Introduction

There will be an estimated 226,160 incident cases of lung cancer and 160,340 deaths in the United States in 2012 (1). Most are directly attributed to tobacco smoking. Tobacco smoke contains benzo[a]pyrene [B(a)P], a polycyclic aromatic hydrocarbon (PAH) compound that is a classic DNA-damaging carcinogen. Bioactivation of B(a)P in vivo generates highly toxic intermediates, such as B(a)P diol epoxide (BPDE) that can irreversibly damage DNA by forming DNA adducts through covalent binding or oxidation (2–5). BPDE–DNA adducts can block the transcription of an essential gene (6), if they are not repaired efficiently by the nucleotide excision repair (NER) pathway (7, 8).

Differential susceptibility to carcinogenesis is suggested by the fact that only a fraction of cigarette smokers develop smoking-related lung cancer (9, 10). This variation has been suggested to be due, in part, to the genetically determined variation in carcinogen metabolism (11, 12) and/or variability in DNA repair capacity (DRC; refs. 13, 14). We have previously reported that suboptimal DRC is a marker for genetic susceptibility to lung cancer (15–17), in which the DRC phenotype was measured in vitro in short-term cultured lymphocytes using the host-cell reactivation assay with a BPDE-treated reporter gene, chloramphenicol acetyltransferase (CAT). There is also a considerable interindividual variation in DRC, which is likely attributed to genetic variation in DNA repair genes (18, 19). Although polymorphisms in DNA repair genes have been reported to be associated with risk of lung cancer (20), genetic determinants of DRC remain to be identified.

We reported a genome-wide association study (GWAS) of histopathologically confirmed non–small cell lung cancer (NSCLC) with genotyping of 317,498 tagging single nucleotide polymorphisms (SNP) in a series of 1,154 ever-smoking lung cancer cases and 1,137 ever-smoking controls in a Texas population of self-reported European descent (21). In this study, we aimed to identify more comprehensively novel loci modulating the DNA repair phenotype. The DRC data were available for 1,774 individuals included in the published lung cancer GWAS that we used as the discovery phase to identify variants predicting the DRC phenotype. Comparable data were available for an additional 1,374 independent cases and controls for the replication phase.

Materials and Methods

Study populations

The study participants were patients with lung cancer and cancer-free healthy control subjects, who were U.S. residents of European ancestry. Between September 1995 and April 2008, patients with histopathologically confirmed lung cancer were accrued for an ongoing and previously described molecular
Table 1. Distribution of demographic characters in discovery and replication sets.

<table>
<thead>
<tr>
<th>Character</th>
<th>Combined (N = 3,148)</th>
<th>Discovery (N = 1,774)</th>
<th>Replication (N = 1,374)</th>
<th>Controls (N = 1,555)</th>
<th>Cases (N = 1,593)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 1,393)</td>
<td>(n = 914)</td>
<td>(n = 695)</td>
<td>(n = 695)</td>
<td>(n = 324)</td>
</tr>
<tr>
<td>Age, y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>31–88</td>
<td>20–188</td>
<td>20–188</td>
<td>35 (29–90)</td>
<td>35 (29–90)</td>
</tr>
<tr>
<td>Sex, (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>986 (55.6)</td>
<td>623 (56.4)</td>
<td>623 (54.3)</td>
<td>389 (54.2)</td>
<td>471 (54.8)</td>
</tr>
<tr>
<td>Female</td>
<td>788 (44.4)</td>
<td>471 (45.6)</td>
<td>302 (45.7)</td>
<td>389 (45.8)</td>
<td>427 (45.2)</td>
</tr>
<tr>
<td>Smoking status, N (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Former</td>
<td>983 (54.3)</td>
<td>586 (46.2)</td>
<td>304 (43.7)</td>
<td>396 (49.0)</td>
<td>389 (42.3)</td>
</tr>
<tr>
<td>Current</td>
<td>811 (45.7)</td>
<td>486 (53.8)</td>
<td>347 (56.3)</td>
<td>390 (50.8)</td>
<td>447 (57.7)</td>
</tr>
</tbody>
</table>

*Six cases and two controls in the replication dataset missed pack-year values.

host-cell reactivation assay

DRC was measured in cultured peripheral lymphocytes using the host-cell reactivation assay with a reporter gene damaged by BPDE (16). Briefly, the assay uses a BPDE-damaged nonreplicating recombinant plasmid (pCMVcat) harboring a CAT reporter gene used in the transfection. Because even a single unrepaired BPDE-DNA adduct can block the CAT transcription, any measurable CAT activity will reflect the ability of the transfected cells to remove BPDE-induced DNA adducts from the plasmids. Before the transfection, the cultured T lymphocytes from peripheral blood are stimulated by phytohemagglutinin so that they can uptake the plasmids. Duplicate transfections with either untreated plasmids or BPDE-treated plasmids are always conducted in parallel. The CAT activity is quantified by adding chloramphenicol and [3H]acetyl-CoA to measure the production of [3H]monoaeylated and [3H]diacylated chloramphenicol with a scintillation counter. DRC is reported as the ratio of the radioactivity of cells transfected with the treated plasmids to that of cells transfected with the untreated plasmids. Assuming that the transfection efficiencies of BPDE-treated and untreated plasmids are equal (22), this ratio reflects the percentage of damaged CAT reporter genes repaired within lymphocytes transfected with the BPDE-treated plasmids.

Genotyping

Genotyping for the discovery set was conducted using the Illumina HumanHap300 BeadChip. We retained 303,669 autosomal SNPs after conducting quality control for the 1,774 subjects using PLINK to exclude SNPs with a call rate less than 95%, with minor allele frequency less than 0.01, and with deviations from Hardy–Weinberg equilibrium (P < 0.0001;
ref. 21). Genotyping for the replication set was conducted using Illumina BeadXpress 384-plex on the 1,374 subjects. The SNPs had a design score more than 0.7 and the calling rate was more than 99%.

**Statistical analysis**

DRC was analyzed as a continuous variable. Student t test was used to compare the differences in DRC between cases and controls in the discovery, replication, and combined datasets. Logistic regression models were used to calculate crude and adjusted ORs and confidence intervals (CI). The median DRC of control subjects was used as the cutoff value; values more than the median were considered to be high DRC and values below the median were considered to be low/suboptimal DRC. The quartiles of DRC in control subjects were used to calculate the DRC dose-effect on lung cancer risk. Adjusted ORs were calculated by fitting unconditional multivariate logistic regression models with adjustment for age, sex, pack-years of smoking, and DRC assay-related variables including blastogenic rate (after the phytohemagglutinin stimulation), cell storage time (the difference between the date when the DRC assay was conducted and date of blood collection), baseline CAT expression levels (for the undamaged plasmids), and assay conducting dates. Interactions between DRC and smoking status/sex/genotypes were tested by using standard unconditional logistic regression models. A more-than-multiplicative interaction was suggested when OR\(_{11}\) > OR\(_{10}\) × OR\(_{01}\), in which OR\(_{11}\) = the OR when both factors were present, OR\(_{01}\) = the OR when only factor 1 was present, and OR\(_{10}\) = the OR when only factor 2 was present. Similarly, a more-than-additive interaction was indicated if OR\(_{11}\) > OR\(_{10}\) + OR\(_{01}\) − 1. First, we evaluated the interactions indicating a more-than-multiplicative effect; when the test for multiplicative interaction was not rejected, further tests for additive interactions were conducted by a bootstrapping test of goodness of fit of the null hypothesis of an additive model with no interaction against an alternative hypothesis that permits an additive interaction. To conduct the hypothesis test for additive models, we implemented bootstrapping by using STATA software (version 10.1, STATA Corporation).

In the DRC phenotype and genotype correlation analysis, a generalized linear model was used with SNPs as predictors and DRC (a continuous variable) as the outcome with adjustment for case–control status in addition to age, sex, pack-years of smoking, and DRC assay-related variables. In the discovery phase, we compared the DRC by genotypes of each autosomal SNP using an additive model. We selected a total of 384 SNPs of the significant findings from the top list of genome-wide scanning (\(P < 10^{-5}\)) and genes involved in the DNA repair pathways from the discovery set (\(P < 0.05\)) for the replication phase. To summarize the results for the discovery set and the replication set, meta-analyses for the most significant findings in the discovery set that were replicated at a \(P < 0.05\) were conducted. The analyses were done with PLINK (version 1.07; ref. 23), STATA software, and SAS 9.2 (SAS Institute Inc.).

The effect of population substructure was assessed by using the principal component analysis in the parent lung cancer GWAS and found to be minimal because there was no evidence of genome-wide inflation of the \(\chi^2\) tests that are expected to arise in the presence of population substructure (21). The linkage disequilibrium (LD) structure of the neighboring region containing loci associated with DRC was inferred by Haploview software (v4.1; ref. 24). The screenshot of all the known genes in the region were obtained from University of California San Francisco (UCSC) genome browser (25).

**Results**

The select characteristics of the study subjects in the discovery, replication, and combined sets as well as stratified by case–control status are shown in Table 1. The mean age was 62 years for both phases with ranges of 31 to 88 and 26 to 90, respectively. The distribution of patients with lung cancer and control subjects between the discovery and replication sets were similar (51.5% vs. 49.4% for cases and 48.5% vs. 50.6% for controls). However, there were more men in the discovery set than those in the replication set (55.6% vs. 45.3%) because only ever smokers were included in the discovery set and men were more likely to be smokers. In the replication set, 586 (42.6%) never smokers were included; therefore, the replication set had a median 13 pack-year smoking history (range 0–189) compared with 42 (range 0.05–294) in the discovery set (Table 1).

We first evaluated the DRC distribution and its association with lung cancer risk in the 3 datasets. In the discovery phase, as shown in Table 2, when DRC was analyzed as a continuous variable, the mean DRC was 8.33% with a range of 2.68 to 18.70% in 914 case patients and 8.88% (range, 2.09–19.91%) in 860 control subjects, representing an average 6% reduction in DRC in patients with lung cancer. Case patients in all subgroup strata consistently exhibited significantly lower mean DRC than did the control subjects (Table 2). As we have previously shown, compared with men, women had significantly lower mean DRC among both case patients (\(P = 0.002\)) and control subjects (\(P = 0.027\)). Compared with former smokers, current smokers exhibited a higher mean DRC, especially in case patients (\(P = 0.038\)). We further evaluated the effect of DRC on risk for lung cancer by logistic regression analysis. As shown in Table 2, when DRC was fit in the model as a continuous predictor variable, without or with adjustment for age, sex, pack-years of smoking, blastogenic rate, cell storage time, baseline CAT expression, and DRC assay dates, the crude and adjusted ORs for lung cancer risk (per one DRC unit decrease) were similar and statistically significantly elevated [\(OR = 1.07 (95\% CI = 1.04–1.11)\) and \(OR = 1.08 (95\% CI = 1.04–1.11)\), respectively]. When DRC values were dichotomized by the median DRC of the control subjects, the crude OR for case status associated with low DRC was 1.32 (95% CI = 1.09–1.59) and the adjusted OR associated with low DRC was 1.29 (95% CI = 1.06–1.56). When the DRC values were further divided by quartile of DRC in control subjects, it was again evident that decreased DRC was associated with increased risk in a dose-dependent manner. By use of the highest quartile of the DRC as the reference, the crude ORs for DRC values lower than values in the 75th, 50th, and 25th were 1.45 (95% CI = 1.10–1.91), 1.48 (95% CI = 1.13–1.94), and 1.74 (95% CI = 1.33–2.27), respectively. The adjusted ORs were nearly identical to the crude ORs [\(OR = 1.48 (95\% CI = 1.12–1.97)\) vs. \(OR = 1.45 (95\% CI = 1.12–1.97)\), respectively].
Table 2. Distribution of DRC between cases and controls and its association with lung cancer in discovery and replication datasets

<table>
<thead>
<tr>
<th>DRC (%)</th>
<th>Discovery (n = 1,774)</th>
<th>Replication (n = 1,374)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Case  (n = 914)</td>
<td>Control  (n = 860)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuous range</td>
<td>2.68–18.70</td>
<td>2.09–19.91</td>
</tr>
<tr>
<td>Mean (95% CI)</td>
<td>8.33 (8.16–8.51)</td>
<td>8.88 (8.69–9.07)</td>
</tr>
<tr>
<td>Overall</td>
<td>8.33 (8.16–8.51)</td>
<td>8.88 (8.69–9.07)</td>
</tr>
<tr>
<td>Male</td>
<td>8.57 (8.34–8.80)</td>
<td>9.08 (8.81–9.35)</td>
</tr>
<tr>
<td>Female</td>
<td>8.02 (7.76–8.28)</td>
<td>8.65 (8.38–8.91)</td>
</tr>
<tr>
<td>By smoking status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Former</td>
<td>8.15 (7.91–8.39)</td>
<td>8.77 (8.51–9.03)</td>
</tr>
<tr>
<td>Current</td>
<td>8.52 (8.26–8.78)</td>
<td>9.04 (8.75–9.32)</td>
</tr>
<tr>
<td>Dichotomized (by median)c</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>&gt;8.6</td>
<td>393 (43.0)</td>
<td>429 (49.9)</td>
</tr>
<tr>
<td>≤8.6</td>
<td>521 (57.0)</td>
<td>431 (50.1)</td>
</tr>
<tr>
<td>Quartiles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;10.7</td>
<td>167 (18.3)</td>
<td>222 (25.8)</td>
</tr>
<tr>
<td>8.7–10.7</td>
<td>226 (24.7)</td>
<td>207 (24.1)</td>
</tr>
<tr>
<td>6.9–8.6</td>
<td>243 (26.6)</td>
<td>218 (25.3)</td>
</tr>
<tr>
<td>≤6.8</td>
<td>278 (30.4)</td>
<td>213 (24.8)</td>
</tr>
<tr>
<td>Trend testb</td>
<td>0.0001</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Abbreviations: DRC, DNA repair capacity; OR, odds ratio; CI, confidence interval; NA, not applicable.

aTwo-sided Student t test or chi-square tests.

bAdjusted in a logistic regression including age (years), sex, pack-year, baseline chloramphenicol acetyltransferase (CAT) activity, blastogenic rate (%), cell storage time (months), and DRC assay dates. Six cases and two controls in the replication dataset missed pack-year values.

cMedian or quartiles of the control subjects’ values.
1.10–1.92), and OR = 1.74 (95% CI = 1.31–2.30), respectively). This trend of an increasing risk with a decreasing DRC was statistically significant for both the crude and adjusted ORs ($P = 0.001$ and $P = 0.004$, respectively).

In the independent replication set from the same study source, as listed in Table 2, the mean DRC was 8.40% with a range of 2.80% to 18.19% in 679 case patients and 8.91% (range, 3.38–18.53%) in 695 control subjects. The stratification analysis on DRC by sex and smoking status also showed consistently lower DRC levels in case patients than those in control subjects (Table 2). However, the differences were statistically significant only in women and never smokers. The trends of DRC among strata by sex and smoking status in case patients and control subjects were similar to those in the discovery set. Never smokers exhibited the lowest DRC, and the significant differences in subgroups were more obvious in case patients ($P = 0.003$ among smoking status and $P = 0.001$ for female vs. male). The association of DRC with risk for lung cancer in 3 logistic regression models with DRC as a continuous variable, dichotomized, or quartiles in the replication set was almost the same as observed in the discovery phase (Table 2).

In the combined dataset of 3,148 study participants (1,593 patients with lung cancer and 1,555 cancer-free controls), the differences in overall and stratified DRC levels between cases and controls, as well as the association of DRC for 3 categories (i.e., continuous, dichotomized, and quartiles) with lung cancer risk were stably consistent with those found in both discovery and replication phases, but the combined results had much narrower CIs (Table 3). As we found and reported in previous relatively small studies (16, 17), smoking seems to upregulate the DRC levels for BPDE-induced DNA damage in both cases and controls but more obviously in the cases, and women seem to have a lower DRC than men, as confirmed in the 2-stage and combined analysis; therefore, we further conducted the tests for interaction between DRC status and smoking status/sex as shown in Supplementary Table S1. Although we observed some significant trends between DRC status by smoking status/sex, we did not find evidence for any multiplicative/additive interaction in this analysis.

We then evaluated the DRC phenotype and genotype correlations in the 3 datasets. In the discovery set, we tested the association of 303,669 autosomal SNPs from the GWAS using Illumina HumanHap300 BeadChip that remained after quality control with DRC in 1,774 subjects. Supplementary Figure S1 shows the Manhattan plot for the GWAS analysis of the DRC phenotype. There was no evidence of a systematic inflation of $P$ values (genomic inflation factor $\lambda = 0.9715$). We further adjusted for residual population structure using the top 5 principal components derived by using the Golden Helix software. The associations observed after the adjustment were

![Table 3. Combined analysis on the association of DRC and lung cancer](data:image/png;base64,iVBORw0KGgoAAAANSUhEUgAAAoAAAAHgCAYAAAA5i8U2AAAAmpDAREH5LZ75BAAAABJRU5ErkJggg==)

Abbreviations: DRC, DNA repair capacity; OR, odds ratio; CI, confidence interval.

*Two-sided Student t test or chi-square tests.

*Adjusted in a logistic regression including age (years), sex, pack-year, baseline chloramphenicol acetyltransferase (CAT) activity, blastogenic rate (%), cell storage time (months), and DRC assay dates. Six cases and two controls in the replication dataset missed pack-year values.

*Median or quartiles of the control subjects’ values.
Table 4. SNPs with the strongest effects on DRC from the joint analysis

<table>
<thead>
<tr>
<th>Gene</th>
<th>Symbol</th>
<th>Chr. Location</th>
<th>Genotype*</th>
<th>MAF</th>
<th>Replication</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>A1/A2</td>
<td>GWAS</td>
<td>Replication</td>
<td>GWAS Replication</td>
</tr>
<tr>
<td>rs13181</td>
<td>ERCC2/XPD</td>
<td>19 Coding</td>
<td>CA</td>
<td>0.3861</td>
<td>0.3167</td>
<td>0.3613</td>
</tr>
<tr>
<td>rs9390123</td>
<td>PHACTR2</td>
<td>6 Intron</td>
<td>AG</td>
<td>0.3957</td>
<td>0.3681</td>
<td>0.3873</td>
</tr>
<tr>
<td>rs7443927</td>
<td>DUSP1</td>
<td>5 Flanking_3UTR</td>
<td>AC</td>
<td>0.1088</td>
<td>0.1159</td>
<td>0.1232</td>
</tr>
</tbody>
</table>

Abbreviation: MAF, minor allele (A1) frequency.

Wild-type homozygote/heterozygote/variant homozygote; 2 were missing for rs13181, 3 were missing for rs9390123, and 2 were missing for rs7443927.

Similar in strength to the unadjusted results, suggesting a minimal effect of population substructure (data not shown).

SNPs with a $P < 10^{-5}$ level in the discovery set were selected for replication in 1,374 individuals with DRC data available. A total of 319 SNPs from the discovery set were found to be significantly associated with the DRC phenotype according to this criterion. We augmented this list with an additional 65 SNPs from candidate genes in the NER pathway (Supplementary Table S2) that were nominally significant ($P < 0.05$) in the GWAS. Genotyping for the replication was conducted using the Illumina BeadXpress 384-plex platform.

As shown in Table 4, rs13181, a coding SNP in the well-known NER gene ERCC2/XPD [xeroderma pigmentosum, complementation group D (MIM 278730)], showed the strongest evidence of association with DRC ($P_{\text{discovery}} = 9.08 \times 10^{-7}$, $P_{\text{replication}} = 0.025$). The second strongest association was observed for rs9390123 in the PHACTR2 [phosphatase and actin regulator 2 (MIM 608724)] gene on chromosome 6 ($P_{\text{discovery}} = 6.68 \times 10^{-5}$, $P_{\text{replication}} = 2.5 \times 10^{-5}$). SNP rs7443927 in the DUSP1 [dual-specificity phosphatase 1 (MIM 600714)] gene on chromosome 5 also showed a consistent association with DRC in both the phases ($P_{\text{discovery}} = 1.76 \times 10^{-4}$, $P_{\text{replication}} = 0.0002$). There was no significant evidence of heterogeneity among the discovery and replication sets (Table 4).

All 3 SNPs are located in regions of low LD, with HapMap database SNPs being in weak to moderate LD ($r^2 < 0.6$). We further conducted imputations for SNPs in the HapMap database within the vicinity of these significant SNPs ($\pm 100$ kb for rs13181 and rs7443927; $\pm 200$ kb for rs9390123 according to LD values) using PLINK (23) and found that these SNPs remained the most significant in the corresponding regions (Supplementary Figs. S2–S4, 3 SNPs with LD plots and $P$ values).

We also conducted a sensitivity analysis, determining the effect of these 3 SNPs in the controls only (Supplementary Table S3). The SNP rs9390123 did not replicate in control subjects, with an (nonsignificant) opposite trend to that in the discovery set, whereas the results for the other 2 SNPs were consistent in control subjects and the whole sample. It could be hypothesized that rs9390123 is relevant for ever-smokers only. Indeed, if only ever-smokers were included in the replication, the trend was the same as in the discovery set, although the replication $P$ value was not significant ($P = 0.2$). Overall, the evidence for association for this SNP is weaker than the other two.

The modification effects of these 3 SNPs on the DRC levels, overall and in ever and never smokers are shown in Table 5. Across all strata, the variant homozygotes of the 3 SNPs exhibited the lowest DRC and the wild-type homozygotes had the highest DRC levels. The trends were statistically significant ($P \leq 0.0001$) in 3,148 study participants and in 2,562 ever-smokers. However, in 586 never smokers, the significant modification effects of these three SNPs on DRC levels was only evident in the ERCC2/XPD rs13181 ($P = 0.0002$). We also examined the modification effects of SNPs on DRC-associated lung cancer risk as presented in Supplementary Table S4. Although we observed some differences in ORs by genotypes, in additional analysis, we did not find evidence of an interaction between the 3 SNPs and DRC status on lung cancer risk in...
The transition of the SNP causes an amino acid change from lysine to glutamine at codon 751 (Lys751Gln). Although Lys751Gln does not reside in a known helicase/ATPase domain (29), it is at an amino acid residue identical in human, mouse, hamster, and fish. The ERCC2/XPD gene is highly conserved in eukaryotes; the amino acid substitution suggests a functional relevance for such a highly evolutionary conserved sequence. A recent meta-analysis of ERCC2/XPD Lys751Gln polymorphism and lung cancer risk from 22 case–control studies showed that C allele was associated with significantly increased risk of lung cancer among Caucasians, especially in smokers (30, 31); whereas we did not find the same trend in the current analysis. However, the evident association between rs13181 and DRC suggests that this SNP is an important predictor for DRC, especially in never smokers who may have some limited exposure but are genetically susceptible to cancer, compared with ever smokers.

Table 5. Modification effects of SNPs on DRC levels, overall and by smoking statusa

<table>
<thead>
<tr>
<th>Gene/SNP</th>
<th>Genotype</th>
<th>n</th>
<th>DRC (%) Mean (95% CI)</th>
<th>Trend test Pb</th>
<th>n</th>
<th>DRC (%) Mean (95% CI)</th>
<th>Trend test Pb</th>
<th>n</th>
<th>DRC (%) Mean (95% CI)</th>
<th>Trend test Pb</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERCC2/XPD</td>
<td>AA</td>
<td>1,290</td>
<td>8.80 (8.65–8.94)</td>
<td>&lt;0.0001</td>
<td>238</td>
<td>8.72 (8.41–9.04)</td>
<td>&lt;0.0001</td>
<td>1,052</td>
<td>8.81 (8.65–8.98)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>/rs13181</td>
<td>AC</td>
<td>1,420</td>
<td>8.61 (8.47–8.75)</td>
<td></td>
<td>264</td>
<td>8.27 (7.97–8.57)</td>
<td></td>
<td>1,156</td>
<td>8.69 (8.53–8.85)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>436</td>
<td>8.13 (7.88–8.39)</td>
<td></td>
<td>83</td>
<td>7.43 (6.90–7.96)</td>
<td></td>
<td>353</td>
<td>8.30 (8.01–8.58)</td>
<td></td>
</tr>
<tr>
<td>/rs9390123</td>
<td>AG</td>
<td>1,510</td>
<td>8.57 (8.43–8.70)</td>
<td></td>
<td>276</td>
<td>8.29 (7.99–8.58)</td>
<td></td>
<td>1,234</td>
<td>8.63 (8.48–8.78)</td>
<td></td>
</tr>
<tr>
<td>DUSP1</td>
<td>CC</td>
<td>2,483</td>
<td>8.72 (8.61–8.83)</td>
<td>&lt;0.0001</td>
<td>466</td>
<td>8.41 (8.18–8.63)</td>
<td>&lt;0.0001</td>
<td>2,017</td>
<td>8.79 (8.67–8.91)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>/rs7443927</td>
<td>AC</td>
<td>622</td>
<td>8.30 (8.08–8.51)</td>
<td></td>
<td>114</td>
<td>8.10 (7.64–8.56)</td>
<td></td>
<td>508</td>
<td>8.34 (8.10–8.58)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>41</td>
<td>7.61 (6.78–8.43)</td>
<td></td>
<td>6</td>
<td>8.07 (6.06–10.08)</td>
<td></td>
<td>35</td>
<td>7.53 (6.62–8.43)</td>
<td></td>
</tr>
</tbody>
</table>

aTwo were missing for rs13181, 3 were missing for rs9390123, and 2 were missing for rs7443927.
bANOVA test.

Discussion

In this two-phase replication study of DRC phenotype–genotype analysis, we confirmed our previous findings that low DRC was associated with significantly increased risk of lung cancer, and we then identified and replicated a number of SNPs associated with the DRC phenotype. The non synonymous SNP rs13181 in the ERCC2/XPD gene was found to be the most significant genetic variant predicting the DRC for removing BPDE-DNA adducts. Although it did not reach the genome-wise significance level after the joint analysis of two-stage data of the study, we thought that this finding is biologically plausible.

The ERCC2/XPD gene is located on 19q13 and is a core gene in the NER pathway. It encodes a protein that functions as an ATP-dependent 5‘-3‘ helicase (27) within the basal transcription factor IHH complex that removes bulky DNA adducts by excising a 24–32 nucleotide single-strand oligomer (27, 28). The A>C base transition of the SNP causes an amino acid change from lysine to glutamine at codon 751 (Lys751Gln). Although Lys751Gln does not reside in a known helicase/ATPase domain (29), it is at an amino acid residue identical in human, mouse, hamster, and fish. The ERCC2/XPD gene is highly conserved in eukaryotes; the amino acid substitution suggests a functional relevance for such a highly evolutionary conserved sequence. A recent meta-analysis of ERCC2/XPD Lys751Gln polymorphism and lung cancer risk from 22 case–control studies showed that C allele was associated with significantly increased risk of lung cancer among Caucasians, especially in smokers (30, 31); whereas we did not find the same trend in the current analysis. However, the evident association between rs13181 and DRC suggests that this SNP is an important predictor for DRC, especially in never smokers who may have some limited exposure but are genetically susceptible to cancer, compared with ever smokers.

The second most significant SNP rs9390123 is located in an intron in PHACTR2 (phosphatase and actin regulator 2) on 6q4. This gene encodes the protein phosphatase and actin regulator 2 that belongs to the PHACTR family containing 4 members (PHACTR1–4), which are abundantly expressed in the nervous system (32). Even though little is known of the proteins’ function, they are suggested to regulate protein phosphatase 1 and to bind to cytoplasmic actin. It has been reported that the intron SNP rs11155313 is associated with risk of Parkinson’s disease (33), and our finding of an association between the SNP rs9390123 (not in LD with any other SNPs based on the current HapMap data) and the DRC phenotype may reveal a new chapter for the PHACTR2 significance.

The rs7443927 was found to be the next significant SNP associated with DRC. This SNP is located in the 3‘ untranslated flanking region of the DUSP1 gene that encodes the dual specificity protein phosphatase 1. The structural features of this protein are similar to members of the nonreceptor type protein–tyrosine phosphatase family. It suppresses the activation of mitogen-activated protein kinase (MAPK) by...
oncogenic ras in extracts of Xenopus oocytes. The gene expression can be induced in human skin fibroblasts by oxidative/heat stress and growth factors. Therefore, DUSP1 may play an important role in the human cellular response to environmental stress as well as in the negative regulation of cellular proliferation (34, 35). The finding in the present study may suggest a new role of this gene in the DNA repair mechanism. The observed correlation between rs9390123 or rs7443927 and DRC suggest that these SNPs may be important predictors for DRC, especially in ever smokers. Intriguingly, their modification effects on DRC-associated lung cancer risk were greater in never smokers than in ever smokers, but no interaction was observed, suggesting that larger studies are warranted to further substantiate this observed difference and its biologic relevance.

Several DNA repair phenotypes have been reported as biomarkers for cancer susceptibility (36–40) and currently available cellular DRC assays with their characteristics have been comprehensively reviewed recently (41). In a series of case–control studies, we have shown that the DRC phenotype, measured in vitro in short-term cultured lymphocytes using the host-cell reactivation assay with a BPDE-treated reporter gene, predicts the risk of developing smoking-related cancers, including lung and head and neck cancer (16, 17, 42). We have also conducted a pilot study using the dimethyl sulfate as a substitute to create alkylating damage for NNK [nicotine-derived nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanoine], a tobacco-specific nitrosamine for lung adenoma and adenocarcinoma. In that study (43), we have reported that DRC for N7- and O6-guanine lesion may have different repair mechanisms because we did not find a correlation between DRC for those non-BPDE-induced adducts and DRC for BPDE-induced damage in 48 patients with lung cancer and 45 cancer-free controls; however, the data did show that simultaneously measuring DRC of these 2 distinct repair pathways would enhance the risk assessment for lung adenocarcinoma. In previously published, relatively small studies, we have also shown that some genetic variants of the DNA repair genes such as XPC, ERCC2/XPD, ERCC5/XPG may be predictors of the DRC phenotype for UV-induced DNA adducts in addition to BPDE-DNA adducts (18, 19, 44–46). For example, in an independent analysis of 333 cancer-free controls, 146 basal cell carcinoma, and 109 squamous cell carcinoma patients, we evaluated the associations between genotypes of the 5 common functional or nonsynonymous SNPs in NER/XP genes [i.e., XPC Ala499Val (C>T, rs2228000), XPC Lys939Gln (A>C, rs2228001), ERCC2/XPD Asp312Asn (G>A, rs179793), ERCC2/XPD Lys751Gln (A>C, rs13181), and ERCC5/XPG His1104Asp (G>C, rs17655)], and we found that all homozygotes of XP minor alleles had the lowest DRC (except for XPG, in the controls and basal cell carcinoma patients) compared with those who had the genotypes of common alleles in the same group. The increased number of variant homozygotes was associated with decreased DRC in a dose–response manner for all groups (44). Other research groups have also reported the correlations between DNA repair genotypes and phenotypes in NER pathway, DNA strand break repair pathway, and their associations with risk of breast or prostate cancer (47–49). However, those studies had either a small number of subjects or investigated only a few candidate SNPs.

In this comprehensive analysis of GWAS data of the DRC phenotype and genotypes with a much large sample size, we have confirmed the dominant role of genetic variation of ERCC2/XPD in predicting the DRC phenotype. However, the SNP rs13181 itself did not predict the risk of lung cancer as the DRC phenotype did, and the genotype–phenotype correlation with DRC was lower for ever smoking patients with lung cancer, suggesting the likelihood of additional genetic variations in predicting cancer risk in smokers in whom the exposure may overwhelm any genotype. Furthermore, in the Illumina HumanHap300 BeadChip, only 3 SNPs out of 22 common (10 tagging) SNPs in ERCC2/XPD (based on the current dbSNP information) were included and only rs13181 reached \( P < 0.05 \) and was replicated in predicting DRC. Therefore, further investigations on all polymorphisms of ERCC2/XPD in predicting DRC, especially in never smokers, are warranted. This is because the genetic variants are amenable to high-throughput analysis, less labor-intensive to measure, and less subjected to misclassification than measurements of the DRC phenotype in molecular epidemiologic studies.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions

Acknowledgments
The authors thank all individuals for their participation in this study.

Grant Support
The study was supported in part by NIH grants (R01CA127219 and R01CA055769 to M.R. Spitz; R01CA121197 and U19CA148127 to C.I. Amos; R01ES011740 and R01CA131274 to Q. Wei; R01CA149462 to O.Y. Gorlova; and P01CA016672 to M.D. Anderson Cancer Center). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 16, 2012; revised September 11, 2012; accepted October 17, 2012; published OnlineFirst October 29, 2012.

References
2. Li D, Firozi PF, Wang LE, Bosken CH, Spitz MR, Hong WK, et al. Sensitivity to DNA damage induced by benzo(a)pyrene diol epoxide...
DNA Repair Phenotype and Genetic Determinants in Lung Cancer


Published OnlineFirst October 29, 2012; DOI: 10.1158/0008-5472.CAN-12-1915
Genome-Wide Association Study Reveals Novel Genetic Determinants of DNA Repair Capacity in Lung Cancer

Li-E Wang, Olga Y. Gorlova, Jun Ying, et al.

Cancer Res  Published OnlineFirst October 29, 2012.

Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-12-1915

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2012/10/30/0008-5472.CAN-12-1915.DC1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.