

Multistage Analysis of Variants in the Inflammation Pathway and Lung Cancer Risk in Smokers

Margaret R. Spitz¹, Ivan P. Gorlov², Qiong Dong³, Xifeng Wu³, Wei Chen⁴, David W. Chang³, Carol J. Etzel³, Neil E. Caporaso⁵, Yang Zhao⁸, David C. Christiani⁸, Paul Brennan⁹, Demetrius Albanes⁷, Jianxin Shi⁶, Michael Thun¹⁰, Maria Teresa Landi⁵, and Christopher I. Amos⁴

Abstract

Background: Tobacco-induced lung cancer is characterized by a deregulated inflammatory microenvironment. Variants in multiple genes in inflammation pathways may contribute to risk of lung cancer.

Methods: We therefore conducted a three-stage comprehensive pathway analysis (discovery, replication, and meta-analysis) of inflammation gene variants in ever-smoking lung cancer cases and controls. A discovery set (1,096 cases and 727 controls) and an independent and nonoverlapping internal replication set (1,154 cases and 1,137 controls) were derived from an ongoing case-control study. For discovery, we used an iSelect BeadChip to interrogate a comprehensive panel of 11,737 inflammation pathway single-nucleotide polymorphisms (SNP) and selected nominally significant ($P < 0.05$) SNPs for internal replication.

Results: There were six SNPs that achieved statistical significance ($P < 0.05$) in the internal replication data set with concordant risk estimates for former smokers and five concordant and replicated SNPs in current smokers. Replicated hits were further tested in a subsequent meta-analysis using external data derived from two published genome-wide association studies (GWAS) and a case-control study. Two of these variants (a *BCL2L14* SNP in former smokers and an SNP in *IL2RB* in current smokers) were further validated. In risk score analyses, there was a 26% increase in risk with each additional adverse allele when we combined the genotyped SNP and the most significant imputed SNP in *IL2RB* in current smokers and a 36% similar increase in risk for former smokers associated with genotyped and imputed *BCL2L14* SNPs.

Conclusions/Impact: Before they can be applied for risk prediction efforts, these SNPs should be subject to further external replication and more extensive fine mapping studies. *Cancer Epidemiol Biomarkers Prev*; 21(7); 1213–21. ©2012 AACR.

Introduction

Tobacco-induced lung cancer is characterized by generation of reactive oxidant species (ROS) leading to tissue destruction and an abundant and deregulated inflammatory microenvironment. Chronic airway inflammation

contributes to alterations in the bronchial epithelium and lung microenvironment, provoking a milieu conducive to pulmonary carcinogenesis (1). Selection has endowed humans with a balance between an appropriately limited inflammatory response that protects the host against infection and an abnormally prolonged or intense inflammatory response that could result in a dysfunctional immune system and create a microenvironment that might promote carcinogenesis (2). Epidemiologic evidence also supports a role of inflammation in lung carcinogenesis (3). For example, besides the well-documented association between lung cancer and obstructive pulmonary disease (with its inflammatory microenvironment), there is reported to be an increased risk of lung cancer among patients with lung infections (e.g., tuberculosis and bacterial pneumonia) as well as in immunosuppressed individuals (3). We and others have previously shown that tobacco-induced chronic obstructive airways disease, likewise characterized by a sustained inflammatory reaction in the airways and lung parenchyma, is a significant contributor to lung cancer risk in smokers (4, 5). However, there is considerable inter-individual variation in susceptibility of long-term smokers to develop chronic

Authors' Affiliations: ¹Dan L. Duncan Cancer Center, Baylor College of Medicine; Departments of ²Genitourinary Medical Oncology, ³Epidemiology, and ⁴Genetics, The University of Texas MD Anderson Cancer Center, Houston, Texas; ⁵Genetic Epidemiology Branch, ⁶Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, ⁷Nutritional Epidemiology Branch, NIH, Department of Health and Human Services, Bethesda, Maryland; ⁸Environmental and Occupational Medicine and Epidemiology Program, Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts; ⁹International Agency for Research on Cancer (IARC), Lyon, France; and ¹⁰Epidemiology & Surveillance Research, American Cancer Society, Atlanta, Georgia

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Author: Margaret R. Spitz, Dan L. Duncan Cancer Center, Baylor College of Medicine, One Baylor Plaza, MS:BCM305, Ste. 450A, Cullen Bldg, Houston, TX 77030. Phone: 713-798-2115; Fax: 713-798-2716; E-mail: spitz@bcm.edu

doi: 10.1158/1055-9965.EPI-12-0352-T

©2012 American Association for Cancer Research.

obstructive airways disease and/or lung cancer. There is also extensive evidence of familial aggregation of both diseases suggesting that genetic components exist. Genetic variants in key inflammation-related genes could alter gene function and cause a shift in balance, resulting in deregulation of the inflammatory response and corresponding modulation of susceptibility to cigarette-induced normal tissue damage (6).

Loza and colleagues (6) have stressed the advantages of conducting pathway-focused analyses, using predefined functional subpathways to evaluate biologically feasible interactions. We therefore conducted an in-depth 3-stage analysis (discovery, replication, and meta-analysis) of gene variants in inflammatory pathways as susceptibility factors for lung cancer in ever-smokers to evaluate their role in the context of relevant covariates. A parallel study in never-smokers has been previously reported (7).

Materials and Methods

Study subjects

Discovery and internal replication sample. The discovery and replication populations were nonoverlapping sets of cases and controls derived from an ongoing multiracial/ethnic lung cancer case-control study at MD Anderson Cancer Center (Houston, TX; refs. 8, 9). Cases were consecutive patients with newly diagnosed, histopathologically confirmed, and previously untreated non-small cell lung cancer with no age, gender, ethnicity, tumor histology, or disease stage restrictions. Medical history, family history of cancer, smoking habits, and occupational history were obtained through an interviewer-administered risk factor questionnaire. Institutional review board approval at MD Anderson Cancer Center was obtained for this study. Case exclusion criteria included a history of prior cancer, prior chemotherapy or radiotherapy for the lung cancer, or recent blood transfusion.

We recruited our control population from the Kelsey-Seybold Foundation, Houston's largest multidisciplinary physician practice (9). Potential controls were first surveyed with a short questionnaire for their willingness to participate in research studies and provide preliminary data for matching demographic characteristics with those of cases (9). Controls were frequency-matched to the cases on the basis of age (± 5 years), sex, smoking status, and ethnicity. Control exclusion criteria for the study included prior chemotherapy or radiotherapy or recent blood transfusion and any previous cancer. To date, the response rate among both the cases and controls has been approximately 75%. Upon receiving informed consent, a 40-mL blood sample was drawn into coded heparinized tubes from all study participants for the assays. Genomic DNA was extracted from peripheral blood lymphocytes and stored at -80°C until use.

This analysis focuses on Caucasian case and control subjects who reported being ever-smokers, that is, had smoked more than 100 cigarettes over a lifetime. Former

smokers were defined as those who had quit smoking more than a year before their diagnosis (cases) or before interview (controls). The internal replication set was composed of the 1,154 ever-smoking Caucasian lung cancer cases and 1,137 controls that was the population used for the published lung cancer genome-wide association studies (GWAS) conducted by Amos and colleagues and who were enrolled into the case-control study from August 1995 through October 2005 (10). The discovery set was based on case and control subjects who were not included in the lung cancer GWAS and who were selected from the entire lung study database through October 2008 (excluding those enrolled in the GWAS) on the basis of histology (non-small cell lung cancer), ethnicity (Caucasian), and ever-smoking status.

Meta-analysis sample. For the third-stage meta-analysis, 3 additional studies (2 GWAS and a case-control study) contributed data using the same inclusion criteria of non-small cell lung cancer in Caucasian ever-smokers. The International Agency for Research on Cancer (IARC) GWAS (ref. 11; 1,426 cases and 1,564 controls) included a lung cancer case-control study conducted in 6 central European countries (Czech Republic, Hungary, Poland, Romania, Russia, and Slovakia). The NCI GWAS (ref. 12; 3,164 cases and 2,983 controls) was drawn from a population-based case-control study, The Environment And Genetics in Lung cancer Etiology (EAGLE) study in Lombardy, Northern Italy, as well as 3 cohort studies; specifically: the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC), a randomized primary prevention trial including nearly 30,000 male smokers enrolled in Finland between 1985 and 1993; the Prostate, Lung, Colon, Ovary Screening Trial (PLCO), a randomized trial including 150,000 individuals enrolled in 10 U.S. study centers between 1992 and 2001; and the Cancer Prevention Study II Nutrition Cohort (CPS-II), of more than 183,000 subjects enrolled by the American Cancer Society (ACS) between 1992 and 2001 across all U.S. states. The third data set for meta-analysis was derived from a case-control study of lung cancer at Massachusetts General Hospital (Boston, MA). Patients were recruited between December 1992 and April 2007. Controls were either case-related (healthy friends and spouses) or case-unrelated (friends or spouses of other hospital patients from oncology or thoracic surgery units). This study included 892 cases and 809 controls for whom genotyping data were available (13). Each study was approved by institutional review boards, and participants signed an informed consent.

Gene and single-nucleotide polymorphism selection. A comprehensive list of candidate genes for the discovery phase was selected as reported previously (7) using the Gene Oncology database (14) and the National Center for Biotechnology Information (NCBI) PubMed (15) to identify inflammation pathway-related genes. We also used the inflammation pathway gene list and functionally defined subpathways as outlined in the study by Loza and colleagues (6). For each gene, we selected tagging single-nucleotide polymorphisms (SNP; tagSNPs) located

within 10-kb upstream of the transcriptional start site or 10-kb downstream of the transcriptional stop site based on data from the International HapMap Project (16). Using the linkage disequilibrium (LD), select program (17), and the UCSC Golden Path Gene Sorter program (18), we further divided identified SNPs into bins based on an r^2 threshold of 0.8 and minor allele frequency (MAF) greater than 0.05 in Caucasians to select tagging SNPs. We also included additional inflammation pathway SNPs located in coding (synonymous and nonsynonymous SNPs) and regulatory regions [promoter, splicing site, 5-untranslated region (UTR), and 3-UTR] of inflammation-related genes. Functional SNPs and SNPs in inflammation pathway genes previously reported to be associated with cancer were also included. The complete set of selected SNPs was submitted to Illumina technical support for Infinium chemistry designability, beadtype analyses, and iSelect Infinium Beadchip synthesis.

Genotyping. A total of 11,930 SNPs mapped to 904 genes that were in or near inflammation pathways (Supplementary Table) were genotyped in the discovery samples using Illumina's Infinium iSelect HD Custom Genotyping BeadChip according to the standard 3-day protocol. Genotypes were autocalled using the BeadStudio Software. We excluded any SNP with a call rate lower than 95% or with MAF = 0. The duplicate sample error rate of 0.137% was derived from 64 duplicate samples. The final set included 11,737 SNPs in the inflammation pathway.

Statistical analysis

For each SNP, Hardy–Weinberg equilibrium was assessed among the controls using a χ^2 test. All subsequent analyses were stratified by smoking status (former and current). Single SNP association tests were conducted using PLINK 1.07 (19). Unconditional logistic regression analysis, implemented using SAS version 9.2, was used to calculate the odds ratios (OR) and 95% confidence intervals (CI) for the association between a single locus and lung cancer risk with and without adjustment for age, sex, or smoking intensity and assuming an additive model on the logistic scale.

We applied the Bayesian false discovery probability (BFDP) test (20) to evaluate the chance of obtaining a false-positive association during the replication and validation stages. This approach calculates the probability of declaring no association given the data and a specified prior on the presence of an association and has a noteworthy threshold that is defined in terms of the costs of false discovery and nondiscovery. We set a level of noteworthiness for BFDP at 0.8, that is, false nondiscovery rate is 4 times as costly as false discovery. We tested priors ranging from 0.01 through 0.07 and ORs from 1.2 through 1.5. For this analysis, we used a prior of 0.05 to determine that the association was unlikely to represent a false-positive result and selected an OR of 1.5 as this was a hypothesis-driven pathway-based, rather than an agnostic, approach.

For the stage 2 analysis, we conducted an internal fast-track replication separately for current and former smokers, by testing those significant SNPs ($P < 0.05$) from the discovery phase in an independent set of cases and controls drawn from the same study population source and that had been included in the published GWAS, for which the HumanHap300 BeadChip was used for genotyping. For imputation of ungenotyped SNPs, we applied the MACH1.0.16 program (21) using HapMap 2 CEU reference data (release 21) for GWAS data plus 1,000 Genomes CEU reference data (March 2010 release) for the candidate regions, that had imputation R^2 values ≥ 0.8 . LD for the top SNPs was visualized using Haploview v. 4.1 (22) to summarize R^2 statistics.

For the meta-analysis, we combined data from the 2 published GWAS and the case–control study using R Software version 1.6-1 (23). Between-study heterogeneity was tested by the Cochran Q test, with $P < 0.05$ as the significance level. If there was evidence of significant heterogeneity, we applied the random-effects model, using the method of DerSimonian and Laird (24). For the fixed-effects model, we used the Mantel–Haenszel method. The significance of the pooled OR was determined by the Z test, and $P < 0.05$ was considered as statistically significant. All the P values reported here were 2-sided.

Results

Discovery phase

The discovery set (Table 1) included 1,096 case patients and 727 control subjects (608 former smoking cases and 325 controls and 488 current smoker cases and 402 controls). The cases (64.7 years) were older than the controls (57.5 years). This exceeds the 5-year matching criterion and reflects ongoing incomplete control recruitment as frequency matching is used. The cases were also heavier smokers, both in terms of cigarettes per day (27.4 vs. 23.7) and years smoked (36.8 vs. 29.5). Former smoker cases were more likely to have quit at an older age than their respective controls. Cases were also more likely to report a prior history of physician-diagnosed emphysema, dust exposure, family history of cancer in first-degree relatives, exposure to asbestos, and less likely to report suffering from hay fever. The distribution of genes by functional subpathway and number of SNPs/gene included in the discovery phase is summarized in the Supplementary Table. Before selection of SNPs for replication, there were 653 SNPs for former smokers and 608 SNPs for current smokers that achieved nominal P values of <0.05 in the discovery set.

Internal replication

The independent replication set included 1,154 cases and 1,137 controls from our GWAS (603 and 657, respectively, former smoking cases and controls; 551 and 480 current smoking cases and controls). The distribution of risk factors between cases and controls in the second phase closely resembled those in the discovery phase.

Table 1. Characteristics of ever-smoking lung cancer cases and controls in discovery and replication populations, MD Anderson Cancer Center

Characteristic	Discovery set		Replication set	
	Cases (N = 1,096)	Controls (N = 727)	Cases (N = 1,154)	Controls (N = 1,137)
Male	638 (58.21)	387 (53.23)	658 (57.02)	644 (56.64)
Female	458 (41.79)	340 (46.77)	496 (42.98)	493 (43.36)
Mean age (SD), y	64.7 (9.7)	57.5 (13.2)	62.1 (10.8)	61.1 (8.9)
Smoking status				
Current	488 (44.5)	402 (55.3)	551 (47.75)	480 (42.22)
Former	608 (55.5)	325 (44.7)	603 (52.25)	657 (57.78)
No. of cigarettes/d ^a				
Mean (SD)	27.4 (13.5)	23.7 (14.4)	28.0 (13.6)	26.6 (14.3)
Years smoked				
Mean (SD)	36.8 (13.0)	29.5 (13.5)	35.9 (12.6)	32.8 (12.7)
Pack-years				
Mean (SD)	52.4 (33.2)	37.3 (29.6)	51.5 (31.4)	44.6 (30.2)
Age stopped smoking ^b				
<42	134 (22.26)	109 (36.33)	141 (23.38)	205 (31.20)
42–53	202 (33.55)	92 (30.67)	188 (31.18)	238 (36.23)
≥54	266 (44.19)	99 (33.00)	274 (45.44)	214 (32.57)
Dust exposure				
Yes	286 (45.76)	177 (30.57)	499 (44.28)	374 (32.89)
Emphysema				
Yes	287 (26.23)	56 (9.52)	268 (23.70)	102 (8.99)
Hay fever				
Yes	96 (16.52)	111 (19.10)	173 (15.34)	245 (21.59)
Asbestos exposure				
Yes	93 (12.81)	63 (9.62)	159 (13.78)	105 (9.23)
Family history of smoking-related cancers				
0	710 (65.38)	535 (76.54)	791 (68.84)	865 (76.28)
1+	376 (34.62)	164 (23.46)	358 (31.16)	269 (23.72)

^aAverage lifetime;
^bFormer smokers only.

From the SNPs that were statistically significant in the discovery phase, we matched 295 SNPs that were directly genotyped from the GWAS database and an additional 200 imputed SNPs. The remaining 741 SNPs could not be evaluated. Using these genotyped and imputed data for the replication phase, we conducted logistic regression analysis assuming an additive model for each available SNP and conducting separate analyses for current and former smokers. On univariate analysis, there were 6 SNPs that achieved statistical significance ($P < 0.05$) in the replication data set with risk estimates concordant for direction for former smokers and 5 SNPs that were concordant for risk estimates and also achieved significance in current smokers (Table 2). For the 2 SNPs in *TGFB1*, r^2 for LD was 0.412 and for the 2 SNPs in *IL2RB*, r^2 for LD was 0.586. All other SNPs that were statistically significant on univariate analysis had r^2 values that were <0.002 .

We also applied the BFDP test (20) to evaluate the likelihood of any of these 11 SNP associations being false-positive associations. On the basis of the criteria outlined above, 10 of the 11 SNPs had BFDP < 0.8 and rs1544669 in former smokers had BFDP = 0.86. Similar results were obtained for OR = 1.2 and priors of 0.03 and 0.07.

In current smokers, the significant SNPs belonged to either the leukocyte (*ICOS*, rs1896286 and *CD86*, rs12106790) or cytokine (*IL2RB*, rs1003694 and rs2235330 and *CSF2RB*, rs20722707) signaling subpathways. In former smokers, besides the leukocyte (*LILRA4*, rs1205316) and cytokine (*TGFB1*, rs 2241715 and rs4803455 and *IL15RA*, rs17146857) signaling pathways, rs4747064 in *PRF1* (ROS/glutathione/cytotoxic granules) and rs1544669 in *BCL2L14* (apoptosis signaling) were also replicated. Of the 741 SNPs significant in the discovery phase that we could not internally replicate, 12 SNPs belonged to genes with replicated SNPs, including 4 in

Table 2. SNPs significant in discovery set and verified in replication set

Chromosome	SNP	BP	Location	Gene name	Minor allele ^a	MAF	OR (95%CI)	P	OR (95%CI)	P
Former smokers										
10	rs17146857 ^b	6,028,077	Flanking 3'-UTR	<i>FBXO18</i> <i>IL15RA</i>	A	0.15	1.34 (1.01–1.77)	0.041	1.39 (1.12–1.74)	0.003
10	rs4747064	72,018,762	Flanking 3'-UTR	<i>PRF1</i>	A	0.26	0.77 (0.62–0.96)	0.018	0.83 (0.69–0.99)	0.034
12	rs1544669^b	12,110,294	Flanking 5'-UTR	<i>BCL2L14</i>	C	0.35	0.82 (0.67–1.00)	0.048	0.81 (0.67–0.97)	0.025
19	rs1205316 ^b	59,531,270	Flanking 3'-UTR	<i>LILRA4</i>	A	0.47	1.32 (1.06–1.64)	0.012	1.38 (1.01–1.88)	0.043
19	rs2241715	46,548,726	Intron	<i>TGFB1</i>	A	0.32	0.75 (0.61–0.93)	0.008	0.77 (0.65–0.91)	0.002
19	rs4803455	46,543,349	Intron	<i>TGFB1</i>	A	0.27	1.24 (1.02–1.51)	0.029	1.22 (1.05–1.43)	0.011
Current smokers										
2	rs1896286	204,540,683	Flanking 3'-UTR	<i>ICOS</i>	C	0.36	1.22 (1.01–1.48)	0.044	1.23 (1.02–1.48)	0.028
3	rs12106790	123,249,744	Flanking 5'-UTR	<i>CD86</i>	C	0.21	0.79 (0.63–1.00)	0.048	0.78 (0.63–0.97)	0.025
22	rs1003694	35,869,074	Intron	<i>IL2RB</i>	A	0.35	0.75 (0.61–0.92)	0.007	0.82 (0.68–0.99)	0.042
22	rs2072707	35,649,027	Intron	<i>CSF2RB</i>	A	0.28	1.24 (1.02–1.51)	0.035	1.22 (1.01–1.48)	0.043
22	rs2235330	35,869,659	Intron	<i>IL2RB</i>	G	0.2	0.79 (0.62–1.00)	0.046	0.77 (0.62–0.95)	0.014

NOTE: Bolded SNPs were replicated in the meta-analysis.

Abbreviation: BP, base position.

^aAllele change/allele frequency based on information in CEU population in HapMap.

^bImputed genotype in replication set.

ICOS, 3 each in *IL2RB* and *PRF1*, and 1 each in *CD86* and *BCL2L14*.

Meta-analysis

We elected to move all 11 SNPs on to phase III replication in the meta-analysis. Three of these SNPs were not included on the 317k chip and had to be imputed in our replication set and in the IARC data set. rs17146857 ($r^2 = 0.84$) and rs1544669 ($r^2 = 0.49$) were included, but we were unable to reliably impute rs1205316.

The meta-analysis results for the 3 external data sets are summarized in Fig. 1A and B for current and former smokers, respectively. Two SNPs were successfully replicated with ORs in similar direction. In current smokers (Fig. 1A), we replicated rs2235330 in *IL2RB* (OR = 0.92; 0.84–1.00) whereas the other *IL2RB* SNP, rs1003694, was borderline significant (OR = 0.95; 0.88–1.02). Both SNPs in *IL2RB* are intronic. For former smokers (Fig. 1B), rs1544669 in *BCL2L14* in a 5'-UTR flanking region was associated with a summary OR of 0.91 (0.83–1.00).

There were no substantial differences when the discovery and replication data were stratified by gender, histology, or smoking intensity. To verify whether there were stronger signals in the 2 identified loci, we conducted SNP imputation in the discovery data set to increase coverage in the regions surrounding rs1003694 and rs2235330 for current smokers and around rs1544669 in former smokers, based on the 1000 genomes March 2010 and June 2010 release CEU reference panels using MACH version 1.016 (21). For *IL2RB*, there were 91 genotyped SNPs 1 Mb from each side of the gene used for imputation. A/T or C/G SNPs were checked for strand orientation and flipped to forward strand as needed before the imputation. There were 480 imputed SNPs with R^2 for imputation quality over 0.8. The most significant SNPs (Fig. 2A) were

rs5995385 (genotyped) and rs5750383 and rs5756527 (both imputed), all with $P = 0.001124$ and in tight LD. Because these SNPs were not included in the GWAS, we were not able to attempt replication. They are all located in the flanking 3'-UTR region of *IL2RB* that is not highly conserved but could tag functional polymorphisms in *IL2RB*. We conducted a similar imputation analysis for *BCL2L14* and identified 351 imputed SNPs with R^2 for imputation quality more than 0.8 (Fig. 2B). The top 3 SNPs, all with P value < 0.004 (rs11054701, rs2075241, and rs2160521) were in complete LD, and only one (rs2075241) was selected for inclusion in the risk score.

Risk score

We computed genetic risk scores by summing the adverse alleles from the replicated and most significant imputed SNPs in *IL2RB* for current smokers and *BCL2L14* SNPs for former smokers (Table 3). We restricted this analysis to our discovery data set for which we had conducted imputation. For current smokers, compared with those carrying 0 or 1 adverse allele, risks increased from 1.19 (0.73–1.94) for those carrying 2 adverse alleles; 1.45 (0.91–2.31) for those carrying 3 adverse alleles; and 2.11 (1.25–3.55) for those with 4 adverse alleles. There was a 26% increase in risk with each additional allele ($P_{\text{trend}} = 0.002$). For former smokers, the risk for those carrying 3 or 4 adverse alleles was 2.84 (1.47–5.48) and there was a 36% increase in risk for each adverse allele ($P_{\text{trend}} = 0.0005$). There was no association between the number of adverse alleles (risk score) and smoking intensity (pack-years) in either current or former smokers. We also conducted logistic regression analyses by including the SNPs as independent variables in the model along with the covariates of age, sex, cigarette pack-years, and smoking-related family history of cancer: for former smokers, the

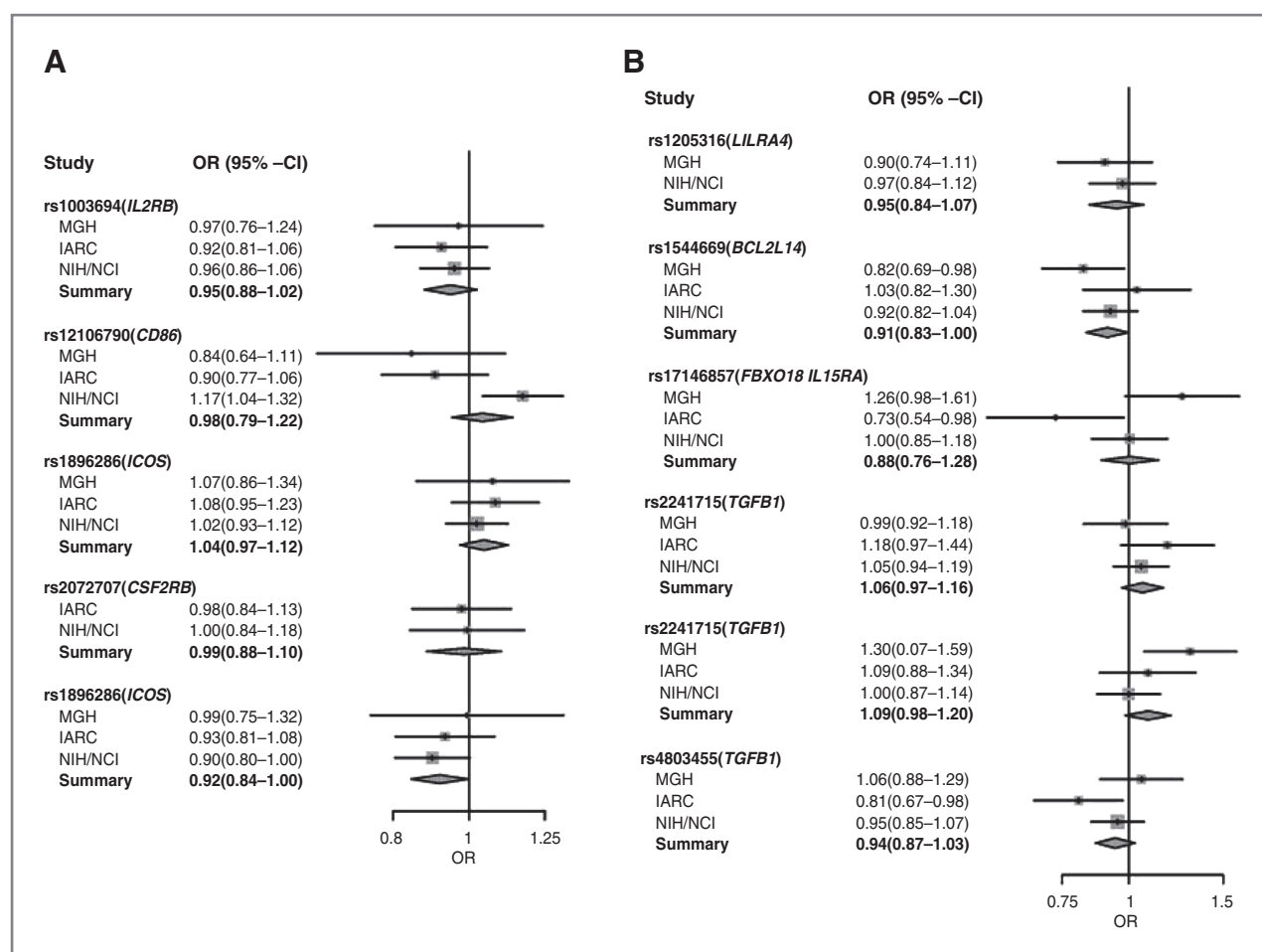


Figure 1. Forest plots for current (A) and former smokers (B). The squares and horizontal lines correspond to the study-specific ORs and 95% CIs. The area of the squares reflects the weight (inverse of the variance). The diamond represents the summary ORs and 95% CIs.

ORs were 1.55 (1.17–2.07, $P = 0.0026$) for rs2075251 and 0.79 (0.64–0.98, $P = 0.0279$) for rs1544669. In current smokers, the comparable ORs were 0.76 (0.61–0.95, $P = 0.0154$) for rs5995385 and 0.84 (0.64–1.11, $P = 0.2214$) for rs1544669 (data not shown).

Discussion

In this multistage analysis, using independent sets of cases and controls for discovery and replication, we were able to successfully replicate 6 SNPs in the inflammation pathways in former smokers and 5 different SNPs in current smokers that were statistically significant in both groups with almost identical risk estimates in the discovery and replication phases. In a subsequent meta-analysis from 3 additional external studies, 2 of these variants achieved statistical significance. These were rs1544669 in *BCL2L14* in former smokers and rs2235330 in *IL2RB* in current smokers.

Inflammation is a physiologic response to cellular and tissue damage. Appropriate response to this damage is tightly regulated through a balance between pro- and anti-

inflammatory cytokines and signaling molecules (26). It is suggested, in fact, that the tumor microenvironment and especially its inflammatory component may be a critical element of carcinogenesis (27). Because common variations in a single gene contribute only modestly to risk, it seems logical that rather than focusing on a few SNPs and/or genes with the strongest evidence of disease association, one considers multiple variants in interacting or related genes in the same pathway to improve the power to detect causal pathways and disease mechanisms (28). Loza and colleagues (6) constructed a comprehensive inflammation pathway gene list and functionally defined subpathways that formed the basis for our own analysis.

The replicated SNP in current smokers was in the interleukin-2 receptor subunit beta (*IL2RB*) gene, a cytokine signaling gene. IL-2 exerts both stimulatory and regulatory functions in the immune system and is a member of the cytokine family that is central to immune homeostasis (29). IL-2 binds to its receptor complex, IL-2R- α , β , and γ chains, and exerts its effect via second messengers, mainly tyrosine kinases, which ultimately are involved in T-cell-mediated immune responses. Local

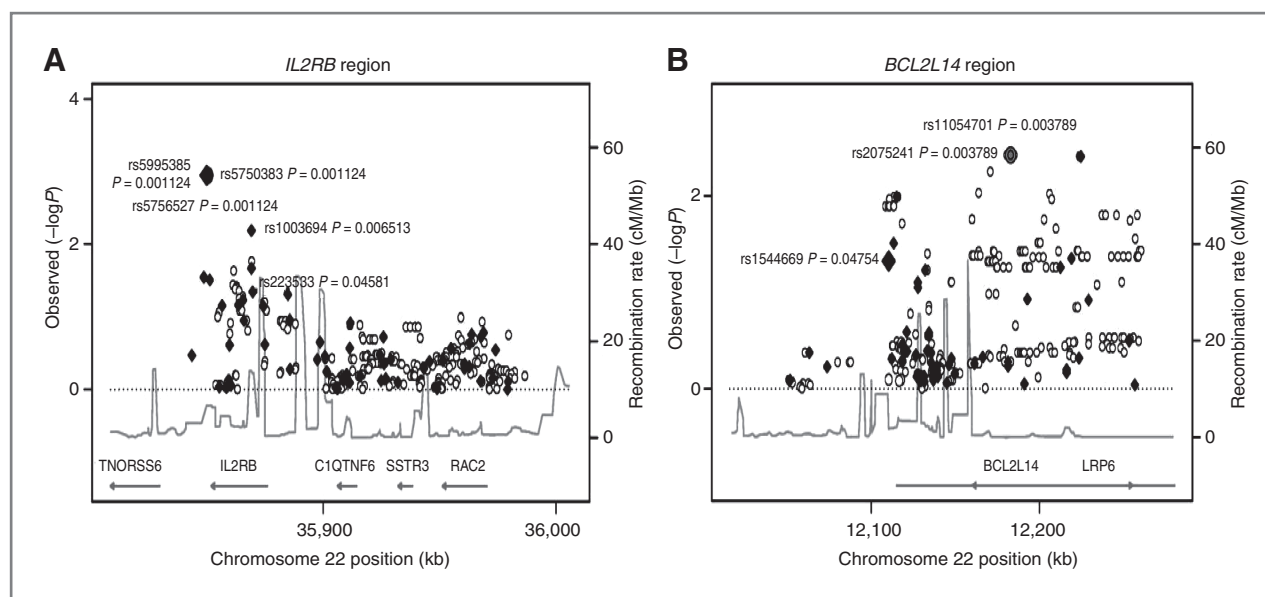


Figure 2. A, association of imputed and genotyped SNPs in the chromosome 22 region around *IL2RB* with lung cancer risk in current smokers. B, association of imputed and genotyped SNPs in the chromosome 12 region around *BCL2L14* with lung cancer risk in former smokers. Chromosomal position is on the x-axis and negative logarithm to the base 10 of the *P* values from logistic regression analysis is on the y-axis. Genotyped SNPs are plotted as filled diamonds and imputed SNPs as open circles. The overall structure of the LD with SNPs in this region is reflected by estimated recombination rates from genetic map of HapMap in build 36 coordinates. The strength of the pairwise correlation between the surrounding markers and the most significant SNPs (rs5995385 in current smokers and rs2075241 in former smokers) is reflected by the size of the symbols: the larger the size, the stronger the LD. LD was calculated from actual genotyped or imputed data using PLINK. Genes in the region are annotated with location, range, and orientation using gene annotations from the UCSC genome browser (downloaded from Broad Institute website). Original files downloaded are in build 35 positions, converted to build 36 positions (25).

blockade of the β -chain of the IL-2R restored an immunosuppressive cytokine milieu that ameliorated both inflammation and airway hyperresponsiveness in experimental allergic asthma (30). IL-2 is the major growth factor for activated T lymphocytes and stimulates clonal expansion and maturation of these lymphocytes.

In former smokers, we replicated an SNP in the *BCL2L14* gene, located on chromosome 12, and that encodes apoptosis facilitator Bcl-2-like protein 14. LOH of the short arm of chromosome 12 is a frequent event in both hematologic malignancies and solid tumors (31). *BCL2L14*, also known as *Bcl-G*, is a proapoptotic member

Table 3. Genetic risk score

No. of adverse alleles	Cases	Controls	Adjusted OR (95%CI) ^b	<i>P</i>
Current smokers^a				
0–1	58 (11.89)	66 (16.42)		
2	121 (24.80)	123 (30.60)	1.19 (0.73–1.94)	0.480
3	190 (38.93)	144 (35.82)	1.45 (0.91–2.31)	0.115
4	119 (24.39)	69 (17.16)	2.11 (1.25–3.55)	0.005
<i>P</i> _{trend}			1.26 (1.09–1.46)	0.002
Former smokers^c				
0	45 (7.40)	31 (9.54)		
1	199 (32.73)	125 (38.46)	1.14 (0.67–1.96)	0.629
2	261 (42.93)	142 (43.69)	1.37 (0.81–2.33)	0.244
3–4	94 (15.46)	26 (8.00)	2.84 (1.47–5.48)	0.002
<i>P</i> _{trend}			1.36 (1.15–1.62)	0.0005

^ars2235330 (*IL2RB*) and rs5995385 (*IL2RB*).

^bAdjusted for age, sex, pack-years, and family history of smoking-related cancers.

^crs2075241 (*BCL2L14*) and rs1544669 (*BCL2L14*).

of the Bcl-2 family which regulates cell death (32). This gene has been shown to be a transcriptional target of TP53 (33) and is a candidate tumor suppressor. Association of this gene with lung cancer has not been previously reported. Because apoptosis plays an essential role in protecting against cellular carcinogenesis, such as one due to oxidative damage from cigarette smoke, it is biologically plausible that this cell death regulator might influence the pathogenesis of lung cancer through the TP53 pathway. It is not surprising that different findings were observed in current versus former smokers. For example, current and former smokers differ with respect to the role that COX-2 plays in maintaining bronchial epithelial proliferation (34), and differences between these 2 groups with respect to bronchial epithelial biology have been reported (35, 36). Different biology is also suggested by the observation that active smokers have fared poorly in large-scale chemoprevention trials, whereas former smokers have exhibited no effect or favorable trends (34).

There are a few published, generally small, candidate gene studies on polymorphisms in inflammation-related genes that have been linked with increased lung cancer risk but with limited replication of the study results, as reviewed by Engels (3). For example, Hart and colleagues (37) studied 11 SNPs in 9 genes in 882 subjects and reported risk by combination of adverse genotypes, but there was no replication phase. Vogel and colleagues (38) included 7 inflammation pathway SNPs in their case-cohort study that included 428 cases. Carriers of a variant allele of *IL-10* and *IL-1B* were at increased risk, although the latter was statistically significant only in current smokers. We (39) have previously published case-control data on more than 1,000 cases and a similar number of controls from the study base and reported a significant association with an SNP in *IL1B*. In the International Lung Cancer Consortium (ILCCO), we conducted a coordinated genotyping study of 10 common variants including this *IL-1B* variant in 4,588 cases and 6,453 controls but found no association with this *IL1B* variant (40), nor was this specific *IL1B* variant included in our discovery analysis.

Our discovery and replication data are derived from a retrospective case-control analysis and therefore we are unable to effectively evaluate any meaningful association with inflammatory marker levels as prediagnostic sera are not available. However, an analysis of prediagnostic

C-reactive protein levels in the PLCO data showed a significant association with subsequent lung cancer risk (41). Another limitation of our study was the lack of genotyping for all of the SNPs included in our discovery study so that we could not validate risk score predictions. Further studies that independently evaluate the risk scores we developed are needed.

There are clear advantages to this pathway-based approach. Because we restricted our analysis to a specific pathway, we have reduced to some extent the issue of false-positive reporting and increased the power of our analyses. Our genes are classified by functional subpathways and thus we were able to evaluate associations in a more biologically driven manner. We cannot exclude the possibility that despite our 3-stage analytic approach, the findings represent false-positives. However, the relatively large sample sizes with dual discovery and replication populations, detailed epidemiologic information, and comprehensive query of genes and SNPs ensure robust power and greater genetic coverage to detect true-positive findings. Follow-up analysis in African-American lung cancer cases and controls is ongoing to confirm and extend these findings.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: M.R. Spitz, I.P. Gorlov, X. Wu, C.J. Etzel, D.C. Christiani, C.I. Amos

Development of methodology: M.R. Spitz, I.P. Gorlov, X. Wu, D.W. Chang, C.J. Etzel, C.I. Amos

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M.R. Spitz, X. Wu, C.J. Etzel, N.E. Caporaso, D.C. Christiani, D. Albanes, M. Thun, M.T. Landi, C.I. Amos

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M.R. Spitz, Q. Dong, W. Chen, N.E. Caporaso, Y. Zhao, J. Shi, C.I. Amos

Writing, review, and/or revision of the manuscript: M.R. Spitz, X. Wu, D. W. Chang, C.J. Etzel, N.E. Caporaso, D.C. Christiani, P.J. Brennan, D. Albanes, M. Thun, M.T. Landi, C.I. Amos

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M.R. Spitz, N.E. Caporaso
Study supervision: M.R. Spitz, X. Wu, N.E. Caporaso

Grant Support

The study was supported by RO1CA55769 and RO1CA127219 (to M.R. Spitz); RO1CA074386 and P01CA090578 (to D.C. Christiani); U19 CA148127, RP100443, and CA121197 (to C.I. Amos).

Received March 22, 2012; revised April 27, 2012; accepted April 27, 2012; published OnlineFirst May 9, 2012.

References

- Lee G, Walser TC, Dubinett SM. Chronic inflammation, chronic obstructive pulmonary disease and lung cancer. *Curr Opin Pulm Med* 2009;15:303-7.
- Forrester JS, Bick-Forrester J. Persistence of inflammatory cytokines cause a spectrum of chronic progressive diseases: implications for therapy. *Med Hypotheses* 2005;65:227-31.
- Engels EA. Inflammation in the development of lung cancer: epidemiological evidence. *Expert Rev Anticancer Ther* 2008;8:605-15.
- Spitz MR, Hong WK, Amos CI, Schabath MB, Dong Q, Shete S, et al. A risk model for prediction of lung cancer. *J Natl Cancer Inst* 2007;99:715-26.
- Brenner DR, McLaughlin JR, Hung RJ. Previous lung diseases and lung cancer risk: a systematic review and meta-analysis. *PLoS One* 2011;6:e17479.
- Loza MJ, McCall CE, Li L, Isaacs WB, Xu J, Chang BL. Assembly of inflammation-related genes for pathway-focused genetic analysis. *PLoS One* 2007;2:e1035.

7. Spitz MR, Gorlov IP, Amos CI, Spitz MR, Dong Q, Chen W, et al. Variants in inflammation genes are implicated in risk of lung cancer in never smokers exposed to second-hand smoke. *Cancer Discov* 2011;1:420–9.
8. Spitz MR, Wei Q, Dong Q, Amos CI, Wu X. Genetic susceptibility to lung cancer: the role of DNA damage and repair. *Cancer Epidemiol Biomarkers Prev* 2003;12:689–98.
9. Hudmon KS, Honn SE, Jiang H, Chamberlain RM, Xiang W, Ferry G, et al. Identifying and recruiting healthy control subjects from a managed care organization: a methodology for molecular epidemiological case-control studies of cancer. *Cancer Epidemiol Biomarkers Prev* 1997;6:565–71.
10. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, 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.
11. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008;452:633–7.
12. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet* 2009;85:679–91.
13. Miller DP, Liu G, De Vivo I, Lynch TJ, Wain JC, Su L, et al. Combinations of the variant genotypes of GSTP1, GSTM1, and p53 are associated with an increased lung cancer risk. *Cancer Res* 2002;62:2819–23.
14. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. [cited 2010 Aug 25]. Available from: <http://www.geneontology.org>.
15. The National Center for Biotechnology Information (NCBI). [cited 2010 Aug 25]. Available from: <http://www.ncbi.nlm.nih.gov/>.
16. International HapMap Consortium. The international HapMap project. [cited 2011 Aug 25]. Available from: <http://hapmap.ncbi.nlm.nih.gov/>.
17. Deborah A. Nickerson, Mark Rieder, Chris Carlson, Qian Yi. Documentation for IdSelect Version 1.0. Seattle, WA: University of Washington. [cited 2010 Aug 25]. Available from: <http://droog.gs.washington.edu/IdSelect.html>.
18. UCSC Genome Bioinformatics Genome Browser. [cited 2010 Aug 25]. Available from: <http://genome.ucsc.edu>.
19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75 Available from: <http://pngu.mgh.harvard.edu/purcell/plink/>
20. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* 2007;81:208–27.
21. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2011;34:816–34.
22. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–5.
23. Guido Schwarzer <sc@imbi.uni-freiburg.de> meta: Meta-Analysis with R. R package version 1.6-1 [cited 2010 Oct 28]. Available from: <http://CRAN.R-project.org/package=meta>.
24. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177–88.
25. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;316:1331–36.
26. Han J, Ulevitch RJ. Limiting inflammatory responses during activation of innate immunity. *Nat Immunol* 2005;6:1198–205.
27. Prendergast GC, Jaffee EM. Cancer immunologists and cancer biologists: why we didn't talk then but need to now. *Cancer Res* 2007;67:3500–4.
28. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genome-wide association studies. *Am J Hum Genet* 2007;81:1278–83.
29. Krieg C, Létourneau S, Pantaleo G, Boyman O. Improved IL-2 immunotherapy by selective stimulation of IL-2 receptors on lymphocytes and endothelial cells. *Proc Natl Acad Sci U S A* 2010;107:11906–11.
30. Doganci A, Karwot R, Maxeiner JH, Scholtes P, Schmitt E, Neurath MF, et al. IL-2 receptor beta-chain signaling controls immunosuppressive CD4+ T cells in the draining lymph nodes and lung during allergic airway inflammation *in vivo*. *J Immunol* 2008;181:1917–26.
31. Montpetit A, Boily G, Sinnett D. A detailed transcriptional map of the chromosome 12p12 tumour suppressor locus. *Eur J Hum Genet* 2002;10:62–71.
32. Guo B, Godzik A, Reed JC. Bcl-G, a novel pro-apoptotic member of the Bcl-2 family. *J Biol Chem* 2001;276:2780–5.
33. Miled C, Pontoglio M, Garbay S, Yaniv M, Weitzman JB. A genomic map of p53 binding sites identifies novel p53 targets involved in an apoptotic network. *Cancer Res* 2005;65:5096–104.
34. Kim ES, Hong WK, Lee JJ, Mao L, Morice RC, Liu DD, et al. Biological activity of celecoxib in the bronchial epithelium of current and former smokers. *Cancer Prev Res* 2010;3:148–59.
35. Lee JJ, Liu D, Lee JS, Kurie JM, Khuri FR, Ibarguen H, et al. Long-term impact of smoking on lung epithelial proliferation in current and former smokers. *J Natl Cancer Inst* 2001;93:1081–8.
36. Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, et al. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci U S A* 2004;101:10143–8.
37. Hart K, Landvik NE, Lind H, Skaug V, Haugen A, Zienolddiny S. A combination of functional polymorphisms in the CASP8, MMP1, IL10 and SEPS1 genes affects risk of non-small cell lung cancer. *Lung Cancer* 2011;71:123–9.
38. Vogel U, Christensen J, Wallin H, Friis S, Nexø BA, Raaschou-Nielsen O, et al. Polymorphisms in genes involved in the inflammatory response and interaction with NSAID use or smoking in relation to lung cancer risk in a prospective study. *Mutat Res* 2008;639:89–100.
39. Engels EA, Wu X, Gu J, Dong Q, Liu J, Spitz MR. Systematic evaluation of genetic variants in the inflammation pathway and risk of lung cancer. *Cancer Res* 2007;67:6520–7.
40. Truong T, Sauter W, McKay JD, Hosgood HD III, Gallagher C, Amos CI, et al. International Lung Cancer Consortium: coordinated association study of 10 potential lung cancer susceptibility variants. *Carcinogenesis* 2010;31:625–33.
41. Chaturvedi AK, Caporaso NE, Katki HA, Wong HL, Chatterjee N, Pine SR, et al. C-reactive protein and risk of lung cancer. *J Clin Oncol* 2010;28:2719–26.