Int J Mol Epidemiol Genet 2012;3(1):1-17
www.ijmeg.org /ISSN1948-1756/IJMEG1111006

Original Article

Germ line variation in nucleotide excision repair genes and lung cancer risk in smokers

Lori C Sakoda[1,2], Melissa M Loomis[1], Jennifer A Doherty[1,2], Liberto Julianto[1], Matt J Barnett[1], Marian L Neuhouser[1], Mark D Thornquist[1], Noel S Weiss[1,2], Gary E Goodman[1], Chu Chen[1,2,3]

[1]*Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA;* [2]*Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA;* [3]*Department of Otolaryngology: Head and Neck Surgery, School of Medicine, University of Washington, Seattle, WA, USA*

Received November 21, 2011; accepted December 18, 2011; Epub February 5, 2012; Published February 28, 2012

Abstract: Since nucleotide excision repair (NER) is primarily responsible for detecting and removing bulky DNA lesions induced by tobacco smoke in the respiratory tract, single nucleotide polymorphisms (SNPs) in NER protein-encoding genes may influence lung cancer risk, particularly in smokers. Studies testing this hypothesis have produced inconsistent results, with most analyzing a few SNPs in relatively small population samples. In a study nested in the Beta-Carotene and Retinol Efficacy Trial, we examined 79 tag and previously reported risk-associated SNPs in the ERCC1, ERCC2, ERCC3, ERCC4, ERCC5, LIG1, POLE, XPA, and XPC genes in 744 lung cancer cases and 1,477 controls, all of whom were non-Hispanic white smokers. Using logistic regression, odds ratios (OR) and 95% confidence intervals (95% CI) were calculated to estimate lung cancer risk associated with SNP genotypes and haplotypes, adjusting for case-control matching factors. Lung cancer risk was modestly associated with LIG1 rs156640 (OR per G allele, 1.23; 95% CI, 1.08-1.40), rs156641 (OR per A allele, 1.23; 95% CI, 1.08-1.40), and rs8100261 (OR per A allele, 0.83; 95% CI, 0.76-0.98); XPA rs3176658 (OR per A allele, 0.83; 95% CI, 0.69-1.00); and ERCC2 rs50871 (OR per C allele, 1.15; 95% CI: 1.01-1.30). Associations with LIG1 and XPA, but not ERCC2, haplotypes were found. The results of this study and others suggest that inherited variants in LIG1 and possibly other NER genes may predispose to smoking-related lung cancer. Given that chance likely accounts for one or more of the associations observed, replication of our findings is needed.

Keywords: Lung cancer, nucleotide excision repair, genetic polymorphism

Introduction

Although cigarette smoking is the predominant risk factor for lung cancer, less than 20% of lifetime smokers develop the disease [1]. Inherited genetic characteristics are presumed to account in part for this interindividual variation in lung cancer susceptibility. Individuals with an affected relative have an increased incidence of lung cancer [2], and a shared tendency to smoke between relatives does not appear to account for the entire extent of this association [3]. Genomewide association (GWA) studies have also consistently identified lung cancer susceptibility loci at 15q24-25.1, 5p15.33, and 6p21.33 [4-11]. However, since few low-penetrance loci have been discovered using comprehensive, but not complete scans of variation within the human genome [12], additional genetic characteristics likely contribute to lung cancer development.

Tobacco smoke and other environmental exposures, such as ultraviolet and ionizing radiation, promote the formation of bulky adducts, crosslinks, and strand breaks in DNA [1]. Higher levels of such DNA damage induced by mutagens, including the tobacco carcinogen benzo(a)pyrene diol epoxide, have been quantified *in vitro* in cultured lymphocytes of persons with lung cancer, relative to healthy persons [13-16]. The extent of smoking-induced DNA damage has been further associated with genotypes of single nucleotide polymorphisms (SNPs) in multiple DNA repair protein-encoding genes [17, 18]. Therefore, polymorphic variation in DNA repair genes could conceivably influence lung cancer risk by modulating DNA repair capacity,

particularly among cigarette smokers.

Of the major DNA repair pathways, nucleotide excision repair (NER) is principally responsible for recognizing and removing bulky chemical adducts [reviewed in [19-21]]. At least thirty proteins, including those encoded by the xeroderma pigmentosum (XP) and excision repair cross-complementing (ERCC) genes, act in NER to coordinate DNA damage detection, helix unwinding, lesion excision, gap filling, and strand ligation. Two NER subpathways exist: transcription-coupled NER (TC-NER), which removes lesions in actively transcribed DNA, and global genome NER (GG-NER), which removes lesions elsewhere in the genome. In TC-NER, DNA damage is sensed when the RNA polymerase II transcription complex is stalled by a lesion, signaling recruitment of the Cockayne syndrome proteins CSA and CSB and other core NER proteins. In GG-NER, DNA damage is instead recognized by the XPC-HHRAD23B heterodimer subcomplex. Thereafter in both subpathways, XPA and replication protein A (RPA) bind, facilitating damage verification and proper assembly of a multiprotein repair complex. ERCC2 (XPD) and ERCC3 (XPB), the two helicase subunits of transcription factor TFIIH, unwind the DNA helix to open a ~30 base segment around the lesion. The endonucleases ERCC5 (XPG) and ERCC1-ERCC4 (ERCC1-XPF) then incise the damaged DNA strand, cleaving 3' and 5' to the lesion, respectively. Using the opposing strand as a template, the resulting gap is filled by DNA polymerase delta (POLD1) or epsilon (POLE), alongside the replication proteins RPA, proliferating cell nuclear antigen (PCNA), and replication factor C (RFC), and sealed by DNA ligase I (LIG1).

Studies of single nucleotide polymorphisms (SNPs) in NER-related genes in relation to lung cancer risk have not provided consistent results. The majority of published studies have evaluated a limited number of SNPs in a few candidate genes in relatively small populations. Also, since these SNPs have been primarily selected on the basis of known or putative function alone, the extent to which variation across these genes contributes to lung cancer susceptibility has not been fully examined. Based on meta-analyses, however, the carriage of some SNPs, including ERCC2 rs13181 (CC versus AA genotype) and XPA rs1800975 (AA versus GA/GG genotype), appear to be associated with increased lung cancer risk [22-25].

In a defined population of non-Hispanic white smokers, including 744 lung cancer cases and 1477 controls, we conducted a more comprehensive evaluation of the hypothesis that germ line variation in NER-related genes predisposes to lung cancer. We examined whether tag and previously reported risk-associated SNPs in the NER protein-encoding genes ERCC1, ERCC2, ERCC3, ERCC4, ERCC5, LIG1, POLE, XPA, and XPC were associated with risk of lung cancer. In addition, we assessed whether any observed SNP associations were modified by host characteristics, including age, sex, smoking quantity, and dietary factors.

Materials and methods

Study population and design

As described previously [26], participants were selected from the β-Carotene and Retinol Efficacy Trial (CARET), a randomized, double-blinded, placebo-controlled trial conducted to assess the safety and efficacy of daily supplementation with 30 mg of β-carotene plus 25,000 IU of retinyl palmitate in reducing lung cancer incidence [27-29]. In this trial, 18,314 high-risk individuals were enrolled at six U.S. study sites from 1985 to 1994. Individuals at high risk were defined as (a) men and women aged 50-69 years who were either former (i.e., quit within six years prior to enrollment) or current smokers with a smoking history of 20+ cigarette pack-years (n=14,254) and (b) men aged 45-69 years who were either former (i.e., quit within fifteen years prior to enrollment) or current smokers with an occupational history of asbestos exposure (n=4,060). At the baseline visit and follow-up visits every two years thereafter, participants were asked to complete a structured questionnaire about health risk factors, including smoking behavior, and a food-frequency questionnaire about dietary intake in the prior year. Administration of the intervention ceased in 1996, after a mean follow-up of four years, based on interim data analyses showing higher lung cancer incidence and overall mortality rates in the intervention versus placebo arm, but follow-up for lung cancer and other outcomes continued until 2005.

Eligibility for this nested case-control study of lung cancer was restricted to CARET partici-

pants who had provided a blood sample from 1994 to 1997 for genetic research use. Those diagnosed with primary lung cancer from the date of blood collection to the end date of CARET follow-up were selected as cases. Of those free of lung cancer who had completed at least one food-frequency questionnaire, two controls were selected per case by matching on age (± 4 years), sex, race/ethnicity, enrollment year (2-year intervals), baseline measures of smoking status (current or former) and asbestos exposure (yes or no), and duration of follow-up. Through this process, 793 cases and 1,586 controls were identified for study inclusion.

Tumor histology was not documented in existing CARET records for all cases. To acquire more complete data on this measure, cases for whom tumor histology had not been recorded (or if deceased, their next-of-kin) were contacted by mail to request permission for medical record access. Retrieved records of consenting participants were systematically abstracted by a medical oncologist (GEG). Histology information was additionally acquired through data record linkages to the cancer registries of California, Oregon, and Washington, the three states in which about 85% of all participants resided at the time of CARET enrollment. Given the high degree of concordance on histological classification of tumors (i.e., non-small cell lung cancer (NSCLC), small cell lung cancer (SCLC)) between data sources, we elected to use the state cancer registry records as the primary data source for histology. We used data contained in CARET records only when registry records were not available or indicated histology as unknown.

Written informed consent was obtained from all participants. All study protocols were approved by the institutional review boards of the Fred Hutchinson Cancer Research Center (Seattle, WA) and the five other participating study sites. Approvals were also obtained from institutional review boards affiliated with the state cancer registries to conduct the data record linkages.

SNP selection

Tag and previously reported risk-associated SNPs were identified to examine common patterns of variation in the ERCC1, ERCC2, ERCC3, ERCC4, ERCC5, LIG1, POLE, XPA, and XPC gene regions. As the majority of participants were non -Hispanic white, tag SNP selection for each gene region (± 2.5 kb of the coding sequence) was based solely on HapMap Phase I and II Centre d'Etude du Polymorphism Humain (Utah residents of northern and western European ancestry; CEU) data (NCBI build 36, dbSNP build 129). The ldSelect algorithm [30] was applied to classify SNPs with a minor allele frequency (MAF) of ≥ 5% into bins with a pairwise linkage disequilibrium (LD) threshold of $r^2 \geq 0.8$. At least one SNP per bin was selected, prioritizing on the basis of SNP function class and predicted genotyping success. With respect to function class, SNPs were ranked in the following order from highest to lowest: non-synonymous SNP; SNP in the 5' promoter region; SNP in the 3' untranslated region, synonymous SNP, intronic SNP in a splice site; intronic enhancer SNP; and intronic SNP with no known function. Bioinformatics tools, including SIFT (http://sift.jcvi.org), PolyDoms (http://polydoms.cchms.org/polydoms), PolyPhen (http://genetics.bwh.harvard.edu/pph), and FastSNP (http://fastsnp.ibms.sinica.edu/tw), aided in annotating SNP function. SNPs with Illumina design scores of ≥ 0.6 were preferentially selected as a means to increase the likelihood of genotyping success.

DNA extraction

For each participant, genomic DNA was extracted from one 2 ml aliquot of stored whole blood using the QIAamp DNA Blood Midi Kit (Qiagen). A second 2 ml aliquot from 84 participants was processed identically and served as duplicate samples for quality control purposes. Eighteen control samples and one duplicate sample were excluded because of limited DNA yield.

Genotyping

Using a custom-designed 384-plex GoldenGate assay (Illumina) and individual pre- or custom-designed TaqMan assays (Applied Biosystems), DNA from 793 cases, 1,568 controls, and 83 duplicate samples were genotyped for 82 of the 84 SNPs selected. The two omitted SNPs (XPC rs13099160 and ERCC2 rs1799788) were tag SNPs with relatively low MAFs in European populations that could not be included in the Illumina panel. Genotyping was conducted by laboratory technicians who were blinded to the case-control status of study samples and the identity of duplicate samples. Along with samples for cases and controls, duplicate samples

were interspersed randomly across genotyping plates. For TaqMan assays, in-house DNA samples of known genotypes and no-template samples prepared in an identical fashion were also tested as positive and negative controls, respectively, on each plate.

Quality control evaluation and data exclusion

Based on the Illumina genotyping results, 10 of the 2,444 samples were excluded due to genotyping failure or gender mismatch error. Of the 69 NER SNPs tested on the Illumina platform, five SNPs failed completely (LIG1 rs274884, LIG1 rs754848, POLE rs14302, POLE rs5744751, XPA rs2808676) and one SNP had < 90% genotyping call success (ERCC1 rs11615). Three of these six poorly performing SNPs (ERCC1 rs11615, LIG1 rs754848, POLE rs14302) were re-examined by TaqMan assay. Three individuals originally identified as cases, who were later determined to have had benign or carcinoid lung cancer, were additionally excluded.

Genotype data for 79 SNPs on 787 cases, 1,562 controls, and 82 blind duplicates were retained. For all 79 SNPs, genotype call success exceeded 95%, and genotype concordance between blind duplicates was 100%. Observed genotype frequencies in non-Hispanic white study controls did not deviate from those expected under Hardy-Weinberg equilibrium ($p > 0.001$ using Fisher's exact test).

Statistical analysis

To minimize the potential for confounding by race/ethnicity, analyses were limited to participants who self-identified as non-Hispanic white (744 cases, 1,477 controls). With the study population being predominantly non-Hispanic white, analyses specific to other race/ethnicity groups were not conducted. Unless otherwise specified, analyses were performed using Stata® 11 (StataCorp).

SNP analysis: To estimate the relative risk of lung cancer associated with SNP genotype, odds ratios (OR) and 95% confidence intervals (95% CIs), adjusted for the case-control matching factors (age, sex, enrollment year, baseline smoking status, and occupational history of asbestos exposure), were calculated using logistic regression. To assess trends in risk, SNP genotype was coded using a three-level ordinal variable that indicated the number of minor alleles carried (0, 1, or 2). The reference group was the most common homozygous genotype among controls.

Subgroup analyses were conducted to explore whether the extent of observed SNP associations with lung cancer differ by specific characteristics, including age at diagnosis (dichotomized using the median value: < 70, ≥ 70 years), sex (male, female), baseline smoking status (former, current), number of pack-years smoked (thirds of distribution among controls: < 40, 40-53, ≥ 54), occupational history of asbestos exposure (yes, no), CARET trial arm assignment (intervention, placebo), tumor histology (NSCLC, SCLC), and diet. Dietary factors included intake level of total fruits, total vegetables, vitamin C, vitamin E, folate, nitrosamine-containing foods, total carotenoids, total polyunsaturated fatty acids, and foods from individual botanical families (Rosaceae, Rutaceae, Cruciferae, Apiaceae, Cucurbitaceae, Leguminosae, Chenopidiaceae, Solonaceae). Levels of each dietary factor were defined by thirds of the intake distribution among controls [as reported in [26]]. To formally test for departure from a multiplicative relation, p-values for the Wald test of the cross-product term between SNP genotype (coded ordinally) and the exposure of interest (coded ordinally for smoking quantity and dietary intake level) were calculated.

Haplotype and diplotype analysis: For each gene region, pairwise LD patterns of genotyped SNPs were visualized using Haploview, version 4.2 [31]. Gene-specific haplotype analyses, including haplotype imputation from genotype data, were conducted using the haplo.stats package (http://mayoresearch.mayo.edu/schaid_lab/software.cfm) in R, version 2.10.1. Haplotypes were inferred from genotype data for tag SNPs only. Haplotype frequencies were estimated using the expectation-maximization algorithm, and case-control differences in haplotype frequencies were assessed using global test score statistics. To evaluate haplotype associations with lung cancer risk, ORs and 95% CIs were estimated for common haplotypes (frequency > 1%), adjusting for the matching factors, under an additive model. The most common haplotype was used as the reference group. Diplotype analyses were conducted post-hoc to determine which SNP genotypes in com-

bination were associated with lung cancer risk.

Gene-level analysis: To address the issue of chance associations arising from multiple testing of individual SNPs with lung cancer risk, set-based tests were conducted for each gene using PLINK version 1.04 [32], with each set comprised of up to 5 independent ($r^2 < 0.5$), nominally significant ($p < 0.05$) SNPs per gene. This approach controlled for the number of SNPs and the extent of LD between SNPs in each gene region, as gene-level association test statistics were first derived by averaging association test statistics of individual SNPs within a set and then permuted 10,000 times to calculate empirical p-values. Max(T) permutation was conducted within clusters defined by the matching variables to preserve existing relations of these variables with case-control status in the study population. Analyses accounting for the total number of genes studied were not performed, due to primary interest in examining variation in genes individually, as opposed to collectively, in relation to lung cancer.

Results

In terms of baseline characteristics, most participants were ages 55 and older, male, and current smokers (Table 1). Cases were slightly older and reported a heavier smoking history than controls.

Table 1. Baseline characteristics of non-Hispanic white study participants, by case-control status

	Cases (n=744)		Controls (n=1,477)	
Characteristic	n	%	n	%
Age, years[a]				
< 50	11	1.5	24	1.6
50-54	132	17.7	364	24.6
55-59	201	27.0	376	25.5
60-64	237	31.8	445	30.1
≥ 65	163	21.9	268	18.1
Sex[a]				
Male	501	67.3	984	66.6
Female	243	32.7	493	33.4
Smoking status[a]				
Former smoker	205	27.5	407	27.6
Current smoker	539	72.5	1070	72.4
No. of pack-years smoked				
< 40	180	24.2	511	34.6
40-53	249	33.5	493	33.4
≥ 54	315	42.3	473	32.0
Occupational asbestos exposure[a]				
Yes	125	16.8	247	16.7
No	619	83.2	1230	83.3
Trial arm assignment				
Intervention	403	54.2	775	52.5
Placebo	341	45.8	702	47.5

[a]Case-control matching variable

Of the 79 SNPs successfully genotyped, 69 were tag SNPs and ten were previously reported risk-associated SNPs. Gene coverage, defined as the proportion of common SNPs captured (through LD) for a given gene in the HapMap Phase I and II CEU populations by those SNPs genotyped, ranged from 92% to 100%. Specifically, coverage of ERCC2, XPA, and XPC was incomplete, since one tag SNP was not genotyped in each of these genes as indicated in the Materials and methods section.

Modest associations of lung cancer risk were detected for five tag SNPs in three of the genes examined (Table 2): ERCC2 rs50871 (OR per C allele, 1.15; 95% CI, 1.01-1.30, p=0.029); LIG1 rs156640 (OR per G allele, 1.23; 95% CI, 1.08-1.40; p=0.001), rs156641 (OR per A allele, 1.23; 95% CI, 1.08-1.40; p=0.002), and rs8100261 (OR per A allele, 0.86; 95% CI, 0.76-0.98; p=0.023); and XPA rs3176658 (OR per A allele, 0.83; 95% CI, 0.69-1.00; p=0.055). For XPA rs3176658, the OR for carriage of the minor allele (i.e., GA/AA vs. GG genotype) was 0.80 (95% CI, 0.65-0.99). These risk estimates remained unchanged after further adjustment for the number of pack-years smoked at baseline (data not shown). None of the previously reported risk-associated SNPs, however, were associated with lung cancer risk.

Some observed SNP associations differed in magnitude by sex and prior history of occupational asbestos exposure (Table 3). The association between XPA rs3176658 and lung cancer risk was more pronounced in women (OR per A allele, 0.62; 95% CI, 0.44-0.87) than men (OR per A allele, 0.95; 95% CI, 0.76-1.19). With regard to asbestos exposure, the association of ERCC2 rs50871 with lung cancer risk was stronger in exposed (OR per C allele, 1.48; 95% CI, 1.09-2.01) than unexposed (OR per C allele, 1.09; 95% CI, 0.95-1.25) persons, while the associations of LIG1 rs156640 and rs156641

Table 2. Association of tag and previously reported risk-associated SNPs in nucleotide excision repair genes with lung cancer risk among non-Hispanic white participants

					No. of cases			No. of controls				
SNP	Gene	Major Allele (A1)	Minor Allele (A2)	MAF[a]	A1/A1	A1/A2	A2/A2	A1/A1	A1/A2	A2/A2	OR per A2 (95% CI)[b]	p-trend
rs735482	ERCC1	A	C	14.3%	564	165	15	1086	360	31	0.90 (0.75-1.08)	0.252
rs3212986	ERCC1	C	A	25.5%	418	284	41	829	537	107	0.97 (0.84-1.11)	0.634
rs16979802[c]	ERCC1	G	C	8.1%	642	99	3	1253	210	14	0.85 (0.67-1.08)	0.192
rs11615	ERCC1	T	C	38.3%	289	359	96	585	653	239	0.95 (0.84-1.08)	0.451
rs2298881	ERCC1	C	A	10.6%	609	124	11	1183	273	20	0.91 (0.74-1.12)	0.379
rs10853773	ERCC2	G	A	28.2%	386	304	53	761	594	118	0.98 (0.85-1.12)	0.742
rs13181	ERCC2	A	C	37.2%	310	330	102	585	681	207	0.95 (0.83-1.08)	0.437
rs1799787	ERCC2	G	A	30.8%	366	304	71	700	639	134	0.97 (0.84-1.11)	0.623
rs238417	ERCC2	G	C	40.8%	254	351	136	513	719	242	1.05 (0.92-1.19)	0.449
rs238416	ERCC2	G	A	34.7%	307	333	103	628	668	178	1.07 (0.93-1.21)	0.339
rs50872	ERCC2	G	A	23.6%	433	261	47	851	554	71	1.02 (0.88-1.18)	0.813
rs50871	ERCC2	A	C	48.2%	166	389	188	400	726	346	1.15 (1.01-1.30)	0.029
rs238404	ERCC2	A	G	44.5%	218	354	167	450	727	289	1.08 (0.95-1.23)	0.218
rs1799793	ERCC2	G	A	35.5%	326	329	89	610	685	182	0.93 (0.82-1.07)	0.316
rs238406	ERCC2	G	T	43.6%	228	357	159	461	745	271	1.07 (0.95-1.22)	0.275
rs3810366	ERCC2	C	G	45.4%	245	353	144	431	744	295	0.92 (0.81-1.04)	0.193
rs1803541	ERCC3	G	A	4.0%	681	58	4	1355	113	2	1.09 (0.80-1.48)	0.580
rs4150506	ERCC3	G	A	22.1%	449	255	37	894	511	71	1.00 (0.86-1.16)	0.995
rs4150471	ERCC3	G	A	29.6%	388	293	62	725	623	125	0.93 (0.81-1.07)	0.316
rs4150454	ERCC3	A	G	38.8%	259	359	121	555	690	225	1.08 (0.95-1.23)	0.235
rs4150403	ERCC3	G	A	8.9%	638	104	2	1223	243	10	0.81 (0.64-1.02)	0.078
rs1799797	ERCC4	T	A	26.7%	391	295	58	796	574	107	1.05 (0.92-1.21)	0.461
rs1800067	ERCC4	G	A	7.1%	635	101	7	1273	199	5	1.09 (0.86-1.39)	0.465
rs3136166	ERCC4	T	G	34.2%	320	340	84	636	672	169	1.00 (0.88-1.14)	0.980
rs1799801[c]	ERCC4	A	G	28.3%	380	302	60	763	592	121	1.02 (0.88-1.17)	0.816
rs2018836[c]	ERCC5	G	A	31.1%	361	320	62	705	624	147	0.95 (0.83-1.09)	0.444
rs2296147	ERCC5	A	G	47.8%	182	385	174	407	723	341	1.07 (0.94-1.21)	0.309
rs7325708	ERCC5	G	C	18.3%	497	225	21	985	439	50	0.97 (0.83-1.15)	0.749
rs1047768	ERCC5	G	A	41.1%	256	378	108	507	722	245	0.96 (0.84-1.09)	0.498
rs1047769[c]	ERCC5	A	G	3.8%	694	45	1	1359	111	1	0.80 (0.56-1.14)	0.214
rs2227869	ERCC5	C	G	3.9%	680	63	1	1362	110	2	1.15 (0.84-1.59)	0.374
rs3759500	ERCC5	G	A	23.0%	445	260	37	879	513	82	0.98 (0.84-1.14)	0.798
rs4150351	ERCC5	A	C	17.0%	499	221	22	1016	417	42	1.07 (0.90-1.26)	0.452
rs4150355	ERCC5	G	A	37.6%	277	375	90	584	672	219	0.98 (0.86-1.12)	0.819
rs732321[c]	ERCC5	A	C	3.9%	680	63	1	1364	111	2	1.14 (0.83-1.56)	0.409
rs4150386	ERCC5	A	C	12.8%	589	150	5	1120	330	24	0.83 (0.68-1.01)	0.060
rs17655[c]	ERCC5	G	C	21.2%	454	261	29	919	489	68	1.01 (0.87-1.18)	0.856
rs873601	ERCC5	A	G	27.1%	392	299	51	783	584	107	1.00 (0.87-1.15)	0.990
rs4150393	ERCC5	A	G	11.1%	567	168	6	1166	291	19	1.10 (0.90-1.34)	0.347
rs274883	LIG1	A	G	18.1%	523	192	29	996	424	55	0.91 (0.77-1.07)	0.247

rs3731014	LIG1	G	A	12.3%	595	138	11	1139	311	26	0.85 (0.70-1.03)	0.102
rs3731007	LIG1	G	A	5.7%	656	86	2	1311	158	5	1.07 (0.82-1.39)	0.630
rs156640	LIG1	C	G	41.0%	217	366	157	499	741	233	1.23 (1.08-1.40)	0.001
rs156641	LIG1	G	A	35.3%	271	352	121	596	709	164	1.23 (1.08-1.40)	0.002
rs754948	LIG1	C	T	4.6%	684	59	1	1345	128	4	0.89 (0.66-1.22)	0.474
rs4987068[c]	LIG1	G	A	2.1%	699	44	0	1411	63	0	1.40 (0.94-2.08)	0.100
rs3730931[c]	LIG1	A	G	12.4%	595	137	11	1137	313	26	0.84 (0.69-1.02)	0.081
rs20580	LIG1	C	A	47.9%	175	381	184	395	744	334	1.11 (0.98-1.26)	0.111
rs8100261	LIG1	G	A	48.3%	227	368	147	396	734	347	0.86 (0.76-0.98)	0.023
rs20579[c]	LIG1	G	A	13.0%	583	141	18	1126	312	36	0.90 (0.75-1.09)	0.278
rs14302	POLE	C	T	41.1%	263	336	144	535	668	272	1.04 (0.92-1.17)	0.551
rs5745066	POLE	G	A	2.1%	709	34	1	1412	63	0	1.11 (0.73-1.68)	0.634
rs7963858	POLE	C	T	31.8%	340	332	72	690	635	152	1.02 (0.89-1.16)	0.809
rs5744990	POLE	G	A	16.1%	512	210	22	1039	400	37	1.08 (0.91-1.28)	0.374
rs5744941	POLE	A	T	8.7%	612	127	5	1230	237	10	1.07 (0.86-1.33)	0.564
rs5744857	POLE	G	A	44.9%	246	362	143	448	723	299	0.92 (0.81-1.05)	0.211
rs5744807	POLE	A	G	4.1%	691	50	2	1356	116	2	0.88 (0.64-1.23)	0.463
rs5744799[c]	POLE	A	T	1.7%	726	18	0	1427	49	0	0.74 (0.43-1.29)	0.287
rs5744769	POLE	C	G	13.7%	551	181	12	1098	354	25	1.00 (0.83-1.20)	0.991
rs5744761	POLE	G	A	4.5%	676	66	1	1345	129	2	1.01 (0.75-1.37)	0.921
rs11147005	POLE	A	G	31.1%	375	300	66	703	627	145	0.91 (0.80-1.05)	0.189
rs3176757	XPA	G	A	19.8%	473	239	32	944	479	52	1.04 (0.89-1.22)	0.610
rs3176748	XPA	A	G	31.5%	335	332	63	679	620	145	0.99 (0.86-1.14)	0.898
rs2808667	XPA	G	A	6.6%	657	83	1	1282	178	8	0.85 (0.65-1.10)	0.220
rs2805835	XPA	G	C	11.3%	597	141	5	1159	301	17	0.89 (0.73-1.10)	0.290
rs3176689	XPA	A	T	16.8%	518	203	23	1024	409	43	0.99 (0.84-1.17)	0.902
rs3176683	XPA	A	G	6.7%	662	77	1	1284	185	7	0.77 (0.59-1.01)	0.056
rs3176658	XPA	G	A	13.9%	581	148	14	1094	348	31	0.83 (0.69-1.00)	0.055
rs3176633	XPA	C	G	15.0%	552	173	19	1064	384	29	0.94 (0.79-1.13)	0.511
rs1800975	XPA	G	A	33.8%	320	326	71	622	621	166	0.95 (0.83-1.09)	0.452
rs1126547	XPC	G	C	13.3%	551	184	9	1105	351	21	1.02 (0.85-1.23)	0.815
rs2228001	XPC	A	C	39.8%	263	379	100	520	739	218	0.96 (0.85-1.10)	0.594
rs3731124	XPC	A	C	24.2%	424	276	44	847	545	84	1.00 (0.87-1.16)	0.973
rs2607737	XPC	G	A	48.5%	177	395	168	381	751	338	1.03 (0.91-1.17)	0.624
rs9653966	XPC	A	C	8.4%	631	104	9	1236	231	8	0.99 (0.79-1.24)	0.921
rs1124303	XPC	T	G	8.0%	626	114	4	1252	214	11	1.02 (0.81-1.28)	0.848
rs3731143	XPC	A	G	6.1%	648	91	5	1307	161	9	1.14 (0.89-1.46)	0.295
rs2228000	XPC	G	A	25.1%	401	299	43	822	566	87	1.06 (0.91-1.22)	0.451
rs2733537	XPC	A	G	33.4%	320	337	84	638	684	150	1.05 (0.91-1.20)	0.515

[a]Minor allele frequency among controls; [b] Adjusted for age, sex, smoking status, occupational asbestos exposure, and enrollment year; [c] non-tag SNP

Table 3. Subgroup-specific odds ratios (OR) and 95% confidence intervals (95% CI) for lung cancer risk associated with selected SNPs in non-Hispanic white study participants

Characteristic Subgroup	*ERCC2* rs50871 OR per C allele (95% CI)[a]	*LIG1* rs156640 OR per G allele (95% CI)[a]	*LIG1* rs156641 OR per A allele (95% CI)[a]	*LIG1* rs8100261 OR per A allele (95% CI)[a]	*XPA* rs3176658 OR per A allele (95% CI)[a]
Age at diagnosis					
<70 years	1.16 (0.98-1.38)	1.31 (1.10-1.56)	1.30 (1.08-1.55)	0.78 (0.65-0.92)	0.76 (0.58-0.99)
≥70 years	1.13 (0.94-1.36)	1.16 (0.96-1.41)	1.16 (0.96-1.41)	0.97 (0.80-1.17)	0.91 (0.69-1.18)
$p_{interaction}$	*0.80*	*0.36*	*0.43*	*0.09*	*0.34*
Sex					
Male	1.20 (1.03-1.40)	1.16 (0.99-1.36)	1.18 (1.00-1.39)	0.89 (0.76-1.04)	0.95 (0.76-1.19)
Female	1.06 (0.85-1.32)	1.39 (1.11-1.74)	1.32 (1.05-1.65)	0.81 (0.66-1.01)	0.62 (0.44-0.87)
$p_{interaction}$	*0.36*	*0.20*	*0.45*	*0.52*	*0.04*
Smoking status					
Former Smoker	1.03 (0.81-1.32)	1.07 (0.83-1.37)	1.07 (0.83-1.38)	0.96 (0.76-1.22)	0.96 (0.67-1.36)
Current Smoker	1.20 (1.03-1.39)	1.30 (1.12-1.52)	1.29 (1.11-1.51)	0.83 (0.71-0.96)	0.79 (0.63-0.98)
$p_{interaction}$	*0.31*	*0.17*	*0.20*	*0.29*	*0.35*
No. of pack-years smoked					
<40	1.34 (1.04-1.71)	1.21 (0.94-1.56)	1.20 (0.93-1.55)	0.81 (0.63-1.03)	0.62 (0.42-0.92)
40-53	1.10 (0.88-1.38)	1.13 (0.91-1.42)	1.24 (0.99-1.55)	0.83 (0.66-1.04)	1.06 (0.77-1.46)
≥54	1.05 (0.86-1.28)	1.31 (1.07-1.61)	1.22 (0.99-1.51)	0.96 (0.78-1.17)	0.79 (0.59-1.07)
$p_{interaction}$	*0.17*	*0.57*	*0.97*	*0.26*	*0.47*
Occupational asbestos exposure					
Yes	1.48 (1.09-2.01)	0.93 (0.68-1.27)	0.91 (0.66-1.27)	0.98 (0.73-1.34)	0.69 (0.42-1.13)
No	1.09 (0.95-1.25)	1.31 (1.13-1.50)	1.30 (1.13-1.50)	0.84 (0.73-0.96)	0.86 (0.70-1.05)
$p_{interaction}$	*0.07*	*0.05*	*0.06*	*0.32*	*0.44*
Trial arm assignment					
Intervention arm	1.09 (0.91-1.30)	1.26 (1.06-1.51)	1.27 (1.06-1.53)	0.84 (0.71-1.00)	0.85 (0.66-1.10)
Placebo arm	1.22 (1.01-1.46)	1.21 (1.00-1.45)	1.17 (0.97-1.42)	0.88 (0.73-1.06)	0.81 (0.62-1.08)
$p_{interaction}$	*0.38*	*0.73*	*0.56*	*0.66*	*0.78*
Histologic subtype					
Non-small cell lung cancer	1.15 (1.00-1.33)	1.18 (1.02-1.37)	1.19 (1.02-1.38)	0.89 (0.78-1.03)	0.80 (0.65-1.00)
Small cell lung cancer	1.16 (0.90-1.49)	1.39 (1.07-1.80)	1.38 (1.06-1.79)	0.76 (0.58-0.98)	0.73 (0.48-1.09)

Total fruits					
≤3.5 servings/wk	1.06 (0.87-1.30)	1.14 (0.92-1.41)	1.13 (0.92-1.40)	0.87 (0.71-1.08)	0.76 (0.55-1.03)
3.6-8.4 servings/wk	1.33 (1.05-1.68)	1.49 (1.18-1.88)	1.58 (1.24-2.01)	0.78 (0.62-0.98)	1.12 (0.80-1.56)
>8.4 servings/wk	1.16 (0.92-1.47)	1.09 (0.86-1.37)	0.98 (0.77-1.25)	0.96 (0.77-1.21)	0.71 (0.50-1.01)
$p_{interaction}$	*0.53*	*0.93*	*0.54*	*0.62*	*0.94*
Total vegetables					
≤8.4 servings/wk	1.09 (0.87-1.36)	1.07 (0.86-1.34)	1.18 (0.93-1.48)	0.92 (0.73-1.15)	0.97 (0.72-1.32)
8.5-13.3 servings/wk	1.14 (0.92-1.43)	1.38 (1.10-1.73)	1.21 (0.96-1.52)	0.78 (0.63-0.97)	0.61 (0.43-0.88)
>13.3 servings/wk	1.21 (0.97-1.51)	1.28 (1.02-1.60)	1.29 (1.02-1.62)	0.88 (0.71-1.10)	0.88 (0.63-1.22)
$p_{interaction}$	*0.50*	*0.26*	*0.56*	*0.80*	*0.56*
Cruciferae[b]					
≤1.0 servings/wk	1.19 (0.97-1.48)	1.14 (0.91-1.42)	1.08 (0.86-1.35)	0.97 (0.78-1.20)	1.07 (0.79-1.47)
1.1-2.3 servings/wk	1.02 (0.81-1.29)	1.57 (1.24-1.99)	1.48 (1.16-1.88)	0.66 (0.52-0.83)	0.75 (0.54-1.05)
>2.3 servings/wk	1.25 (1.00-1.57)	1.07 (0.85-1.34)	1.16 (0.93-1.46)	0.96 (0.77-1.18)	0.66 (0.47-0.95)
$p_{interaction}$	*0.84*	*0.70*	*0.65*	*0.99*	*0.04*
Nitrosamine-containing foods					
≤1.2 servings/wk	1.13 (0.90-1.44)	1.56 (1.23-1.98)	1.43 (1.12-1.82)	0.71 (0.57-0.90)	0.72 (0.50-1.03)
1.3-3.6 servings/wk	0.98 (0.79-1.22)	1.14 (0.91-1.42)	1.16 (0.93-1.44)	0.93 (0.75-1.16)	0.92 (0.68-1.25)
>3.6 servings/wk	1.34 (1.08-1.66)	1.11 (0.89-1.39)	1.14 (0.91-1.44)	0.91 (0.73-1.13)	0.82 (0.59-1.15)
$p_{interaction}$	*0.26*	*0.04*	*0.18*	*0.14*	*0.59*
Total carotenoids[c]					
≤6890.4 mcg/d	1.14 (0.91-1.42)	1.41 (1.12-1.77)	1.44 (1.14-1.81)	0.73 (0.58-0.92)	0.80 (0.58-1.11)
6890.5-10802.1 mcg/d	1.24 (0.99-1.54)	1.22 (0.97-1.53)	1.21 (0.96-1.53)	0.89 (0.71-1.11)	1.00 (0.72-1.38)
>10802.1 mcg/d	1.11 (0.89-1.39)	1.10 (0.88-1.38)	1.07 (0.85-1.34)	0.96 (0.77-1.18)	0.71 (0.51-1.00)
$p_{interaction}$	*0.90*	*0.14*	*0.08*	*0.10*	*0.60*

[a]Adjusted for case-control matching factors, as appropriate; [b]Broccoli, cauliflower, brussel sprouts, cole slaw, cabbage, sauerkraut, mustard greens, turnip greens, and collards; [c]Alpha-carotene, beta-carotene, beta-crytoxanthin, lutein, zeaxanthin, and lycopene

with lung cancer risk were stronger in unexposed (OR per G allele, 1.31; 95% CI, 1.13-1.50; OR per A allele, 1.30; 95% CI, 1.13-1.50) than exposed (OR per G allele, 0.93; 95% CI, 0.68-1.27; OR per A allele, 0.91; 95% CI, 0.66-1.27) persons. Although the observed associations did not vary appreciably by age at diagnosis, smoking history, trial arm assignment, or histology, associations for the LIG1 and XPA SNPs were slightly stronger among persons diagnosed with lung cancer at earlier than later ages and among current than former smokers. For most of the dietary factors examined, the observed SNP associations varied little according to intake level [data not shown]. Monotonic trends in risk per minor allele were evident only for XPA rs3176658 by Cruciferae intake, with a moderately stronger association among persons with higher intake, and for LIG1 rs156640 and rs156641 by intake of nitrosamine-containing foods and all three LIG1 SNPs by intake of total carotenoids, each with a moderately stronger association among persons with lower intake.

Haplotype frequencies for LIG1 (global p-value=0.008) and XPA (global p-value=0.01), but not ERCC2 (global p-value=0.30), differed between cases and controls. Relative to the most common LIG1 haplotype, two LIG1 haplotypes were associated with increased lung cancer risk (Table 4): the second most common haplotype (OR, 1.22; 95% CI, 1.06-1.41), which contained the putative risk-conferring alleles for rs156640, rs156641, and rs8100261, and the least common haplotype (OR, 1.88; 95% CI, 1.13-3.14), which contained the putative risk-conferring alleles for rs156640 and rs8100261. In the analysis of LIG1 rs156640, rs156641, and rs8100261 diplotypes, lung cancer risk was greatest (OR: 1.69, 95% CI: 1.24-2.30) for persons carrying the homozygous minor (versus major) genotypes of rs156640 and rs156641 and homozygous major (versus minor) genotype of rs8100261 (Table 4). These three SNPs were moderately correlated with one another (r^2 = 0.50-0.78) among controls. For XPA, one haplotype, which contained all major alleles, was associated with elevated lung cancer risk (OR, 1.47; 95% CI, 1.14-1.91), relative to the most common haplotype, which contained the minor allele for rs3176748 and the major alleles for the other eight SNPs (Table 4). Only a suggestive decrease in risk was found for the two XPA haplotypes containing the minor (A) allele of rs3176658.

Gene-level analyses supported an association of lung cancer risk with LIG1 only. Except for LIG1 (empirical p-value=0.01) and ERCC2 (empirical p-value=0.18), the empirical p-values for XPA and the other six genes were essentially 1.00, since none of the per-allele ORs for individual SNPs in these genes attained nominal statistical significance.

Discussion

In this nested case-control study, which was comprised entirely of former or current smokers, lung cancer risk was modestly associated with tag SNPs in the ERCC2, LIG1, and XPA genes. The associations of ERCC2 rs13181 and XPA rs1800975 SNPs with lung cancer risk seen in meta-analyses [22-24], however, were not confirmed. Of the SNP associations detected, some varied in magnitude by sex and occupational asbestos exposure history, while none varied appreciably by age at diagnosis, smoking quantity, and lung cancer histology. Diet also appeared to exert minimal influence on genetic susceptibility, given the lack of effect modification by most of the dietary characteristics studied.

The most robust evidence to support our hypothesis was found for the LIG1 gene: rs156640 and rs156641, two intronic SNPs of unknown function located about 645 base pairs apart, were each associated with a 23% increase in risk per minor allele, and rs8100261, a predicted intronic enhancer, was associated with a 14% decrease in risk per minor allele. Homozygous carriage of the putative risk (versus non-risk) allele for all three markers combined was associated with a 69% increase in lung cancer risk. LIG1 is a key nuclear enzyme that maintains genomic integrity by joining Okazaki fragments during DNA replication and sealing single-strand breaks in both nucleotide and base excision repair processes [33-35]. Missense mutations in the human LIG1 gene result in extreme sensitivity to DNA-damaging substances, including alkylating agents, ionizing radiation, and ultraviolet light [36]. In mice, cells lacking LIG1 display normal DNA repair capacity, but less genome stability [37].

The results of prior studies examining LIG1 variants in relation to lung cancer risk [38-45] and ours, although not entirely consistent, point to LIG1 or another gene at 19q13.2-q13.3 as a

Table 4. Relation of *LIG1* haplotypes and diplotypes and *XPA* haplotypes to lung cancer risk in non-Hispanic white participants

LIG1 Haplotypes

rs274883	rs3731014	rs3731007	rs156640	rs156641	rs754948	rs20580	rs8100261	Cases (%)	Controls (%)	OR (95% CI)[a]
A	G	G	C	G	C	C	A	43.1	46.3	1.00 (reference)
A	G	G	G	A	C	A	G	39.7	35.0	1.22 (1.06-1.41)
G	A	G	C	G	C	A	G	9.3	10.4	0.95 (0.76-1.18)
G	G	A	G	G	T	C	G	3.9	4.4	0.97 (0.70-1.35)
G	A	G	C	G	C	A	A	1.4	1.8	0.79 (0.47-1.33)
G	G	A	G	G	C	C	G	1.9	1.1	1.88 (1.13-3.14)
*	*	*	*	*	*	*	*	0.7	1.0	0.71 (0.30-1.70)

global p-value = 0.008

LIG1 Diplotypes

rs156640	rs156641	rs8100261	Cases (n)	Controls (n)	OR (95% CI)[a]
CC	GG	AA	146	340	1.00 (reference)
CG	AG	AG	267	523	1.17 (0.92-1.50)
GG	AA	GG	119	163	1.69 (1.24-2.30)
CC	GG	AG	61	137	1.00 (0.70-1.44)
CG	AG	GG	47	118	0.92 (0.62-1.35)
CG	GG	AG	38	69	1.27 (0.81-1.98)
GG	AG	GG	35	65	1.24 (0.79-1.96)
Other genotype combinations			25	50	1.14 (0.68-1.92)

XPA Haplotypes

rs3176757	rs3176748	rs2808667	rs2805835	rs3176689	rs3176683	rs3176658	rs3176633	rs1800975	Cases (%)	Controls (%)	OR (95% CI)[a]
G	G	G	G	A	A	G	C	G	31.6	31.7	1.00 (reference)
G	A	G	G	T	A	G	C	G	16.7	16.7	1.00 (0.83-1.21)
G	A	G	C	A	A	G	C	G	10.2	11.3	0.92 (0.73-1.15)
G	A	G	G	A	A	G	C	G	9.1	6.3	1.47 (1.14-1.91)
G	A	G	G	A	A	A	C	A	6.1	7.2	0.86 (0.65-1.13)
G	A	A	G	A	A	A	C	A	5.7	6.5	0.89 (0.68-1.18)
A	A	G	G	A	A	G	C	A	6.4	5.0	1.31 (0.99-1.75)
A	A	G	G	A	A	G	G	A	8.6	8.0	1.08 (0.85-1.38)
A	A	G	G	A	G	G	G	A	5.4	6.7	0.79 (0.60-1.06)
*	*	*	*	*	*	*	*	*	0.2	0.6	0.97 (0.17-5.68)

global p-value = 0.01

[a] Adjusted for age, sex, smoking status, occupational asbestos exposure, and enrollment year

susceptibility locus for lung cancer. Among these studies, different sets of SNPs have been analyzed, and all except for one by Hung et al. [41], in which participants (1,604 cases, 2,053 controls) were enrolled at multiple sites across Central and Eastern Europe, have included fewer cases (range: 138-599). Of those examining rs156640, rs156641, rs8100261, or other strongly correlated SNPs [40, 42-45], one has likewise found a ~20% increase in lung cancer risk per minor allele of rs10500298, a SNP in perfect LD with rs156641, among Caucasian smokers with a ≥ 10 pack-year history [45]. Although the other studies did not identify these individual SNPs as putative risk markers, they were relatively less powered to detect effects of low penetrance variants [40, 42-44]. Similar to others [38, 39, 42, 44], we observed no association between rs20580 and lung cancer risk. In contrast, we did not find associations with rs3730931 and rs20579, two correlated SNPs for which Hung et al. [41] reported a 19% increase in risk per minor allele. In a preceding study of early onset lung cancer also conducted in Europe, the heterozygous (versus major homozygous) genotype for each of these SNPs was associated with a 73% increase in risk [40]. Among U.S.-based studies, one found no relation with rs20579 in a multiethnic population [44], while one found a decreased risk per minor allele for rs20579 in African Americans, but not Latinos [43]. In the latter [43], a modest association with rs439132, a common SNP present in non-Europeans only, was also detected in African Americans.

All studies that have examined LIG1 haplotypes have found associations with lung cancer risk, particularly in smokers, despite using different SNPs to construct haplotypes [42-44]. In the most analogous study to ours, Michiels et al. [42] similarly inferred three common LIG1 haplotypes (although based on 22 SNPs) in a small population of Caucasian smokers. The haplotype frequencies among controls were 45%, 33%, and 15% in their study, compared to 46%, 35%, and 10%, respectively, in ours. However, these investigators observed a decreased risk of lung cancer associated with the third most (versus most) common haplotype, which we did not, and we observed an increased risk of lung cancer associated with the second most (versus most) common haplotype, which they did not. This discrepancy does not appear to be explained by differences in the coverage of LIG1 variation between studies, since the SNPs that we genotyped tag (at $r^2 > 0.8$) all 22 SNPs that Michiels et al. included in their haplotype analysis.

In the absence of gene-level associations for XPA and ERCC2, the SNP and haplotype associations observed for these genes are likely to be spurious. The lack of association for XPA at the gene level, however, may be partly attributed to the fact that SNP associations were evaluated strictly under the additive model in the set-based tests conducted. At the nominal level of statistical significance, XPA rs3176658 was associated with lung cancer risk under the dominant, but not additive, model. The XPA SNP rs1800975 has been primarily examined in prior candidate gene studies of lung cancer. In a meta-analysis of six studies [25], the rs1800975 AA (versus GA/GG) genotype was associated with a 26% increased risk of lung cancer in Caucasians, a relation in the opposite direction to what we observed. In a pooled analysis of 2803 lung cancer cases and 3452 controls from three European studies (two were included in the aforementioned meta-analysis), no association was evident [46]. Mutations in the human XPA gene have been shown to inhibit the interaction of XPA with ERCC1, a critical event in the NER process [47]. XPA-deficient mice have also been found to develop lung tumors after benzo(a)pyrene exposure [48]. Taken together, a closer inspection of the underlying genetic architecture in the XPA gene region (chromosome 9q22.3) in relation to lung cancer may be warranted.

Our inability to confirm the results of meta-analyses associating carriage of the ERCC2 rs13181 CC (versus AA) genotype with a ~25% increased risk of lung cancer in Caucasians [23, 24] may stem from studying primarily long-term, heavy smokers. Some studies, including the largest conducted in the U.S., have found a stronger positive association in never smokers than ever smokers, along with a possible inverse association in heavy smokers [49, 50]. Due to the marked effect of smoking on DNA damage, the extent to which this non-synonymous variant affects DNA repair and thereby lung cancer risk, especially if modest, may be only apparent when examining never and light smokers. With regard to ERCC2, we observed an association between rs50871 and lung cancer risk, particularly in persons exposed

to asbestos. This intronic SNP of unknown function was unrelated to overall lung cancer risk in a smaller study conducted in China [51].

Among the ~315,000 SNPs analyzed in the earliest GWA studies of lung cancer [4-7] were some that we examined, including LIG1 rs156641 and ERCC2 rs13181. It should be noted, however, that by employing a tag SNP approach, we were able to more extensively survey common patterns of variation in specific gene regions and identify possibly novel lung cancer susceptibility markers in persons of European ancestry. For example, eleven LIG1 SNPs were genotyped in the present study, with at least one SNP selected from each of the eight tag bins identified based on HapMap CEU data. In comparison, five LIG1 SNPs were genotyped using the Illumina HumanHap 300 array in those GWA studies, which capture SNPs in only four of the bins, excluding the putative risk-bearing LIG1 SNPs rs156640 and rs8100261. Although the Illumina HumanHap 300 array does include one SNP in strong LD with rs8100261 ($r^2 = 0.90$) that lies outside of LIG1, none of the commercial genomewide arrays presently assess rs156640 or its proxies (within ± 500 kb).

Our study was limited by common constraints inherent to candidate gene association studies. Due to the size and composition of our study population, only SNPs with ≥ 5% MAF in persons of European ancestry were examined, and SNP associations, both overall and subgroup-specific (e.g., by race, smoking and other lifestyle factors, and histology), may have been missed. In addition, the associations that we did observe could be attributed to highly correlated variants, or alternatively, given the number of SNPs and subgroups examined, to chance alone. In the case of LIG1, SNPs in strong LD with rs156640, rs156641, and rs8100261 appear to cluster primarily in the 19q13.2-q13.3 region where LIG1 resides. While GWA studies have identified 19q13.2 as a locus associated with smoking behavior, where two genes encoding for the nicotine-metabolizing enzymes CYP2A6 and CYP2B6 lie [52, 53], we did find that associations between LIG1 SNPs and lung cancer risk persisted after adjusting for the number of pack-years smoked. Our results are also suggestive that individuals carrying the putative risk-conferring LIG1 genotypes with diets low in carotenoids might be more susceptible to developing lung cancer. However, with the unexpected finding of some LIG1 associations being more pronounced in those with lower (rather than higher) intake of nitrosamine-containing foods, such indications of diet as a modifying factor should be interpreted cautiously.

Although subject to a number of caveats, our results, along with others, suggest that inherited variation in LIG1 and possibly other NER protein-encoding genes predispose to lung cancer in smokers. Further research to validate these observed associations in large and well-characterized populations is needed.

Acknowledgements

The authors thank the CARET participants, CARET investigators and staff at the participating institutions, and staff at the California, Oregon, and Washington state cancer registries for their contributions to this research. This work is supported by the National Cancer Institute (NCI) of the National Institutes of Health [R01 CA111703, U01 CA63673, K05 CA092002, T32 CA009168]. The collection of cancer incidence data from California was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the NCI's Surveillance, Epidemiology, and End Results Program under contract N01-PC-35136 awarded to the Northern California Cancer Center, contract PC-35139 awarded to the University of Southern California, and contract N01-PC-54404 awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries (CDC-NPCR), under agreement 1U58DP00807-01 awarded to the Public Health Institute. The collection of cancer incidence data from Oregon and Washington was also supported by the CDC-NPCR. The contents herein are solely the responsibility of the authors and do not necessarily represent the official views of the NIH, CDC, and the States of California, Oregon, and Washington or their Contractors and Subcontractors.

Address correspondence to: Dr. Chu Chen, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, P.O. Box 19024, Mailstop M5-C800, Seattle, WA 98109-1024, USA Tel: 1-206-667-6644; Fax: 1-206-667-2537; E-mail: cchen@fhcrc.org

References

[1] Hecht SS. Tobacco smoke carcinogens and

lung cancer. J Natl Cancer Inst 1999; 91: 1194 -1210.
[2] Matakidou A, Eisen T and Houlston RS. Systematic review of the relationship between family history and lung cancer risk. Br J Cancer 2005; 93: 825-833.
[3] Lorenzo BJ, Hemminki K. Familial lung cancer and aggregation of smoking habits: a simulation of the effect of shared environmental factors on the familial risk of cancer. Cancer Epidemiol Biomarkers Prev 2005; 14: 1738-1740.
[4] Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, Chen C, Goodman G, Field JK, Liloglou T, Xinarianos G, Cassidy A, McLaughlin J, Liu G, Narod S, Krokan HE, Skorpen F, Elvestad MB, Hveem K, Vatten L, Linseisen J, Clavel-Chapelon F, Vineis P, Bueno-de-Mesquita HB, Lund E, Martinez C, Bingham S, Rasmuson T, Hainaut P, Riboli E, Ahrens W, Benhamou S, Lagiou P, Trichopoulos D, Holcatova I, Merletti F, Kjaerheim K, Agudo A, Macfarlane G, Talamini R, Simonato L, Lowry R, Conway DI, Znaor A, Healy C, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M and Brennan P. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 2008; 452: 633-637.
[5] Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tsai YY, Chen WV, Shete S, Spitz MR and Houlston RS. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet 2008; 40: 616-622.
[6] Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, Stacey SN, Bergthorsson JT, Thorlacius S, Gudmundsson J, Jonsson T, Jakobsdottir M, Saemundsdottir J, Olafsdottir O, Gudmundsson LJ, Bjornsdottir G, Kristjansson K, Skuladottir H, Isaksson HJ, Gudbjartsson T, Jones GT, Mueller T, Gottsater A, Flex A, Aben KK, de VF, Mulders PF, Isla D, Vidal MJ, Asin L, Saez B, Murillo L, Blondal T, Kolbeinsson H, Stefansson JG, Hansdottir I, Runarsdottir V, Pola R, Lindblad B, van Rij AM, Dieplinger B, Haltmayer M, Mayordomo JI, Kiemeney LA, Matthiasson SE, Oskarsson H, Tyrfingsson T, Gudbjartsson DF, Gulcher JR, Jonsson S, Thorsteinsdottir U, Kong A and Stefansson K. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 2008; 452: 638-642.
[7] McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, McLaughlin J, Shepherd F, Montpetit A, Narod S, Krokan HE, Skorpen F, Elvestad MB, Vatten L, Njolstad I, Axelsson T, Chen C, Goodman G, Barnett M, Loomis MM, Lubinski J, Matyjasik J, Lener M, Oszutowska D, Field J, Liloglou T, Xinarianos G, Cassidy A, Vineis P, Clavel-Chapelon F, Palli D, Tumino R, Krogh V, Panico S, Gonzalez CA, Ramon QJ, Martinez C, Navarro C, Ardanaz E, Larranaga N, Kham KT, Key T, Bueno-de-Mesquita HB, Peeters PH, Trichopoulou A, Linseisen J, Boeing H, Hallmans G, Overvad K, Tjonneland A, Kumle M, Riboli E, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M and Brennan P. Lung cancer susceptibility locus at 5p15.33. Nat Genet 2008; 40: 1404-1406.
[8] Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WV, Spitz MR, Eisen T, Amos CI and Houlston RS. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat Genet 2008; 40: 1407-1409.
[9] Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, Mirabello L, Jacobs K, Wheeler W, Yeager M, Bergen AW, Li Q, Consonni D, Pesatori AC, Wacholder S, Thun M, Diver R, Oken M, Virtamo J, Albanes D, Wang Z, Burdette L, Doheny KF, Pugh EW, Laurie C, Brennan P, Hung R, Gaborieau V, McKay JD, Lathrop M, McLaughlin J, Wang Y, Tsao MS, Spitz MR, Wang Y, Krokan H, Vatten L, Skorpen F, Arnesen E, Benhamou S, Bouchard C, Metsapalu A, Vooder T, Nelis M, Valk K, Field JK, Chen C, Goodman G, Sulem P, Thorleifsson G, Rafnar T, Eisen T, Sauter W, Rosenberger A, Bickeboller H, Risch A, Chang-Claude J, Wichmann HE, Stefansson K, Houlston R, Amos CI, Fraumeni JF Jr, Savage SA, Bertazzi PA, Tucker MA, Chanock S and Caporaso NE. 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: 679-691.
[10] Broderick P, Wang Y, Vijayakrishnan J, Matakidou A, Spitz MR, Eisen T, Amos CI and Houlston RS. Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. Cancer Res 2009; 69: 6633-6641.
[11] Hsiung CA, Lan Q, Hong YC, Chen CJ, Hosgood HD, Chang IS, Chatterjee N, Brennan P, Wu C, Zheng W, Chang GC, Wu T, Park JY, Hsiao CF, Kim YH, Shen H, Seow A, Yeager M, Tsai YH, Kim YT, Chow WH, Guo H, Wang WC, Sung SW, Hu Z, Chen KY, Kim JH, Chen Y, Huang L, Lee KM, Lo YL, Gao YT, Kim JH, Liu L, Huang MS, Jung TH, Jin G, Caporaso N, Yu D, Kim CH, Su WC, Shu XO, Xu P, Kim IS, Chen YM, Ma H, Shen M, Cha SI, Tan W, Chang CH, Sung JS, Zhang M, Yang TY, Park KH, Yuenger J, Wang CL, Ryu JS, Xiang Y, Deng Q, Hutchinson A, Kim

JS, Cai Q, Landi MT, Yu CJ, Park JY, Tucker M, Hung JY, Lin CC, Perng RP, Boffetta P, Chen CY, Chen KC, Yang SY, Hu CY, Chang CK, Fraumeni JF Jr, Chanock S, Yang PC, Rothman N and Lin D. The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. PLoS Genet 2010; 6: e1001051.
[12] Frazer KA, Murray SS, Schork NJ and Topol EJ. Human genetic variation and its contribution to complex traits. Nat Rev Genet 2009; 10: 241-251.
[13] Wu X, Delclos GL, Annegers JF, Bondy ML, Honn SE, Henry B, Hsu TC and Spitz MR. A case-control study of wood dust exposure, mutagen sensitivity, and lung cancer risk. Cancer Epidemiol Biomarkers Prev 1995; 4: 583-588.
[14] Wei Q, Gu J, Cheng L, Bondy ML, Jiang H, Hong WK and Spitz MR. Benzo(a)pyrene diol epoxide-induced chromosomal aberrations and risk of lung cancer. Cancer Res 1996; 56: 3975-3979.
[15] Li D, Firozi PF, Wang LE, Bosken CH, Spitz MR, Hong WK and Wei Q. Sensitivity to DNA damage induced by benzo(a)pyrene diol epoxide and risk of lung cancer: a case-control analysis. Cancer Res 2001; 61: 1445-1450.
[16] Spitz MR, Wei Q, Dong Q, Amos CI and Wu X. Genetic susceptibility to lung cancer: the role of DNA damage and repair. Cancer Epidemiol Biomarkers Prev 2003; 12: 689-698.
[17] Lin J, Swan GE, Shields PG, Benowitz NL, Gu J, Amos CI, de Andrade M, Spitz MR and Wu X. Mutagen sensitivity and genetic variants in nucleotide excision repair pathway: genotype-phenotype correlation. Cancer Epidemiol Biomarkers Prev 2007; 16: 2065-2071.
[18] Rondelli CM, El-Zein RA, Wickliffe JK, Etzel CJ and bdel-Rahman SZ. A comprehensive haplotype analysis of the XPC genomic sequence reveals a cluster of genetic variants associated with sensitivity to tobacco-smoke mutagens. Toxicol Sci 2010; 115: 41-50.
[19] de Laat WL, Jaspers NG and Hoeijmakers JH. Molecular mechanism of nucleotide excision repair. Genes Dev 1999; 13: 768-785.
[20] de Boer J, Hoeijmakers JH. Nucleotide excision repair and human syndromes. Carcinogenesis 2000; 21: 453-460.
[21] Friedberg EC. How nucleotide excision repair protects against cancer. Nat Rev Cancer 2001; 1: 22-33.
[22] Vineis P, Manuguerra M, Kavvoura FK, Guarrera S, Allione A, Rosa F, Di GA, Polidoro S, Saletta F, Ioannidis JP and Matullo G. A field synopsis on low-penetrance variants in DNA repair genes and cancer susceptibility. J Natl Cancer Inst 2009; 101: 24-36.
[23] Kiyohara C, Takayama K and Nakanishi Y. Lung cancer risk and genetic polymorphisms in DNA repair pathways: a meta-analysis. J Nucleic Acids 2010; 2010: 701760.
[24] Zhang J, Gu SY, Zhang P, Jia Z and Chang JH. ERCC2 Lys751Gln polymorphism is associated with lung cancer among Caucasians. Eur J Cancer 2010; 46: 2479-2484.
[25] Qian B, Zhang H, Zhang L, Zhou X, Yu H and Chen K. Association of genetic polymorphisms in DNA repair pathway genes with non-small cell lung cancer risk. Lung Cancer 2011; 73: 138-146.
[26] Sakoda LC, Loomis MM, Doherty JA, Neuhouser ML, Barnett MJ, Thornquist MD, Weiss NS, Goodman GE and Chen C. Chromosome 15q24-25.1 variants, diet, and lung cancer susceptibility in cigarette smokers. Cancer Causes Control 2011; 22: 449-461.
[27] Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, Keogh JP, Meyskens FL, Valanis B, Williams JH, Barnhart S and Hammar S. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. N Engl J Med 1996; 334: 1150-1155.
[28] Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, Keogh JP, Meyskens FL Jr, Valanis B, Williams JH Jr, Barnhart S, Cherniack MG, Brodkin CA and Hammar S. Risk factors for lung cancer and for intervention effects in CARET, the Beta-Carotene and Retinol Efficacy Trial. J Natl Cancer Inst 1996; 88: 1550-1559.
[29] Goodman GE, Thornquist MD, Balmes J, Cullen MR, Meyskens FL Jr, Omenn GS, Valanis B and Williams JH Jr. The Beta-Carotene and Retinol Efficacy Trial: incidence of lung cancer and cardiovascular disease mortality during 6-year follow-up after stopping beta-carotene and retinol supplements. J Natl Cancer Inst 2004; 96: 1743-1750.
[30] Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L and Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 2004; 74: 106-120.
[31] Barrett JC, Fry B, Maller J and Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005; 21: 263-265.
[32] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ and Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559-575.
[33] Lindahl T, Barnes DE. Mammalian DNA ligases. Annu Rev Biochem 1992; 61: 251-281.
[34] Levin DS, McKenna AE, Motycka TA, Matsumoto Y and Tomkinson AE. Interaction between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base-excision repair. Curr Biol 2000; 10: 919-922.
[35] Aboussekhra A, Biggerstaff M, Shivji MK, Vilpo JA, Moncollin V, Podust VN, Protic M, Hubscher

U, Egly JM and Wood RD. Mammalian DNA nucleotide excision repair reconstituted with purified protein components. Cell 1995; 80: 859-868.
[36] Barnes DE, Tomkinson AE, Lehmann AR, Webster AD and Lindahl T. Mutations in the DNA ligase I gene of an individual with immunodeficiencies and cellular hypersensitivity to DNA-damaging agents. Cell 1992; 69: 495-503.
[37] Bentley DJ, Harrison C, Ketchen AM, Redhead NJ, Samuel K, Waterfall M, Ansell JD and Melton DW. DNA ligase I null mouse cells show normal DNA repair activity but altered DNA replication and reduced genome stability. J Cell Sci 2002; 115: 1551-1561.
[38] Shen H, Spitz MR, Qiao Y, Zheng Y, Hong WK and Wei Q. Polymorphism of DNA ligase I and risk of lung cancer-a case-control analysis. Lung Cancer 2002; 36: 243-247.
[39] Sobti RC, Kaur P, Kaur S, Janmeja AK, Jindal SK, Kishan J and Raimondi S. No association of DNA ligase-I polymorphism with the risk of lung cancer in north-Indian population. DNA Cell Biol 2006; 25: 484-489.
[40] Landi S, Gemignani F, Canzian F, Gaborieau V, Barale R, Landi D, Szeszenia-Dabrowska N, Zaridze D, Lissowska J, Rudnai P, Fabianova E, Mates D, Foretova L, Janout V, Bencko V, Gioia-Patricola L, Hall J, Boffetta P, Hung RJ and Brennan P. DNA repair and cell cycle control genes and the risk of young-onset lung cancer. Cancer Res 2006; 66: 11062-11069.
[41] Hung RJ, Baragatti M, Thomas D, McKay J, Szeszenia-Dabrowska N, Zaridze D, Lissowska J, Rudnai P, Fabianova E, Mates D, Foretova L, Janout V, Bencko V, Chabrier A, Moullan N, Canzian F, Hall J, Boffetta P and Brennan P. Inherited predisposition of lung cancer: a hierarchical modeling approach to DNA repair and cell cycle control pathways. Cancer Epidemiol Biomarkers Prev 2007; 16: 2736-2744.
[42] Michiels S, Danoy P, Dessen P, Bera A, Boulet T, Bouchardy C, Lathrop M, Sarasin A and Benhamou S. Polymorphism discovery in 62 DNA repair genes and haplotype associations with risks for lung and head and neck cancers. Carcinogenesis 2007; 28: 1731-1739.
[43] Chang JS, Wrensch MR, Hansen HM, Sison JD, Aldrich MC, Quesenberry CP, Jr., Seldin MF, Kelsey KT, Kittles RA, Silva G and Wiencke JK. Nucleotide excision repair genes and risk of lung cancer among San Francisco Bay Area Latinos and African Americans. Int J Cancer 2008; 123: 2095-2104.
[44] Lee YC, Morgenstern H, Greenland S, Tashkin DP, Papp J, Sinsheimer J, Cao W, Hashibe M, You NC, Mao JT, Cozen W, Mack TM and Zhang ZF. A case-control study of the association of the polymorphisms and haplotypes of DNA ligase I with lung and upper-aerodigestive-tract cancers. Int J Cancer 2008; 122: 1630-1638.
[45] Brenda D, Buch S, Nukui T, Day R, Siegfried J, Weissfeld J and Romkes M. Variation in base excision repair genes and risk of smoking-related non-small cell lung cancer [abstract LB-414]. In: Proceedings of the 101st Annual Meeting of the American Association for Cancer Research, AACR, Philadelphia, 2010.
[46] Hung RJ, Christiani DC, Risch A, Popanda O, Haugen A, Zienolddiny S, Benhamou S, Bouchardy C, Lan Q, Spitz MR, Wichmann HE, LeMarchand L, Vineis P, Matullo G, Kiyohara C, Zhang ZF, Pezeshki B, Harris C, Mechanic L, Seow A, Ng DP, Szeszenia-Dabrowska N, Zaridze D, Lissowska J, Rudnai P, Fabianova E, Mates D, Foretova L, Janout V, Bencko V, Caporaso N, Chen C, Duell EJ, Goodman G, Field JK, Houlston RS, Hong YC, Landi MT, Lazarus P, Muscat J, McLaughlin J, Schwartz AG, Shen H, Stucker I, Tajima K, Matsuo K, Thun M, Yang P, Wiencke J, Andrew AS, Monnier S, Boffetta P and Brennan P. International Lung Cancer Consortium: pooled analysis of sequence variants in DNA repair and cell cycle pathways. Cancer Epidemiol Biomarkers Prev 2008; 17: 3081-3089.
[47] Li L, Peterson CA, Lu X and Legerski RJ. Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair. Mol Cell Biol 1995; 15: 1993-1998.
[48] Ide F, Iida N, Nakatsuru Y, Oda H, Tanaka K and Ishikawa T. Mice deficient in the nucleotide excision repair gene XPA have elevated sensitivity to benzo[a]pyrene induction of lung tumors. Carcinogenesis 2000; 21: 1263-1265.
[49] Hou SM, Falt S, Angelini S, Yang K, Nyberg F, Lambert B and Hemminki K. The XPD variant alleles are associated with increased aromatic DNA adduct level and lung cancer risk. Carcinogenesis 2002; 23: 599-603.
[50] Zhou W, Liu G, Miller DP, Thurston SW, Xu LL, Wain JC, Lynch TJ, Su L and Christiani DC. Gene-environment interaction for the ERCC2 polymorphisms and cumulative cigarette smoking exposure in lung cancer. Cancer Res 2002; 62: 1377-1381.
[51] Yin J, Vogel U, Ma Y, Qi R and Wang H. HapMap-based study of the DNA repair gene ERCC2 and lung cancer susceptibility in a Chinese population. Carcinogenesis 2009; 30: 1181-1185.
[52] Tobacco, Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet 2010; 42: 441-447.
[53] Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, Sulem P, Rafnar T, Esko T, Walter S, Gieger C, Rawal R, Mangino M, Prokopenko I, Magi R, Keskitalo K, Gudjonsdottir IH, Gretarsdottir S, Stefansson H, Thompson JR, Aulchenko YS, Nelis M, Aben KK, den HM, Dirksen A, Ashraf H, Soranzo N, Valdes AM, Steves C, Uitterlinden AG, Hofman A, Tonjes A, Kovacs P, Hottenga JJ, Willemsen G, Vogelzangs N, Doring A, Dahmen N, Nitz B, Pergadia

ML, Saez B, De Diego V, Lezcano V, Garcia-Prats MD, Ripatti S, Perola M, Kettunen J, Hartikainen AL, Pouta A, Laitinen J, Isohanni M, Huei-Yi S, Allen M, Krestyaninova M, Hall AS, Jones GT, van Rij AM, Mueller T, Dieplinger B, Haltmayer M, Jonsson S, Matthiasson SE, Oskarsson H, Tyrfingsson T, Kiemeney LA, Mayordomo JI, Lindholt JS, Pedersen JH, Franklin WA, Wolf H, Montgomery GW, Heath AC, Martin NG, Madden PA, Giegling I, Rujescu D, Jarvelin MR, Salomaa V, Stumvoll M, Spector TD, Wichmann HE, Metspalu A, Samani NJ, Penninx BW, Oostra BA, Boomsma DI, Tiemeier H, van Duijn CM, Kaprio J, Gulcher JR, McCarthy MI, Peltonen L, Thorsteinsdottir U and Stefansson K. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nat Genet 2010; 42: 448-453.