Published in final edited form as:
*Cancer Epidemiol Biomarkers Prev*. 2010 January ; 19(1): 240–244. doi:10.1158/1055-9965.EPI-09-0710.

# A Rigorous and Comprehensive Validation: Common Genetic Variations and Lung Cancer

**Ping Yang**[1], **Yafei Li**[1], **Ruoxiang Jiang**[1], **Julie M. Cunningham**[2], **Fang Zhang**[1], and **Mariza de Andrade**[1]

[1] Department of Health Sciences Research, Mayo Clinic College of Medicine

[2] Department of Laboratory Medicine and Pathology, Mayo Clinic College of Medicine

## Abstract

**BACKGROUND—**Multiple recent genome-wide studies of single nucleotide polymorphisms (SNPs) reported associations between candidate chromosome loci and lung cancer susceptibility. We evaluated five of the top candidate SNPs (rs402710, rs2736100, rs4324798, rs16969968, and rs8034191) for their effects on lung cancer risk and overall survival.

**METHODS—**Over 1,700 cases and 2,200 controls were included in this study. Seven independent, complementary case-control datasets were tested for risk assessment encompassing cigarette smokers and never smokers, using unrelated controls and unaffected full-sibling controls. Five patient groups were tested for survival prediction stratified by smoking status, histology subtype, and treatment.

**RESULTS—**After considering a history of chronic obstructive pulmonary disease (COPD) as a risk factor altering lung cancer risk and comparing to sibling controls, none of the five SNPs was significant. However, the variant, rs4324798, was significant in predicting overall survival (hazard ratio HR=0.46, 95% CI: 0.30–0.73, p=0.001) in small cell lung cancer (SCLC).

**CONCLUSIONS—**None of the five candidate SNPs in lung cancer risk can be confirmed in our study. The previously reported association could be explained by disparity in tobacco smoke exposure and COPD history between cases and controls. Instead, we found rs4324798 to be an independent predictor in SCLC survival, warranting further elucidation of the underlying mechanisms.

### Keywords

GWAS; lung cancer; single nucleotide polymorphisms

## INTRODUCTION

Recently, eight genome-wide association SNP studies (GWAS) have reported association between chromosome loci and lung cancer susceptibility (1–8). Of the five top candidate SNPs, some reside in known genes (rs402710, rs2736100, and rs16969968) and some are *de-novo* (rs8034191 and rs4324798), have been validated (9,10) or are under validation. rs402710 and rs2736100 are located in chromosome 5p15.33, containing two known genes: the human telomerase reverse transcriptase (TERT) gene and the cleft lip and palate transmembrane 1 like (CLPTM1L, alias CRR9) gene. rs2736100 is located in intron 1 of the TERT gene; and rs402710 is in a region of high linkage disequilibrium (LD) that includes the promoter regions of TERT and the entire coding region of the CLPTM1L gene (intron 16). The overall estimated allelic odds ratios (ORs) for rs402710 and rs2736100 are 1.18 and 1.14, respectively (4).

Corresponding Author: Ping Yang, M.D., Ph.D., Mayo Clinic, 200 First Street SW, Rochester, Minnesota, 55905; yang.ping@mayo.edu.

rs8034191 and rs16969968 are located in chromosome 15q25, containing six known genes, three of which encode nicotinic acetylcholine receptor subunits (CHRNA5, cholinergic receptor nicotinic α 5; CHRNA3, cholinergic receptor nicotinic α 3; and CHRNB4, cholinergic receptor nicotinic β 4). The remaining three are IREB2 (iron-responsive element-binding protein 2), PSMA4 (implicated in DNA repair), and LOC123688 (unknown function). rs16969968, a non-synonymous variant in CHRNA5 and rs8034191, an unknown locus demonstrated association with lung cancer susceptibility [allelic ORs at 1.30 (1.23–1.38) and 1.32 (1.21–1.45) respectively] (2). rs4324798 is located in chromosome 6p21.33 within an extended region of high LD near the major histocompatibility complex containing more than 20 genes (2); genotyping of rs4324798 in five validation studies provided evidence of association with lung cancer risk, at an OR of 1.28 (1.16–1.40) (2).

However, these GWAS were primarily conducted in cigarette smokers except for one study, which did not support the observed association in never smokers (6). We report results of validating the five SNPs in relationship to lung cancer susceptibility when they were separately evaluated in smokers and never smokers using unrelated and full-sibling controls. Importantly, history of chronic obstructive pulmonary disease (COPD) was carefully considered as a confounding factor because the well-established shared etiology with lung cancer from tobacco smoking and genetic susceptibility to both diseases (11–13), and studies since 1980 have shown COPD to be an independent risk factor for lung cancer (14). Moreover, we also evaluated their prognostic value for lung cancer survival.

## METHODS

### Study Subjects

Lung cancer patients were identified and enrolled at Mayo Clinic between 1997 and 2006. The research protocol and consent form were approved by the Mayo Clinic Institutional Review Board; detailed study design and procedure were reported previously (11). Unrelated controls were selected from community residents who were identified by having had a general medical examination and a leftover blood sample from routine clinical tests (11). All full siblings, who were free of cancer and who donated a blood sample, were recruited as controls through lung cancer cases (11).

### Data Collection

Demographic and other risk information was obtained from all subjects via a combination of a structured interview, self-administered questionnaire, and medical records (11,15). Never smokers were defined as having smoked fewer than 100 cigarettes during their lifetimes and second hand smoking history was collected as previously reported(16). Cigar or pipe smokers were excluded. Ever smoking included current and/or previous use. History of COPD was determined based on explicit diagnosis recorded in the medical history. Family history of lung cancer in first-degree relatives (parents, siblings, and children) included vital status and age at diagnosis.

### SNPs Selection and Allele Typing

Five candidate SNPs were selected from eight recent GWAS: (1–8) rs402710 (G->A); rs2736100 (C->A); rs4324798 (G->A); rs16969968 (G->A); and rs8034191 (T->C). The LD structure of each SNP constructed by Haploview (17) illustrates the known genes or candidate locus regions (Supplementary Figure S1 A–C). Genotyping, performed in the Mayo Clinic Genomic Shared Resource, used TaqMan (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. Primers and probes were Assay-by Design (Applied Biosystems). Quality control procedures of genotyping tests are in Supplementary material and Table S1.

### Analytical Strategy and Statistical Models

Our strategy to rigorously and comprehensively evaluate the role of top SNPs was accomplished by testing the specified hypothesis in the targeted subgroup while best controlling for the strong confounding effect of cigarette smoking history in risk assessment (case-control study) and treatment in survival outcome (patient follow-up study). Other potential confounders included age, sex, COPD history, lung cancer stage and histology, and progression or recurrence. Cases and controls are described in Supplementary Table S2 where eight datasets were defined and respective hypotheses specified. The three main categories are total cases (1,735) and controls (2,242), cigarette smokers (1,406 cases; 1,053 controls), and never smokers (329 cases; 757 controls); under each main category, two control groups were used, unrelated community residents and unaffected full-siblings of cases (11). The sixth group had a limited sample size and was not analyzed further. Survival analysis included 1,742 consecutive patients who were diagnosed between 1997 and 2006. The following contrasting groups were analyzed separately: 1,418 cigarette smokers and 324 never smokers. Among smokers, NSCLC and SCLC were further separated according to surgical resection: 849 surgically resected NSCLC, 334 NSCLC without surgery; and 235 SCLC (all without surgery). Adjustments for covariates are specified in the footnotes of the results tables.

We tested for association between each SNP and lung cancer status using unconditional logistic regression for cases and unrelated controls and conditional logistic regression for cases and sibling controls (18). We also tested the association of each SNP with survival time, defined as the time from lung cancer diagnosis to last follow-up, using Cox Proportional hazards regression analysis (19). Significant covariates were selected and identified through forward and backward variable selection procedures. The level of $p$=0.05 was chosen as our threshold for statistical significance. Multiple comparison correction was not applied because our goal was to validate each SNP independently. All analyses were performed using SAS software (SAS Institute Inc. SAS®/STAT User's Guide, v9. Cary, NC: 2008.)

## RESULTS

### Lung cancer risk assessment

Seven predefined case-control sets are provided in Supplementary Table S2, and basic descriptions of age, sex, cigarette pack-year smoking, and prior medical history of COPD are in Supplementary Table S3. As an initial step to replicate published results in a comparable design, each of the five SNPs were assessed in our total cases and unrelated controls (Table 1, Dataset 1). Four of the five SNPs were significantly associated with lung cancer risk in univariate models; however, after accounting for previously adjusted risk factors, only rs402710 remained significant; and the significant association holds after further adjusting for COPD, suggesting individuals with the minor allele have a 21% reduced risk. Although all subjects were self-reported Caucasians, an alternative design using full sibling controls was applied to avoid sub-population stratification; the results do not support an association of any tested SNPs and lung cancer risk (Table 1, Dataset 2).

Next, the five candidate SNPs and lung cancer risk in smokers and never smokers were assessed separately. Only two SNPs on chromosome 5 showed some significant results (Table 2): rs402710 was significant in all smokers, but was not significant when tested in *heavy* smokers only; whereas, rs2736100 was only significant in never smokers. When compared to siblings, none of the five SNPs were significant. Specific to rs402710, the estimated OR was 0.78 (p=0.002) when cases were compared to unrelated controls among all smokers, attenuated to 0.85 when restricted to heavy smokers (p=0.117), and diminished to 0.91 (p=0.211) when compared to siblings. A similar pattern was observed with rs2736100.

### Lung cancer overall survival outcome

The prognostic role of each SNP was also tested for lung cancer overall survival, as shown in Table 3; more detailed description of patients' characteristics, which are included in the multivariable Cox models, is provided in Supplementary Table S4. rs432478 is the only SNP that showed a significant effect in SCLC patients, with a minor allele associated with longer survival.

## DISCUSSION

We rigorously evaluated the five top candidate SNPs that have been revealed to alter lung cancer susceptibility from multiple GWAS and a few validation studies. Our initial validation results, using a similar design as in the published studies, confirmed results for two SNPs on chromosome 5 (rs402710 and rs2736100). However, more rigorous evaluation by controlling for the effect of pre-existing COPD attenuated the association; using sibling controls diminished the association. Specifically noted is that for rs2736100, the estimated effect in never smokers was significant (OR=1.23) but attenuated to 1.14 for related controls, no longer significant. These findings could be due to the small sample size; three alternative explanations are postulated: First, any genetic effects were dampened in the presence of heavy exposure to environmental carcinogens. Second, findings from previous studies were confounded by variable degrees of tobacco smoke exposure, likely a residual effect even after adjusting for cigarette smoking history (20). Indeed, in one of the previous validation studies, when never smokers were analyzed independently, rather than in the midst of smokers, no association with the top SNPs remained significant (6). As nicotine dependence phenotype confounds carcinogen exposure, the authors of one GWAS (1,6) had interpreted their finding of chromosome 15q24/25.1 with no consensus as to the relative impact of the variants on the propensity to smoke versus a direct carcinogenic effect. Third, the association between these SNPs and lung cancer may, in part, be confounded by COPD (21); future studies should carefully adjust for COPD and evaluate the dual-effects of the at-risk SNPs in both COPD and lung cancer.

Finally, we have revealed one SNP, rs4324798, as an independent prognostic factor for overall survival in SCLC patients, calling for further validation by other studies. The prognostic value of this SNP and its context gene and region in treatment response and toxicities need to be evaluated.

In conclusion, three unique strengths of our study are the dual-control design, multiple independent subsets, and the consideration of medical history of COPD. Although none of the five candidate SNPs were found to be significant, we did obtain comparable results when we chose a more liberal design that mimicked previously published studies, specifically, without adjusting for COPD and only using unrelated controls. Results are subject to limited sample size, calling for effective multi-center collaborations.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to thank Susan Ernst, M.A., for her technical assistance with the manuscript.

## 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;40:616–22. [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;452:633–7. [PubMed: 18385738]
3. Liu P, Vikis HG, Wang D, et al. Familial aggregation of common sequence variants on 15q24–25.1 in lung cancer. J Natl Cancer Inst 2008;100:1326–30. [PubMed: 18780872]
4. McKay JD, Hung RJ, Gaborieau V, et al. Lung cancer susceptibility locus at 5p15.33. Nat Genet 2008;40:1404–6. [PubMed: 18978790]
5. Rafnar T, Sulem P, Stacey SN, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nat Genet 2009;41:221–7. [PubMed: 19151717]
6. Spitz MR, Amos CI, Dong Q, Lin J, Wu X. 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;100:1552–6. [PubMed: 18957677]
7. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 2008;452:638–42. [PubMed: 18385739]
8. Wang Y, Broderick P, Webb E, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat Genet 2008;40:1407–9. [PubMed: 18978787]
9. Shiraishi K, Kohno T, Kunitoh H, et al. Contribution of nicotine acetylcholine receptor polymorphisms to lung cancer risk in a smoking-independent manner in the Japanese. Carcinogenesis 2009;30:65–70. [PubMed: 19005185]
10. Zienolddiny S, Skaug V, Landvik NE, et al. The TERT-CLPTM1L lung cancer susceptibility variant associates with higher DNA adduct formation in the lung. Carcinogenesis 2009;30:1368–71. [PubMed: 19465454]
11. Yang P, Sun Z, Krowka MJ, et al. Alpha-1 Antitrypsin Deficiency Carriers, Tobacco Smoke, Chronic Obstructive Pulmonary Disease, and Lung Cancer Risk. Archives of Internal Medicine 2008;168:1097–103. [PubMed: 18504338]
12. Punturieri A, Szabo E, Croxton TL, Shapiro SD, Dubinett SM. Lung cancer and chronic obstructive pulmonary disease: needs and opportunities for integrated research. Journal of the National Cancer Institute 2009;101:554–9. [PubMed: 19351920]
13. Cohen BH. Chronic obstructive pulmonary disease: A challenge in genetic epidemiology. American Journal of Epidemiology 1980;112(2):274–88. [PubMed: 6968157]
14. Wu AH, Fontham ETH, Reynolds P, et al. Previous lung disease and risk of lung cancer among lifetime nonsmoking women in the United States. American Journal of Epidemiology 1995;141:1023–32. [PubMed: 7771438]
15. Yang P, Allen MS, Aubry MC, et al. Clinical Features of 5,628 Primary Lung Cancer Patients: Experience at Mayo Clinic from 1997–2003. Chest 2005;128:452–62. [PubMed: 16002972]
16. de Andrade M, Ebbert JO, Wampfler JA, et al. Environmental Tobacco Smoke Exposure in Women with Lung Cancer. Lung Cancer 2004;43:127–34. [PubMed: 14739032]
17. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263–5. [PubMed: 15297300]
18. Breslow, NE.; Day, NE. Statistical Methods in Cancer Research. Lyon, France: IARC Scientific Publications No. 32; 1980. The analysis of case-control studies. The analysis of case-control studies.
19. Therneau, T.; Grambsch, P. Modeling Survival Data: Extending the Cox Model. In: Dietz, K., editor. Statistics for Biology and Health. New York: Springer-Verlag; 2000.
20. Sellers TA, Weaver TW, Phillips B, Altmann M, Rich SS. Environmental factors can confound identification of a major gene effect: results from a segregation analysis of a simulated population of lung cancer families. Genet Epidemiol 1998;15:251–62. [PubMed: 9593112]
21. Young RP, Hopkins RJ, Hay BA, et al. Lung cancer gene associated with COPD: triple whammy or possible confounding effect? Eur Respir J 2008;32:1158–64. [PubMed: 18978134]

**Table 1**

Association analysis of 5 candidate SNPs in all cases and controls with and without adjustment of COPD

| Datasets (cases/controls) and Models | rs402710 | | rs2736100 | | rs4324798 | | rs16969968 | | rs8034191 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | OR(95%CI) | P | OR(95%CI) | p | OR(95%CI) | p | OR(95%CI) | p | OR(95%CI) | p |
| **Set 1. (1735/1036) Cases vs. unrelated controls** | | | | | | | | | | |
| Univariate model | **0.86(0.76–0.96)** | **0.010** | 1.11(0.99–1.24) | 0.063 | **1.24(1.04–1.49)** | **0.020** | **1.14(1.02–1.28)** | **0.020** | **1.17(1.04–1.31)** | **0.007** |
| Multivariate model | | | | | | | | | | |
| Without COPD [a] | **0.79(0.67–0.93)** | **0.005** | 1.10(0.94–1.29) | 0.249 | 1.24(0.95–1.61) | 0.112 | 1.09(0.93–1.28) | 0.309 | 1.08(0.92–1.27) | 0.364 |
| With COPD [b] | **0.83(0.69–0.98)** | **0.033** | 1.15(0.97–1.36) | 0.117 | 1.26(0.96–1.67) | 0.099 | 1.04(0.88–1.23) | 0.671 | 1.04(0.87–1.23) | 0.691 |
| **Set 2. (658/1206) Cases vs. unaffected full sibling controls** | | | | | | | | | | |
| Univariate model | 0.90(0.80–1.01) | 0.062 | 1.05(0.94–1.18) | 0.372 | 1.00(0.84–1.19) | 0.973 | 1.08(0.97–1.21) | 0.178 | 1.08(0.96–1.20) | 0.201 |
| Multivariate model | | | | | | | | | | |
| Without COPD [a] | 0.93(0.81–1.06) | 0.253 | 1.08(0.95–1.22) | 0.252 | 1.04(0.86–1.27) | 0.698 | 1.11(0.98–1.26) | 0.112 | 1.10(0.97–1.25) | 0.129 |
| With COPD [b] | 0.95(0.84–1.08) | 0.447 | 1.06(0.94–1.21) | 0.351 | 1.06(0.87–1.28) | 0.579 | 1.07(0.94–1.21) | 0.305 | 1.07(0.94–1.21) | 0.310 |

[a] Adjusted for age at diagnosis, sex, pack year history of smoking, and life time second hand smoking.

[b] Adjusted for age at diagnosis, sex, pack year history of smoking, life time second hand smoking, and history of COPD.

**Table 2**

Association analysis of 2 candidate SNPs in five cases and controls subsets

| Datasets (cases/controls) | rs402710 | | | | rs2736100 | | | |
|---|---|---|---|---|---|---|---|---|
| | Without COPD [b] | | With COPD [a] | | Without COPD [b] | | | With COPD [a] |
| | OR(95%CI) | P | OR(95%CI) | P | OR(95%CI) | P | OR(95%CI) | P |
| **All smokers** | | | | | | | | |
| Set 3: Cases vs. unrelated controls (1406/412) | **0.78(0.66–0.91)** | **0.002** | **0.81(0.68–0.97)** | **0.018** | 1.08(0.92–1.26) | 0.351 | 1.12(0.95–1.32) | 0.184 |
| Set 4: Cases vs. unaffected full sibling controls (415/641) | 0.91(0.79–1.06) | 0.211 | 0.94(0.81–1.09) | 0.435 | 1.07(0.93–1.23) | 0.328 | 1.05(0.91–1.21) | 0.486 |
| **Heavy smokers**[†] | | | | | | | | |
| Set 5: Cases vs. unrelated Controls (771/260) | 0.85(0.69–1.04) | 0.117 | 0.92(0.74–1.16) | 0.500 | 0.99(0.81–1.21) | 0.902 | 0.97(0.78–1.21) | 0.776 |
| **Never smokers** | | | | | | | | |
| Set 7: Cases vs. unrelated controls (329/624) | 1.00(0.82–1.22) | 0.961 | 1.00(0.82–1.24) | 0.966 | 1.19(0.98–1.44) | 0.079 | **1.23(1.01–1.50)** | **0.036** |
| Set 8: Cases vs. unaffected full sibling controls (82/133) | 0.86(0.61–1.23) | 0.413 | 0.84(0.59–1.20) | 0.338 | 1.13(0.81–1.58) | 0.477 | 1.14(0.81–1.61) | 0.439 |

[a] Adjusted for age at diagnosis, sex, and history of COPD.

[b] Adjusted for age at diagnosis and sex.

[†] >= 20 pack years

**Table 3**

Summary of Allele Type Effects of Five Candidate SNPs on Lung Cancer Survival

| Patient group (number of cases) | rs402710 | | rs2736100 | | rs4324798 | | rs16969968 | | rs8034191 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | HR(95%CI) | p- value | HR(95%CI) | p- value | HR(95%CI) | p- value | HR(95%CI) | p- value | HR(95%CI) | p- value |
| **All SCLC**[a] (235) | 0.91(0.71–1.17) | 0.479 | 1.01(0.81–1.25) | 0.944 | **0.46(0.30–0.73)** | **0.001** | 1.00(0.79–1.26) | 0.99 | 1.02(0.82–1.27) | 0.880 |
| **NSCLC Ever Smokers** | | | | | | | | | | |
| With surgery [b] (849) | 1.05(0.90–1.23) | 0.522 | 0.91(0.79–1.06) | 0.227 | 1.18(0.97–1.43) | 0.102 | 0.93(0.80–1.08) | 0.336 | 0.91(0.78–1.05) | 0.202 |
| No surgery [c] (334) | 1.04(0.87–1.25) | 0.643 | 0.96(0.80–1.14) | 0.623 | 1.28(0.93–1.76) | 0.136 | 1.05(0.88–1.25) | 0.596 | 0.96(0.80–1.15) | 0.663 |
| **NSCLC Never Smokers** | | | | | | | | | | |
| With surgery [d] (211) | 1.16(0.79–1.71) | 0.448 | 1.21(0.84–1.76) | 0.311 | 1.12(0.60–2.10) | 0.730 | 1.26(0.89–1.79) | 0.186 | 1.15(0.81–1.62) | 0.435 |
| No surgery [d] (113) | 0.80(0.56–1.14) | 0.220 | 0.75(0.54–1.04) | 0.086 | 1.40(0/79,2.46) | 0.248 | 1.14(0.84–1.55) | 0.407 | 1.16(0.86–1.58) | 0.332 |

[a] Adjusted for age at diagnosis, sex, smoking status, years quit smoking, stages of lung cancer, performance status, treatment modality and disease progression/recurrence.

[b] Adjusted for age at diagnosis, sex, history of COPD, smoking status, pack year history of smoking, years quit smoking, stages and histological types of lung cancer, treatment modality and disease progression/recurrence.

[c] Adjusted for age at diagnosis, sex, history of COPD, pack year history of smoking, years quit smoking, stages and histological types of lung cancer, treatment modality and disease progression/recurrence.

[d] Adjusted for age at diagnosis, sex, second hand smoke, stages and histological types of lung cancer, treatment modality and disease progression/recurrence.