Inherited Predisposition of Lung Cancer: A Hierarchical Modeling Approach to DNA Repair and Cell Cycle Control Pathways

Rayjean J. Hung,[1,2] Meili Baragatti,[1,3] Duncan Thomas,[4] James McKay,[1] Neonila Szeszenia-Dabrowska,[5] David Zaridze,[6] Jolanta Lissowska,[7] Peter Rudnai,[8] Eleonora Fabianova,[9] Dana Mates,[10] Lenka Foretova,[11] Vladimir Janout,[12] Vladimir Bencko,[13] Amelie Chabrier,[1] Norman Moullan,[1] Federico Canzian,[14] Janet Hall,[15] Paolo Boffetta,[1] and Paul Brennan[1]

[1]IARC, Lyon, France; [2]University of California at Berkeley, Berkeley, California; [3]Ecole Nationale de la Statistique et de l'Analyse de l'Information, Bruz, France; [4]Department of Preventive Medicine, University of Southern California, Los Angeles, California; [5]Department of Epidemiology, Institute of Occupational Medicine, Lodz, Poland; [6]Institute of Carcinogenesis, Cancer Research Centre, Moscow, Russia; [7]Department of Cancer Epidemiology and Prevention, Cancer Center and Maria Sklodowska-Curie Institute of Oncology, Warsaw, Poland; [8]National Institute of Environmental Health, Fodor József National Center for Public Health, Budapest, Hungary; [9]Specialized Institute of Hygiene and Epidemiology, Banska Bystrica, Slovakia; [10]Institute of Public Health, Bucharest, Romania; [11]Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic; [12]Department of Preventive Medicine, Faculty of Medicine, Palacky University, Olomouc, Czech Republic; [13]Charles University of Prague, First Faculty of Medicine, Institute of Hygiene and Epidemiology, Prague, Czech Republic; [14]German Cancer Research Center-Deutsches Krebsforschungszentrum, Heidelberg, Germany; and [15]Institut National de la Sante et de la Recherche Medicale U612, Institute Curie-Recherche, Orsay, France

Abstract

The DNA repair systems maintain the integrity of the human genome and cell cycle checkpoints are a critical component of the cellular response to DNA damage. We hypothesized that genetic variants in DNA repair and cell cycle control pathways will influence the predisposition to lung cancer, and studied 27 variants in 17 DNA repair enzymes and 10 variants in eight cell cycle control genes in 1,604 lung cancer patients and 2,053 controls. To improve the estimation of risks for specific variants, we applied a Bayesian approach in which we allowed the prior knowledge regarding the evolutionary biology and physicochemical properties of the variant to be incorporated into the hierarchical model. Based on the estimation from the hierarchical modeling, subjects who carried *OGG1* 326C/326C homozygotes, *MGMT* 143V or 178R, and *CHEK2* 157I had an odds ratio of lung cancer equal to 1.45 [95% confidence interval (95% CI), 1.05-2.00], 1.18 (95% CI, 1.01-1.40), and 1.58 (95% CI, 1.14-2.17). The association of *CHEK2* 157I seems to be overestimated in the conventional analysis. Nevertheless, this association seems to be robust in the hierarchical modeling. None of the pathways seem to have a prominent effect. In general, our study supports the notion that sequence variation may explain at least some of the variation of inherited susceptibility. In particular, further investigation of *OGG1, MGMT*, and *CHEK2* focusing on the genetic regions where the present markers are located or the haplotype blocks tightly linked with these markers might be warranted. (Cancer Epidemiol Biomarkers Prev 2007;16(12):2736–44)

Background

The familial relative risk of lung cancer was reported to be ~2-fold from several registry-based studies (1-3). Although shared environmental factors may explain part of the elevated familial relative risk, a meta-analysis reported a 1.5-fold elevated risk of lung cancer among never-smoking probands with affected first-degree relatives (4), which provides evidence for a genetic component in lung carcinogenesis. Despite a recent linkage analysis of 52 high-risk pedigrees that localized a lung cancer susceptibility locus at chromosome 6q23-25 (5), the exact inheritance mechanism of lung cancer is still largely undefined. The candidate gene approach has guided research on inherited susceptibility to lung cancer in the past decade. However, most of the associations in such studies have not been replicated. The likely reasons for this include lack of study power and false positives due to multiple comparisons, which may be exacerbated by publication bias. We therefore conducted a large-scale genetic association study on lung cancer to ensure adequate statistical power and conducted analyses using hierarchical modeling to include the prior knowledge of sequence variation into the modeling. The present investigation is focused on the sequence variations in DNA repair and cell cycle control pathways, two critical defense mechanisms against human carcinogenesis.

Received 6/5/07; revised 9/12/07; accepted 9/27/07.

Grant support: National Cancer Institute R01 grant (contract no. CA092039-04A1), R03 grant (contract no. CA 119704-01), and an Association for International Cancer Research grant (contract 03-281).

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.

Requests for reprints: Rayjean J. Hung, IARC, 150 cours Albert Thomas, F-69372 Lyon Cedex 08, France. Phone: 33-4-7273-8023; Fax: 33-4-7273-8342. E-mail: hung@iarc.fr

Copyright © 2007 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-07-0494

Downloaded from cebp.aacrjournals.org on November 29, 2021. © 2007 American Association for Cancer Research.

Table 1. Prior matrix for hierarchal modeling

Gene variant	dbNSP rs no.	Risk genotype/ haplotype	SIFT covariate	BER	NER	MMR/DR	DSB	Cell cycle
APEX Q51H, D148E	rs1048945, rs3136820	51H, 148E	0.585	1	0	0	0	0
LIG3 Ex21-250C>T	rs1052536	250T	0	1	0	0	0	0
OGG1 S326C	rs1052133	326C/326C	0.78	1	0	0	0	0
PARP V762A	rs1136410	762A	0.02	1	0	0	0	0
XRCC1 R194W, R280H, R399Q	rs1799782, rs25489, rs25487	194R, 280H, 399Q	0.563	1	0	0	0	0
LIG1 7C>T, ivs 9-21A>G	rs20579, rs3730931	7T- int9G	0	1	1	0	0	0
ERCC1 8092C>A, 354T>C	rs3212986, rs11615	8092A, 354C	0.5	0	1	0	0	0
XPA 23G>A	rs1800975	23A	0	0	1	0	0	0
XPC K939Q	rs2228001	939Q	0	0	1	0	0	0
XPD D312N	rs1799793	312N	0.42	0	1	0	0	0
XPD K751Q	rs13181	751Q	0.16	0	1	0	0	0
XPF P379S, R415Q	rs1799802, rs1800067	379S, 415Q	0.855	0	1	0	0	0
XPG D1104H	rs17655	1104H	0	0	1	0	0	0
MGMT 171C>T, L84F	rs1803965, rs12917	171T, 84F	0.145	0	0	1	0	0
MGMT I143V, K178R	rs2308321, rs2308327	143V, 178R	0.39	0	0	1	0	0
MLH1 I219V	rs1799977	219V	0.53	0	0	1	0	0
XRCC2 R188H	rs3218536	188H	0.8	0	0	0	1	0
XRCC3 T241M	rs861539	241T	0.94	0	0	0	1	0
XRCC4 S298N	rs1805377	298N	0.79	0	0	0	1	0
CHEK2 I157T	rs17879961	157I	0.96	0	0	0	1	1
CCND1 G870A	rs603965	870A	0	0	0	0	0	1
MDM2 162A>G	rs769412	162G	0	0	0	0	0	1
MDM2 309T>G	rs2279744	309G	0	0	0	0	0	1
MMP1 insG	rs1799750	insG	0	0	0	0	0	1
P16/CDKN2A A148T	rs3731249	148T	0.58	0	0	0	0	1
P21/CDKN1A S31R	rs1801270	31R	0.2	0	0	0	0	1
TP53 R72P, intron3 16bp repeats	rs1042522 (no rs no. for intron 3)	2 copies of Int3A2 -72P	0.355	0	0	0	0	1
P73 14C>T	rs1801173	14T	0	0	0	0	0	1

Abbreviations: BER, base-excision repair; NER, nucleotide excision repair; MMR, mismatch repair; DR, direct reversal pathway; DSB, double strand break.

The DNA repair system maintains the integrity of the human genome by removing DNA damage, reducing the mutation frequency of cancer-related genes, minimizing replication errors, and minimizing deleterious rearrangements arising via aberrant recombination (6). DNA repair processes are classified into four main pathways responsible for repairing different classes of DNA damage. First, base excision repair occurs through the excision of a modified base, followed by fill-in repair synthesis using the opposite strand as a template. Second, nucleotide excision repair removes photoproducts formed by UV radiation and other bulky adducts formed by a variety of chemicals. Third, DNA double-strand breaks are repaired either through the nonhomologous end-joining pathway or by homologous recombination. Finally, mismatch repair operates on base mismatches and small loop-outs that arise during replication by misincorporation or slippage on the template strand (7) and direct reversal of methylation is operated by enzymes such as O_6-methylguanine-DNA methyltransferase (MGMT). Cells with damaged DNA must either pause in the cell cycle to allow for repair or succumb to elimination by apoptosis and the activation of cell cycle checkpoints is a critical component of the cellular response to DNA damage (8).

We hypothesized that genetic variants in DNA repair and cell cycle control pathways will influence the predisposition to lung cancer, and studied 27 variants in 17 DNA repair enzymes and 10 variants in eight cell cycle control genes in 1,604 lung cancer patients and 2,053 controls. The majority of variants were selected from published association studies or chosen from the SNP500 project[1] on the basis of their location (such as promoter or coding regions) and minor allele frequencies (>5%). Our previous exploratory analysis among lung cancer patients with young onset also served as the basis of hypothesis generation for the current investigation (9).

To improve the estimation of risks for specific variants, we applied a Bayesian approach in which we allowed the prior knowledge to be incorporated in a hierarchical model. To develop a coherent scale of prior probability of whether the variant is likely to be deleterious or neutral, we took an evolutionary perspective, which takes account of evolutionary biology and physicochemical properties of the variant. The underlying idea is that the ability of a missense substitution to alter the function of the protein is related to two factors: (*a*) the degree of sequence conservation observed at the position of the missense substitution in the candidate genes and (*b*) the chemical difference between the missense substitution and the canonical residue at its corresponding position in the protein (10, 11). Such an approach has been implemented in the classification algorithm SIFT (Sorting Intolerant from Tolerant),[2] so we adopted this homology-based tool into our risk assessment of lung cancer predisposition.

[1] http://snp500cancer.nci.nih.gov
[2] http://blocks.fhcrc.org/sift/SIFT.html

Downloaded from cebp.aacrjournals.org on November 29, 2021. © 2007 American Association for Cancer Research.

Materials and Methods

Study Population. The study was conducted in 15 centers in six countries of Central and Eastern Europe, including Czech Republic (Prague, Olomouc, Brno), Hungary (Borsod, Heves, Szabolcs, Szolnok, Budapest), Poland (Warsaw, Lodz), Romania (Bucharest), Russia (Moscow), and Slovakia (Banska Bystrica, Bratislava, Nitra). The study details have been previously described (12). Briefly, each center followed an identical protocol and recruited a consecutive group of newly diagnosed cases of lung cancer and a comparable group of controls between 1998 and 2002. Eligible subjects were resident in the study area for at least 1 year. All lung cancer diagnoses had a histologic or cytologic confirmation. Consent for participation was required from both the patient and the patient's physician. This study was approved by the institutions at all study centers, and ethical approval was obtained from the IARC (Lyon, France), the coordinating center.

Controls in all centers except Warsaw were chosen among subjects admitted as in-patients or out-patients in the same hospital as the cases, and were frequency matched with the case group by sex, age (±3 years), center, and referral (or residence) area. The eligible diseases for controls were non–tobacco-related diseases, including minor surgical conditions, benign disorders, common infections, eye conditions (except cataract or diabetic retinopathy), common orthopedic diseases (except osteoporosis), etc. In Warsaw, population controls were selected by random sampling from the Polish Electronic List of Residents.

Both cases and controls underwent an identical face-to-face interview during which they completed a detailed questionnaire including sections on (*a*) demographic variables, (*b*) medical history, (*c*) family history of cancer, (*d*) tobacco smoking and involuntary smoking, (*e*) alcohol drinking, (*f*) consumption of selected food items, and (*g*) occupational history.

Laboratory Techniques. Blood samples were collected at the time of interview. In total, 2,250 cases and 2,289 controls have provided blood samples. Genomic DNA was extracted from blood samples by automated Gentra Systems; DNA concentrations were measured by using PicoGreen dsDNA Quantification kits (Molecular Probes). Genomic DNA was extracted from blood samples with the use of a QIAamp 96 DNA Blood Kit (Qiagen), or with Puregene chemistry (Gentra Systems) on an Autopure instrument (Gentra Systems). Samples that yielded an insufficient amount of DNA at extraction were subjected to whole-genome amplification by use of a phi29-based protocol (GenomiPhi, Amersham Biosciences) or reextracted using Gentra technology. DNA concentrations were measured by using PicoGreen dsDNA quantification kits (Molecular Probes).

A total of 37 single nucleotide polymorphisms (SNP) were genotyped, including *APEX* Q51H, D148E, *LIG3* Ex21-250C>T, *OGG1* S326C, *PARP* V762A, *XRCC1*, R194W, R280H, R399Q, *LIG1* 7C>T, ivs 9-21A>G, *ERCC1* 8092C>A, 354T>C, *XPA* 23G>A, *XPC* K939Q, *XPD* D312N, K751Q, *XPF* P379S, R415Q, *XPG* D1104H, *MGMT* 171C>T, L84F, I143V, K178R, *MLH1* I219V, *XRCC2* R188H, *XRCC3* T241M, *XRCC4* S298N, *CHEK2* I157T, *CCND1* G870A, *MDM2* 162A>G, *MDM2* 309T>G, *MMP1* insG, p16/*CDKN2A* A148T, *p21/CDKN1A* S31R, *TP53* R72P, intron 3 16-bp repeats and *P73* 14C>T (rs numbers are listed in Table 1). Thirty-three of them were analyzed by the 5 exonuclease assay (i.e., TaqMan assay), one using the allele-specific PCR-based Ampli-fluor assay (*XRCC1* R399Q rs25487), and two using the MGB Eclipse Probe System (*P73* C14T, rs1801173: *MLH1* I219V, rs179977). In addition, the presence of the *TP53* intron 3 16-bp duplication was assessed using a gel-based discrimination assay. Designs of genotyping assays for the SNPs were either taken from the Web site of the SNP500 project[1] or were designed by Applied Biosystems, Proligo, or in-house using the Genbank databases and Primer Express (Applied Biosystems). PCR primers and TaqMan probes were synthesized by Applied Biosystems or Proligo. Eclipse probes were obtained from Nanogen and Ampliflour probes were from Serologicals Corporation. Briefly, 10 ng aliquots of DNA were placed into separate wells of a 96-well or 384-well PCR plate along with a PCR cocktail that included fluorescently labeled allele-specific probes. The fluorescence of the PCR products was then plotted and genotype was determined according to the signal of the two probes. Details of primers, probes, and protocols for all genotyping are available upon request. Ten percent of the subjects were randomly selected and regenotyped for each polymorphism to examine the reliability of the genotyping. In total, 1,604 cases (71.3% of all cases) and 2,053 (70.8% of all controls) had complete genetic data for all makers in DNA repair and cell cycle control pathways, and were included in the final hierarchical regression. Primary data of 9 of these 37 variants included in the hierarchical regression have been previously reported, including variants of *CHEK2, XRCC1, OGG1, TP53, CCND1*, and *p16/CDKN2A* (12-14).

Statistical Analysis. The frequency distribution of demographic variables and putative risk factors of lung cancer, including country of residence, age, sex, education, and smoking was examined for cases and controls. Tobacco smoking included cigarettes, pipes, and/or cigars. Former smokers were defined as smokers who quit smoking at least 2 years before interview or diagnosis. Cumulative tobacco consumption was calculated as the product of smoking duration and intensity, and expressed as pack-years. Hardy-Weinberg equilibrium of allele distributions was tested in cases and in controls separately.

Genotypes were dichotomized into two categories based on the prior knowledge or frequency distribution of the SNPs as risk versus nonrisk genotypes. Minor alleles were assumed to confer an increased risk unless there was prior evidence indicating otherwise. For markers in linkage disequilibrium, we conducted haplotype analyses based on the expectation-maximization algorithms using the TagSNPs software (15). When there was no evidence of a specific haplotype conferring the primary association with lung cancer risk, we considered subjects who carried any risk genotype for the markers in linkage disequilibrium as the subjects at risk. On the other hand, when there was evidence that a specific haplotype conferred the primary association with lung cancer risk, we used the haplotype dosage as the input variable in the first stage. For example, two variants in

Downloaded from cebp.aacrjournals.org on November 29, 2021. © 2007 American Association for Cancer Research.

ligase 1 (7C>T and inv9-21A>G) are in linkage disequilibrium ($D' = 0.83$), and haplotype analysis showed that 7T-inv9G was the primary haplotype associated with lung cancer risk [odds ratio (OR), 1.19; 95% confidence interval (95% CI), 1.02-1.38]. In this case, the haplotype dosage of 7T-inv9G was used as the input value for ligase 1 risk genotype. Similar results were observed for P53 R72P and intron 3 16-bp repeats.

Hierarchical Modeling. The hierarchical model we used has been described previously (16). Briefly, we applied a two-stage model in which we used the conventional logistic regression as the first stage, and added a second stage with the goal of improving the estimation of the coefficients. The prior matrix in the second-stage model was built on our prior knowledge of the biological function of the gene and the possible functional significance of the variants. The genetic variants were broadly classified into five pathways related to lung carcinogenesis, based on the biological functions of the genes: (*a*) base excision repair, (*b*) nucleotide excision repair, (*c*) mismatch repair and direct reversal, (*d*) double strand break, and (*e*) cell cycle control genes. For each pathway, a score was assigned to each genetic marker according to the biological function of the gene and the functional significance of its polymorphism. A score of 0 was assigned to the genetic polymorphism if the gene is not involved in such a pathway (the coefficient of the factor was expected to be zero for such a pathway and the corresponding gene effect was not shrunk toward the other genes in that pathway either). Genes assigned to the same pathway were assumed to be exchangeable; that is, for their effects to arise from a common distribution. The five pathways were not mutually exclusive.

In addition to its role in different pathways, whether a nucleotide change is likely to be deleterious would also influence the prior probability of how likely that the variant would have an effect on cancer risk. One method to predict whether a nucleotide change is likely to be functionally important is to measure how conserved is the region where the nucleotide change occurs: a high conservation level of a genetic region across species indicates that the region is likely to be fundamentally important. Thus, amino acid changes that occur in a more conserved genetic region would have a higher likelihood to be functionally important. We have subsequently expanded the hierarchical modeling by incorporating information on sequence conservation of the variants into the prior model (second stage). Sequence conservation of the variants were obtained by SIFT,[2] which searches for the sequences of the same gene across species and returns the probability of such an amino acid appearing at the specific position (denoted as P_{ca}). This probability is used as an indication of how tolerant an amino acid substitution is at the given position (11, 17). For each gene variant, the importance of its functional changes is weighted by the probability that these amino acid substitutions are intolerant or deleterious (which is $1 - P_{ca}$, denoted as D_{ca}). This estimation can only be obtained for variants that lead to an amino acid substitution; other variants were assigned a D_{ca} of 0.

The prior matrix used in the second-stage model is reported in Table 1. In summary, the biological implication of this prior matrix is that *APEX1, LIG3, OGG1, PARP, XRCC1*, and *LIG1* play a role in base excision repair pathway, and, *LIG1, ERCC1, XPA, XPC, XPD, XPF*, and *XPG* are involved in the nucleotide excision repair pathway, whereas *MGMT* and *MLH1* encode enzymes that are part of the direct reversal or mismatch repair pathway, and *XRCC2, XRCC3, XRCC4*, and *CHEK2* are part of the double strand break pathway or downstream from it. Their designated risk genotypes or haplotypes decrease the DNA repair capacity through various mechanisms. Furthermore, *CHEK2, CCND1, MDM2, MMP1, CDKN2A, CDKN1A, TP53*, and *P73* are crucial in the cell cycle control pathway, and their designated risk genotype or haplotype are hypothesized to lead to aberrant cell cycle control. As described above, D_{ca} served as an additional covariate in the prior matrix for nucleotide changes that lead to amino acid changes. For markers with multiple genotype or haplotypes, we used the average of D_{ca} as the covariates.

We applied the hierarchical modeling with the empirical-Bayes approach (allowing τ^2, the variance of the residual effect to be estimated from the data). When the estimate of τ^2 was set to 0, which may occur due to model overparameterization, we subsequently took the semi-Bayes approach and specified τ^2 as 0.05 (which would allow ~2.4-fold of variation in the effect of the gene variants).

We conducted stratified analyses by histology to investigate the effect of each genetic polymorphism on each major histologic type. We estimated the effect of sequence variants separately for subjects with no family history of any cancer, with family history of lung or

Table 2. Frequency distribution of demographic variables and putative factors

	Case, *n* (%)	Control, *n* (%)
Total	1,604	2,053
Country		
Romania	116 (7)	123 (6)
Hungary	268 (17)	195 (10)
Poland	523 (33)	593 (29)
Russia	293 (18)	610 (30)
Slovakia	217 (14)	125 (6)
Czech Republic	187 (12)	407 (20)
Sex		
Male	1,244 (78)	1,483 (72)
Female	360 (22)	570 (28)
Age		
≤40	20 (1)	65 (3)
41-50	250 (16)	342 (17)
51-60	508 (32)	658 (32)
61-70	605 (38)	715 (35)
71+	221 (14)	273 (13)
Education		
High	224 (14)	456 (22)
Medium	1,098 (69)	1,370 (67)
Low	278 (17)	223 (11)
Smoking		
Never	123 (8)	712 (35)
Former	310 (19)	544 (27)
Current	1,167 (73)	789 (39)
Histology		
Squamous cell carcinoma	384 (24)	
Small-cell carcinoma	667 (42)	
Adenocarcinomas	236 (15)	
Others/mixed	317 (20)	

Downloaded from cebp.aacrjournals.org on November 29, 2021. © 2007 American Association for Cancer Research.

Table 3. Main effect of sequence variants on lung cancer risk estimated by hierarchical models

	Case, *n*	Control, *n*	Conventional 1,* OR (95% CI)	Conventional 2,† OR (95% CI)	Pathway HM (EB),‡ OR (95% CI)	SIFT HM (EB),§ OR (95% CI)	SIFT HM (SB),‖ OR (95% CI)
APEX 148E, 51H	1,144	1,518	0.85 (0.75-0.98)	0.90 (0.76-1.06)	0.95 (0.81-1.10)	0.95 (0.81-1.11)	0.93 (0.79-1.08)
LIG3 250T	1,232	1,523	1.10 (0.96-1.27)	1.13 (0.95-1.34)	1.13 (0.97-1.32)	1.13 (0.97-1.32)	1.13 (0.96-1.33)
OGG1 326C/326C	84	63	**1.57 (1.14-2.17)**	**2.03 (1.39-2.95)**	**1.40 (1.03-1.90)**	**1.45 (1.05-2.00)**	**1.59 (1.17-2.18)**
PARP 762A	500	668	0.98 (0.86-1.12)	1.03 (0.88-1.21)	1.05 (0.91-1.21)	1.04 (0.90-1.21)	1.04 (0.89-1.21)
XRCC1 194R, 280H, 399Q	1,591	2,042	0.60 (0.28-1.25)	0.49 (0.20-1.19)	0.97 (0.64-1.45)	0.98 (0.63-1.51)	0.97 (0.62-1.50)
LIG1 7T- int9G	155	185	**1.19 (1.02-1.38)**	1.19 (0.99-1.42)	1.17 (0.99-1.37)	1.16 (0.99-1.37)	1.17 (0.99-1.39)
ERCC1 8092A, 354C	10	9	**1.90 (1.02-3.56)**	1.34 (0.63-2.86)	1.06 (0.75-1.51)	1.09 (0.75-1.57)	1.11 (0.74-1.68)
XPA 23A	959	1,215	1.00 (0.89-1.14)	1.03 (0.88-1.19)	1.02 (0.89-1.17)	1.02 (0.89-1.17)	1.03 (0.89-1.18)
XPC 939Q	1,039	1,313	1.04 (0.91-1.18)	1.02 (0.88-1.19)	1.02 (0.89-1.17)	1.02 (0.88-1.17)	1.02 (0.88-1.18)
XPD 312N	1,022	1,315	1.01 (0.89-1.15)	0.99 (0.81-1.22)	0.99 (0.84-1.16)	0.99 (0.84-1.17)	0.99 (0.83-1.19)
XPD 751Q	1,026	1,332	0.98 (0.86-1.11)	0.96 (0.78-1.18)	0.97 (0.83-1.14)	0.97 (0.82-1.14)	0.97 (0.81-1.16)
XPF 379S, 415Q	211	285	0.97 (0.81-1.16)	0.95 (0.77-1.18)	0.97 (0.81-1.16)	0.98 (0.81-1.19)	0.97 (0.80-1.19)
XPG 1104H	1,521	1,950	1.05 (0.79-1.40)	0.95 (0.68-1.33)	0.98 (0.78-1.24)	0.98 (0.77-1.24)	0.97 (0.74-1.28)
MGMT 171T, 84F	461	614	0.98 (0.86-1.12)	0.95 (0.81-1.12)	0.97 (0.84-1.13)	0.97 (0.83-1.12)	0.96 (0.82-1.12)
MGMT 143V, 178R	373	412	1.09 (0.94-1.26)	**1.23 (1.03-1.47)**	**1.18 (1.00-1.39)**	**1.18 (1.01-1.40)**	**1.20 (1.02-1.43)**
MLH1 219V	827	1,078	0.99 (0.88-1.12)	1.02 (0.88-1.18)	1.02 (0.89-1.17)	1.02 (0.89-1.18)	1.02 (0.89-1.18)
XRCC2 188H	179	215	1.02 (0.84-1.23)	1.08 (0.85-1.36)	1.12 (0.91-1.36)	1.11 (0.91-1.36)	1.10 (0.89-1.37)
XRCC3 241T	1,429	1,797	1.11 (0.91-1.35)	1.18 (0.94-1.49)	1.18 (0.97-1.44)	1.19 (0.97-1.45)	1.19 (0.96-1.47)
XRCC4 298N	347	451	1.04 (0.90-1.21)	1.02 (0.85-1.22)	1.06 (0.90-1.25)	1.06 (0.90-1.25)	1.05 (0.88-1.24)
CHEK2 157I	1,562	1,945	**2.26 (1.57-3.25)**	**2.15 (1.44-3.23)**	**1.55 (1.13-2.13)**	**1.58 (1.14-2.17)**	**1.72 (1.24-2.39)**
CCND1 870A	1,161	1,458	0.98 (0.86-1.13)	1.02 (0.87-1.20)	1.04 (0.89-1.20)	1.03 (0.89-1.20)	1.03 (0.88-1.20)
MDM2 162G	187	241	1.07 (0.88-1.29)	0.96 (0.76-1.21)	1.01 (0.84-1.22)	1.01 (0.83-1.22)	0.99 (0.80-1.22)
MDM2 309G	955	1,205	0.96 (0.85-1.09)	1.01 (0.87-1.18)	1.03 (0.90-1.18)	1.02 (0.89-1.17)	1.02 (0.88-1.18)
MMP1 insG	1,160	1,458	1.02 (0.89-1.17)	1.09 (0.92-1.28)	1.09 (0.94-1.26)	1.09 (0.94-1.26)	1.09 (0.93-1.27)
CDKN2A 148T	95	114	1.03 (0.79-1.35)	0.98 (0.71-1.35)	1.04 (0.83-1.31)	1.06 (0.83-1.35)	1.04 (0.80-1.37)
CDKN1A 31R	243	297	1.16 (0.97-1.39)	1.10 (0.89-1.35)	1.10 (0.92-1.31)	1.10 (0.92-1.32)	1.11 (0.91-1.34)
TP53 Int3A2 -72P	39	24	**1.83 (1.14-2.92)**	**1.96 (1.10-3.51)**	1.26 (0.90-1.77)	1.29 (0.91-1.83)	1.38 (0.95-2.00)
P73 14T	428	516	1.07 (0.93-1.23)	1.08 (0.92-1.28)	1.09 (0.94-1.27)	1.09 (0.94-1.27)	1.09 (0.93-1.28)
Estimated τ^2					0.0188	0.0208	Set to 0.05
Pathway estimation							
BER					1.10 (0.95-1.28)	1.09 (0.93-1.29)	1.09 (0.86-1.37)
NER					1.01 (0.89-1.14)	1.00 (0.87-1.14)	0.99 (0.82-1.20)
MMR/DR					1.05 (0.88-1.26)	1.03 (0.83-1.28)	1.02 (0.76-1.38)
DSB					1.18 (0.98-1.42)	1.12 (0.83-1.52)	1.12 (0.74-1.68)
Cell cycle					1.10 (0.97-1.24)	1.09 (0.96-1.24)	1.11 (0.93-1.32)
SIFT						1.06 (0.80-1.40)	1.09 (0.75-1.60)

*Logistic regression with single marker in each regression.
† Logistic regression with all markers in one single model.
‡ Hierarchical modeling with empirical Bayes approach and pathway indicators only.
§Hierarchical modeling with the empirical Bayes approach, pathway indicators and a SIFT covariate.
‖Hierarchal modeling with semi-Bayes approach, pathway indicators and a SIFT covariate.

upper aerodigestive tract cancers, or with family history of lung cancer. The rationale for analyzing cases with family history separately is that one would expect genetic factors to have a stronger effect on cancer development among such subjects. We also evaluated the modulating effects of tobacco consumption, by stratifying and comparing the stratum-specific risk estimates.

Matching variables and potential confounders, such as country of residence, age (continuous), sex, and smoking pack-year (continuous), were included in the multivariate logistic regression at the first-stage model. Hierarchical modeling analyses were conducted with SAS IML code in conjunction with GLIMMIX macro (18).

Results

The frequency distribution of the demographic factors and the putative risk factors among subjects with complete genetic data are shown in Table 2. Comparing cases and controls on demographic variables, controls were slightly more educated than cases. As expected, smoking prevalence was higher among cases. The percentages of subjects from each country ranged from 6% to 33%, but the distributions were slightly different between cases and controls ($P < 0.01$) due to differences in DNA availability from different countries. None of the SNP allele frequencies in the control group deviated from the Hardy-Weinberg equilibrium at $P < 0.01$.

Table 3 shows the main effect of each genetic variant, as well as the estimated effects of each pathway. For the sake of comparison, we presented five estimates for the main effects from (*a*) a conventional analysis with single marker in each logistic regression, (*b*) conventional analysis with all markers in one single logistic regression, (*c*) hierarchical modeling with empirical Bayes approach and pathway indicators only, (*d*) hierarchical modeling with the empirical Bayes approach and a SIFT covariate D_{ca}, and (*e*) hierarchal modeling with semi-Bayes approach and a SIFT covariate D_{ca}. The single marker analysis suggested that five genes were associated with lung cancer risk (*OGG1, LIG1, ERCC1, CHEK2*, and *TP53*), among which two risk estimates were no longer

Downloaded from cebp.aacrjournals.org on November 29, 2021. © 2007 American Association for Cancer Research.

Table 4. Effect of risk variants on lung cancer risk by histology using hierarchal modeling*

	Squamous cell carcinoma		Adenocarcinoma		Small-cell carcinoma	
	Cases	OR (95% CI)	Cases	OR (95% CI)	Cases	OR (95% CI)
APEX 148E, 51H	483	0.97 (0.79-1.19)	275	1.05 (0.82-1.33)	168	1.03 (0.77-1.39)
LIG3 250T	523	1.17 (0.94-1.45)	274	1.01 (0.80-1.29)	183	1.14 (0.86-1.51)
OGG1 326C/326C	27	1.20 (0.80-1.79)	24	**1.54 (1.03-2.30)**	13	1.41 (0.87-2.28)
PARP 762A	206	1.01 (0.83-1.23)	119	1.04 (0.83-1.31)	72	0.99 (0.76-1.29)
XRCC1 194R, 280H, 399Q	663	0.94 (0.52-1.68)	381	1.11 (0.68-1.82)	234	1.00 (0.52-1.92)
LIG1 7T- int9G	63	1.15 (0.92-1.44)	38	1.08 (0.84-1.41)	23	1.07 (0.80-1.45)
ERCC1 8092A, 354C	7	1.25 (0.77-2.03)	1	1.04 (0.65-1.68)	1	0.89 (0.26-3.06)
XPA 23A	387	1.00 (0.83-1.20)	244	1.15 (0.92-1.43)	140	0.99 (0.78-1.25)
XPC 939Q	438	1.08 (0.89-1.30)	244	0.92 (0.74-1.15)	156	0.99 (0.78-1.27)
XPD 312N	428	0.97 (0.78-1.22)	248	1.01 (0.78-1.30)	160	1.13 (0.84-1.52)
XPD 751Q	430	0.98 (0.78-1.23)	249	1.01 (0.79-1.30)	154	0.95 (0.72-1.25)
XPF 379S, 415Q	81	0.93 (0.71-1.22)	59	1.08 (0.81-1.45)	27	1.02 (0.70-1.48)
XPG 1104H	635	1.10 (0.78-1.54)	364	0.94 (0.65-1.35)	219	0.86 (0.60-1.24)
MGMT 171T, 84F	186	0.93 (0.75-1.14)	119	1.09 (0.86-1.38)	63	0.93 (0.70-1.23)
MGMT 143V, 178R	151	1.17 (0.94-1.46)	98	1.25 (0.98-1.61)	60	1.19 (0.88-1.61)
MLH1 219V	358	1.11 (0.92-1.34)	188	0.96 (0.77-1.20)	115	0.97 (0.75-1.26)
XRCC2 188H	63	1.03 (0.77-1.38)	44	1.07 (0.78-1.46)	40	**1.61 (1.15-2.24)**
XRCC3 241T	595	1.21 (0.92-1.59)	336	1.02 (0.75-1.38)	218	**1.59 (1.10-2.28)**
XRCC4 298N	146	1.09 (0.87-1.36)	76	0.91 (0.70-1.19)	55	1.28 (0.94-1.75)
CHEK2 157I	657	**2.00 (1.16-3.47)**	367	1.06 (0.71-1.59)	229	1.56 (0.98-2.47)
CCND1 870A	487	1.10 (0.90-1.35)	280	1.05 (0.83-1.32)	171	1.03 (0.80-1.32)
MDM2 162G	77	0.96 (0.74-1.26)	43	0.93 (0.69-1.26)	27	0.96 (0.70-1.32)
MDM2 309G	395	1.02 (0.85-1.23)	234	1.01 (0.81-1.25)	127	0.88 (0.69-1.14)
MMP1 insG	483	1.09 (0.90-1.33)	276	1.02 (0.81-1.29)	165	0.97 (0.76-1.25)
CDKN2A 148T	32	0.99 (0.69-1.41)	23	1.03 (0.71-1.49)	20	1.26 (0.86-1.84)
CDKN1A 31R	106	1.16 (0.91-1.47)	51	0.99 (0.75-1.32)	38	1.11 (0.83-1.48)
TP53 Int3A2 -72P	22	1.48 (0.92-2.38)	8	1.16 (0.75-1.80)	4	1.19 (0.72-1.96)
P73 14T	182	1.17 (0.96-1.43)	103	0.96 (0.76-1.22)	69	1.10 (0.85-1.42)
Estimated τ^2		0.0416		Set to 0.05		0.0217
Pathway estimation						
BER		1.07 (0.85-1.34)		1.07 (0.83-1.39)		1.05 (0.84-1.33)
NER		1.04 (0.86-1.25)		0.98 (0.79-1.21)		0.96 (0.80-1.15)
MMR/DR		1.06 (0.79-1.43)		1.04 (0.74-1.45)		0.95 (0.71-1.28)
DSB		1.22 (0.79-1.88)		0.88 (0.55-1.42)		1.26 (0.81-1.97)
Cell cycle		1.15 (0.96-1.37)		1.00 (0.82-1.22)		1.02 (0.85-1.22)
SIFT		1.00 (0.67-1.48)		1.17 (0.76-1.81)		1.21 (0.80-1.82)

*Prior matrix included pathway indicators and a SIFT covariate.

significant when we included all markers at once. The conventional analysis with all genes in one model suggested that four genes (*OGG1, MGMT, CHEK2*, and *TP53*) might be associated with lung cancer risk, three of which were supported by the empirical-Bayes hierarchical modeling (*OGG1, MGMT*, and *CHEK2*). When we allowed the risk to vary to a larger extent using a prespecified τ^2 of 0.05 (i.e., 2.4-fold variation), *OGG1, MGMT*, and *CHEK2* remained associated with lung cancer risk. Based on the estimation from the hierarchical modeling, subjects who carried *OGG1* 326C/326C genotype, *MGMT* 143V or 178R, and *CHEK2* 157I allele had ORs of lung cancer of 1.45 (95% CI, 1.05-2.00), 1.18 (95% CI, 1.01-1.40), and 1.58 (95% CI, 1.14-2.17), respectively. The association of *CHEK2* 157I allele, which has been previously reported and independently replicated (14), seems to be robust in the hierarchical modeling. Although *TP53* A2-72P was shown to be associated with increased risk of lung cancer in the conventional analysis, the association was not supported by the hierarchal modeling. Based on the data we have for each pathway, we could estimate pathway effects, although none appeared to be prominent. Stratification by family history of cancer did not seem to be informative, which might be due to small numbers.

Table 4 shows the effect of genetic markers on different histologic subtypes. A semi-Bayes approach was adopted for adenocarcinoma, as the τ^2 was estimated to be zero in the empirical-Bayes approach. *CHEK2* 157I allele seemed to have a stronger effect on squamous cell carcinoma with OR of 2.00 (95% CI, 1.16-3.47). On the other hand, the *XRCC2* 188H allele and *XRCC3* 214T allele seemed to be more important for small cell carcinoma with an OR of 1.61 (95% CI, 1.15-2.24), and 1.59 (95% CI, 1.10-2.28), respectively. *OGG1* seemed to be positively associated with adenocarcinoma (OR, 1.54; 95% CI, 1.03-2.30).

In the analysis stratified by smoking status, we observed a positive association between lung cancer risk and *XRCC2* 188H and *CHEK2* 157I among current smokers, with ORs of 1.30 (95% CI, 1.03-1.65) and 1.50 (95% CI, 1.07-2.11), respectively (see Table 5).

Discussion

In this study, we applied hierarchical modeling in a Bayesian framework to incorporate evolutionary biology into the risk estimates. In general, our study supports the notion that sequence variation may explain at least some

Downloaded from cebp.aacrjournals.org on November 29, 2021. © 2007 American Association for Cancer Research.

of the variation of inherited susceptibility. In particular, sequence variants in *CHEK2*, *MGMT*, and *OGG1* were suggested to be positively associated with lung cancer risk overall.

MGMT is a DNA repair protein that acts on O^6-methylguanine, one of the most potent premutagenic lesions (19), by transferring the methyl group from the O^6-methylguanine moieties to a cysteine residue located at the active site of the MGMT protein. The expression level of *MGMT* in humans varies up to 40-fold (20), and studies have shown that genetic components might explain part of the interindividual variation (21). Several studies have been conducted to investigate the association between *MGMT* polymorphisms and lung cancer risk, but the results are inconsistent, mainly due to lack of statistical power, because the majority of the studies had under 400 case-control pairs (22-27). I143V and K178R polymorphisms are in strong linkage disequilibrium. K178R was suggested to be unlikely to lead to altered AGT activity, as human AGT can be truncated from the COOH-terminal to the position 176 without loss of activity (28); nevertheless, the I143V variant was suggested to be functionally important as it is located adjacent to the alkyl acceptor ^{145}Cys of the active site (29) and was therefore hypothesized to modulate its repair activity and increase cancer risk. The association we observed with lung cancer might reflect the functional consequence of I143V, although further confirmation through experimental studies would be required.

The primary analysis of *CHEK2* I157T and *OGG1* S326C polymorphisms was reported previously (12, 30), in which we identified *CHEK2* 157T to be a protective allele and *OGG1* 326C to be a risk allele. The risk estimate using hierarchical modeling confirmed the positive associations reported in the previous publication, for both the overall main effect and the modifying effect of histology subgroup. Specifically, the prominent associations of *CHEK2* I157T variant with squamous cell carcinoma, and of *OGG1* S326C with adenocarcinoma were both supported by the risk estimates from hierarchical modeling. However, we did observe a higher point estimate from the conventional model compared with the estimates derived from the hierarchical model. This can be due to either an overestimation of the conventional model or inaccurate assumption of the prior distribution in the hierarchical model. In addition, the precision of the estimates from the hierarchical modeling was substantially higher than in the conventional model: For example, the *OGG1* 326C variant conferred an OR of 2.24 (95% CI, 1.30-3.86) for

Table 5. Effect of sequence variants on lung cancer risk by smoking status using hierarchical modeling*

	Never smokers			Former smokers			Current smokers		
	Cases	Controls	OR (95% CI)	Cases	Controls	OR (95% CI)	Cases	Controls	OR (95% CI)
APEX 148E, 51H	87	538	0.94 (0.65-1.37)	221	414	0.91 (0.68-1.20)	834	564	1.06 (0.88-1.28)
LIG3 250T	96	525	1.16 (0.82-1.64)	233	392	1.10 (0.84-1.42)	900	602	1.09 (0.90-1.31)
OGG1 326C/326C	13	26	1.27 (0.68-2.36)	12	21	0.97 (0.64-1.47)	58	16	1.33 (0.90-1.96)
PARP 762A	39	234	1.05 (0.76-1.46)	104	177	1.06 (0.84-1.36)	355	254	1.00 (0.84-1.19)
XRCC1 194R, 280H, 399Q	120	706	0.86 (0.43-1.75)	307	542	0.90 (0.45-1.82)	1160	786	1.06 (0.63-1.80)
LIG1 7T- int9G	13	66	1.13 (0.77-1.68)	31	53	1.03 (0.77-1.36)	111	65	1.11 (0.91-1.36)
ERCC1 8092A, 354C	1	2	0.89 (0.13-6.06)	3	3	1.01 (0.62-1.67)	7	4	1.03 (0.69-1.54)
XPA 23A	77	410	1.15 (0.83-1.61)	175	341	0.90 (0.71-1.15)	705	460	1.04 (0.88-1.22)
XPC 939Q	74	458	0.99 (0.73-1.35)	217	356	1.05 (0.83-1.34)	745	493	1.01 (0.85-1.19)
XPD 312N	79	447	0.90 (0.58-1.40)	204	361	0.96 (0.75-1.22)	737	503	1.02 (0.84-1.22)
XPD 751Q	89	449	1.25 (0.78-2.01)	202	359	0.99 (0.78-1.25)	734	519	0.92 (0.77-1.12)
XPF 379S, 415Q	15	112	0.88 (0.55-1.41)	33	65	0.90 (0.63-1.27)	163	107	1.02 (0.81-1.29)
XPG 1104H	118	671	1.08 (0.68-1.72)	292	520	0.91 (0.63-1.30)	1107	751	0.97 (0.75-1.26)
MGMT 171T, 84F	35	202	1.13 (0.79-1.62)	87	168	0.96 (0.72-1.27)	337	242	0.96 (0.80-1.15)
MGMT 143V, 178R	33	145	1.27 (0.86-1.89)	77	103	1.13 (0.83-1.52)	262	163	1.06 (0.87-1.29)
MLH1 219V	63	383	1.05 (0.74-1.47)	168	290	1.01 (0.79-1.28)	593	403	1.03 (0.87-1.22)
XRCC2 188H	11	80	0.88 (0.55-1.42)	23	60	0.97 (0.66-1.42)	145	75	**1.30 (1.03-1.65)**
XRCC3 241T	105	609	0.96 (0.63-1.45)	281	484	1.12 (0.79-1.59)	1040	697	1.20 (0.95-1.52)
XRCC4 298N	26	162	0.94 (0.64-1.38)	66	113	1.08 (0.80-1.45)	255	176	1.11 (0.90-1.36)
CHEK2 157I	120	682	1.05 (0.55-2.01)	298	513	1.19 (0.77-1.85)	1140	743	**1.50 (1.07-2.11)**
CCND1 870A	87	513	0.99 (0.72-1.37)	216	361	1.08 (0.86-1.35)	854	579	0.97 (0.81-1.16)
MDM2 162G	15	79	1.03 (0.71-1.50)	50	64	1.14 (0.85-1.53)	121	97	0.93 (0.73-1.18)
MDM2 309G	66	414	0.95 (0.70-1.29)	182	315	1.04 (0.84-1.30)	703	470	1.05 (0.89-1.24)
MMP1 insG	87	510	0.99 (0.72-1.36)	226	393	1.01 (0.79-1.29)	844	552	1.10 (0.92-1.30)
CDKN2A 148T	7	36	0.97 (0.59-1.61)	18	31	0.97 (0.68-1.39)	70	47	1.02 (0.76-1.35)
CDKN1A 31R	19	110	1.01 (0.70-1.44)	36	90	0.90 (0.65-1.25)	188	96	1.22 (0.98-1.53)
TP53 Int3A2 -72P	2	9	1.01 (0.53-1.91)	7	6	1.12 (0.68-1.83)	30	9	1.19 (0.83-1.71)
P73 14T	37	178	1.11 (0.80-1.54)	77	147	0.98 (0.76-1.26)	312	190	1.11 (0.93-1.32)
Estimated τ^2			0.025			0.0133			0.0139
Pathway estimation									
BER			1.11 (0.84-1.45)			1.04 (0.84-1.28)			1.08 (0.92-1.27)
NER			1.06 (0.84-1.33)			0.98 (0.83-1.17)			0.98 (0.86-1.12)
MMR/DR			1.19 (0.82-1.71)			1.06 (0.81-1.40)			0.98 (0.80-1.21)
DSB			1.04 (0.59-1.85)			1.17 (0.75-1.84)			1.15 (0.84-1.58)
Cell cycle			1.02 (0.81-1.29)			1.04 (0.88-1.23)			1.06 (0.94-1.21)
SIFT			0.89 (0.53-1.50)			0.90 (0.60-1.35)			1.09 (0.81-1.46)

*Prior matrix included pathway indicators and a SIFT covariate.

Downloaded from cebp.aacrjournals.org on November 29, 2021. © 2007 American Association for Cancer Research.

adenocarcinoma under conventional regression model, whereas the hierarchical model obtained a more precise estimate of 1.54 (95% CI, 1.03-2.30). The biological implications of the association between *CHEK2, OGG1* and lung cancer risk were discussed in the previous reports.

XRCC2 188H was shown to increase risk of lung cancer among current smokers, and the effect was mainly on small cell carcinoma. XRCC2 protein is an essential component of homologous recombination involved in the double-strand breaks. *In vitro* studies showed that cells with a mutation or deletion of amino acid at codon 188 had a lower survival fraction after mitomycin C–inducted DNA damage (31). The common polymorphism R188H was also shown to have a small effect on XRCC2 function and cell survival after DNA damage (31). Only limited studies have been conducted to assess the association between *XRCC2* R188H and lung cancer risk and the results have been inconsistent (26, 32, 33). Our observations of *XRCC2* 188H allele among current smokers and small-cell carcinoma are compatible to the previous experimental evidence.

TP53 variants, on the other hand, were suggested to increase lung cancer risk in conventional analysis, but the risk estimate was no longer significant in the hierarchical model. This could indicate that the results from the conventional analyses were overly estimated, or that the effect of TP53 variants may be derived from a different distribution compared with the other variants in the cell cycle pathways. To further clarify the role of TP53 variants and lung cancer risk, a more detailed analysis of TP53 based on haplotype tagging SNPs is currently under way.

Recent genetic association studies have been hampered by failures to replicate results, which could be due to multiple comparison or selective reporting. We showed that hierarchical modeling may improve the analyses of genetic data by increasing the precision of the risk estimates (e.g., as evidenced by narrower confidence intervals), and reducing the likelihood of false positive via shrinkage toward to the prior mean. The hierarchical modeling approach also has an appealing advantage of providing the risk estimate of each pathway investigated, although none of the pathways investigated have seemed to be have a more prominent effect than the others. This might be due to the low number of markers included in each pathway.

There are several limitations of our study. The first limitation to be considered is the possible violation of the exchangeability assumption, as the true effect size of the variants in the same pathway might not arise from the same distribution. Some variants in the pathway may have a larger effect, whereas others in the same pathway have very modest effects; or, in a more extreme scenario, the alleles specified in the same pathway may confer opposite effects. In addition, we dichotomized the genotypes into risk and nonrisk genotypes to minimize the number parameters to be modeled. Given that the previous literature on some of these variants remains inconclusive, the assumption of the minor alleles being the risk allele, unless there is prior evidence indicating otherwise, could also potentially lead to the violation of exchangeability. When the assumption of exchangeability is violated, the effect estimation would be shrunk toward a wrong prior mean. Nevertheless, previous sensitivity analyses have shown that hierarchical modeling is relatively robust to the alteration of the prior, and it should provide a more precise risk estimates (compared with the conventional model) under a reasonable specified prior (16, 34, 35). Second, despite our attempts to incorporate the evolutionary biology perspectives into our models in addition to the pathway indicators, we can only obtain prior information on nonsynonymous variants systematically based on the available tool such as SIFT. However, SIFT scores vary depending on depth and the species for which sequence alignment of specific variants is available. Thus, it is important to recognize that these scores are not the absolute measurements of how conserved the variants of interests are. Nevertheless, it provides reasonable representation of the evolution biology and physicochemical properties of the variants for the purpose of our analysis. Finally, we consider it reasonable to assume that the prior probability of a synonymous variant conferring a true effect to be very low. Nevertheless, it is possible that this type of variants can be functional through other mechanisms such as alternation of the splicing site, which currently is difficult to predict. The prior matrix can be further refined when such information become available.

In summary, our data provided robust positive signals of genetic regions that might harbor the inherited predisposition of lung cancer through a Bayesian framework and conservative estimation. To confirm the association found in this study, independent replication is required. Further investigation on *OGG1, MGMT*, and *CHEK2* focusing on the genetic regions where the present markers are located or the haplotype blocks tightly linked with these markers might be warranted.

Acknowledgments

We thank Valérie Gaborieu for the technical assistance.

References

1. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. J Natl Cancer Inst 1994;86:1600–8.
2. Amundadottir LT, Thorvaldsson S, Gudbjartsson DF, et al. Cancer as a complex phenotype: pattern of cancer distribution within and beyond the nuclear family. PLoS Med 2004;1:e65.
3. Hemminki K, Czene K. Attributable risks of familial cancer from the Family-Cancer Database. Cancer Epidemiol Biomarkers Prev 2002;11:1638–44.
4. Matakidou A, Eisen T, Houlston RS. Systematic review of the relationship between family history and lung cancer risk. Br J Cancer 2005;93:825–33.
5. 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.
6. Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. Nature 2001;411:366–74.
7. Mohrenweiser HW, Jones IM. Variation in DNA repair is a factor in cancer susceptibility: a paradigm for the promises and perils of individual and population risk estimation? Mutat Res 1998;400:15–24.
8. Kastan MB, Bartek J. Cell-cycle checkpoints and cancer. Nature 2004;432:316–23.
9. Landi S, Gemignani F, Canzian F, et al. DNA repair and cell cycle control genes and the risk of young onset lung cancer. Carcinogenesis. In press 2006.
10. Ware MD, DeSilva D, Sinilnikova OM, Stoppa Lyonnet D, Tavtigian SV, Mazoyer S. Does nonsense-mediated mRNA decay explain the ovarian cancer cluster region of the BRCA2 gene? Oncogene 2006;25:323–8.
11. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. Genome Res 2002;12:436–46.

Downloaded from cebp.aacrjournals.org on November 29, 2021. © 2007 American Association for Cancer Research.

12. Hung RJ, Brennan P, Canzian F, et al. Large-scale investigation of base excision repair genetic polymorphisms and lung cancer risk in a multicenter study. J Natl Cancer Inst 2005;97:567–76.
13. Hung RJ, Boffetta P, Canzian F, et al. Sequence variants in cell cycle control pathway, x-ray exposure, and lung cancer risk: a multicenter case-control study in Central Europe. Cancer Res 2006;66:8280–6.
14. Brennan P, McKay J, Moore L, et al. Uncommon CHEK2 mis-sense variant and reduced risk of tobacco-related cancers: case control study. Hum Mol Genet 2007;16:1794–801.
15. Stram DO. Software for tag single nucleotide polymorphism selection. Hum Genomics 2005;2:144–51.
16. Hung RJ, Brennan P, Malaveille C, et al. Using hierarchical modeling in genetic association studies with multiple markers: application to a case-control study of bladder cancer. Cancer Epidemiol Biomarkers Prev 2004;13:1013–21.
17. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res 2001;11:863–74.
18. Witte JS, Greenland S, Kim LL. Software for hierarchical modeling of epidemiologic data. Epidemiology 1998;9:563–6.
19. Ishikawa T, Ide F, Qin X, et al. Importance of DNA repair in carcinogenesis: evidence from transgenic and gene targeting studies. Mutat Res 2001;477:41–9.
20. Pegg AE, Dolan ME, Moschel RC. Structure, function, and inhibition of O6-alkylguanine-DNA alkyltransferase. Prog Nucleic Acid Res Mol Biol 1995;51:167–223.
21. Heighway J, Margison GP, Santibanez-Koref MF. The alleles of the DNA repair gene O^6-alkylguanine-DNA alkyltransferase are expressed at different levels in normal human lung tissue. Carcinogenesis 2003;24:1691–4.
22. Kaur TB, Travaline JM, Gaughan JP, Richie JP Jr., Stellman SD, Lazarus P. Role of polymorphisms in codons 143 and 160 of the O^6-alkylguanine DNA alkyltransferase gene in lung cancer risk. Cancer Epidemiol Biomarkers Prev 2000;9:339–42.
23. Cohet C, Borel S, Nyberg F, et al. Exon 5 polymorphisms in the O^6-alkylguanine DNA alkyltransferase gene and lung cancer risk in non-smokers exposed to second-hand smoke. Cancer Epidemiol Biomarkers Prev 2004;13:320–3.
24. Yang M, Coles BF, Caporaso NE, Choi Y, Lang NP, Kadlubar FF. Lack of association between Caucasian lung cancer risk and O^6-methylguanine-DNA methyltransferase-codon 178 genetic polymorphism. Lung Cancer 2004;44:281–6.
25. Krzesniak M, Butkiewicz D, Samojedny A, Chorazy M, Rusin M. Polymorphisms in TDG and MGMT genes—epidemiological and functional study in lung cancer patients from Poland. Ann Hum Genet 2004;68:300–12.
26. Zienolddiny S, Campa D, Lind H, et al. Polymorphisms of DNA repair genes and risk of non-small cell lung cancer. Carcinogenesis 2005;27:260–7.
27. Chae MH, Jang JS, Kang HG, et al. O^6-alkylguanine-DNA alkyltransferase gene polymorphisms and the risk of primary lung cancer. Mol Carcinog 2006;45:239–49.
28. Hazra TK, Roy R, Biswas T, Grabowski DT, Pegg AE, Mitra S. Specific recognition of O^6-methylguanine in DNA by active site mutants of human O^6-methylguanine-DNA methyltransferase. Biochemistry 1997;36:5769–76.
29. Ford BN, Ruttan CC, Kyle VL, Brackley ME, Glickman BW. Identification of single nucleotide polymorphisms in human DNA repair genes. Carcinogenesis 2000;21:1977–81.
30. Brennan P, McKay J, Moore L, et al. An uncommon CHEK2 polymorphism and the risk of lung and UADT cancer. Human Molecular Genetics 2007;16:1794–801.
31. Rafii S, O'Regan P, Xinarianos G, et al. A potential role for the XRCC2 R188H polymorphic site in DNA-damage repair and breast cancer. Hum Mol Genet 2002;11:1433–8.
32. Matullo G, Dunning AM, Guarrera S, et al. DNA repair polymorphisms and cancer risk in non-smokers in a cohort study. Carcinogenesis 2006;27:997–1007.
33. Rudd MF, Webb EL, Matakidou A, et al. Variants in the GH-IGF axis confer susceptibility to lung cancer. Genome Res 2006;16:693–701.
34. Greenland S. Principles of multilevel modelling. Int J Epidemiol 2000;29:158–67.
35. Greenland S, Poole C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. Arch Environ Health 1994;49:9–16.

Downloaded from cebp.aacrjournals.org on November 29, 2021. © 2007 American Association for Cancer Research.

Cancer Epidemiology, Biomarkers & Prevention

AACR American Association for Cancer Research

Inherited Predisposition of Lung Cancer: A Hierarchical Modeling Approach to DNA Repair and Cell Cycle Control Pathways

Rayjean J. Hung, Meili Baragatti, Duncan Thomas, et al.

Cancer Epidemiol Biomarkers Prev 2007;16:2736-2744.

Updated version	Access the most recent version of this article at: http://cebp.aacrjournals.org/content/16/12/2736

Cited articles	This article cites 33 articles, 8 of which you can access for free at: http://cebp.aacrjournals.org/content/16/12/2736.full#ref-list-1
Citing articles	This article has been cited by 2 HighWire-hosted articles. Access the articles at: http://cebp.aacrjournals.org/content/16/12/2736.full#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, use this link http://cebp.aacrjournals.org/content/16/12/2736. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.

Downloaded from cebp.aacrjournals.org on November 29, 2021. © 2007 American Association for Cancer Research.