of interest.

Contributions: TJV, EJTT and XL conceived the analysis; PB, CIA, RJH, JM and DCC conducted the cancer studies; TJV, KA, YH, VG and YVW analyzed the data; TJV drafted the manuscript; KA, MRS, SS, XW, JM, RJH, PB, CIA, DCC and XL provided critical review.

The authors declare no competing interests.

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. *Nat Genet*. 2010;42(5):436-440.

7. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Tobacco and Genetics Consortium. *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. *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.

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, M.J. 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 Statist 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.

Table 1. Characteristics of study populations

	MGH		MD Anderson		IARC		Toronto	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
Total Sample	1836	1452	2827	2345	1867	2463	333	501
Number Ever Smokers	1671	943	2365	1836	1728	1584	238	281
Number Never Smokers	165	509	462	509	139	879	95	220
Average Cigarettes per Day	25.42	13.97	23.51	8.46	18.06	10.83	20.65	9.75
Smoking Duration	38.5	18.93	30.91	10.50	35.74	20.38	34.44	12.01
Age	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%
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%
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%

Table 2. Direct and indirect effects on lung cancer and assessment of variant-smoking interaction for rs8034191

	Original Study		Replication						Meta-Analysis		
	MGH		MD Anderson		IARC		Toronto		Pooled OR		
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	P-Value
w/o SNPxSmoking Interaction											
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	$1.8x10^{-15}$
Indirect Effect	1.01	0.99, 1.02	1.01	0.99, 1.03	1	0.99, 1.02	1	0.99, 1.01	1.01	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	$6.7x10^{-16}$
Proportion Mediated (%)		3.6		6.8		0.2		0.4		3.2	
w/ SNPxSmoking Interaction											
T/C vs. TT Direct Effect	1.31	1.15, 1.49	1.08	0.97, 1.21	1.24	1.11, 1.38	1.3	0.99, 1.71	1.2	1.12, 1.28	$5.6x10^{-8}$
T/C vs. TT Indirect Effect	1.01	0.99, 1.02	1.01	0.99, 1.03	1	0.99, 1.02	1	0.99, 1.01	1.01	1.00, 1.01	0.09
T/C vs. TT Total Effect	1.32	1.16, 1.50	1.1	0.98, 1.23	1.24	1.11, 1.39	1.3	0.99, 1.71	1.21	1.13, 1.29	$1.9x10^{-8}$
Proportion Mediated (%)		4.2		13		0.3		0.7		4.2	
C/C vs. TT Direct Effect	1.72	1.34, 2.21	1.18	0.96, 1.47	1.53	1.23, 1.92	1.7	0.99, 2.93	1.44	1.27, 1.63	$2.3x10^{-8}$
C/C vs. TT Indirect Effect	1.03	0.99, 1.07	1.03	0.99, 1.07	1	0.97, 1.04	1	0.97, 1.05	1.02	1.00, 1.04	0.07
C/C vs. TT 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	$4.4x10^{-9}$
Proportion Mediated (%)		6.3		15.2		0.3		2.3		5.6	
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	$2.3x10^{-10}$
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

Table 3. Direct and indirect effects on lung cancer and assessment of variant-smoking interaction for rs1051730

	Original Study		Replication						Meta-Analysis		
	MGH		MD Anderson		IARC		Toronto		Pooled		
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	P-Value
w/o SNPxSmoking Interaction											
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}
Proportion Mediated (%)		3.5		4.9		0		0.4		2.3	
w/ SNPxSmoking Interaction											
A/G vs. GG 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}
A/G vs. GG 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
A/G vs. GG 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}
Proportion Mediated (%)		4.1		9.3		0		0.5		3	
A/A vs. GG 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}
A/A vs. GG 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
A/A vs. GG 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}
Proportion Mediated (%)		5.7		11.1		0		0.7		4	
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

Table 2. Direct and indirect effects on lung cancer and assessment of variant-smoking interaction for rs8034191

	Original Study		Replication								Meta-Analysis	
	MGH		MD Anderson		IARC		Toronto		Overall Replication			
	OR	P-Value	OR	P-Value	OR	P-Value	OR	P-Value	P-Value		Pooled OR	P-Value
w/o SNPxSmoking Interaction												
Direct Effect	1.35 (1.21,1.52)	$3.1x10^{-7}$	1.18 (1.08, 1.29)	$2.0x10^{-4}$	1.26 (1.14,1.40)	$5.4x10^{-6}$	1.33 (1.02,1.73)	0.04	$4.7x10^{-10}$		1.26 (1.19,1.33)	$1.8x10^{-15}$
Indirect Effect	1.01 (0.99,1.02)	0.15	1.01 (0.99,1.03)	0.13	1.00 (0.99,1.02)	0.95	1.00 (0.99,1.01)	0.87	0.23		1.01 (1.00,1.01)	0.09
Total Effect	1.36 (1.22,1.54)	$1.5x10^{-7}$	1.19 (1.09,1.30)	$8.8x10^{-5}$	1.26 (1.14,1.40)	$6.4x10^{-6}$	1.33 (1.02,1.74)	0.03	$2.6x10^{-10}$		1.26 (1.19,1.34)	$6.7x10^{-16}$
Proportion Mediated	3.6%		6.8%		0.2%		0.4%				3.2%	
w/ SNPxSmoking Interaction												
T/C vs. TT Direct Effect	1.31 (1.15,1.49)	$3.1x10^{-5}$	1.08 (0.97,1.21)	0.15	1.24 (1.11,1.38)	$1.8x10^{-4}$	1.30 (0.99,1.71)	0.06	$8.4x10^{-5}$		1.20 (1.12,1.28)	$5.6x10^{-8}$
T/C vs. TT Indirect Effect	1.01 (0.99,1.02)	0.15	1.01 (0.99,1.03)	0.13	1.00 (0.99,1.02)	0.95	1.00 (0.99,1.01)	0.78	0.22		1.01 (1.00,1.01)	0.09
T/C vs. TT Total Effect	1.32 (1.16,1.50)	$1.6x10^{-5}$	1.10 (0.98,1.23)	0.10	1.24 (1.11,1.39)	$2.0x10^{-4}$	1.30 (-.99,1.71)	0.06	$4.8x10^{-5}$		1.21 (1.13,1.29)	$1.9x10^{-8}$
Proportion Mediated	4.2%		13.0%		0.3%		0.7%				4.2%	
C/C vs. TT Direct Effect	1.72 (1.34,2.21)	$2.1x10^{-5}$	1.18 (0.96,1.47)	0.11	1.53 (1.23,1.92)	$1.5x10^{-4}$	1.70 (0.99,2.93)	0.06	$4.8x10^{-5}$		1.44 (1.27,1.63)	$2.3x10^{-8}$
C/C vs. TT Indirect Effect	1.03 (0.99,1.07)	0.16	1.03 (0.99,1.07)	0.13	1.00 (0.97,1.04)	0.95	1.00 (0.97,1.05)	0.61	0.2		1.02 (1.00,1.04)	0.07
C/C vs. TT Total Effect	1.77 (1.38,2.26)	$6.7x10^{-6}$	1.22 (0.98,1.51)	0.07	1.54 (1.23,1.92)	$1.7x10^{-4}$	1.72 (1.00,2.94)	0.05	$2.1x10^{-5}$		1.46 (1.29,1.66)	$4.4x10^{-9}$
Proportion Mediated	6.3%		15.2%		0.3%		2.3%				5.6%	
Additive Interaction	0.06 (0.02,0.10)	$1x10^{-3}$	0.05 (0.03,0.08)	$1.8x10^{-5}$	0.07 (0.03,0.10)	$1.0x10^{-4}$	0.04 (-0.06,0.15)	0.48	$5.6x10^{-8}$		0.06 (0.04,0.08)	$2.3x10^{-10}$
Multiplicative Interaction	1.04 (0.98,1.09)	0.17	1.05 (1.01,1.09)	0.01	1.03 (0.96,1.09)	0.42	1.04 (0.93,1.15)	0.51	0.03		1.04 (1.01,1.07)	0.01

Table 3. Direct and indirect effects on lung cancer and assessment of variant-smoking interaction for rs1051730

	Original Study		Replication					
	MGH		MD Anderson		IARC		Toronto	
	OR	P-Value	OR	P-Value	OR	P-Value	OR	P-Value
w/o SNPxSmoking Interaction								
Direct Effect	1.32 (1.18,1.48)	$1.8x10^{-6}$	1.19 (1.09,1.30)	$1.0x10^{-4}$	1.27 (1.14,1.40)	$4.1x10^{-6}$	1.42 (1.09,1.86)	0.01
Indirect Effect	1.01 (0.99,1.02)	0.19	1.01 (0.99,1.02)	0.26	1.00 (0.98,1.01)	0.87	1.00 (0.99,1.01)	0.82
Total Effect	1.33 (1.19,1.50)	$1.0x10^{-6}$	1.20 (1.10,1.31)	$5.8x10^{-5}$	1.27 (1.14,1.40)	$5.9x10^{-6}$	1.42 (1.09,1.86)	0.01
Proportion Mediated	3.5%		4.9%		0.0%		0.4%	
w/ SNPxSmoking Interaction								
A/G vs. GG Direct Effect	1.29 (1.14,1.46)	$7.6x10^{-5}$	1.09 (0.98,1.22)	0.12	1.24 (1.10,1.38)	$2.0x10^{-4}$	1.41 (1.07,1.87)	0.01
A/G vs. GG Indirect Effect	1.01 (0.99,1.02)	0.19	1.01 (0.99,1.02)	0.26	1.00 (0.98,1.01)	0.87	1.00 (0.99,1.01)	0.81
A/G vs. GG Total Effect	1.30 (1.15,1.47)	$4.2x10^{-5}$	1.10 (0.98,1.23)	0.09	1.24 (1.10,1.38)	$2.7x10^{-4}$	1.42 (1.07,1.87)	0.01
Proportion Mediated	4.1%		9.3%		0.0%		0.5%	
A/A vs. GG Direct Effect	1.66 (1.30,2.12)	$5.8x10^{-5}$	1.20 (0.97,1.48)	0.10	1.53 (1.22,1.91)	$1.8x10^{-4}$	2.00 (1.15,3.48)	0.01
A/A vs. GG Indirect Effect	1.02 (0.99,1.06)	0.20	1.02 (0.98,1.06)	0.26	1.00 (0.96,1.03)	0.88	1.00 (0.97,1.04)	0.82
A/A vs. GG Total Effect	1.70 (1.33,1.17)	$2.3x10^{-5}$	1.22 (0.99,1.51)	0.07	1.53 (1.22,1.91)	$2.4x10^{-4}$	2.01 (1.16,3.47)	0.01
Proportion Mediated	5.7%		11.1%		0.0%		0.7%	
Additive Interaction	0.05 (0.02,0.09)	$4x10^{-3}$	0.05 (0.03,0.08)	$1.8x10^{-5}$	0.07 (0.04,0.11)	$6.0x10^{-5}$	0.01 (-0.10,0.13)	0.85
Multiplicative Interaction	1.03 (0.98,1.08)	0.26	1.05 (1.01, 1.09)	0.01	1.03 (0.97,1.10)	0.34	1.01 (0.90,1.12)	0.95

Meta-Analysis			
Overall Replication P-Value	Pooled		P-Value
7.7×10^{-11}	1.26 (1.19,1.33)		1.1×10^{-15}
0.45	1.00 (0.99,1.01)		0.22
5.6×10^{-11}	1.26 (1.19,1.34)		4.4×10^{-16}
		2.3%	
3.4×10^{-5}	1.20 (1.12,1.28)		3.7×10^{-8}
0.45	1.01 (0.99,1.01)		0.22
1.8×10^{-5}	1.21 (1.13,1.29)		1.2×10^{-8}
		3.0%	
1.8×10^{-5}	1.44 (1.27,1.64)		1.4×10^{-8}
0.43	1.01 (0.99,1.03)		0.2
1.2×10^{-5}	1.46 (1.29,1.66)		5.1×10^{-9}
		4.0%	
1.1×10^{-7}	0.06 (0.04,0.08)		1.4×10^{-9}
0.03	1.04 (1.00,1.07)		0.01