Published in final edited form as:
Cancer. 2010 July 15; 116(14): 3458–3462. doi:10.1002/cncr.25085.

MEDIATING EFFECTS OF SMOKING AND CHRONIC OBSTRUCTIVE AIRWAY DISEASE ON THE RELATIONSHIP BETWEEN THE CHRNA5-A3 GENETIC LOCUS AND LUNG CANCER RISK

Jian Wang, **Margaret R. Spitz**, **Christopher I. Amos**, **Anna V. Wilkinson**, **Xifeng Wu**, and **Sanjay Shete**
Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

Abstract

Background—Recent genome-wide association (GWA) studies of lung cancer have shown that the CHRNA5-A3 region on chromosome 15q24-25.1 is strongly associated with an increased risk of lung cancer and nicotine dependence, and thought to be associated with chronic obstructive airways disease as well. However, it has not been established whether the association between genetic variants and lung cancer risk is a direct one or one mediated by nicotine dependence.

Methods—In this paper we applied a rigorous statistical approach, mediation analysis, to examine the mediating effect of smoking behavior and self-reported physician-diagnosed emphysema (chronic obstructive pulmonary disease [COPD]) on the relationship between the CHRNA5-A3 region genetic variant rs1051730 and the risk of lung cancer.

Results—Our results showed that rs1051730 is directly associated with lung cancer risk, but that it is also associated with lung cancer risk through its effect on both smoking behavior and COPD. Furthermore, we showed that COPD is a mediating phenotype that explains part of the effect of smoking behavior on lung cancer. Our results also suggested that smoking behavior is a mediator of the relationship between rs1051730 and COPD risk.

Conclusions—Smoking behavior and COPD are mediators of the association between the SNP rs1051730 and the risk of lung cancer. Also, COPD is a mediator of the association between smoking behavior and lung cancer. Finally, smoking behavior also has mediating effects on the association between the SNP and COPD.

Keywords

Lung Cancer; COPD; Mediation analysis; smoking behavior; genetic variants

INTRODUCTION

Recent genome-wide association (GWA) studies of lung cancer have shown that the CHRNA5-A3 region on chromosome 15q24-25.1 is strongly associated with an increased risk of lung cancer.[1–3] However, different etiological explanations for the association

Address for correspondence and reprints: Dr. Sanjay Shete, Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, 1155 Pressler Street, Unit 1340, Houston, TX 77030, USA; Telephone: (713) 745-2483; Fax: (713) 792-8261; sshete@mdanderson.org.

between genetic variants in this region and lung cancer have been provided. Thorgeirsson et al.3 suggested that the genetic variants were only associated with smoking behavior, while, Hung et al.2 found no association with smoking behavior. Our initial study[1] suggested a weak association with nicotine dependence (based on number of cigarettes smoked per day and pack-years of exposure) but a stronger direct association with lung cancer risk *per se*. Further analysis conducted by Spitz et al.4 confirmed the dual pathway between the genetic variants and lung carcinogenesis. In an accompanying editorial, Wacholder et al.5 stressed the need for sophisticated statistical analyses to dissect these relationships. In this paper we have applied a rigorous statistical approach, mediation analysis, to identify the mediation effect of smoking behavior on the relationship between the CHRNA5-A3 region single nucleotide polymorphism (SNP), rs1051730, and the risk of lung cancer. Mediation models for analyzing indirect effects between initial and outcome variables have been widely applied.6 In this paper, we conducted the mediation analysis based on the procedure described by Baron and Kenny.7 In addition to smoking behavior, we also evaluated the potential role of self-reported physician-diagnosed emphysema (chronic obstructive pulmonary disease [COPD]) as a mediator for CHRNA5-A3 region association with risk of lung cancer.

METHODS

Case and Control Subjects

This analysis included 1154 lung cancer case subjects who were current or former smokers and 1136 control subjects frequency-matched to the cases by age, sex, and smoking status. All the case and control subjects were Caucasian. Lung cancer cases were accrued at The University of Texas M. D. Anderson Cancer Center and were histologically confirmed. Controls were ascertained through a multi-specialty physician practice from the same area. Questionnaire data were obtained by personal interview. This study was approved by the institutional review board at the University of Texas M. D. Anderson Cancer Center and all participants provided written informed consent.

Phenotype

To study the possible mediating effects of smoking behavior, we selected pack-years (PKYRS) smoked as the measurement of smoking intensity and nicotine dependence. In this study, PKYRS was computed as the number of cigarettes smoked per day divided by 20 and multiplied by the number of years smoked. All the cases and controls also self-reported whether a physician had ever diagnosed them with COPD, categorized as present or absent.

Statistical Model

We implemented mediation analysis to identify the potential mediation effects of PKYRS and COPD on the relationship between rs1051730 and lung cancer risk. Most often, mediation analyses are conducted statistically using the procedure described by Baron and Kenny.[7] The total effect of the initial variable (e.g., SNP) on the outcome (e.g., lung cancer) can be expressed as the sum of their direct and indirect effects, where the indirect effect is obtained as the product of the effect of the initial variable on the mediator (e.g., smoking behavior or COPD) and the effect of the mediator on the outcome while controlling for the initial variable. The total, direct, and indirect effects are obtained from the coefficients of the regression models. According to the general notations in the mediation analysis [7,8], we denote the total effect of the initial variable on the outcome variable as *c*, which is the regression coefficient of the outcome on the initial variable. We denote the effect of the initial variable on the mediator as *a*, which is the regression coefficient of the mediator on the initial variable. We denote the effect of the mediator on the outcome as *b*, which is the regression coefficient of the outcome on the mediator, while controlling for the initial

variable and we denote the direct effect of the initial variable on the outcome as *c'*, which is obtained by regressing the outcome on the initial variable controlling for the mediator. Finally, the significance of the indirect effect (mediating effect *ab*) is tested using the Sobel test.[9,10] The Sobel test estimates the standard error of *ab* as the square root of $b^2s_a^2 + a^2s_b^2$, where s_a and s_b are the standard errors of *a* and *b*, respectively. The ratio of *ab* and its standard error is treated as a standard normal variate. To perform a formal mediation analysis, the coefficients *c*, *a*, and *b* need to be statistically significant, according to the steps outlined by Baron and Kenny.

We studied four different models: (i) the mediating effect of smoking behavior on the relationship between the SNP and lung cancer; (ii) the mediating effect of COPD on the relationship between the SNP and lung cancer; (iii) the mediating effect of COPD on the relationship between smoking behavior and lung cancer; and (iv) the mediating effect of smoking behavior on the relationship between the SNP and COPD. In some models, outcome or mediator variables are dichotomous, we used logistic regressions in our analyses. Although the mediation analysis was originally developed for linear regression models, it has been shown to be valid for logistic regression as well.[8,11] The coefficients obtained from logistic regression were standardized to make them comparable before the Sobel test was applied for the mediation analysis.[8,11,12] Specifically, we multiplied each coefficient by the standard deviation of the predictor and divided by the standard deviation of the response variable in the corresponding equation. For more detail regarding these computations, please refer to the studies of Kenny and Herr.[8,11] Because our cases and controls were matched for sex, this factor was not significant in the analyses. However, age was found to be significant. Therefore, all analyses were adjusted for age. The genetic variant (rs1051730) was coded assuming an additive model as in our initial study.[1]

RESULTS

In our data, the average ages were 62.1 (standard deviation (SD) = 10.8) in cases, 61.1 (SD = 8.9) in controls. There were 658 males and 496 females in cases and 643 males and 493 females in controls. In cases, 603 individuals were former smokers, and 551 individuals were current smokers. In controls, 656 individuals were former smokers, and 480 individuals were current smokers. The overall minor allele frequency (MAF) of the SNP rs1051730 was 36.65%. In the cases, the MAF was 39.77%, and in the controls, the MAF was 33.49%. PKYRS is not normally distributed, therefore, we performed a Box-Cox transformation before mediation analyses. A maximum likelihood approach[13] was used to choose the optimal power transformation parameter λ. The estimated value of λ was equal to 0.5. All the results from the mediation analysis are reported in Table 1.

In the first model, the initial variable is the SNP, the outcome variable is lung cancer, and the mediator variable is PKYRS. Following the Baron and Kenny steps[7], the total effect of the SNP on lung cancer risk was statistically significant ($c = 0.1056$, SE = 0.0232, $Z = 4.5428$, $p < 0.0001$). Both the effects of the SNP on PKYRS and of PKYRS on lung cancer were also statistically significant ($a = 0.0682$, SE = 0.0205, $Z = 3.3317$, $p = 0.0009$; $b = 0.1178$, SE = 0.0239, $Z = 4.9379$, $p < 0.0001$). Therefore, the first three steps for mediation analysis as described by Baron and Kenny were satisfied. Our analysis based on the Sobel test concluded that smoking behavior (PKYRS) was a mediator of the relationship between the SNP and lung cancer ($ab = 0.0080$, SE = 0.0029, $Z = 2.7619$, $p = 0.0057$). The direct effect of the SNP on lung cancer ($c' = 0.0980$, SE = 0.0232), controlling for the mediator (PKYRS), was still statistically significant ($Z = 4.2167$, $p < 0.0001$), suggesting that PKYRS only partially explains the effect (and only a small portion of the effect) of the SNP on lung cancer ($ab/(ab+c') = 7.6\%$).[14] Therefore, the SNP is both directly and indirectly associated with lung cancer.

In the second model, we studied the effect of COPD on the association between the SNP and lung cancer. The initial and outcome variables remained the SNP and lung cancer respectively, while the mediator was COPD. The total effect of the SNP on lung cancer was the same as in the first model. The effects of the SNP on COPD and of COPD on lung cancer were statistically significant ($a = 0.1047$, SE = 0.0315, $Z = 3.3258$, $p = 0.0009$; $b = 0.1116$, SE = 0.0128, $Z = 8.7319$, $p < 0.0001$). After controlling for COPD, the direct effect of the SNP on lung cancer ($c' = 0.0899$, SE = 0.0237) was statistically significant ($Z = 3.7906$, $p = 0.0002$). The result from the Sobel test showed that COPD was a more significant mediator than PKYRS on the effect between the SNP and lung cancer ($ab = 0.0117$, SE = 0.0038, $Z = 3.1080$, $p = 0.0019$) and that it accounts for a higher proportion of the association ($ab/(ab+c') = 11.5\%$). Therefore, it can be concluded that the SNP is a direct risk factor associated with lung cancer and COPD is a significant mediator of the effect of the SNP on lung cancer.

The third model tested the effect of COPD on the relationship between smoking behavior (PKYRS) and lung cancer. In the mediation analysis, the total effect of PKYRS on lung cancer ($c = 0.1246$, SE = 0.0239) was statistically significant ($Z = 5.2129$, $p < 0.0001$). The effects of PKYRS on COPD and of COPD on lung cancer were also statistically significant ($a = 0.3426$, SE = 0.0331, $Z = 10.3404$, $p < 0.0001$; $b = 0.1057$, SE = 0.0130, $Z = 8.1256$, $p < 0.0001$). COPD was a mediator for smoking behavior and lung cancer, as evidenced by the highly significant p value obtained from the Sobel test ($ab = 0.0362$, SE = 0.0057, $Z = 6.3890$, $p < 0.0001$). Furthermore, COPD explained about one-third of the effect of smoking behavior on lung cancer ($ab/(ab+c') = 32.1\%$), and we observed that the direct effect of smoking behavior on lung cancer, controlling for COPD ($c' = 0.0767$, SE = 0.0247), was still statistically significant ($Z = 3.1093$, $p = 0.0019$). Therefore, we can conclude that COPD is a mediator of the relationship between smoking behavior and lung cancer according to Baron and Kenny steps.

We also investigated the mediating effect of both smoking behavior and COPD on the association between the SNP and lung cancer. Since smoking behavior and COPD are highly correlated (Kendall Tau b test p value < 0.0001), it is not recommended to perform an analysis with both as mediators.[8]

Recently, it was suggested[15] that CHRNA5-A3 variants are also directly associated with the risk of COPD. The aim of the last model in our study was to examine the indirect effect of smoking behavior (PKYRS) on the relationship between the SNP and COPD. In this situation, the initial and outcome variables were the SNP and COPD, and the mediator was PKYRS. In addition to age, all the regressions in this model were also adjusted for lung cancer status. The total effect of the SNP on COPD ($c = 0.0827$, SE = 0.0321) was statistically significant ($Z = 2.5726$, $p = 0.0101$). The effects of the SNP on PKYRS and of PKYRS on COPD were also statistically significant ($a = 0.0585$, SE = 0.0205, $Z = 2.8584$, $p = 0.0043$; $b = 0.3287$, SE = 0.0339, $Z = 9.6944$, $p < 0.0001$). While controlling for PKYRS, the direct effect of the relationship between the SNP and COPD ($c' = 0.0621$, SE = 0.0311) was still statistically significant ($Z = 1.9966$, $p = 0.0459$). The result of the Sobel test showed that PKYRS is a mediator of the effect of the SNP on COPD ($ab = 0.0192$, SE = 0.0070, $Z = 2.7417$, $p = 0.0061$), and that it explains about one-fourth of the association between the SNP and COPD ($ab/(ab+c') = 23.6\%$).

DISCUSSION

Recent GWA studies have provided different possible explanations for the association between the CHRNA5-A3 region on chromosome 15q24-25.1 and the increased risk of lung cancer: a direct association, an indirect association mediated through nicotine dependence,

or dual pathways. In this paper, we performed mediation analysis to investigate the mediating effect of smoking behavior and COPD on the relationship between the SNP, rs1051730, and the risk of lung cancer. The mediation analysis was conducted according to the Baron and Kenny steps.[8]

Four different models were investigated in this paper. In the first model, we showed that the SNP is directly associated with lung cancer risk, and it is also associated with lung cancer through smoking behavior. However, smoking behavior only explains a small portion of this relationship between the SNP and the risk of lung cancer (7.6%). This result supports our initial analysis[1,4], which suggested a weaker association of the SNP and smoking behavior, but a stronger association between the SNP and the risk of lung cancer.

One may argue that the best way to identify the association between the SNP and lung cancer would be to examine the effect of the SNP among never smokers. In 452 never smoking cases and 487 never smoking controls, we found no significant effect of the SNP on the risk of lung cancer (p value = 0.0697, OR = 0.83, 95% confidence interval CI = [0.69 to 1.02]), further confirming the mediating effect of smoking on the association between the SNP and lung cancer risk. Importantly, the mediation analyses showed that the association of the SNP on lung cancer risk is only marginally explained by smoking. In other words, smoking behavior is necessary-but-not-sufficient for the association between the SNP and lung cancer risk.

COPD is an independent risk factor of lung cancer and may be a stronger risk factor compared to smoking behavior.[16] Our analyses also showed that COPD is a mediating phenotype that explains about one-third of the effect of smoking behavior on lung cancer. Furthermore, we found that COPD is a stronger mediator than smoking behavior to explain the relationship between the SNP and lung cancer risk (11.5% versus 7.6%).

We further explored the effect of this lung-cancer-related SNP on COPD risk. The result suggested that smoking behavior is also a mediator of the relationship between the SNP and COPD and that the SNP is also associated with COPD through smoking behavior. One of the limitations of this analysis is that our study was not designed to study COPD. However, we adjusted for lung cancer status while performing mediation analysis with COPD as outcome variable. One may argue that the use of self-reported physician-diagnosed emphysema as a COPD measure might result in misclassification of the disease and that documentation through pulmonary function tests or radiological evidence are the gold standard. However, previous studies have shown that the questionnaire-based approach to defining COPD is quite accurate for epidemiologic studies.[17–19] For example, the study of Barr et al.[17] showed that 78% of questionnaire-based self-reported COPD cases were validated by a blinded medical records review. Another validation study of Eisner et al.[19] indicated that about 90% of participants who had spirometry data had airflow obstruction.

It is important to note that, in this paper, when we refer to the direct association between the SNP and the risk of lung cancer, we mean the association between the SNP and the risk of lung cancer that is not mediated through smoking behavior or COPD. However, this association between the SNP and the risk of lung cancer, which is not explained by smoking behavior or COPD, could still be explained by other, unmeasured risk factors, or attributed to measurement error with smoking behavior and level of carcinogen exposure. Also a recent study by Le Marchand et al.[20] suggested that the risk variant rs1051730 is associated with higher levels of tobacco-specific metabolites after controlling for smoking intensity.

In summary, our findings showed that the SNP rs1051730 has a direct association with lung cancer risk, as well as an indirect (and lesser) association through its effect on both smoking behavior and COPD. Furthermore, our results suggested that smoking behavior is also a

mediator of the association of the same SNP with risk for COPD. Further, COPD is a mediating phenotype of the relationship between smoking behavior and lung cancer risk. Importantly, our study findings show direct role of variant(s) in CHRNA5-A3 gene in smoking, lung cancer and COPD. These study findings suggest that statistical models can help tease out the complex relationship between the candidate SNP, smoking behavior, COPD and lung cancer risk, and therefore can offer a direction for future research, either in laboratory experiments or in statistical risk modeling, to obtain more precise insights into these complex relationships.

Acknowledgments

Grand support: 1R01CA131324, 5R01CA55769, 5R01CA127219 and K07CA126988.

References

1. 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; 40:616–622. [PubMed: 18385676]
2. 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; 452:633–637. [PubMed: 18385738]
3. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature. 2008; 452:638–642. [PubMed: 18385739]
4. 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; 100:1552–1556. [PubMed: 18957677]
5. 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; 100:1488–1491. [PubMed: 18957674]
6. Shrout PE, Bolger N. Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychol Methods. 2002; 7:422–445. [PubMed: 12530702]
7. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986; 51:1173–1182. [PubMed: 3806354]
8. Kenny, DA. Mediation. [accessed 6-12-2009]. http://davidakenny.net/cm/mediate.htm#SE
9. Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav Res Methods Instrum Comput. 2004; 36:717–731. [PubMed: 15641418]
10. Sobel, ME. Asymptotic confidence intervals for indirect effects in structural equation models. In: Leinhart, S., editor. Sociological Methodology. San Francisco: Jossey-Bass; 1982. p. 290-312.
11. Herr, NR. Mediation with dichotomous outcomes. [accessed 6-12-2009]. http://nrherr.bol.ucla.edu/Mediation/logmed.html
12. MacKinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. Evaluation Rev. 1993; 17:144–158.
13. Draper, N.; Smith, H. Applied regression analysis. Second ed.. New York: Wiley; 1981.
14. MacKinnon DP, Lockwood CM, Brown CH, et al. The intermediate endpoint effect in logistic and probit regression. Clin Trials. 2007; 4:499–513. [PubMed: 17942466]
15. Young RP, Hopkins RJ, Hay BA, et al. Lung cancer gene associated with COPD: triple whammy or possible confounding effect? Eur Respir J. 2008; 32:1158–1164. [PubMed: 18978134]
16. Spitz MR, Hong WK, Amos CI, et al. A risk model for prediction of lung cancer. J Natl Cancer Inst. 2007; 99:715–726. [PubMed: 17470739]
17. Barr RG, Herbstman J, Speizer FE, et al. Validation of self-reported chronic obstructive pulmonary disease in a cohort study of nurses. Am J Epidemiol. 2002; 155:965–971. [PubMed: 11994237]
18. Straus SE, McAlister FA, Sackett DL, et al. Accuracy of history, wheezing, and forced expiratory time in the diagnosis of chronic obstructive pulmonary disease. J Gen Intern Med. 2002; 17:684–688. [PubMed: 12220364]

19. Eisner MD, Trupin L, Katz PP, et al. Development and validation of a survey-based COPD severity score. Chest. 2005; 127:1890–1897. [PubMed: 15947299]

20. 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; 68:9137–9140. [PubMed: 19010884]

Table 1

Results of Mediation Analysis

Model	Effect	Estimate	SE	*Z*	p value
Initial: SNP	Total: SNP on LC *(c)*	0.1056	0.0232	4.5428	<0.0001
Mediator: PKYRS	SNP on PKYRS *(a)*	0.0682	0.0205	3.3317	0.0009
Outcome: LC	PKYRS on LC given SNP *(b)*	0.1178	0.0239	4.9379	<0.0001
	Indirect: SNP on LC *(ab)*	0.0080	0.0029	2.7619	0.0057
	Direct: SNP on LC given PKYRS *(c')*	0.0980	0.0232	4.2167	<0.0001
Initial: SNP	Total: SNP on LC *(c)*	0.1056	0.0232	4.5428	<0.0001
Mediator: COPD	SNP on COPD *(a)*	0.1047	0.0315	3.3258	0.0009
Outcome: LC	COPD on LC given SNP *(b)*	0.1116	0.0128	8.7319	<0.0001
	Indirect: SNP on LC *(ab)*	0.0117	0.0038	3.1080	0.0019
	Direct: SNP on LC given COPD *(c')*	0.0899	0.0237	3.7906	0.0002
Initial: PKYRS	Total: PKYRS on LC *(c)*	0.1246	0.0239	5.2129	<0.0001
Mediator: COPD	PKYRS on COPD *(a)*	0.3426	0.0331	10.3404	<0.0001
Outcome: LC	COPD on LC given PKYRS *(b)*	0.1057	0.0130	8.1256	<0.0001
	Indirect: PKYRS on LC *(ab)*	0.0362	0.0057	6.3890	<0.0001
	Direct: PKYRS on LC given COPD *(c')*	0.0767	0.0247	3.1093	0.0019
Initial: SNP	Total: SNP on COPD *(c)*	0.0827	0.0321	2.5726	0.0101
Mediator: PKYRS	SNP on PKYRS *(a)*	0.0585	0.0205	2.8584	0.0043
Outcome: COPD	PKYRS on COPD given SNP *(b)*	0.3287	0.0339	9.6944	<0.0001
	Indirect: SNP on COPD *(ab)*	0.0192	0.0070	2.7417	0.0061
	Direct: SNP on COPD given PKYRS *(c')*	0.0621	0.0311	1.9966	0.0459

All the results were adjusted for age. In the last model, the results were also adjusted for lung cancer case control status.

SNP = single nucleotide polymorphisms
PKYRS = pack-years
COPD = chronic obstructive pulmonary disease

Cancer. Author manuscript; available in PMC 2011 July 15.

LC = lung cancer

c represents the total effect of initial variable on outcome variable

a represents the effect of initial variable on mediator variable

b represents the effect of mediator variable on outcome variable, controlling for initial variable

c' represents the direct effect of initial variable on outcome variable, controlling for mediator variable

ab represents the indirect effect of initial variable on outcome variable through mediator variable