



Cancer Research

A Susceptibility Locus on Chromosome 6q Greatly Increases Lung Cancer Risk among Light and Never Smokers

Christopher I. Amos, Susan M. Pinney, Yafang Li, et al.

Cancer Res 2010;70:2359-2367. Published OnlineFirst March 9, 2010.

Updated version	Access the most recent version of this article at: doi: 10.1158/0008-5472.CAN-09-3096
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2010/03/08/0008-5472.CAN-09-3096.DC1.html

Cited Articles	This article cites by 28 articles, 7 of which you can access for free at: http://cancerres.aacrjournals.org/content/70/6/2359.full.html#ref-list-1
Citing articles	This article has been cited by 3 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/70/6/2359.full.html#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org .

A Susceptibility Locus on Chromosome 6q Greatly Increases Lung Cancer Risk among Light and Never Smokers

Christopher I. Amos¹, Susan M. Pinney², Yafang Li¹, Elena Kupert², Juwon Lee², Mariza A. de Andrade³, Ping Yang³, Ann G. Schwartz⁴, Pam R. Fain⁵, Adi Gazdar⁶, John Minna⁶, Jonathan S. Wiest⁷, Dong Zeng¹, Henry Rothschild⁸, Diptasri Mandal⁸, Ming You⁹, Teresa Coons¹⁰, Colette Gaba¹¹, Joan E. Bailey-Wilson¹², and Marshall W. Anderson²

Abstract

Cigarette smoking is the major cause for lung cancer, but genetic factors also affect susceptibility. We studied families that included multiple relatives affected by lung cancer. Results from linkage analysis showed strong evidence that a region of chromosome 6q affects lung cancer risk. To characterize the effects that this region of chromosome 6q region has on lung cancer risk, we identified a haplotype that segregated with lung cancer. We then performed Cox regression analysis to estimate the differential effects that smoking behaviors have on lung cancer risk according to whether each individual carried a risk-associated haplotype or could not be classified and was assigned unknown haplotypic status. We divided smoking exposures into never smokers, light smokers (<20 pack-years), moderate smokers (20 to <40 pack-years), and heavy smokers (≥40 pack-years). Comparing results according to smoking behavior stratified by carrier status, compared with never smokers, there was weakly increasing risk for increasing smoking behaviors, with the hazards ratios being 3.44, 4.91, and 5.18, respectively, for light, moderate, or heavy smokers, whereas among the individuals from families without the risk haplotype, the risks associated with smoking increased strongly with exposure, the hazards ratios being, respectively, 4.25, 9.17, and 11.89 for light, moderate, and heavy smokers. The never smoking carriers had a 4.71-fold higher risk than the never smoking individuals without known risk haplotypes. These results identify a region of chromosome 6q that increases risk for lung cancer and that confers particularly higher risks to never and light smokers. *Cancer Res*; 70(6); 2359-67. ©2010 AACR.

Introduction

More than 40 years ago, Tokuhata and Lilienfeld (1) provided clear epidemiologic evidence for familial aggregation of lung cancer after accounting for personal smoking, suggesting the possible interaction of genes and smoking behavior in the etiology of lung cancer. The familial effects were most pronounced among smoking relatives, for which the case relatives showed a 2.4-fold higher risk compared with smoking

relatives of controls. A later study by Ooi and colleagues (2) similarly showed a 2.5-fold higher risk in the relatives of cases compared with controls after conditioning on smoking behavior and age. In case-control studies, a positive family history has consistently been found to be a risk factor for lung cancer (3–5). The study of Jonsson and colleagues (5) used a population-based approach and obtained familial risks of 2.69 comparing parents of cases with controls, and this relative risk increased to 3.48 comparing parents of cases younger than 60 with age-matched controls.

Genetic modeling studies have suggested that at least some of the observed familial aggregation of lung cancer may be due to inheritance of strongly acting genetic factors. Sellers and colleagues (6) performed segregation analyses on the families studied by Ooi and colleagues (2) and found results that were compatible with Mendelian codominant inheritance of a rare major autosomal gene that acts in conjunction with cigarette smoking to produce earlier age of onset of the cancer (6). Under this model, average smoking heterozygotes had relative risks of 14, 11.8, and 6.2 at ages 50, 60, and 70, respectively, compared with average smoking noncarriers. However, the model that was fitted could not allow for a possible interaction between unmeasured genetic effects and the measured environmental factor, tobacco smoke, and could not evaluate the potential effects of multiple genetic factors. Gauderman and colleagues (7) applied a Gibbs sampling method to examine gene-environment

Authors' Affiliations: ¹Department of Epidemiology, University of Texas M.D. Anderson Cancer Center, Houston, Texas; ²University of Cincinnati, Cincinnati, Ohio; ³Mayo Clinic College of Medicine, Rochester, Minnesota; ⁴Karmanos Cancer Institute, Wayne State University, Detroit, Michigan; ⁵University of Colorado, Denver, Colorado; ⁶University of Texas Southwestern Medical Center, Dallas, Texas; ⁷National Cancer Institute, NIH, Bethesda, Maryland; ⁸Louisiana State University Health Sciences Center, New Orleans, Louisiana; ⁹Washington University, St. Louis, Missouri; ¹⁰Saccomanno Research Institute, Grand Junction, Colorado; ¹¹Medical College of Ohio, Toledo, Ohio; and ¹²National Human Genome Research Institute, Baltimore, Maryland

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Corresponding Author: Christopher I. Amos, Department of Epidemiology, University of Texas M.D. Anderson Cancer Center, 1155 Pressler Street, Unit 1340, Houston, TX 77005. Phone: 713-792-3020; Fax: 713-792-0807; E-mail: camos@mdanderson.org.

doi: 10.1158/0008-5472.CAN-09-3096

©2010 American Association for Cancer Research.

interaction models on the same lung cancer data set and found evidence for a dominant major locus with significant effects of smoking and weak evidence of gene-environmental statistical interaction.

The Genetic Epidemiology of Lung Cancer Consortium (GELCC) identified a region of chromosome 6q that cosegregates with lung cancer susceptibility in families that included four or more individuals affected with lung cancer (8). The heterogeneity LOD (HLOD) score associating lung cancer with the chromosome 6q region increased from 2.79 when families that included three or more relatives affected by lung cancer families were studied to 3.47 when studying families with four or more affected individuals, and the HLOD score increased to 4.26 when multigenerational families with five or more affected lung cancer relatives were analyzed. The HLOD score represents the \log_{10} ratio of the data from a model including linkage to the model without linked markers assuming genetic heterogeneity, that is assuming that only a proportion, α , of the families are linked to the region and that the other $1-\alpha$ families do not show evidence of linkage to this region. Hence, the best model for these multigenerational families allowing for linkage in the presence of genetic heterogeneity was $\sim 18,000$ times more likely than a model excluding linkage, and this result yields a P value of $< 1 \times 10^{-5}$ (9).

These data suggest segregation of a dominant major gene in a subset of families that show excess lung cancer. Preliminary evaluation of smoking behavior in the families studied by Bailey-Wilson and colleagues (8) provided some evidence for a differential effect of smoking among individuals who are carriers of the chromosome 6q susceptibility locus compared with noncarriers. The goal of this study is to further charac-

terize the effect that smoking behavior has on susceptibility to lung cancer according to whether or not a family is segregating a risk allele at the 6q susceptibility locus and to provide updated results from linkage analyses including an additional 41 families that have been genotyped since our first report.

Materials and Methods

The methods for sample collection have been summarized by Bailey-Wilson and colleagues (8). Samples and data have been collected by the familial lung cancer recruitment sites of the GELCC: University of Cincinnati, University of Colorado, Karmanos Cancer Institute, Louisiana State University Health Sciences Center, Mayo Clinic, Johns Hopkins University, and Medical College of Ohio. Of the 28,085 lung cancer cases screened at GELCC sites for use in this report, 23.7% had at least one first-degree relative with lung cancer (details by data collection site shown in Table 1). Families were identified from the Mayo Clinic and Karmanos Cancer Institute as a part of ongoing case series based in these hospitals. All other sites accrued patients by physician referral, and in addition, some patients were self-referred to the Johns Hopkins, Karmanos Cancer Institute, and University of Cincinnati sites. All sites accrued participants to Institutional Review Board (IRB)-approved protocols and obtained informed consent from each participant, and the analytic site at the University of Texas M.D. Anderson Cancer Center also maintained an IRB-approved protocol for analysis of the data.

The pedigree development process began at all GELCC sites by screening lung cancer cases for family history (focusing on number of first-degree relatives affected with lung

Table 1. Total number of lung cancer cases and families accrued by the data collection sites

Site	Cases screened	Families for development (no. affected persons)			Potential lung cancer families identified	Families actively developed	Families formally reviewed by GELCC	Submitted to CIDR for genotyping
		2	3	≥ 4				
University of Cincinnati	7,037	705	222	161	1,088	171	24	12
University of Colorado*	—	127	47	42	216	69	40	13
Karmanos Cancer Institute	1,400	51	87	31	213	169	30	9
Saccomanno Research Institute†	3,170	59	22	10	91	38	15	12
Louisiana State University	4,250	363	79	86	528	82	18	12
Mayo	7,885	750	183	83	1,016	226	47	19
Medical College of Ohio	4,000	625	40	61	625	116	30	14
Johns Hopkins University‡	343	30	17	3	50			2
Total	28,085	2,710	697	477	3,827	871	204	93

*Colorado receives referrals of familial lung cancer families rather than probands; cases not included in % with one or more first-degree relative.

†Saccomanno Research Institute joined the consortium in 2001.

‡Johns Hopkins University was a part of the consortium until 1999.

cancer). After the initial screening process, we collected additional data from 3,827 willing probands or their family representatives about additional cancer-affected persons in the extended family, vital status of cancer-affected individuals, availability of archival tissue, and willingness of family members to participate in the study. We then initiated full pedigree development and biospecimen collection on 871 families, most with three or more affected relatives. We eliminated the majority of these families from further study because they did not contain enough family members with lung cancer from whom blood samples or nontumor tissues could be obtained for genotyping or, if affected member(s) was deceased, who had children willing to participate in the study, from whom the genotype of the affected parent could be deduced. To date, 93 families that include genetic information for at least two lung cancer-affected relatives have been genotyped, representing 0.3% of the cases we screened and 2.4% of the potential families that were identified (Table 1).

Data on tumors in the families have been obtained by requesting pathology reports, death certificates, and original tumor blocks and slides, where available. When tumor blocks or slides could be obtained, they were transmitted to the tumor pathology core, headed by Adi Gazdar at the University of Texas Southwestern Medical Center. Otherwise, tumor histology was assigned according to pathology report or death certificate. Cancer diagnoses could not be verified for 72 of the 489 subjects from the 93 families who were reported by relatives in the families to have had cancers.

Sample preparation and genotyping. Blood, buccal cells, and archival biospecimens have been used as sources of DNA for genotyping family members of the lung cancer kindreds. DNA isolated from blood has been genotyped at the Center for Inherited Disease Research (CIDR; a NIH-supported core research facility), and DNA from buccal cells, archival tissue, or sputum was genotyped at the University of Cincinnati.

DNA from archival tissue for genotyping was obtained from ten 10- μ m paraffin sections containing normal tissue. The archival tissue blocks were examined at the University of Texas Southwestern, and sections of normal tissue were prepared for genotyping at the University of Cincinnati. We required the specimen to have at least 50% normal cells for genotyping to ensure the germline rather than tumor genotype was observed. DNA was isolated from paraffin sections and sputum samples by a modified Wright and Manos [10,10] procedure, performed by incubating the tissue with 0.5 μ g/ μ L of proteinase K in 1 \times PCR buffer with NP40 and Tween 20 for 1 h at 55°C. This is followed by a 95°C incubation for 10 min to inactivate the proteinase K and then treatment of the isolated DNA with 24:1 (v/v) chloroform/isoamyl alcohol. DNA was isolated from the buccal cells and from whole blood using the Puregene kit (Gentra Systems, Inc.) following the manufacturer's protocols.

The CIDR global genotyping set consisted of 392 markers (15 families genotyped from 1998 to 2000) or 388 markers (78 families). PCR amplifications, using the primer set for each of the markers, were performed at CIDR and the University of Cincinnati. The standard protocol for PCR at CIDR can be found on the CIDR Web site.¹³ Conditions for genotyping

markers using archived DNA were similar to the protocol of CIDR but with a modification of increasing the number of amplification cycles to 35. All samples were amplified in an MJ Research Thermocycler. Briefly, the cycles were as follows: 95°C for 12 min, 94°C for 45 s, 55°C for 1 min, and 72°C for 1 min for an initial 10 cycles, and then 89°C for 1 min, 55°C for 1 min, and 72°C for 1 min for an additional 20 cycles, followed by a final extension at 72°C for 10 min. PCR amplifications were performed using a single fluorescently labeled primer obtained from CIDR. Following the reactions, PCR products were resolved on an ABI 3100 automated DNA sequencer and analyzed with genotype software. Due to the reduced amounts of genomic DNA in the archived samples, none of the amplification products was pooled before loading onto the 96 wells of a plate for subsequent analysis.

Integrating genotype data across platforms and quality control procedures. Assignment of alleles generated at CIDR and the University of Cincinnati was accomplished by genotyping several samples in common for each gel (or plate) at both facilities. These common samples included CEPH controls 1331-01 and 1331-02 as well as several lymphocyte DNA samples from members of the lung cancer families.

Our first step in evaluating the genetic data was to appropriately bin the allele lengths. To allow us to jointly analyze data across different platforms used at CIDR versus the University of Cincinnati, we first compared the raw allele lengths for 16 subjects who had been genotyped on both platforms. We next generated a linear regression to predict CIDR lengths from the UC data while identifying any errors in the data as alleles that failed to satisfy the criterion: distance = $\text{abs}(\cos(\arctan(b)) * (\hat{y} - y)) < 1$, where \hat{y} is the predicted value of a point. The prediction of allele lengths between centers routinely yielded an R^2 value of >99% for all but two markers (which had R^2 values of 97% and 98%, respectively). However, the intercepts were routinely different from 0, indicating a shift in allele lengths between labs, and the slope often varied from 1, indicating that, without regression adjustment, alleles at the extremes could have been misclassified.

The programs Relative (10) and PREST (11) were used to verify relationships among individuals in the data. SIBPAIR (12) and PEDCHECK (13) were used to check for Mendelian inconsistencies. All such errors were corrected by eliminating the genotypes indicated to have been most likely to cause errors. Once verification of pedigree structures and elimination of marker inconsistencies had been completed, we estimated allele frequencies for the chromosome 6 linkage analysis using maximum likelihood methods as discussed by Boehnke (14). To perform this analysis we used the Fastlink program, which is a module of Fastlink (15). To allow for both genotyping heterogeneity and racial heterogeneity in allele frequencies, we estimated allele lengths separately for Caucasians and non-Caucasians and by genotyping set (three sets of samples were separately analyzed by CIDR).

¹³ <http://www.cidr.jhmi.edu>

LOD score analyses and haplotyping. Our primary analytic approach in analysis of data from the GELCC assumed a model with 10% penetrance in carriers and 1% penetrance in the noncarriers. This analytic approach weights information only from the affected subjects (16) and so provides an essentially model-free analysis. To obtain linkage results, we used SIMWALK2 (17) and calculated HLOD scores (18) from the output using Perl scripts we have developed. In this analysis, we estimated the evidence for linkage from each family separately using the Markov chain Monte Carlo (MCMC) provided by SIMWALK2. MCMC analysis was used to estimate LOD scores because the pedigrees were too large to permit exact multipoint computation of the likelihood of the data. The LOD scores from each family were then combined, allowing for an additional heterogeneity parameter, which models the effect on the LOD score that will occur if not all of the families are linked to a specific region. We performed all analyses separately within genotyping set and within racial group to avoid any issues that might arise if marker alleles were not faithfully mapped among studies. Results from each LOD score analysis were then summed across study and ethnicity to obtain the final results.

To obtain haplotypes for the linked region of chromosome 6q, we used a feature of SIMWALK2 that assigns marker genotypes to haplotypes using markers in the linked region, including D6S2436 and D6S1035, covering the region from 155 to 165 cM on chromosome 6q. We then integrated the haplotype data onto pedigree drawings that we developed using Progeny. Finally, where possible, in 40 multigenerational pedigrees, we visually identified haplotypes that cosegregate with disease susceptibility by tracing the segregation of haplotypes with disease in families. This tracing algorithm was only possible in families that supported evidence for linkage and included multiple generations. To assign phase, it was helpful to have more than one generation available for study. In addition, to assign a haplotype indicating risk, we required that the family provide positive support for linkage. There were two families that seemed to segregate two risk haplotypes because of bilineality in the family (i.e., the inheritance of disease susceptibility seemed to segregate from both parents of the proband). Then, conditional on the carrier status and smoking behavior of subjects, we performed Kaplan-Meier analyses and Cox regression analysis to assess the relationship between smoking behavior and lung cancer risk, according to the carrier status of the subjects we were studying. We defined never smokers to be individuals who smoked <100 cigarettes, light smokers as individuals who reported having smoked <20 pack-years, moderate smokers had ≥ 20 but <40 pack-years exposure, and heavy smokers had >40 pack-years exposure.

To adjust for nonrandom sampling of individuals into our study, we also used a previously developed approach that weights the cases and controls according to population-based incidence rates of cancer (19). Specifically, we obtained incidences of lung cancer for 5-y age intervals from statistics compiled by the American Cancer Society,¹⁴ averaging rates

for males and females. These age-specific incidences were then used to obtain sampling fractions to weight the proportion of cases versus controls in each age stratum. The weights so derived are presented in Supplementary Table S1 and show that there were more cases observed in the sample than expected from a population-based sample. Therefore, for all intervals, we downweighted the information from cases compared with controls in the subsequent analysis. Although this weighted Cox regression approach eliminated bias in simulation studies, it also had reduced power compared with unweighted analysis. Therefore, we present results from both analyses.

Results

The 93 families that have been studied include 489 persons affected with lung cancer, of whom 45 are unrelated (marrying-in to the pedigree) and 444 are related to other affected family members, and informative for linkage analysis. From these families, we have accrued 1,156 blood samples, 24 buccal cell samples, 58 sputum samples, and 274 archival blocks containing normal tissue. Archival tumor blocks of lung cancer-affected subjects have been collected from 186 persons and 88 blocks from other tissues. When other sources of DNA were not available, we used archival tissue blocks for genotyping. Where possible, because we are interested in studying the coinheritance of lung cancer with genetic markers present in the germline, we have performed analyses on tumor blocks from non-lung cancer specimens. Otherwise, when lung cancers have been studied, one of us (A.G.) has retrieved normal tissue from the tumor margins. Of the 93 families, three are African-American and 1 family has mixed racial composition (African-American, Creole, and Caucasian); the remaining 89 families are Caucasian.

Lung cancer-affected individuals are 63.4% male, 81.8% deceased at the time of data collection, and 86.3% ever smokers, with a median value of 50 pack-years. For the unaffected individuals who reported cigarette smoking history data, 73.2% were ever smokers with median pack-year value of 26, a generally higher level of smokers than in the general population (20). However, because these persons come from families with a strong history of lung cancer among smoking relatives, and smoking aggregates in families, they are more likely to be smokers. Smoking histories for deceased individuals were obtained from surrogates. Numerous studies have reported that surrogate reported data are about 90% to 95% accurate for smoking status but usually underestimate pack-years (21–25). Cancer status has been verified with medical records, cancer registry data, or death certificates on 417 (85.3%) of the 489 lung cancer-affected persons. Pathology reports were obtained whenever possible (i.e., the tissue sample was obtained for diagnosis, medical records could be located, and patient or family had signed a medical record release). The distribution of cell type of lung cancer was similar to that reported in the past for the general population (26). In 59 families studied by Bailey-Wilson and colleagues (8) of 224 lung cancer-affected persons on whom we have pathology reports, 75 (33.5%) had adenocarcinoma, 69

¹⁴ Cancer facts and figures 2008. Available from: <http://www.cancer.org/>.

Table 2. Number of lung cancer-affected individuals in families, having at least a third-degree relationship to each other

No. affected in each pedigree	No. pedigrees	Total no. affected	Total no. affected genotyped	Total no. unaffected genotyped
2	2	4	4	12
3	25	75	39	210
4	19	76	32	174
5	20	100	35	125
6	11	66	19	153
7	12	84	19	130
≥8	4	39	31	192
Total	93	444	179	996

(30.8%) had squamous cell carcinoma, and 22 (9.85%) had small cell carcinoma. Seven families presented with predominantly either adenocarcinoma ($n = 3$) or squamous cell carcinoma ($n = 4$).

Two pedigree characteristics that affect informativeness for linkage analysis are the number of affected persons in the family and number of generations with affected persons (Table 2). For assessing informativeness, we count only affected persons who have at least a third-degree relationship to another affected. In bilineal families, we count only those in the predominant lineage with lung cancer when both parents are lung cancer affected (Table 2). Because at least some of the families with only two and three affected relatives may not segregate effects from a major susceptibility factor but may rather reflect chance clustering of lung cancer, we have separated this group into subset 1. Similarly, families that include five or more affected relatives in two or more generations are most likely to segregate a dominantly inherited locus that increases susceptibility and these families have been denoted as subset 2. Families that include four or more relatives in a sibship are denoted subset 3. The median number of affected persons per family is 5. In the 93 families, there are 66 families with four or more affected persons and 57 of these families have affected persons in more than one generation (subset 4; results in Supplementary Figures).

Of the 50 families with five or more affected persons, 47 have affected persons in multiple generations. Linkage analyses of chromosome 6 show that families with five or more affected persons in multiple generations exhibited linkage to chromosome 6q.

Linkage and haplotype analyses of risk. Maximal HLOD scores from genome-wide linkage analyses are presented in Table 3. Results from linkage analyses are presented in Fig. 1 and Supplementary Fig. S1. In Fig. 1, we present the results of linkage analysis on chromosome 6, whereas Supplementary Fig. S1 provides results for other chromosomes that yielded a HLOD score of >1.0 in any subset. The proportions of families estimated from HLOD score analysis were 0.53 for the entire data set, and for subsets 1 to 3, the heterogeneity estimates were 0.74, 1.0, and 0.35, respectively. Of the entire set of 93 families, 10 had a LOD score of >0.3 on chromosome 6q at 158 cM.

Further analysis of the effect of smoking on risk for cancer was carried out as indicated above by first defining carrier status and then by performing Cox regression modeling treating the intensity of smoking as an ordinal variable. There were 292 individuals who carried a risk haplotype, 441 who were in families segregating a risk haplotype who were noncarriers of that haplotype, and 2,248 individuals for whom carrier status could not be derived and were

Table 3. Maximum HLOD scores of >1.0 in linkage analysis of any subset

Chromosome	Combined		Subset 1		Subset 2		Subset 3	
	Maximum HLOD	Position (cM)	Maximum HLOD	Position (cM)	Maximum HLOD	Position (cM)	Max HLOD	Position (cM)
1	0.337	126.0	1.113	126.0	0.424	180.8	1.202	180.8
4	0.656	133.9	1.083	73.5	0.475	158.0	0.359	198.9
6p	0.876	65.0	0.430	9.0	1.607	61.2	0.211	61.2
6q	2.384	158.0	0.239	112.0	4.668	158.0	0.322	155.0
8	0.090	119.0	0.0506	41.2	0.450	1.00	1.050	119.0
9	0.672	4.0	0.038	44.0	1.354	4.0	0.974	143.7
20	1.088	34.2	0.281	25.0	0.693	39.0	1.123	62.0

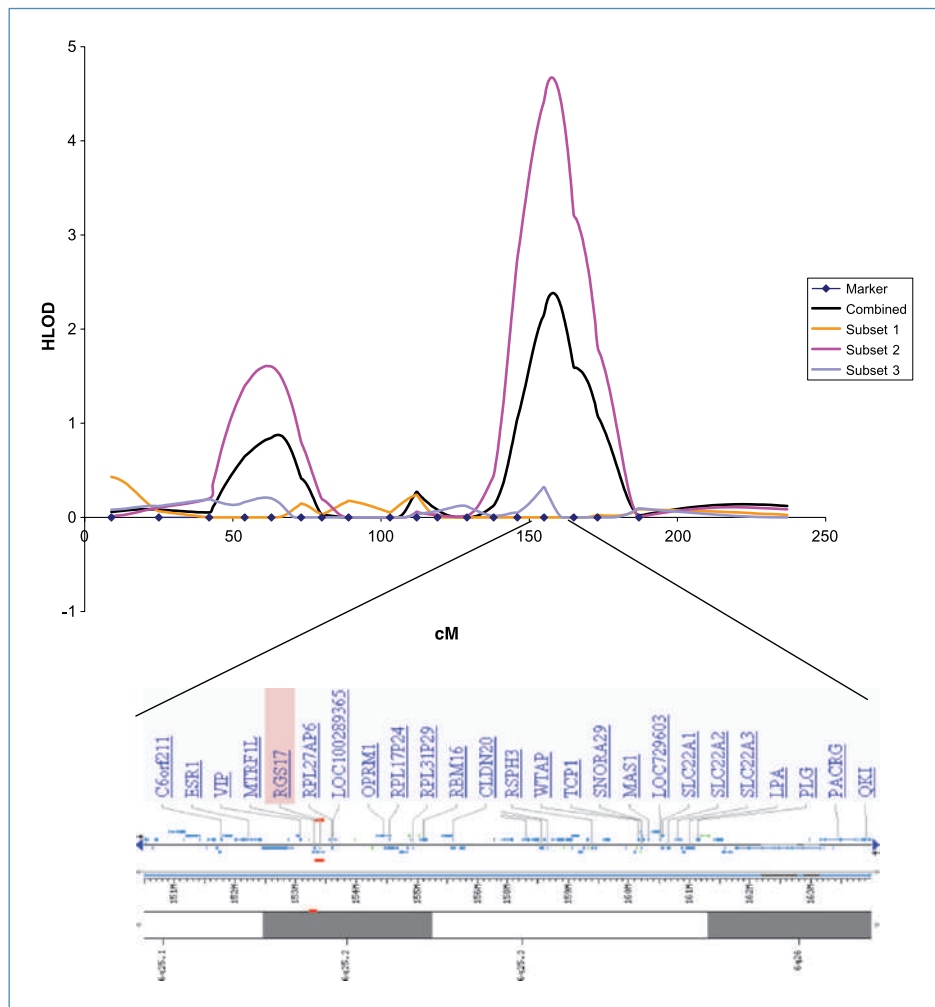


Figure 1. HLOD scores from analysis of chromosome 6 for 93 families selected to include multiple relatives with lung cancer. Subset 1 includes families with two or three individuals affected by lung cancer, subset 2 includes families with five or more individuals in two or more generations, and subset 3 comprises individuals with four or more individuals in a sibship who had lung cancer.

classified as unknown carrier status. Figure 2 results from Kaplan-Meier analysis showing that, among carriers, the overall risk for lung cancer was higher than among noncarriers. There is also significantly higher risk for lung cancer among ever compared with never smokers, as assessed by the log-rank test. However, among smoking carriers, there was no evidence for increasing risk with an increasing expo-

sure level to cigarette smoke ($P = 0.36$). On the other hand, among noncarriers ($P = 0.085$) and individuals with unknown carrier status ($P = 0.0008$), a more usual dose-effect relationship between smoking and lung cancer risk is observed. These findings suggest that any level of tobacco exposure increases risk among those with inherited lung cancer susceptibility, suggesting that such individuals should be heavily

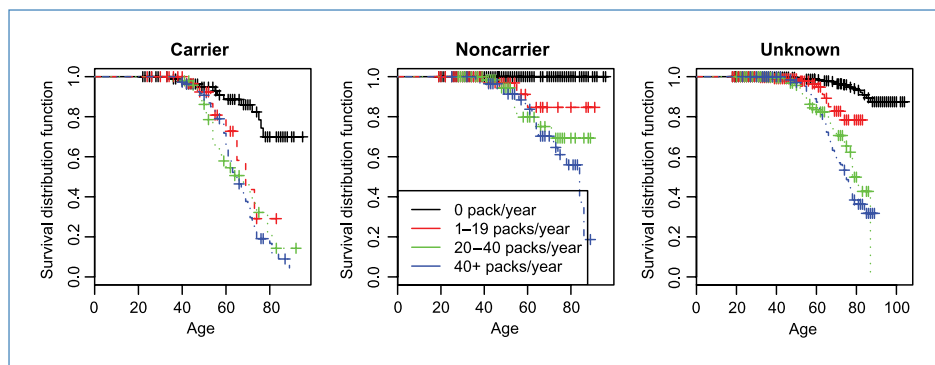


Figure 2. Time to lung cancer among carriers (left), noncarriers (middle), and individuals with unknown carrier status (right). Smoking strata are shown with the black line reserved for nonsmokers, the red line for light smokers (1–19 pack-years), the green line for 20–39 pack-years, and the blue line for heavier smokers (≥ 40 pack-years). Tick marks on lines indicate ages at censoring due to either currently alive without lung cancer or death from a competing cause.

targeted for smoking prevention and monitored by early detection procedures. Compared with the risk in never smokers (Table 4A), carriers had higher hazard ratios of 3.44 [95% confidence interval (95% CI), 1.40–8.48; $P = 0.007$] for light smokers, 4.91 (95% CI, 2.46–9.8; $P < 0.0001$) for moderate smokers, and 5.18 (95% CI, 2.81–9.56; $P < 0.0001$) for heavy smokers. Among noncarriers, no events occurred in never smokers so that hazard ratios could not be estimated. For unknown carrier status, there was a much stronger effect of smoking, with all groups having highly significant differences from never smokers ($P < 0.0001$). For those light smokers with unknown carrier status, the hazard ratio compared with never smokers was 4.25 (95% CI, 2.11–8.54), for moderate smokers the hazard ratio was 9.77 (95% CI, 5.9–16.20), and for heavy smokers the hazard ratio was 11.89 (95% CI, 7.59–18.61). When the analyses were adjusted for excess selection for affected individuals (Table 4B), we found very little trend in carriers, with the hazard ratios in carriers being 2.67 (95% CI, 1.22–5.86), 2.34 (95% CI, 1.37–3.98), and 2.75 (95% CI, 1.74–4.37) in light, moderate, and heavy smokers, respectively, whereas for those with unknown carrier status the hazard ratios were 3.00 (95% CI, 1.64–5.88), 5.20 (95% CI, 3.67–7.58), and 7.32 (95% CI, 5.28–10.14), respectively, for light, moderate, and heavy smokers.

An alternative approach to evaluating risk compares risk among carrier groups, conditioning on smoking behavior (Supplementary Table S2A). Using individuals with unknown carrier status as the referent, for never smokers the hazard ratio for noncarriers was 0 (no events; $P = 0.99$) and 4.71 for carriers (95% CI, 2.35–9.43; $P < 0.0001$). For light smokers, the hazard ratio was 1.08 for noncarriers (95% CI, 0.31–3.83; $P = 0.90$) and 4.34 for carriers (95% CI, 1.76–10.7; $P = 0.0001$). For moderate smokers, the hazard ratios were 0.83 for noncarriers (95% CI, 0.41–1.165; $P = 0.59$) and 2.51 for carriers (95% CI, 1.53–4.13; $P = 0.0003$). For heavy smokers, the hazard ratio was 0.83 for noncarriers (95% CI, 0.54–1.29; $P = 0.41$) and 2.21 for carriers (95% CI, 1.65–2.97; $P < 0.0001$). Thus, comparing noncarriers and those with no known haplotype (unknown carrier status), there is no significant difference in risk between these two groups according to smoking behavior. However, among those who are carriers, the increased risk is most prominent in never smokers. However, as shown in Fig. 2, any degree of smoking confers a marked increase in risk beyond this baseline. Decreasing hazard ratios according to increasing smoking reflect the higher risks among the noncarriers of risk haplotypes according to increased effects from smoking, but comparable risks for lung cancer among carriers who have any degree of smoking exposure.

riers (95% CI, 0.41–1.165; $P = 0.59$) and 2.51 for carriers (95% CI, 1.53–4.13; $P = 0.0003$). For heavy smokers, the hazard ratio was 0.83 for noncarriers (95% CI, 0.54–1.29; $P = 0.41$) and 2.21 for carriers (95% CI, 1.65–2.97; $P < 0.0001$). Thus, comparing noncarriers and those with no known haplotype (unknown carrier status), there is no significant difference in risk between these two groups according to smoking behavior. However, among those who are carriers, the increased risk is most prominent in never smokers. However, as shown in Fig. 2, any degree of smoking confers a marked increase in risk beyond this baseline. Decreasing hazard ratios according to increasing smoking reflect the higher risks among the noncarriers of risk haplotypes according to increased effects from smoking, but comparable risks for lung cancer among carriers who have any degree of smoking exposure.

Discussion

Inclusion of additional families collected and analyzed since our 2004 report continues to support evidence for linkage in the 6q region. Among 22 new families that have been collected, 4 showed substantive evidence for linkage (LOD > 0.3), with 1 family yielding a LOD score of 0.826, whereas for the entire set of 93 families, 10 showed substantive evidence for linkage (LOD > 0.3). Interestingly, analysis including all the families now shows a bifurcation in the linkage signal around D6S1048. Aside from the linkage studies we report here, there has also been a report of linkage of mesothelioma susceptibility to the same region of chromosome 6q from a family study in an area of Turkey exposed to mineral fibers (27).

Further association analysis of the chromosome 6q region identified one locus that influences lung cancer susceptibility (28). The identified gene, *RGS17*, is a signaling protein with

Table 4. Contrasts in risk between light, moderate, and heavy smokers according to carrier status

Smoker	Carrier		Noncarrier		Unknown	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
A. Comparison of risks for light, moderate, and heavy smokers versus nonsmokers stratified by carrier status, without adjustment for sampling through multiple affected relatives						
Light	3.44 (1.40–8.48)	0.0072	∞ (0 to ∞)	~1	4.25 (2.11–8.54)	<0.0001
Moderate	4.91 (2.46–9.80)	<0.0001	∞ (0 to ∞)	~1	9.77 (5.90–16.20)	<0.0001
Heavy	5.18 (2.81–9.56)	<0.0001	∞ (0 to ∞)	~1	11.89 (7.59–18.61)	<0.0001
B. Comparison of risks for light, moderate, and heavy smokers versus nonsmokers stratified by carrier status, with adjustment for sampling through multiple affected relatives						
Light	2.67 (1.22–5.86)	0.0014	∞ (0 to ∞)	~1	3.10 (1.64–5.88)	0.00051
Moderate	2.34 (1.37–3.98)	0.0018	∞ (0 to ∞)	~1	5.20 (3.56–7.58)	<0.0001
Heavy	2.75 (1.74–4.37)	1.7e-05	∞ (0 to ∞)	~1	7.32 (5.28–10.14)	<0.0001
C. Comparison of risks for cigarette use treated as an ordinal variable (0 = never smokers, 1 = light smokers, 2 = moderate smokers, 3 = heavy smokers), without and with adjustment for sampling through multiple affected relatives						
Without weights	1.598 (1.349–1.892)	<0.0001	2.719 (1.863–3.969)	<0.0001	2.143 (1.887–2.434)	<0.0001
With weights	1.325 (1.154–1.522)	<0.0001	2.640 (1.801–3.868)	<0.0001	1.874 (1.693–2.074)	<0.0001

homology to opioid receptors that has an oncogenic effect in cell culture. Although genetic analysis has shown a strong effect of this locus in selected high-risk families that we have studied, its effects remain insufficient to explain the high penetrance observed in Fig. 2. Therefore, additional variability either in the promoter region of RGS17 or in additional linked loci seems likely to explain the high penetrance observed in these families. It is possible that the region on 6q harbors more than a single genetic locus influencing susceptibility to lung cancer. Toward the aim of fully querying the region of chromosome 6q, the GELCC is performing a comprehensive resequencing effort for all of the loci within a 10-Mb region of the linkage peak. Because we cannot yet fully identify all of the risk alleles for lung cancer that exist on chromosome 6q, we have used a haplotype-based approach to identify individuals who are at increased lung cancer risk.

Statistical modeling of the risk for cancer among those carrying a haplotype associated with increased lung cancer risk showed evidence for an interaction between exposure to smoking and inherent susceptibility to lung cancer. Among those with inherited susceptibility to lung cancer, the risk for lung cancer among never smokers was higher than never smokers who did not inherit susceptibility. However, the more dramatic observation from our analysis was the finding that any degree of smoking yielded a similar and substantive increase in risk for developing lung cancer among carriers of inherited susceptibility, whereas there was a quantitative increase in risk according to the increasing level of smoking among individuals who we did not infer to carry a lung cancer susceptibility haplotype. The observation that environmental factors can have striking effects on individuals with inherited susceptibility to disease parallels many observations in medical genetics. For example, individuals with metabolic deficiencies in phenylalanine hydroxylase or porphobilinogen deaminase are greatly adversely affected by exposure to even small levels of, respectively, phenylalanine (29) or barbiturates or other drugs (30). Therefore, adverse response to even small amounts of exogenous compounds such as those present in tobacco smoke may be a particular effect of the genetic locus we have identified on

chromosome 6q. Because we are not able to obtain detailed information about passive smoking in the family study we conducted, we do not know what level of exposure never smokers have had in our families, but it is possible that the elevated risks we observe in never smokers in carriers reflect in part their exposure patterns.

Although our ongoing studies of families collected by the GELCC continue to support effects on risk of a locus on chromosome 6q, we also are here reporting additional evidence for loci on chromosomes 6p, 1q, 8q, and 9p in subsets of families that have multiple affected relatives with lung cancer. To characterize more fully these regions of linkage and to refine the region of linkage on chromosome 6q, further efforts in identifying familial lung cancer cases and families are under way. To identify highly penetrant causal genetic factors, families that include multiple affected relatives are informative (31). The GELCC is pursuing initial genome-wide single-nucleotide polymorphism-based association studies as well as resequencing. Initially, the resequencing efforts by GELCC have targeted selected regions that showed linkage as well as known candidate loci, but we anticipate that as more global resequencing become cost-effective, we will seek to adopt this strategy. The continued collection of families with multiple affected relatives will allow us to identify additional loci through both linkage and association studies.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

NIH grants UO1CA076293, P30ES06096, P30CA016772, R01CA133996, R01CA060691, R01CA87895, P30ES007789, P50CA70907, and NO1PC35145 and the intramural programs of the National Cancer Institute and the National Human Genome Research Institute, NIH.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 08/21/2009; revised 11/30/2009; accepted 12/03/2009; published OnlineFirst 03/09/2010.

References

1. Tokuhata GK, Lilienfeld AM. Familial aggregation of lung cancer in humans. *J Natl Cancer Inst* 1963;30:289-312.
2. Ooi WL, Elston RC, Chen VW, Bailey-Wilson JE, Rothschild H. Increased familial risk for lung cancer. *J Natl Cancer Inst* 1986;76:217-22.
3. Shaw GL, Falk RT, Pickle LW, Mason TJ, Buffler PA. Lung cancer risk associated with cancer in relatives. *J Clin Epidemiol* 1991;44:429-37.
4. Etzel CJ, Amos CI, Spitz MR. Risk for smoking-related cancer among relatives of lung cancer patients. *Cancer Res* 2003;63:8531-5.
5. Jonsson S, Thorsteinsdottir U, Gudbjartsson DF, et al. Familial risk of lung carcinoma in the Icelandic population. *JAMA* 2004;292:2977-83.
6. Sellers TA, Bailey-Wilson JE, Elston RC, et al. Evidence for mendelian inheritance in the pathogenesis of lung cancer. *J Natl Cancer Inst* 1990;82:1272-9.
7. Gauderman WJ, Morrison JL, Carpenter CL, Thomas DC. Analysis of gene-smoking interaction in lung cancer. *Genet Epidemiol* 1997;14:199-214.
8. Bailey-Wilson JE, Amos CI, Pinney SM, et al. A major lung cancer susceptibility locus maps to chromosome 6q23-25. *Am J Hum Genet* 2004;75:460-74.
9. Ott J. Analysis of human genetic linkage. Baltimore (MD): Johns Hopkins University Press; 1999.
10. Goring HH, Ott J. Relationship estimation in affected sib pair analysis of late-onset diseases. *Eur J Hum Genet* 1997;5:69-77.
11. McPeck MS, Sun L. Statistical tests for detection of misspecified relationships by use of genome-screen data. *Am J Hum Genet* 2000;66:1076-94.
12. SIB-PAIR [computer program]. Brisbane: Queensland Institute for Medical Research; 2009.

13. O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998;63:259–66.
14. Boehnke M. Allele frequency estimation from data on relatives. *Am J Hum Genet* 1991;48:22–5.
15. Cottingham RW, Jr., Idury RM, Schaffer AA. Faster sequential genetic linkage computations. *Am J Hum Genet* 1993;53:252–63.
16. Speer MC. Use of LINKAGE programs for linkage analysis. *Curr Protoc Hum Genet* 2006;Chapter 1:Unit 1.7.
17. Elston RC, Sobel E. Sampling considerations in the gathering and analysis of pedigree data. *Am J Hum Genet* 1979;31:62–9.
18. Ott J. Complex traits on the map. *Nature* 1996;379:772–3.
19. Antoniou AC, Goldgar DE, Andrieu N, et al. A weighted cohort approach for analysing factors modifying disease risks in carriers of high-risk susceptibility genes. *Genet Epidemiol* 2005;29:1–11.
20. Schoenborn CA, Adams PF, Schiller JS. Summary health statistics for the U.S. population: National Health Interview Survey, 2000. *Vital Health Stat* 2003;10:1–83.
21. Woo JG, Pinney SM. Retrospective smoking history data collection for deceased workers: completeness and accuracy of surrogate reports. *J Occup Environ Med* 2002;44:915–23.
22. Hansen KS. Validity of occupational exposure and smoking data obtained from surviving spouses and colleagues. *Am J Ind Med* 1996;30:392–7.
23. Lerchen ML, Samet JM. An assessment of the validity of questionnaire responses provided by a surviving spouse. *Am J Epidemiol* 1986;123:481–9.
24. McLaughlin N. Smoking bans, restrictions increase at country's healthcare facilities. *Mod Healthc* 1987;17:92.
25. Hyland A, Cummings KM, Lynn WR, Corle D, Giffen CA. Effect of proxy-reported smoking status on population estimates of smoking prevalence. *Am J Epidemiol* 1997;145:746–51.
26. Alberg AJ, Samet JM. Epidemiology of lung cancer. *Chest* 2003;123:21S–49S.
27. Below JE, Pluznikov A, Aquino-Michaels K, et al. Localization of a dominant genetic susceptibility factor in familial malignant mesothelioma. Presented at the annual meeting of The American Society of Human Genetics; 2008 Nov 12; Philadelphia, Pennsylvania. <http://www.ashg.org/2008meeting/abstracts/fulltext>.
28. You M, Wang D, Liu P, et al. Fine mapping of chromosome 6q23-25 region in familial lung cancer families reveals RGS17 as a likely candidate gene. *Clin Cancer Res* 2009;15:2666–74.
29. Paine RS. The variability in manifestations of untreated patients with phenylketonuria (phenylpyruvic aciduria). *Pediatrics* 1957;20:290–302.
30. Tschudy DP, Valsamis M, Magnussen CR. Acute intermittent porphyria: clinical and selected research aspects. *Ann Intern Med* 1975;83:851–64.
31. Jones S, Hruban RH, Kamiyama M, et al. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science* 2009;324:217.