Published in final edited form as:

Mol Carcinog. 2012 October ; 51(Suppl 1): E11–E20. doi:10.1002/mc.20858.

Genetic variability in DNA repair and cell cycle control pathway genes and risk of smoking-related lung cancer

Shama C. Buch^{1,*}, Brenda Diergaarde^{2,*}, Tomoko Nukui¹, Roger S. Day³, Jill M. Siegfried⁴, Marjorie Romkes¹, and Joel L. Weissfeld^{2,†}

¹University of Pittsburgh Cancer Institute and Center for Clinical Pharmacology, University of Pittsburgh School of Medicine

²Department of Epidemiology, University of Pittsburgh Graduate School of Public Health

³Department of Biomedical Informatics, University of Pittsburgh School of Medicine

⁴Department of Pharmacology and Chemical Biology, University of Pittsburgh School of Medicine

Abstract

DNA repair and cell cycle control play an important role in the repair of DNA damage caused by cigarette smoking. Given this role, functionally relevant single nucleotide polymorphisms (SNPs) in genes in these pathways may well affect the risk of smoking-related lung cancer. We examined the relationship between 240 SNPs in DNA repair and cell cycle control pathway genes and lung cancer risk in a case-control study of white current and ex-cigarette smokers (722 cases and 929 controls). Additive, dominant and recessive genetic models were evaluated for each SNP. A genetic risk summary score was also constructed. Odds ratios (OR) for lung cancer risk and 95% confidence intervals (95% CI) were estimated using logistic regression models. Thirty-eight SNPs were associated with lung cancer risk in our study population at $P < 0.05$. The strongest associations were observed for rs2074508 in *GTF2H4* ($P_{\text{additive}} = 0.003$), rs10500298 in *LIG1* ($P_{\text{recessive}} = 2.7 \times 10^{-4}$), rs747658 and rs3219073 in *PARP1* (rs747658: $P_{\text{additive}} = 5.8 \times 10^{-5}$; rs3219073: $P_{\text{additive}} = 4.6 \times 10^{-5}$), and rs1799782 and rs3213255 in *XRCC1* (rs1799782: $P_{\text{dominant}} = 0.006$; rs3213255: $P_{\text{recessive}} = 0.004$). Compared to individuals with first quartile (lowest) risk summary scores, individuals with third and fourth quartile summary score results were at increased risk for lung cancer (OR: 2.21, 95% CI: 1.66–2.95 and OR: 3.44, 95% CI: 2.58–4.59, respectively; $P_{\text{trend}} < 0.0001$). Our data suggests that variation in DNA repair and cell cycle control pathway genes is associated with smoking-related lung cancer risk. Additionally, combining genotype information for SNPs in these pathways may assist in classifying current and ex-cigarette smokers according to lung cancer risk.

Keywords

SNP; case-control; lung cancer

Introduction

Lung cancer causes approximately 160,000 deaths annually in the United States [1]. Although cigarette smoking can explain over 80% of this risk [2], lung cancer also occurs in never smokers [3] and fewer than one of five smokers develop lung cancer [4]. Plausible explanations for variable risk in smokers, occurrence in never smokers and aggregation in

[†]To whom correspondence should be addressed. Tel: +1-412-623-3313; Fax: +1-412-623-3303; jwepid@pitt.edu.

*These authors contributed equally to this work.

families include inherited differences in genes intimately involved in carcinogenesis by way of carcinogen metabolism, DNA repair, cell cycle control, apoptosis or inflammation [5].

Approximately 160 human genes mediate DNA repair through processes that include nucleotide excision repair (NER), base excision repair (BER), mismatch repair, homologous recombination, and non-homologous end-joining, the latter two pathways responsible for repairing double strand DNA breaks [6]. The NER pathway repairs DNA damage caused by the tobacco-related carcinogen benzo(a)pyrene, while the BER pathway repairs DNA damage caused by reactive oxygen species produced by cigarette smoke [7]. Given these roles, functionally relevant single nucleotide polymorphisms (SNPs) in genes in these pathways may well affect the risk of smoking-related lung cancer. In addition, cell cycle control proteins, including p53, p21, cyclin D, and cyclin E, mediate DNA damage responses that lead either to apoptosis or cell cycle arrest [8]. Therefore, variation in relevant cell cycle control pathway genes could magnify or attenuate cumulative effects from deficiencies in DNA repair.

Early studies typically evaluated associations with lung cancer risk for a limited number of SNPs in selected NER genes, mainly *ERCC1*, *ERCC2*, *ERCC5*, *XPA* and *XPC* [9,10], in selected BER genes, mainly *APEX1*, *OGG1* and *XRCC1* [9], or in selected cell cycle control genes *TP53* (p53) [9,11], *CDKN1A* (p21) [11], and *CCND1* (cyclin D1) [11,12]. Unfortunately, this approach has not yet identified a single gene variant useful for lung cancer risk stratification. This result motivates the search for stronger and more useful patterns of variation that involve multiple candidate genes representing more than one cancer-relevant biological pathway. Therefore, we used a custom-designed 384-SNP microarray to examine the relationship between common variation in NER, BER, and cell cycle control pathway genes and risk of smoking-related lung cancer in a large case-control study of white current and ex-cigarette smokers (722 cases and 929 controls).

Materials and Methods

Study population

Cases—The case group includes patients enrolled between 1990 and 2008 within one year of a thoracic surgery procedure for lung cancer diagnosis, staging or treatment at a University of Pittsburgh Medical Center hospital. Additional eligibility criteria for the current study included: 1) non-missing sex, 2) current or ex-cigarette smoker, 3) 10 pack-year cumulative cigarette dose exposure, 4) 45–85 years old at time of lung cancer diagnosis, 5) pathologically verified lung cancer diagnosis (excluding carcinoid), and 6) DNA sample available for genotyping. Cases previously enrolled in a computed tomography (CT) lung cancer screening study (Pittsburgh Lung Screening Study, PLuSS) were excluded. In total, 837 cases fulfilled these criteria.

Controls—Controls for this study were selected from PLuSS. Between 2002 and 2005, PLuSS enrolled 50–79 year-old current and ex-cigarette smokers of at least one-half pack/day for at least 25 years and excluded individuals who 1) quit smoking more than 10 years earlier, 2) reported a history of lung cancer, or 3) reported chest CT within one year of enrollment. For the current study, we randomly selected 1000 controls from among the 3463 white or black race CT-screened PLuSS participants with DNA available and no interval lung cancer diagnosis as of September 2008. All cases and controls provided written informed consent and were enrolled under University of Pittsburgh Institutional Review Board-approved protocols.

Gene and SNP selection—SNPs in key genes in the NER, BER and cell cycle control pathways were included on the 384-SNP microarray (See Supplemental Tables 1 and 2 for a

listing of all included SNPs). The SNP selection strategy incorporated six features: 1) functional significance as determined by amino acid substitution; 2) potential for regulatory change; 3) alteration in protein stability; 4) evolutionary conservation across species, 5) disease association reported in the literature; and, 6) tagSNP in either Utah residents with Northern or Western European ancestry (HapMap CEU) or Yoruban Africans from Ibadan, Nigeria (HapMap YRI). We additionally included 10 ancestry informative markers (AIMs) (rs1426654, rs16891982, rs1871534, rs2814778, rs3827760, rs6003, rs722098, rs723632, rs7349 and rs930072), previously validated against self-reported race in a study population including Caucasian and African-American subjects.

DNA isolation and genotyping—Genomic DNA was isolated using Gentra Systems Inc. isolation kits and analyzed using an Illumina® GoldenGate custom-designed 384-SNP microarray. Analyses serially excluded: 1) assigned genotypes by Illumina® GenomeStudio 2010.1 using a GenTrain score <0.45 or a ClusterSep score <0.25, 2) individuals with genotyping call rates <90%, and 3) SNPs that failed in more than 2% of the samples. Forty cases and eight controls had call rates <90% and were excluded leaving 797 cases and 992 controls. Seven SNPs failed completely and 21 SNPs failed in > 2% of the samples, these SNPs were excluded. SNPs with minor allele frequency (MAF) <0.02 in the control group were excluded as well ($N=62$, mainly HapMap YRI tagSNPs), as were 19 SNPs that had been included on the 384-SNP microarray for a separate study and were located in genes not directly involved in DNA repair or cell cycle control (11 in *ABCB1* and 8 in *GSTP1*). Finally, 25 SNPs that deviated from Hardy-Weinberg equilibrium (HWE; $P<0.01$) were also excluded from the analyses. Supplemental Table 1 lists all 134 SNPs that were excluded. Supplemental Table 2 shows HWE P -values and control group MAFs for the remaining 240 non-AIM SNPs. The 240 SNP set consists of 185 SNPs in 20 NER pathway genes (gene: number of SNPs – *ERCC1*: 5, *ERCC2*: 12, *ERCC3*: 10, *ERCC4*: 8, *ERCC5*: 15, *ERCC6*: 1, *ERCC8*: 9, *GTF2H1*: 10, *GTF2H3*: 6, *GTF2H4*: 6, *GTF2H5*: 2, *LIG1*: 21, *MNAT1*: 8, *RAD23A*: 1, *RAD23B*: 9, *RPA1*: 17, *RPA2*: 3, *RPA3*: 25, *XPA*: 6, *XPC*: 11), 29 SNPs in four BER pathway genes (*MPG*: 2, *OGG1*: 2, *PARP1*: 12, *XRCC1*: 13), and 26 SNPs in seven cell cycle control genes (*CDKN2A*: 5, *CCND1*: 5, *CCNE1*: 2, *CCNH*: 2, *CDK7*: 4, *E2F1*: 2, *TP53/P53*: 6).

Statistical analyses

Using data from the 10 AIMs for the 797 cases and 992 controls with 90% call rates and a two-cluster solution in Structure 2.3.1 [13], we inferred genetic ancestry using the estimated white genome fraction cutoff that correctly classified 929 (99.5%) of 934 and 58 (100%) of 58 eligible control subjects self-reporting white and black race, respectively. To avoid spurious candidate genotype-disease association due to population stratification [14], all subsequent data analyses used the 722 cases (677 self-reported white and 45 no reported race) and the 929 controls from the white genetic ancestry cluster.

Deviation from HWE was examined in the control population for each SNP using the χ^2 goodness-of-fit test. We also used χ^2 tests to evaluate case-control differences with respect to sex, age, enrollment year, and cumulative cigarette dose (pack-year) exposure, and to test independence, in the control group, between genotype and other factors (sex, age, enrollment year, and pack-year). We used logistic regression (SAS 9.2, SAS Institute, Cary, NC) to estimate crude and adjusted odds ratio (OR) for lung cancer risk and corresponding 95% confidence interval (95% CI). Additive, dominant and recessive genetic models were evaluated for each SNP. Adjusted models contained parameters for sex, age, year enrolled, and pack-year. Fitting logistic regression models containing terms for genotype group and exposure group (sex, age, year enrolled, or pack-year) and terms for the genotype group by exposure group cross classification, screens for gene-environment interaction applied the

log-likelihood ratio test (at $P<0.05$) to the cross classification terms. Case-case comparisons, comparing 258 squamous cell carcinoma cases with 301 adenocarcinoma cases, were conducted to evaluate heterogeneity in genetic risk factors for these two non-small cell lung cancer subsets.

We used HapMap phase 1 & 2 data (<http://hapmap.ncbi.nlm.nih.gov/>) and Haploview 4.1 [15] to determine, for each gene, the fraction of all successfully genotyped (HWE P -value 0.01) common-variant (MAF 0.05) SNPs in the HapMap-CEU population captured at r^2 0.8 by the subset of SNPs included in our panel and screened for association with lung cancer risk. The 240 SNP set captured all (100%) common-variant HapMap-CEU SNPs in five genes (*E2F1*, *GTF2H1*, *LIG1*, *MPG*, and *TP53/P53*), most (80–99%) common-variant HapMap-CEU SNPs in 12 genes (*CCND1*, *CDKN2A*, *ERCC1*, *ERCC3*, *ERCC4*, *ERCC5*, *GTF2H3*, *GTF2H4*, *GTF2H5*, *MNAT1*, *PARP1*, and *RPA2*), many (60–79%) common-variant HapMap-CEU SNPs in 9 genes (*CCNH*, *CDK7*, *ERCC2*, *OGG1*, *RPA1*, *RPA3*, *XPA*, *XPC*, and *XRCC1*), and comparatively few (<60%) common-variant HapMap-CEU SNPs in five genes (*CCNE1*, *ERCC6*, *ERCC8*, *RAD23A*, and *RAD23B*).

A compound statistic [16,17] was used to rank SNPs according to statistical association with lung cancer risk. That is, SNPs were ranked according to the minimum, unadjusted P -value obtained under the additive, dominant, and recessive genetic models (min 3p). Results from the additive and recessive models were ignored in sample sets containing fewer than 10 minor allele homozygotes.

Using the most frequent haplotype as referent and ignoring haplotypes with a control group frequency <0.01, whole gene haplotype analyses used HPlus 3.2 (Fred Hutchinson Cancer Research Center, WA, <http://qge.fhcrc.org/software.php>) to identify lung cancer risk-associated haplotypes. Using selected lung cancer risk-associated SNPs in low linkage disequilibrium, we additionally summarized genetic lung cancer risk by summing genotypes coded as shown in Table 4, footnote 2.

Results

Characteristics of the study population are presented in Table 1. Compared with the control group ($N=929$), the lung cancer case group ($N=722$) included more men, more subjects in the two oldest age categories, and more subjects in the highest pack-year category.

Thirty-eight SNPs (15.8% of 240) were associated with risk of lung cancer at $P<0.05$ (unadjusted). Table 2 shows genotype distribution, best genetic model (*i.e.*, model with smallest P -value), rank based on min 3p, unadjusted and adjusted ORs and 95% CIs and P -values for these SNPs. Eleven SNPs had unadjusted P -values <0.01, and P -values remained <0.01 after adjustment for sex, age, year enrolled, and pack-years for six of these SNPs (Table 2). This latter set includes: rs2074508 in *GTF2H4* (OR: 0.73, 95% CI: 0.59–0.90, $P=0.003$, additive model), rs10500298 in *LIG1* (OR: 1.73, 95% CI: 1.29–2.32, $P=2.7\times10^{-4}$, recessive model), rs747658 and rs3219073 in *PARP1* (two SNPs in high linkage disequilibrium, r^2 0.8; rs747658: OR: 0.62, 95% CI: 0.49–0.78, $P=5.8\times10^{-5}$; rs3219073: OR: 0.61, 95% CI: 0.48–0.77, $P=4.6\times10^{-5}$; additive model for both), and rs1799782 (Arg194Trp) and rs3213255 in *XRCC1* (rs1799782: OR: 0.53, 95% CI: 0.34–0.84, $P=0.006$, dominant model; rs3213255: OR: 1.52, 95% CI: 1.14–2.01, $P=0.004$, recessive model). Similar results were observed when the case group was limited to non-small cell lung cancer only (data not shown).

Evaluation of gene-environment interactions was limited to the 38 SNPs associated with risk of lung cancer at $P<0.05$ (Table 2). Being homozygous for the rare allele of SNP rs7783714 in *RPA3* was associated with a reduced risk of lung cancer among women (OR: 0.41, 95%

CI: 0.23–0.74), but not among men (OR: 0.91, 95% CI: 0.55–1.49) ($P_{\text{interaction}}=0.04$). Otherwise, analyses stratified by sex, age, enrollment year, or pack-year did not identify meaningful instances of SNP-lung cancer risk associations modified by co-factor level (data not shown).

Case-case comparisons conducted to evaluate heterogeneity in genetic risk factors for the two largest lung cancer subsets (squamous cell carcinomas and adenocarcinomas) were limited to the same 38 SNP set. *PARP1* rs2048424 minor allele homozygotes were more common among adenocarcinoma cases than among squamous cell carcinoma cases (12.0% vs. 7.0%, $P=0.049$); among controls, 14.1% were homozygous for the rs2048424 minor allele. *LIG1* rs10500298 minor allele homozygotes were also more common among adenocarcinoma cases than among squamous cell carcinoma cases (28.0% vs. 19.9%, $P=0.027$); among controls, 13.9% were homozygous for the rs10500298 minor allele. Frequencies for the other SNPs did not differ significantly between the two tumor subsets (data not shown).

Compared to the most common haplotype, rarer haplotypes in five different genes (*ERCC2*, *GTF2H4*, *PARP1*, *XRCC1* and *CCND1*) were associated with a decreased risk of lung cancer and in one gene (*GTF2H1*) with an increased risk of lung cancer ($P<0.01$; Table 3). Results were qualitatively similar after adjustment for sex, age, enrollment year, and pack-year (Table 3). As indicated in Table 3, the rare *ERCC2* haplotype associated with decreased lung cancer risk contained the rs238405 common allele, the *GTF2H4* haplotype contained the rs2074508 minor allele, the *PARP1* haplotypes contained the rs8679 common allele and the most rare *PARP1* haplotype additionally contained the minor alleles from rs747658, rs3219073, rs3219090, rs2048424 and rs2666428, one of the *XRCC1* haplotypes contained the rs1799782 minor allele and the common alleles from rs762507 and rs3213255 and the other *XRCC1* haplotype contained the rs1799782 common allele and the rs3213266 minor allele, and the *CCND1* haplotype contained the rs603965 common allele and the minor alleles from rs3918298 and rs678653. The rare *GTF2H1* haplotype associated with increased risk of lung cancer contained the rs4150667 minor allele (Table 3).

We constructed a genotype risk summary score that used best genetic model results for 31 of the 38 SNPs that were statistically significantly associated with lung cancer risk at $P<0.05$ (see Table 2). The 7 SNPs that were excluded (*RPA1* rs4281767 and rs7503173, *RPA2* rs7356, *RPA3* rs13227585, and *PARP1* rs747658, rs2048424 and rs2666428) were in high linkage disequilibrium ($r^2 \geq 0.80$) with one or more of the 31 SNPs included in the summary score (data not shown). Summary scores were stratified according to quartile distribution in the control group. Individuals with third and fourth quartile summary score results were at increased risk for lung cancer compared to individuals with first quartile (lowest) summary scores (OR: 2.21, 95% CI: 1.66–2.95 and OR: 3.44, 95% CI: 2.58–4.59, respectively; $P_{\text{trend}} < 0.0001$; Table 4). Adjustment for sex, age, enrollment year, and pack-year attenuated these associations minimally (Table 4). Summary score quartile predicted lung cancer with a 0.636 area under the receiver operating characteristic (ROC) curve.

Discussion

In this case-control study, we evaluated associations of 240 common SNPs in genes involved in NER, BER or cell cycle control with risk of smoking-related lung cancer. Thirty-eight SNPs, 24 in NER, 10 in BER and four in cell cycle control genes, were associated with lung cancer risk at $P<0.05$ in our study population. The strongest associations with risk of lung cancer were observed for rs2074508 in *GTF2H4*, rs10500298 in *LIG1*, rs747658 and rs3219073 in *PARP1*, and rs1799782 and rs3213255 in *XRCC1*. In

addition, haplotypes in six different genes (*ERCC2*, *GTF2H1*, *GTF2H4*, *PARP1*, *XRCC1* and *CCND1*) were found associated with lung cancer risk.

GTF2H4 codes for the p52 subunit of NER transcription factor IIIH (TFIIH). The TFIIH complex has ATPase and helicase activities and opens DNA at sites of DNA distorting damage. The p52 subunit may regulate the ATPase activity of the TFIIH subunit (XPB), a protein coded by *ERCC3* [18]. In our study, single SNP analyses found the minor allele of rs2074508 in *GTF2H4* significantly associated with a reduced risk of lung cancer (additive model). The common *GTF2H4* haplotype associated with reduced lung cancer risk captured this minor allele. Wang *et al.* [19] recently reported associations between cervical dysplasia and two SNPs near *GTF2H4*, rs2894054, located 3.7 kb upstream, and rs6926723, located 0.9 kb downstream (within valyl-tRNA synthetase, *VARSL*) from *GTF2H4*. Published studies of *GTF2H4* genetic variation and cancer risk appear otherwise not to exist [19]. Providing a first comprehensive survey of common variation in *GTF2H4* and lung cancer susceptibility, the *GTF2H4* SNPs evaluated in our study captured 9 (90%) of the 10 HapMap phase 1 & 2 common-variant (MAF 0.05) CEU SNPs.

DNA ligase I (encoded by *LIG1*) joins Okazaki fragments in the lagging strand during DNA synthesis and completes BER and NER [20]. In our study, case and control genotype distributions differed significantly for one of the 21 examined *LIG1* SNPs. *LIG1* rs10500298 minor allele homozygotes were more common among lung cancer cases than controls and more common among adenocarcinomas than squamous cell carcinomas. A literature search identified five published studies on lung cancer risk and *LIG1* genetic variability [study #1: one SNP, 530 and 570 non-Hispanic white cases and controls [21]; study #2, 34 SNPs, 143 and 172 French Caucasian cases and controls [22]; study #3: four SNPs, ~440 and ~790 mixed race (~60% Caucasian) cases and controls [23]; study #4: five SNPs, 113 and 299 Latino American cases and controls [24], and study #5: five SNPs, 255 and 280 African American cases and controls [24]]. The French Caucasian, mixed race, and African American studies reported significant lung cancer associations involving either single *LIG1* SNPs and/or haplotypes. Our SNP panel included two SNPs (rs20581, rs20579) associated with risk of lung cancer in the mixed race study. However, neither SNP was found to be associated with lung cancer risk in our study. SNP rs10500298 was not in strong linkage disequilibrium with a third mixed race lung cancer risk-associated SNP (rs20580) nor with three French Caucasian lung cancer risk-associated SNPs (rs3730994, rs3786763, rs3730912).

Upon detection of DNA strand breaks, poly (ADP-ribose) polymerase 1 (PARP1) adds poly(ADP-ribose) to nuclear proteins, recruits XRCC1, and initiates BER [25]. Interestingly, inhibition of PARP1 creates a state of synthetic lethality in cells that are unable to complete homologous recombination as a result of *BRCA1* or *BRCA2* loss and PARP1 inhibitors show promise in treating BRCA-deficient breast and ovarian cancer [25,26]. In our study, the minor alleles of two *PARP1* SNPs, rs747658 and rs3219073 (in high linkage disequilibrium), reduced lung cancer risk, per minor allele, by one third. Consistent with this, the relatively common *PARP1* haplotype that contained the minor alleles from these two SNPs was strongly statistically significantly associated with a lower risk of lung cancer as well. Thus far, only a few studies on *PARP1* variability and lung cancer risk have been published. Among Japanese, having at least one minor allele of SNP rs3219145 (Lys940Arg), a SNP absent in HapMap CEU whites and not evaluated by us, was associated with increased lung cancer risk (OR 1.40, 95% CI 1.04–1.90) [27]. Lockett *et al.* [28] reported that poly (ADP-ribose) polymerase activity is lower in lymphocytes collected from individuals with at least one 762Ala allele (Val762Ala; rs1136410) and in vitro experiments confirmed lower enzymatic activity of PARP1-Ala762 [29]. Consistent with this, among Han Chinese, having at least one minor allele of SNP rs1136410 was associated with an

increased lung cancer risk (OR 1.26, 95% CI 1.05–1.52) [30]. However, no significant association was observed in Korean [31] and Japanese study populations [27], and in our study population the association was not statistically significant either (OR 1.18, 95% CI 0.96–1.46). The *PARP1* SNPs evaluated in our study captured 46 (96%) of 48 phase 1 & 2 common-variant HapMap CEU SNPs.

Commonly described as a scaffold protein, XRCC1 may coordinate BER through interactions not only with PARP1, but also with DNA ligase III (LIG3), DNA polymerase β (POLB), apurinic/apyrimidinic (AP) nuclease (APEX1), and polynucleotide kinase 3'-phosphatase (PNKP) [32]. In our study population, two SNPs in *XRCC1*, rs1799782 (Arg194Trp) and rs3213255, were strongly associated with risk of lung cancer. Many published studies on *XRCC1* variability and lung cancer risk have evaluated rs1799782 and two other non-synonymous SNPs, rs25489 (Arg280His) and rs25487 (Arg399Gln) (for example, [33–35]). However, four recent meta-analyses, Kiyohara *et al.* [36], Wang *et al.* [37], Zheng *et al.* [38], and Vineis *et al.* [39], did not find any one of these three commonly studied non-synonymous coding SNPs associated with lung cancer risk in whites. In analyses adjusted for age and restricted to smokers with high pack-year exposures, a large white-only case-control study [34] also observed significantly lower lung cancer risk in association with Trp194 [34]. Therefore, our observation with respect to rs1799782 may reflect a study population enriched with heavy smokers combined with a Trp194 lung cancer protective effect that depends on a history of heavy smoking. The *XRCC1* SNPs evaluated in our study captured 18 (75%) of 24 phase 1 & 2 common-variant (MAF 0.05) HapMap CEU SNPs.

The effect of genetic variation in DNA repair and cell cycle control genes on lung cancer risk may become detectable only in the presence of certain environmental factors that cause DNA damage such as cigarette smoking. Our study population consisted of former and current smokers only and different results may be observed among never smokers. It should also be noted that multiple testing may have led to chance findings and that associations will need to be confirmed in additional studies. However, SNPs rs747658 and rs3219073 in *PARP1* did remain associated with lung cancer risk at $P < 0.05$ after applying the conservative Bonferroni correction.

Although our study population was relatively large, it is possible that some associations were not detected due to insufficient power. Specifically, testing under a log-additive model at a 0.05 two-sided significance level, our case-control study (722 cases and 929 controls) had under 90% power for very common variants (population MAF 0.25) with lung cancer effects (OR) less than 1.3 and for common variants (population MAF=0.05) with lung cancer effects (OR) less than 1.6 [40]. Haplotype-based analyses may have improved our power for detecting genes with lung cancer association involving uncommon or multiple susceptibility alleles [41]. Evaluations of gene-environment interactions and case-case comparisons were limited to the 38 SNPs associated with lung cancer risk at $P < 0.05$ and it is possible that important interactions and/or associations were not identified due to the limited number of SNPs investigated.

Unfortunately, our study population contained only a relatively small number of non-white subjects which limited our ability to evaluate SNP-lung cancer risk associations in non-white race groups. However, using 45 case and 58 control subjects with African genetic ancestry, we observed that four SNPs (*ERCC3* rs4150407, *LIG1* rs175628, *RPA2* rs7356, and *RPA3* rs13246995) were statistically significantly ($P < 0.05$) associated with lung cancer risk, and replicated directional associations observed in subjects with Caucasian genetic ancestry.

Contributing to an expanding knowledge base about the role of genetic variation in DNA repair and cell cycle pathway genes in lung carcinogenesis, our findings may help identify biological processes behind lung cancer susceptibility. However, knowing the genotype of a single common SNP can not be used usefully to assign smokers to very high or low risk groups. This limitation motivates the search for multi-SNP indices capable of stratifying smokers into more meaningful risk groups [42]. As an illustration of one uncomplicated approach [42], we formed a 31-SNP genotype risk summary score that identified second, third, and fourth quartile groups at 1.04-fold, 2.21-fold, and 3.44-fold increased risk, respectively, relative to a first, lowest risk, quartile reference group.

To conclude, our data suggests that common variation in DNA repair and cell cycle control pathway genes is associated with smoking-related risk of lung cancer. Specifically, we observed strong associations with lung cancer risk for SNPs in *GTF2H4*, *LIG1*, *PARP1* and *XRCC1*. An illustrative genotype risk summary score that combined genotype information for 31 SNPs in 15 genes risk-stratified current and ex-cigarette smokers over a 3.4-fold range. If confirmed in additional studies, combining genotype information for SNPs known to be associated with lung cancer risk may assist in classifying current and former smokers according to risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant support: This work was supported by National Cancer Institute grant 5P50 CA090440 to J. M. S. and Cancer Center Core Grant 2P30 CA047904 to the University of Pittsburgh Cancer Institute.

Abbreviations

CI	confidence interval
CT	computed tomography
MAF	minor allele frequency
OR	odds ratio
PLuSS	Pittsburgh Lung Screening Study
SNP	single nucleotide polymorphism

References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. *CA Cancer J Clin*. 2009; 59(4):225–249. [PubMed: 19474385]
2. The Health Consequences of Smoking: A report of the Surgeon General. Atlanta, Ga: Dept. of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; Washington, D.C: U.S. G.P.O.; 2004. The impact of smoking on disease and the benefits of smoking reduction (Table 7.3). For sale by the Supt. of Docs
3. Samet JM, Avila-Tang E, Boffetta P, Hannan LM, Olivo-Marston S, Thun MJ, Rudin CM. Lung cancer in never smokers: Clinical epidemiology and environmental risk factors. *Clin Cancer Res*. 2009; 15(18):5626–5645. [PubMed: 19755391]
4. Villeneuve PJ, Mao Y. Lifetime probability of developing lung cancer, by smoking status, Canada. *Can J Public Health*. 1994; 85(6):385–388. [PubMed: 7895211]

5. Rudin CM, Avila-Tang E, Harris CC, et al. Lung cancer in never smokers: Molecular profiles and therapeutic implications. *Clin Cancer Res.* 2009; 15(18):5646–5661. [PubMed: 19755392]
6. Wood RD, Mitchell M, Lindahl T. Human DNA repair genes, 2005. *Mutat Res.* 2005; 577(1–2): 275–283. [PubMed: 15922366]
7. Asami S, Hirano T, Yamaguchi R, Tomioka Y, Itoh H, Kasai H. Increase of a type of oxidative DNA damage, 8-hydroxyguanine, and its repair activity in human leukocytes by cigarette smoking. *Cancer Res.* 1996; 56(11):2546–2549. [PubMed: 8653695]
8. el-Deiry WS, Harper JW, O'Connor PM, et al. WAF1/CIP1 is induced in p53-mediated G1 arrest and apoptosis. *Cancer Res.* 1994; 54(5):1169–1174. [PubMed: 8118801]
9. Hung RJ, Christiani DC, Risch A, et al. International Lung Cancer Consortium: pooled analysis of sequence variants in DNA repair and cell cycle pathways. *Cancer Epidemiol Biomarkers Prev.* 2008; 17(11):3081–3089. [PubMed: 18990748]
10. Shen M, Berndt SI, Rothman N, et al. Polymorphisms in the DNA nucleotide excision repair genes and lung cancer risk in Xuan Wei, China. *Int J Cancer.* 2005; 116(5):768–773. [PubMed: 15849729]
11. Wang W, Spitz MR, Yang H, Lu C, Stewart DJ, Wu X. Genetic variants in cell cycle control pathway confer susceptibility to lung cancer. *Clin Cancer Res.* 2007; 13(19):5974–5981. [PubMed: 17908995]
12. Pabalan N, Bapat B, Sung L, Jarjanazi H, Francisco-Pabalan O, Ozcelik H. Cyclin D1 Pro241Pro (CCND1-G870A) polymorphism is associated with increased cancer risk in human populations: a meta-analysis. *Cancer Epidemiol Biomarkers Prev.* 2008; 17(10):2773–2781. [PubMed: 18843022]
13. Pritchard J, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000; 155:949–959.
14. Thomas DC, Witte JS. Point: Population stratification: A problem for case-control studies of candidate-gene associations? *Cancer Epidemiol Biomarkers Prev.* 2002; 11:505–512. [PubMed: 12050090]
15. Barrett J, Fry B, Maller J, Daly M. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21(2):263–265. [PubMed: 15297300]
16. Friedlin B, Zheng G, Li Z, Gastwirth J. Trend tests for case-control studies of genetic markers: Power, sample size and robustness. *Hum Hered.* 2002; 53:146–152. [PubMed: 12145550]
17. Kuo C-L, Feingold E. What's the best statistic for a simple test of genetic association in a case-control study? *Genet Epidemiol.* 2010; 34:246–253. [PubMed: 20025064]
18. Oksenyuk V, Coin F. The long unwinding road: XPB and XPD helicases in damaged DNA opening. *Cell Cycle.* 2010; 9(1):90–96. [PubMed: 20016270]
19. Wang SS, Gonzalez P, Yu K, et al. Common genetic variants and risk for HPV persistence and progression to cervical cancer. *PLoS One.* 2010; 5(1):e8667. [PubMed: 20084279]
20. Lindahl T, Barnes DE. Mammalian DNA ligases. *Annu Rev Biochem.* 1992; 61:251–281. [PubMed: 1497311]
21. Shen H, Spitz MR, Qiao Y, Zheng Y, Hong WK, Wei Q. Polymorphism of DNA ligase I and risk of lung cancer--a case-control analysis. *Lung Cancer.* 2002; 36(3):243–247. [PubMed: 12009232]
22. Michiels S, Danoy P, Dessen P, et al. Polymorphism discovery in 62 DNA repair genes and haplotype associations with risks for lung and head and neck cancers. *Carcinogenesis.* 2007; 28(8): 1731–1739. [PubMed: 17494052]
23. Lee Y-CA, Morgenstern H, Greenland S, et al. 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(7):1630–1638. [PubMed: 18059021]
24. Chang JS, Wrensch MR, Hansen HM, et al. Nucleotide excision repair genes and risk of lung cancer among San Francisco Bay Area Latinos and African Americans. *Int J Cancer.* 2008; 123(9): 2095–2104. [PubMed: 18709642]
25. Rouleau M, Patel A, Hendzel MJ, Kaufmann SH, Poirier GG. PARP inhibition: PARP1 and beyond. *Nat Rev Cancer.* 2010; 10(4):293–301. [PubMed: 20200537]
26. Fong PC, Boss DS, Yap TA, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *New Engl J Med.* 2009; 361(2):123–134. [PubMed: 19553641]

27. Sakiyama T, Kohno T, Mimaki S, et al. Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1 and LIG4 with lung cancer risk. *Int J Cancer*. 2005; 114(5): 730–737. [PubMed: 15609317]
28. Lockett KL, Hall MC, Xu J, et al. The ADPRT V762A genetic variant contributes to prostate cancer susceptibility and deficient enzyme function. *Cancer Res*. 2004; 64(17):6344–6348. [PubMed: 15342424]
29. Wang X-G, Wang Z-Q, Tong W-M, Shen Y. PARP1 Val762Ala polymorphism reduces enzymatic activity. *Biochem Biophys Res Commun*. 2007; 354(1):122–126.
30. Zhang X, Miao X, Liang G, et al. Polymorphisms in DNA base excision repair genes ADPRT and XRCC1 and risk of lung cancer. *Cancer Res*. 2005; 65(3):722–726. [PubMed: 15705867]
31. Choi JE, Park SH, Jeon H-S, et al. No association between haplotypes of three variants (codon 81, 284, and 762) in poly(ADP-ribose) polymerase gene and risk of primary lung cancer. *Cancer Epidemiol Biomarkers Prev*. 2003; 12(9):947–949. [PubMed: 14504211]
32. Horton JK, Watson M, Stefanick DF, Shaughnessy DT, Taylor JA, Wilson SH. XRCC1 and DNA polymerase beta in cellular protection against cytotoxic DNA single-strand breaks. *Cell Res*. 2008; 18(1):48–63. [PubMed: 18166976]
33. Cote ML, Yoo W, Wenzlaff AS, et al. Tobacco and estrogen metabolic polymorphisms and risk of non-small cell lung cancer in women. *Carcinogenesis*. 2009; 30(4):626–635. [PubMed: 19174490]
34. 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(8):567–576. [PubMed: 15840879]
35. Zhou W, Liu G, Miller DP, et al. Polymorphisms in the DNA repair genes XRCC1 and ERCC2, smoking, and lung cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2003; 12(4):359–365. [PubMed: 12692111]
36. Kiyohara C, Takayama K, Nakanishi Y. Association of genetic polymorphisms in the base excision repair pathway with lung cancer risk: a meta-analysis. *Lung Cancer*. 2006; 54(3):267–283. [PubMed: 16982113]
37. Wang Y, Yang H, Li H, Li L, Wang H, Liu C, Zheng Y. Association between X-ray repair cross complementing group 1 codon 399 and 194 polymorphisms and lung cancer risk: a meta-analysis. *Cancer Lett*. 2009; 285(2):134–140. [PubMed: 19481337]
38. Zheng H, Wang Z, Shi X, Wang Z. XRCC1 polymorphisms and lung cancer risk in Chinese populations: a meta-analysis. *Lung Cancer*. 2009; 65(3):268–273. [PubMed: 19329222]
39. Vineis P, Manuguerra M, Kavvoura FK, et al. A field synopsis on low-penetrance variants in DNA repair genes and cancer susceptibility. *J Natl Cancer Inst*. 2009; 101(1):24–36. [PubMed: 19116388]
40. Gauderman, W.; Morrison, J. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies. 2006. <http://hydra.usc.edu/gxe/>
41. Morris RW, Kaplan NL. On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. *Genet Epidemiol*. 2002; 23:221–233. [PubMed: 12384975]
42. Young RP, Hopkins RJ, Hay BA, et al. A gene-based risk score for lung cancer susceptibility in smokers and ex-smokers. *Postgrad Med J*. 2009; 85(1008):515–524. [PubMed: 19789190]

Table 1

Characteristics of lung cancer case and control subjects

Characteristic	Case group <i>N</i> _{total = 722}		Control group <i>N</i> _{total = 929}		<i>P</i> -value
	<i>N</i>	%	<i>N</i>	%	
Sex					0.08
men	398	55.1	472	50.8	
women	324	44.9	457	49.2	
Age					<0.0001
45–54	70	9.7	271	29.2	
55–64	201	27.8	458	49.3	
65–74	293	40.6	179	19.3	
75–84	158	21.9	21	2.3	
Pack-year					<0.0001
<35	178	24.7	264	28.4	
35–49	138	19.1	233	25.1	
50–74	219	30.3	307	33.0	
75+	187	25.9	125	13.5	
Histology					
Adenocarcinoma	301	41.7			
Squamous cell	258	35.7			
Non-small cell (NOS) [*]	60	8.3			
Adenosquamous	23	3.2			
Small cell	21	2.9			
Large cell	20	2.8			
BAC [†]	19	2.6			
Neuroendocrine	12	1.7			
Other	8	1.1			
Stage group [‡]					
I	311	44.4			
II	104	14.8			

Characteristic	Case group N _{total} = 722		Control group N _{total} = 929		P-value
	N	%	N	%	
III	215	30.7			
IV	39	5.6			
Missing	32	4.6			

* NOS: not otherwise specified

[‡] BAC – bronchioalveolar carcinoma

[‡] Excluding small cell

Table 2

SNPs statistically (unadjusted $P < 0.05$) associated with lung cancer risk, by pathway and gene.

Gene	SNP	Alleles*	Case group [†] N _{total} =722	Control group [†] N _{total} =929	Best genetic model [‡]	Rank [§]	Unadjusted			Adjusted [¶]		
							OR	95% CI	P-value	OR	95% CI	P-value
Nucleotide excision repair pathway												
ERCC1	rs3212955	C/T	389/281/51/721	521/364/43/928	Recessive	31	1.57	1.03–2.38	0.035	1.66	1.01–2.73	0.045
ERCC2	rs238405	A/T	218/349/143/710	287/485/146/918	Recessive	24	1.33	1.03–1.72	0.027	1.33	0.99–1.81	0.061
ERCC5	rs4150351	C/A	477/221/24/722	668/229/32/929	Dominant	14	1.31	1.07–1.62	0.011	1.43	1.11–1.84	0.005
GTF2H1	rs4150667	T/C	378/290/54/722	443/396/90/929	Additive	28	0.85	0.73–0.99	0.032	0.88	0.73–1.05	0.164
GTF2H3	rs2340693	G/A	673/46/3/722	887/42/0/929	Dominant	35	1.54	1.01–2.35	0.047	1.28	0.78–2.10	0.327
GTF2H4	rs2074508	A/G	496/200/26/722	571/308/49/928	Additive	6	0.76	0.64–0.91	0.002	0.73	0.59–0.90	0.003
LIG1	rs10500298	C/T	287/257/172/716	371/429/129/929	Recessive	1	1.96	1.52–2.53	1.8×10 ^{−7}	1.73	1.29–2.32	2.7×10 ^{−4}
RAD23B	rs6477569	G/T	586/116/20/722	756/161/12/929	Recessive	29	2.18	1.06–4.48	0.035	2.00	0.87–4.58	0.103
RPA1	rs4281767	C/T	521/189/12/722	666/232/31/929	Recessive	32	0.49	0.25–0.96	0.038	0.47	0.22–1.03	0.059
RPA1	rs7503173	T/G	522/187/12/721	662/235/32/929	Recessive	26	0.47	0.24–0.93	0.030	0.46	0.21–0.99	0.048
RPA1	rs7501623	G/A	523/187/12/722	656/239/34/929	Recessive	18	0.44	0.23–0.87	0.017	0.45	0.21–0.96	0.040
RPA1	rs4790838	G/A	377/276/64/717	435/410/83/928	Dominant	22	0.80	0.65–0.97	0.022	0.76	0.60–0.96	0.020
RPA2	rs7356	C/T	239/358/125/722	340/462/127/929	Additive	30	1.17	1.01–1.35	0.035	1.11	0.94–1.32	0.215
RPA2	rs6700473	A/G	243/351/125/719	346/459/124/929	Recessive	23	1.37	1.04–1.79	0.024	1.19	0.86–1.64	0.290
RPA3	rs3735461	G/A	215/369/137/721	248/463/217/928	Additive	27	0.86	0.75–0.99	0.032	0.91	0.77–1.08	0.293
RPA3	rs3735460	A/G	595/126/1/722	764/155/10/929	Recessive	38	0.13	0.02–1.00	0.050	0.14	0.02–1.15	0.067
RPA3	rs3757527	G/A	329/303/85/717	367/432/128/927	Dominant	13	0.77	0.63–0.94	0.011	0.77	0.61–0.98	0.032
RPA3	rs6974232	T/C	264/342/114/720	385/411/132/928	Dominant	36	1.22	1.00–1.50	0.047	1.14	0.90–1.44	0.290
RPA3	rs12668835	T/C	269/347/105/721	400/402/126/928	Dominant	19	1.27	1.04–1.55	0.018	1.18	0.94–1.50	0.158
RPA3	rs7783714	G/T	352/323/46/721	489/350/90/929	Recessive	17	0.64	0.44–0.92	0.016	0.71	0.46–1.09	0.118
RPA3	rs12538804	G/A	437/249/36/722	517/356/55/928	Dominant	37	0.82	0.67–1.00	0.050	0.84	0.66–1.06	0.137
RPA3	rs6974820	T/C	438/263/21/722	592/284/53/929	Recessive	11	0.50	0.30–0.83	0.008	0.53	0.29–0.97	0.038
RPA3	rs13227585	T/C	433/267/21/721	589/288/52/929	Recessive	12	0.51	0.30–0.85	0.010	0.52	0.28–0.96	0.035
RPA3	rs13246995	C/T	194/370/156/720	300/423/206/929	Dominant	20	1.29	1.04–1.60	0.019	1.25	0.97–1.61	0.085
Base excision repair pathway												
PARP1	rs8679	G/A	393/276/51/720	557/319/51/927	Additive	21	1.21	1.03–1.42	0.020	1.17	0.97–1.41	0.111

Gene	SNP	Alleles [*]	Case group [†] N _{total} =722	Control group [†] N _{total} =929	Best genetic model [±]	Rank [§]	Unadjusted			Adjusted [‡]		
							OR	95% CI	P-value	OR	95% CI	P-value
<i>PARP1</i>	rs747658	G/A	543/171/8/722	627/268/34/929	Additive	3	0.67	0.55–0.82	6.9×10 ⁻⁵	0.62	0.49–0.78	5.8×10 ⁻⁵
<i>PARP1</i>	rs3219090	T/C	332/331/56/719	397/417/115/929	Recessive	7	0.60	0.43–0.84	0.003	0.64	0.43–0.95	0.027
<i>PARP1</i>	rs3219073	C/G	543/170/8/721	627/268/34/929	Additive	2	0.67	0.55–0.81	5.9×10 ⁻⁵	0.61	0.48–0.77	4.6×10 ⁻⁵
<i>PARP1</i>	rs2048424	C/G	330/322/70/722	396/402/131/929	Recessive	10	0.65	0.48–0.89	0.007	0.72	0.50–1.04	0.078
<i>PARP1</i>	rs2666428	T/C	331/335/56/722	394/420/114/928	Recessive	8	0.60	0.43–0.84	0.003	0.64	0.43–0.95	0.027
<i>XRCC1</i>	rs1799782	T/C	682/362/720	839/83/6/928	Dominant	5	0.53	0.35–0.78	0.001	0.53	0.34–0.84	0.006
<i>XRCC1</i>	rs762507	T/C	241/317/163/721	320/437/172/929	Recessive	33	1.29	1.01–1.64	0.041	1.25	0.94–1.67	0.130
<i>XRCC1</i>	rs3213266	A/G	625/89/8/722	754/168/7/929	Dominant	9	0.67	0.51–0.88	0.004	0.67	0.49–0.93	0.017
<i>XRCC1</i>	rs3213255	G/A	219/318/184/721	319/441/168/928	Recessive	4	1.55	1.22–1.96	2.8×10 ⁻⁴	1.52	1.14–2.01	0.004
<i>Cell cycle control pathway</i>												
<i>CCND1</i>	rs603965	G/A	200/347/173/720	281/472/176/929	Recessive	15	1.35	1.07–1.71	0.012	1.25	0.94–1.66	0.120
<i>CCND1</i>	rs3918298	A/G	685/360/721	854/71/4/929	Dominant	16	0.60	0.40–0.90	0.014	0.53	0.33–0.87	0.011
<i>CCND1</i>	rs678653	C/G	304/317/100/721	346/452/130/928	Dominant	34	0.82	0.67–0.99	0.044	0.73	0.58–0.93	0.009
<i>CDKN2A</i>	rs11515	C/G	512/198/12/722	676/221/32/929	Recessive	25	0.47	0.24–0.93	0.029	0.44	0.19–1.02	0.056

^{*} Minor/Common alleles

[†] Common allele homozygotes/Heterozygotes/Minor allele homozygotes/Total genotyped

[±] OR, 95% CI, and P-value, according to best genetic model: Additive model – lung cancer risk in heterozygotes relative to common allele homozygotes or in minor allele homozygotes relative to heterozygotes; Dominant model – lung cancer risk in heterozygotes or minor allele homozygotes relative to common allele homozygotes; Recessive model – lung cancer risk in minor allele homozygotes relative to common allele homozygotes or heterozygotes

[§] Ranking by min 3p.

[‡] Adjusted for sex (two groups), age (four groups), year enrolled (three groups), and pack-year (four groups)

Table 3

Haplotypes associated with risk of lung cancer at $p < 0.01$. Underlines identify loci associated with lung cancer risk in single SNP analyses (Table 2). SNP labels identify common ("0") and minor ("1") alleles.

Gene	Haplotype*	Frequencies		Unadjusted			Adjusted [†]		
		Cases	Controls	OR	95% CI	P-value	OR	95% CI	P-value
ERCC2	00000000 <u>1</u> 111	0.236	0.224	reference			reference		
	111110000 <u>1</u> 00	0.006	0.017	0.29	0.11–0.74	0.010	0.31	0.11–0.90	0.031
GTF2H1	00000000 <u>1</u> 00	0.279	0.307	reference			reference		
	0100000000	0.21	0.185	1.31	1.07–1.60	0.009	1.38	1.10–1.73	0.005
GTF2H4	000000	0.432	0.414	reference			reference		
	111110	0.168	0.214	0.72	0.61–0.86	3.0×10 ^{−4}	0.70	0.58–0.85	4.0×10 ^{−4}
PARP1	<u>1</u> 000000000000	0.249	0.217	reference			reference		
	<u>0</u> 010100000000	0.171	0.196	0.74	0.60–0.92	0.006	0.74	0.59–0.94	0.013
XRCC1	<u>0</u> 10000101011	0.125	0.179	0.60	0.48–0.74	4.6×10 ^{−6}	0.55	0.43–0.71	5.4×10 ^{−6}
	<u>0</u> 001001000000	0.306	0.315	reference			reference		
CCND1	<u>1</u> 000000000000	0.024	0.048	0.47	0.32–0.70	1.6×10 ^{−4}	0.47	0.31–0.70	2.4×10 ^{−4}
	<u>0</u> 0000000000100	0.002	0.011	0.11	0.03–0.47	0.003	0.14	0.05–0.40	3.2×10 ^{−4}
	<u>0</u> 1000	0.458	0.413	reference			reference		
	10111	0.022	0.040	0.54	0.35–0.84	0.006	0.51	0.32–0.81	0.005

* SNP positions in haplotype

ERCC2: rs13181-rs1799790-rs238418-rs1052555-rs1799787-rs3916874-rs238414-rs50872-rs238405-rs238406-rs1618536-rs1799783

GTF2H1: rs4150550-rs4150563-rs4150606-rs4150610-rs4237721-rs4150642-rs4150655-rs4150667-rs4150673-rs4596

GTF2H4: rs2074510-rs1052693-rs2074509-rs2074508-rs3218815-rs1264307

PARP1: rs8679-rs747658-rs3219142-rs1136410-rs3219110-rs2027440-rs3219090-rs1805410-rs3219073-rs2117114-rs2048424-rs2666428

XRCC1: rs1799782-rs731420-rs762507-rs1001581-rs2854496-rs939461-rs12611088-rs3213282-rs2854510-rs2854509-rs3213266-rs2854506-rs3213255

CCND1: rs3862792-rs603965-rs3918298-rs649392-rs678653

[†] Adjusted for sex (two groups), age (four groups), year enrolled (three groups), and pack-year (four groups)

Table 4

Lung cancer case ($N_{\text{total}}=722$) and control ($N_{\text{total}}=929$) subjects distributed according to genotype risk summary score ^{*,†}.

Summary score	Cases, N	Controls, N	Unadjusted			Adjusted [±]		
			OR	95% CI	P-value	OR	95% CI	P-value
-8.5 to -2.5	107	259	Reference			Reference		
-2.0 to -1.0	90	210	1.04	0.74-1.45	0.830	0.96	0.65-1.41	0.820
-0.5 to 1.0	232	254	2.21	1.66-2.95	<.0001	2.21	1.57-3.10	<.0001
1.5 to 9.5	293	206	3.44	2.58-4.59	<.0001	3.12	2.23-4.38	<.0001

^{*}The genotype risk summary score used genotype results for 20 NER pathway SNPs (*ERCC1* rs3212955, *ERCC2* rs238405, *ERCC5* rs4150351, *GTF2H1* rs4150667, *GTF2H2* rs2340693, *GTF2H4* rs2074508, *LIG1* rs10500298, *RAD23B* rs6477569, *RPA1* rs7501623 and rs4790838, *RPA2* rs6700473, and *RPA3* rs3735461, rs3735460, rs3757527, rs6974232, rs12668835, rs7783714, rs12538804, rs6974820 and rs13246995), seven BER pathway SNPs (*PARP1* rs8679, rs3219090, and rs3219073 and *XRCC1* rs1799782, rs762507, rs3213266, and rs3213255), and four cell cycle pathway SNPs (*CCND1* rs603965, rs3918298, and rs678653 and *CDKN2A* rs11515).

[†]The genotype risk summary score summed SNP genotype results scored in the manner shown below.

Best model	Minor allele associated with higher risk			Minor allele associated with lower risk		
	Common allele homozygote	Heterozygote	Minor allele homozygote	Common allele homozygote	Heterozygote	Minor allele homozygote
Additive	0.0	0.5	1.0	0.0	-0.5	-1.0
Dominant	0.0	1.0	1.0	0.0	-1.0	-1.0
Recessive	0.0	0.0	1.0	0.0	0.0	-1.0

[±] Adjusted for sex (two groups), age (four groups), year enrolled (three groups), and pack-year (four groups)