

Genome-Wide Association Study Reveals Novel Genetic Determinants of DNA Repair Capacity in Lung Cancer

Li-E Wang¹, Olga Y. Gorlova¹, Jun Ying¹, Yawei Qiao¹, Shih-Feng Weng¹, Annette T. Lee⁴, Peter K. Gregersen⁴, Margaret R. Spitz³, Christopher I. Amos^{1,2}, and Qingyi Wei¹

Abstract

Suboptimal cellular DNA repair capacity (DRC) has been shown to be associated with enhanced cancer risk, but genetic variants affecting the DRC phenotype have not been comprehensively investigated. In this study, with the available DRC phenotype data, we analyzed correlations between the DRC phenotype and genotypes detected by the Illumina 317K platform in 1,774 individuals of European ancestry from a Texas lung cancer genome-wide association study. The discovery phase was followed by a replication in an independent set of 1,374 cases and controls of European ancestry. We applied a generalized linear model with single nucleotide polymorphisms as predictors and DRC (a continuous variable) as the outcome. Covariates of age, sex, pack-years of smoking, DRC assay-related variables, and case-control status of the study participants were adjusted in the model. We validated that reduced DRC was associated with an increased risk of lung cancer in both independent datasets. Several suggestive loci that contributed to the DRC phenotype were defined in *ERCC2/XPD*, *PHACTR2*, and *DUSP1*. In summary, we determined that DRC is an independent risk factor for lung cancer, and we defined several genetic loci contributing to DRC phenotype. *Cancer Res*; 73(1); 256–64. ©2012 AACR.

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

Authors' Affiliations: Departments of ¹Epidemiology and ²Genetics, The University of Texas M.D. Anderson Cancer Center; ³The Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas; and ⁴The Feinstein Institute for Medical Research Center for Genomics and Human Genetics, Manhasset, New York

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

Current address for C.I. Amos: Department of Community and Family Medicine, Geisel College of Medicine, Dartmouth College, Hanover, NH

L.-E. Wang and O.Y. Gorlova contributed equally to this work.

Corresponding Author: Qingyi Wei, Department of Epidemiology, Unit 1365, The University of Texas MD Anderson Cancer Center, 1155 Pressler Boulevard, Houston, TX 77030. Phone: 713-745-2481; Fax: 713-563-0999; E-mail: qwei@mdanderson.org

doi: 10.1158/0008-5472.CAN-12-1915

©2012 American Association for Cancer Research.

Table 1. Distribution of demographic characters in discovery

	Discovery (n = 1,774)		Replication (n = 1,374)		Combined (N = 3,148)		
	Cases (n = 914)	Controls (n = 860)	Cases (n = 679)	Controls (n = 695)	All	Cases (n = 1,593)	Controls (n = 1,555)
Age, y, median (range)	62 (31–88)	62 (32–86)	63 (26–90)	62 (28–90)	62 (26–90)	62 (26–88)	62 (28–90)
Pack-year ^a , median (range)	46 (0.1–210)	38 (0.05–294)	20 (0–188)	10 (0–189)	35 (0–294)	40 (0–210)	30 (0–294)
Case-control status, (%)	(51.5)	(48.5)	(49.4)	(50.6)	(100)	(50.6)	(49.4)
Sex, N (%)							
Male	986 (55.6)	471 (54.8)	302 (44.5)	321 (46.2)	1,609 (51.1)	817 (51.3)	792 (50.9)
Female	788 (44.4)	389 (45.2)	377 (55.5)	374 (53.8)	1,539 (48.9)	776 (48.7)	763 (49.1)
Smoking status, N (%)							
Never	0	0	282 (41.5)	304 (43.7)	586 (18.6)	282 (17.7)	304 (19.5)
Former	963 (54.3)	496 (57.7)	201 (29.6)	197 (28.4)	1,361 (43.2)	668 (41.9)	693 (44.6)
Current	811 (45.7)	364 (42.3)	196 (28.9)	194 (27.9)	1,201 (38.2)	643 (40.4)	558 (35.9)

^aSix cases and 2 controls in the replication dataset missed pack-year values.

epidemiologic study (21) on susceptibility markers for lung cancer at The University of Texas M.D. Anderson Cancer Center (Houston, TX). There were no age, sex, or stage restrictions. Healthy controls without a previous diagnosis of cancer (except for nonmelanoma skin cancer) were recruited from the Kelsey-Seybold Clinics, Houston's largest private multispecialty physician group that has a network of 23 clinics and more than 300 physicians. This National Institute of Cancer-funded research was approved by the M.D. Anderson Cancer Center and Kelsey-Seybold Institutional Review Boards (Houston, TX). The discovery-phase subjects, all ever smokers, were used in the primary analysis from a published GWAS of lung cancer (21). For the replication phase, we included subjects from the same study source, who were not included in the GWAS, and for whom comparable data on DRC were available. Unlike the discovery population, the replication-phase subjects included never smokers, defined as those who smoked less than 100 cigarettes in their lifetimes. After data quality control, DRC data and complete demographic information were available for 1,774 individuals in the discovery set (914 NSCLC cases and 860 controls) and for 1,374 individuals in the replication set (679 cases and 695 controls; Table 1). Informed consent had been obtained from all study participants before the collection of epidemiologic data and blood samples by trained MD Anderson staff interviewers.

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]monoacetylated and [³H]diacetylated chloramphenicols 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 $OR_{11} > OR_{10} \times OR_{01}$, 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 ($P < 10^{-3}$) 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 χ^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], OR = 1.45 (95% CI =

Table 2. Distribution of DRC between cases and controls and its association with lung cancer in discovery and replication datasets

DRC (%)	Discovery (n = 1,774)				Replication (n = 1,374)				
	Case (n = 914)	Control (n = 860)	P ^a	Adjusted ^b OR (95% CI)	Case (n = 679)	Control (n = 695)	P ^a	Crude OR (95% CI)	Adjusted ^b OR (95% CI)
Continuous range	2.68–18.70	2.09–19.91			2.80–18.19	3.38–18.53			
Mean (95% CI)	8.33 (8.16–8.51)	8.88 (8.69–9.07)	<0.0001	1.07 (1.04–1.11)	8.40 (8.20–8.59)	8.91 (8.72–9.09)	0.0002	1.08 (1.04–1.13)	1.08 (1.04–1.13)
Overall									
By sex									
Male	8.57 (8.34–8.80)	9.08 (8.81–9.35)	0.0055	1.06 (1.02–1.11)	8.76 (8.46–9.06)	9.04 (8.77–9.31)	0.169	1.04 (0.98–1.11)	1.04 (0.98–1.11)
Female	8.02 (7.76–8.28)	8.65 (8.38–8.91)	0.0011	1.09 (1.04–1.15)	8.11 (7.85–8.36)	8.79 (8.54–9.05)	0.0002	1.11 (1.05–1.18)	1.12 (1.06–1.19)
By smoking status									
Never	NA	NA			8.00 (7.70–8.29)	8.67 (8.39–8.94)	0.0012	1.11 (1.04–1.19)	1.11 (1.04–1.19)
Former	8.15 (7.91–8.39)	8.77 (8.51–9.03)	0.0005	1.08 (1.04–1.13)	8.61 (8.26–8.96)	9.10 (8.73–9.47)	0.057	1.08 (1.00–1.16)	1.08 (1.00–1.18)
Current	8.52 (8.26–8.78)	9.04 (8.75–9.32)	0.0086	1.07 (1.04–1.12)	8.75 (8.36–9.14)	9.09 (8.75–9.42)	0.201	1.05 (0.97–1.14)	1.04 (0.95–1.13)
Dichotomized (by median) ^c	n (%)	n (%)			n (%)	n (%)			
>8.6	393 (43.0)	429 (49.9)	0.0037	1.00 (referent)	292 (43.0)	356 (51.2)	0.0023	1.00 (referent)	1.00 (referent)
≤8.6	521 (57.0)	431 (50.1)		1.32 (1.09–1.59)	387 (57.0)	339 (48.8)		1.39 (1.13–1.72)	1.39 (1.11–1.73)
Quartiles ^c									
>10.7	167 (18.3)	222 (25.8)	0.0007	1.00 (referent)	128 (18.9)	161 (23.2)	0.0089	1.00 (referent)	1.00 (referent)
8.7–10.7	226 (24.7)	207 (24.1)		1.45 (1.10–1.91)	164 (24.1)	195 (28.1)		1.06 (0.78–1.44)	1.16 (0.84–1.60)
6.9–8.6	243 (26.6)	218 (25.3)		1.48 (1.13–1.94)	190 (28.0)	185 (26.6)		1.29 (0.95–1.76)	1.38 (1.00–1.90)
≤6.8	278 (30.4)	213 (24.8)		1.74 (1.33–2.27)	197 (29.0)	154 (22.1)		1.61 (1.18–2.20)	1.65 (1.19–2.29)
Trend test ^b				0.0001				0.0009	0.0013

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 narrow 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

DRC (%)	Case (n = 1,593)	Control (n = 1,555)	P^a	Crude OR (95% CI)	Adjusted ^b OR (95% CI)
Continuous, Range	2.68–18.70	2.09–19.91			
Mean (95% CI)					
Overall	8.36 (8.23–8.49)	8.89 (8.76–9.03)	<0.0001	1.08 (1.05–1.11)	1.08 (1.05–1.11)
By sex					
Male	8.64 (8.46–8.82)	9.06 (8.87–9.26)	0.002	1.06 (1.02–1.10)	1.06 (1.02–1.10)
Female	8.06 (7.88–8.25)	8.72 (8.53–8.90)	<0.0001	1.10 (1.06–1.15)	1.11 (1.06–1.15)
By smoking status					
Never	8.00 (7.70–8.29)	8.67 (8.39–8.94)	0.0012	1.11 (1.04–1.19)	1.19 (1.04–1.11)
Former	8.29 (8.09–8.49)	8.86 (8.65–9.08)	<0.0001	1.08 (1.04–1.12)	1.08 (1.04–1.13)
Current	8.59 (8.38–8.80)	9.05 (8.83–9.27)	0.0032	1.06 (1.02–1.11)	1.07 (1.02–1.12)
Dichotomized (by median) ^c	n (%)	n (%)			
>8.6	685 (43.0)	785 (50.5)	<0.0001	1.00 (referent)	1.00 (referent)
≤8.6	908 (57.0)	770 (49.5)		1.35 (1.17–1.56)	1.36 (1.18–1.57)
Quartiles ^c					
>10.7	295 (18.5)	383 (24.6)	<0.0001	1.00 (referent)	1.00 (referent)
8.7–10.7	390 (24.5)	402 (25.9)		1.26 (1.03–1.55)	1.30 (1.06–1.60)
6.9–8.6	433 (27.2)	403 (25.9)		1.40 (1.14–1.71)	1.42 (1.15–1.74)
≤6.8	475 (29.8)	367 (23.6)		1.68 (1.37–2.06)	1.73 (1.41–2.14)
Trend test ^b				<0.0001	<0.0001

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

^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.

Table 4. SNPs with the strongest effects on DRC from the joint analysis

SNP	Gene	Symbol	Chr.	Location	A1/A2	MAF			Genotype ^a						
						GWAS	Replication	GWAS	Replication	GWAS	Replication	Combined			
rs13181	<i>ERCC2/XPD</i>	C/A	19	Coding	C/A	0.3613	0.3681	727/812/235	563/608/201	1290/1420/436	Additive	0.02549	5.39E-06	9.08E-07	0.3167
rs9390123	<i>PHACTR2</i>	A/G	6	Intron	A/G	0.3957	0.3873	639/866/269	518/644/209	1157/1510/478	Additive	0.000025	0.02364	6.68E-06	0.1661
rs7443927	<i>DUSP1</i>	A/C	5	Flanking_3UTR	A/C	0.1088	0.1159	1411/340/23	1072/282/18	2483/622/41	Additive	0.001225	0.03202	0.000176	0.2565

Abbreviation: MAF, minor allele (A1) frequency.

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

^b*P* for heterogeneity.

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

SNPs with a $P < 10^{-3}$ 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 ($P_{\text{joint}} = 9.08 \times 10^{-7}$, $P_{\text{discovery}} = 0.025$; $P_{\text{replication}} = 5.39 \times 10^{-6}$). The second strongest association was observed for rs9390123 in the *PHACTR2* [phosphatase and actin regulator 2 (MIM 608724)] gene on chromosome 6 ($P_{\text{joint}} = 6.68 \times 10^{-6}$, $P_{\text{discovery}} = 2.5 \times 10^{-5}$, $P_{\text{replication}} = 0.024$). 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 ($P_{\text{joint}} = 1.76 \times 10^{-4}$; $P_{\text{discovery}} = 1.23 \times 10^{-3}$, $P_{\text{replication}} = 0.032$). 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 (± 100 kb for rs13181 and rs7443927; ± 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

Table 5. Modification effects of SNPs on DRC levels, overall and by smoking status^a

Gene/SNP	Genotype	All (N = 3,148)		Never smokers (n = 586)		Ever smokers (n = 2,562)	
		n	DRC (%) Mean (95% CI)	n	DRC (%) Mean (95% CI)	n	DRC (%) Mean (95% CI)
ERCC2/XPD /rs13181	AA	1,290	8.80 (8.65–8.94)	238	8.72 (8.41–9.04)	1,052	8.81 (8.65–8.98)
	AC	1,420	8.61 (8.47–8.75)	264	8.27 (7.97–8.57)	1,156	8.69 (8.53–8.85)
	CC	436	8.13 (7.88–8.39)	83	7.43 (6.90–7.96)	353	8.30 (8.01–8.58)
	Trend test P ^b		<0.0001		0.0002		0.009
PHACTR2 /rs9390123	GG	1,157	8.86 (8.71–9.02)	223	8.58 (8.25–8.91)	934	8.93 (8.76–9.11)
	AG	1,510	8.57 (8.43–8.70)	276	8.29 (7.99–8.58)	1,234	8.63 (8.48–8.78)
	AA	478	8.20 (7.96–8.44)	85	7.92 (7.39–8.46)	393	8.26 (7.99–8.53)
	Trend test P ^b		<0.0001		0.102		0.0002
DUSP1 /rs7443927	CC	2,483	8.72 (8.61–8.83)	466	8.41 (8.18–8.63)	2,017	8.79 (8.67–8.91)
	AC	622	8.30 (8.08–8.51)	114	8.10 (7.64–8.56)	508	8.34 (8.10–8.58)
	AA	41	7.61 (6.78–8.43)	6	8.07 (6.06–10.08)	35	7.53 (6.62–8.43)
	Trend test P ^b		0.0001		0.490		0.0001

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

^bANOVA test.

the overall analysis and in the stratification by smoking status, which may be simply due to a limited study power.

To further support biologic plausibility of our observations, we queried the gene expression database in 90 Caucasian parents and children (CEU) at SNPexp—A web tool using the SNP genotypes from HapMap2 release 23, or HapMap3 release 3, and Genevar—the gene expression levels of lymphoblastoid cell lines derived from the same individuals (26). We found that there was a consistent and significantly decreased trend in the expression levels of the *ERCC2/XPD* gene when SNP rs13181 was fitting in the additive linear regression model [$P = 0.0006541$; 8.007 ± 0.2099 (mean \pm SD) for AA genotypes, 7.885 ± 0.1899 for AC, and 7.791 ± 0.1236 for CC]. In contrast, the trend between genotypes and expression levels of the corresponding genes in the HapMap CEU population was not significant for the SNP rs9390123 in *PHACTR2* and the SNP rs7443927 in *DUSP1*.

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 nonsynonymous 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-wide 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 IIIH 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 6q24. 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-butanone], 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, rs1799793), *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

Conception and design: L.-E. Wang, O.Y. Gorlova, M.R. Spitz, C.I. Amos, Q. Wei
Development of methodology: L.-E. Wang, O.Y. Gorlova, Y. Qiao, Q. Wei
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): L.-E. Wang, A.T. Lee, P.K. Gregersen, M.R. Spitz, C.I. Amos, Q. Wei

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L.-E. Wang, O.Y. Gorlova, J. Ying, S.-F. Weng, A.T. Lee, M.R. Spitz, C.I. Amos, Q. Wei

Writing, review, and/or revision of the manuscript: L.-E. Wang, O.Y. Gorlova, J. Ying, A.T. Lee, M.R. Spitz, C.I. Amos, Q. Wei

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): L.-E. Wang, Y. Qiao, A.T. Lee, C.I. Amos, Q. Wei

Study supervision: L.-E. Wang, O.Y. Gorlova, M.R. Spitz, Q. Wei

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 P30CA016672 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

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012;62:10–29.
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

- and risk of lung cancer: a case-control analysis. *Cancer Res* 2001;61:1445–50.
3. MacLeod MC, Tang MS. Interactions of benzo(a)pyrene diol-epoxides with linear and supercoiled DNA. *Cancer Res* 1985;45:51–6.
 4. Gelboin HV. Benzo[alpha]pyrene metabolism, activation and carcinogenesis: role and regulation of mixed-function oxidases and related enzymes. *Physiol Rev* 1980;60:1107–66.
 5. Phillips DH. Fifty years of benzo(a)pyrene. *Nature* 1983;303:468–72.
 6. Tang MS, Pierce JR, Doisy RP, Nazimiec ME, MacLeod MC. Differences and similarities in the repair of two benzo[a]pyrene diol epoxide isomers induced DNA adducts by uvrA, uvrB, and uvrC gene products. *Biochemistry (Mosc)* 1992;31:8429–36.
 7. Sancar A. DNA repair in humans. *Annu Rev Genet* 1995;29:69–105.
 8. Braithwaite E, Wu X, Wang Z. Repair of DNA lesions induced by polycyclic aromatic hydrocarbons in human cell-free extracts: involvement of two excision repair mechanisms *in vitro*. *Carcinogenesis* 1998;19:1239–46.
 9. Mattson ME, Pollack ES, Cullen JW. What are the odds that smoking will kill you? *Am J Public Health* 1987;77:425–31.
 10. Woloshin S, Schwartz LM, Welch HG. The risk of death by age, sex, and smoking status in the United States: putting health risks in context. *J Natl Cancer Inst* 2008;100:845–53.
 11. Caporaso N, Landi MT, Vineis P. Relevance of metabolic polymorphisms to human carcinogenesis: evaluation of epidemiologic evidence. *Pharmacogenetics* 1991;1:4–19.
 12. Wogan GN, Hecht SS, Felton JS, Conney AH, Loeb LA. Environmental and chemical carcinogenesis. *Semin Cancer Biol* 2004;14:473–86.
 13. Wei Q, Spitz MR. The role of DNA repair capacity in susceptibility to lung cancer: a review. *Cancer Metastasis Rev* 1997;16:295–307.
 14. Pavanello S, Clonfero E. [Individual susceptibility to occupational carcinogens: the evidence from biomonitoring and molecular epidemiology studies]. *G Ital Med Lav Ergon* 2004;26:311–21.
 15. Shen H, Spitz MR, Qiao Y, Guo Z, Wang LE, Bosken CH, et al. Smoking, DNA repair capacity and risk of nonsmall cell lung cancer. *Int J Cancer* 2003;107:84–8.
 16. Wei Q, Cheng L, Amos CI, Wang LE, Guo Z, Hong WK, et al. Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: a molecular epidemiologic study. *J Natl Cancer Inst* 2000;92:1764–72.
 17. Wei Q, Cheng L, Hong WK, Spitz MR. Reduced DNA repair capacity in lung cancer patients. *Cancer Res* 1996;56:4103–7.
 18. Qiao Y, Spitz MR, Guo Z, Hadeyati M, Grossman L, Kraemer KH, et al. Rapid assessment of repair of ultraviolet DNA damage with a modified host-cell reactivation assay using a luciferase reporter gene and correlation with polymorphisms of DNA repair genes in normal human lymphocytes. *Mutat Res* 2002;509:165–74.
 19. Qiao Y, Spitz MR, Shen H, Guo Z, Shete S, Hadeyati M, et al. Modulation of repair of ultraviolet damage in the host-cell reactivation assay by polymorphic XPC and XPD/ERCC2 genotypes. *Carcinogenesis* 2002;23:295–9.
 20. Schwartz AG, Prysak GM, Bock CH, Cote ML. The molecular epidemiology of lung cancer. *Carcinogenesis* 2007;28:507–18.
 21. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 2008;40:616–22.
 22. Cheng L, Bucana CD, Wei Q. Fluorescence in situ hybridization method for measuring transfection efficiency. *Biotechniques* 1996;21:486–91.
 23. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
 24. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–5.
 25. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res* 2002;12:996–1006.
 26. Holm K, Melum E, Franke A, Karlsen TH. SNPexp - A web tool for calculating and visualizing correlation between HapMap genotypes and gene expression levels. *BMC Bioinformatics* 2010;11:600.
 27. Lindahl T, Karran P, Wood RD. DNA excision repair pathways. *Curr Opin Genet Dev* 1997;7:158–69.
 28. Friedberg EC. How nucleotide excision repair protects against cancer. *Nat Rev Cancer* 2001;1:22–33.
 29. Shen MR, Jones IM, Mohrenweiser H. Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans. *Cancer Res* 1998;58:604–8.
 30. Zhan P, Wang Q, Wei SZ, Wang J, Qian Q, Yu LK, et al. ERCC2/XPD Lys751Gln and Asp312Asn gene polymorphism and lung cancer risk: a meta-analysis involving 22 case-control studies. *J Thorac Oncol* 2010;5:1337–45.
 31. Christiani DC. ERCC2/XPD polymorphisms and lung cancer risk. *J Thorac Oncol* 2011;6:233.
 32. Allen PB, Greenfield AT, Svenningsson P, Haspelslagh DC, Greengard P. Phactr 1–4: a family of protein phosphatase 1 and actin regulatory proteins. *Proc Natl Acad Sci U S A* 2004;101:7187–92.
 33. Wider C, Lincoln SJ, Heckman MG, Diehl NN, Stone JT, Haugavoll K, et al. Phactr2 and Parkinson's disease. *Neurosci Lett* 2009;453:9–11.
 34. Keyse SM, Emslie EA. Oxidative stress and heat shock induce a human gene encoding a protein-tyrosine phosphatase. *Nature* 1992;359:644–7.
 35. Sun H, Charles CH, Lau LF, Tonks NK. MKP-1 (3CH134), an immediate early gene product, is a dual specificity phosphatase that dephosphorylates MAP kinase *in vivo*. *Cell* 1993;75:487–93.
 36. Kennedy DO, Agrawal M, Shen J, Terry MB, Zhang FF, Senie RT, et al. DNA repair capacity of lymphoblastoid cell lines from sisters discordant for breast cancer. *J Natl Cancer Inst* 2005;97:127–32.
 37. Machella N, Terry MB, Zipprich J, Gurvich I, Liao Y, Senie RT, et al. Double-strand breaks repair in lymphoblastoid cell lines from sisters discordant for breast cancer from the New York site of the BCFR. *Carcinogenesis* 2008;29:1367–72.
 38. Li D, Wang M, Cheng L, Spitz MR, Hittelman WN, Wei Q. *In vitro* induction of benzo(a)pyrene diol epoxide-DNA adducts in peripheral lymphocytes as a susceptibility marker for human lung cancer. *Cancer Res* 1996;56:3638–41.
 39. Bau DT, Mau YC, Ding SL, Wu PE, Shen CY. DNA double-strand break repair capacity and risk of breast cancer. *Carcinogenesis* 2007;28:1726–30.
 40. Bau DT, Fu YP, Chen ST, Cheng TC, Yu JC, Wu PE, et al. Breast cancer risk and the DNA double-strand break end-joining capacity of nonhomologous end-joining genes are affected by BRCA1. *Cancer Res* 2004;64:5013–9.
 41. Decordier I, Loock KV, Kirsch-Volders M. Phenotyping for DNA repair capacity. *Mutat Res* 2010;705:107–29.
 42. Wang LE, Hu Z, Sturgis EM, Spitz MR, Strom SS, Amos CI, et al. Reduced DNA repair capacity for removing tobacco carcinogen-induced DNA adducts contributes to risk of head and neck cancer but not tumor characteristics. *Clin Cancer Res* 2010;16:764–74.
 43. Wang L, Wei Q, Shi Q, Guo Z, Qiao Y, Spitz MR. A modified host-cell reactivation assay to measure repair of alkylating DNA damage for assessing risk of lung adenocarcinoma. *Carcinogenesis* 2007;28:1430–6.
 44. Wang LE, Li C, Strom SS, Goldberg LH, Brewster A, Guo Z, et al. Repair capacity for UV light induced DNA damage associated with risk of nonmelanoma skin cancer and tumor progression. *Clin Cancer Res* 2007;13:6532–9.
 45. Shi Q, Wang LE, Bondy ML, Brewster A, Singletary SE, Wei Q. Reduced DNA repair of benzo[a]pyrene diol epoxide-induced adducts and common XPD polymorphisms in breast cancer patients. *Carcinogenesis* 2004;25:1695–700.
 46. Spitz MR, Wu X, Wang Y, Wang LE, Shete S, Amos CI, et al. Modulation of nucleotide excision repair capacity by XPD polymorphisms in lung cancer patients. *Cancer Res* 2001;61:1354–7.
 47. Santella RM, Gammon M, Terry M, Senie R, Shen J, Kennedy D, et al. DNA adducts, DNA repair genotype/phenotype and cancer risk. *Mutat Res* 2005;592:29–35.
 48. Gu J, Ye Y, Spitz MR, Lin J, Kiemeny LA, Xing J, et al. A genetic variant near the PMAIP1/Noxa gene is associated with increased bleomycin sensitivity. *Hum Mol Genet* 2011;20:820–6.
 49. Rinckel AE, Surowy HM, Luedeke M, Varga D, Schrader M, Hoegel J, et al. The prostate cancer risk locus at 10q11 is associated with DNA repair capacity. *DNA Repair (Amst)* 2012;11:693–701.