

High-Order Interactions among Genetic Variants in DNA Base Excision Repair Pathway Genes and Smoking in Bladder Cancer Susceptibility

Maosheng Huang,¹ Colin P. Dinney,² Xin Lin,¹ Jie Lin,¹ H. Barton Grossman,² and Xifeng Wu¹

Departments of ¹Epidemiology and ²Urology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

Abstract

Cancer is a common multifactor human disease resulting from complex interactions between many genetic and environmental factors. In this study, we used a multifaceted analytic approach to explore the relationship between eight single nucleotide polymorphisms in base excision repair (BER) pathway genes, smoking, and bladder cancer susceptibility in a hospital-based case-control study. Overall, we did not find an association between any single BER gene single nucleotide polymorphism and bladder cancer risk. However, in stratified analysis, the OGG1 S326C variant genotypes in ever smokers (odds ratio, 0.74; 95% confidence interval, 0.56-0.99) and ADP-ribosyltransferase (ADPRT) V762A variant genotypes in never smokers (odds ratio, 0.58; 95% confidence interval, 0.37-0.91) conferred a significantly reduced risk. Using logistic regression, we observed that there was a two-way interaction between ADPRT V762A and smoking status. We next used classification and regression tree analysis to explore high-order gene-gene

and gene-environment interactions. We found that smoking is the most important influential factor for bladder cancer risk. Consistent with the above findings, we found that the ADPRT V762A was only significantly involved in bladder cancer risk in never smokers and the OGG1 S326C was only significantly involved in ever smokers. We also observed gene-gene interactions among OGG1 S326C, XRCC1 R194W, and MUTYH H335Q in ever smokers. Using multifactor dimensionality reduction approach, the four-factor model, including smoking status, OGG1 S326C (rs1052133), APEX1 D148E (rs3136820), and ADPRT762 (rs1136410), had the best ability to predict bladder cancer risk with the highest cross-validation consistency (100%) and the lowest prediction error (37.02%; $P < 0.001$). These results support the hypothesis that genetic variants in BER genes contribute to bladder cancer risk through gene-gene and gene-environmental interactions. (Cancer Epidemiol Biomarkers Prev 2007;16(1):84-91)

Introduction

It has become increasingly clear that cancer is a common complex multifactorial human disease and can be considered neither purely genetic nor purely environmental. It mainly results from complex interactions between many genetic and environmental factors, particularly for the sporadic forms of cancer (1). Human cancer can be initiated by DNA damage caused by UV, ionizing radiation, tobacco exposure, and environmental chemical agents. However, humans have developed a set of complex DNA repair systems to safeguard the integrity of genome by defending harmful consequences of DNA damage (2, 3). It is generally believed that carcinogenesis is a multistep, multigenic, multicausal process (4). Single gene or single environmental factor studies are likely to provide limited information in predicting risk; therefore, recently, many studies of cancer research have focused on the interaction between genes and environment in the same causal mechanism (5-7).

It was estimated that in 2006, bladder cancer would be the fourth most frequently diagnosed cancer in men and the ninth in women in the United States (8). Cigarette smoking is an established risk factor for bladder cancer (9). Occupational exposures to 4-aminobiphenyl, 2-naphthylamine, benzidine (10), and aromatic amines, such as *o*-toluidine (11), also play an important role in the initiation of bladder cancer. These

exposures lead to DNA damage that, if remained damaged, may result in unregulated cell growth and even cancer. DNA damage repair and cell cycle checkpoints facilitate cellular responses to DNA damage from endogenous and exogenous mutagenic exposures to maintain genomic integrity. The base excision repair (BER) pathway is one of the four major DNA repair pathways in human cells. The proteins in the BER pathway mainly work on damaged DNA bases arising from endogenous oxidative and hydrolytic decay of DNA. Base damage and DNA single-strand breaks are mainly repaired through the BER pathway (12).

This pathway is a multistep process that requires the activity of several proteins (12, 13). Cigarette smoke is a rich source of reactive oxygen species that can induce a variety of DNA damages, some of which are repaired by the BER pathway.

In this study, we estimated the frequency of eight SNPs from seven BER pathway genes, including *MBD4* Glu³⁴⁶Lys, *MUTYH* Gln³³⁵His, *OGG1* Ser³²⁶Cys, *APEX1* Glu¹⁴⁸Asp, *XRCC1* Arg¹⁹⁴Trp, *XRCC1* Arg³⁹⁹Gln, *ADP-ribosyltransferase (ADPRT)* Val⁷⁶²Ala, and *POLD1* Arg¹¹⁹His in bladder cancer cases and controls. We applied several statistical approaches to evaluate BER pathway gene-gene and gene-environment interactions in bladder cancer susceptibility.

Materials and Methods

Study Population. Beginning in July 1999, incident urinary bladder cancer cases were accrued from The University of Texas M. D. Anderson Cancer Center and Baylor College of Medicine in Houston, TX. All cases were histologically confirmed and untreated previously with chemotherapy or radiotherapy. There were no age, gender, ethnic, or cancer stage restrictions. M. D. Anderson Cancer Center staff

Received 8/22/06; revised 10/23/06; accepted 11/10/06.

Grant support: National Cancer Institute grants CA 74880 and CA 91846.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Xifeng Wu, Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Unit 1340, 1155 Pressler Boulevard, Houston, TX 77030. Phone: 713-745-2485; Fax: 713-792-4657. E-mail: xwu@mdanderson.org

Copyright © 2007 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-06-0712

interviewers identified bladder cancer cases through a daily review of computerized appointment schedules for the Departments of Urology and Genitourinary Medical Oncology. Each new patient was screened with a brief eligibility questionnaire that assessed prior cancer therapy and willingness to participate in the epidemiologic study. If the patient was willing to participate, the interviewer accompanied the study participant to a private room to conduct the interview after obtaining informed consent. Healthy control subjects without a history of cancer, except nonmelanoma skin cancer, were recruited from an ongoing collaboration with the Kelsey-Seybold clinics, Houston's largest private multispecialty physician group. The controls were frequency matched to the cases on age (± 5 years), sex, and ethnicity. The potential control subjects were first surveyed by using a short questionnaire to elicit willingness to participate in the study and to provide preliminary demographic data for matching. The potential control subjects were contacted by telephone at a later date to confirm their willingness to participate and to schedule an interview appointment at a Kelsey-Seybold clinical site convenient to the participant.

Epidemiologic Data. After informed consent was obtained, all study participants completed a 90-min in-person interview that was given by M. D. Anderson Cancer Center staff interviewers. The interview elicited information on demographics and smoking history. The questionnaire consisted of a fixed script and included introductory and transitional statements. All interviewers were trained for the use of probes. At the conclusion of the interview, a 40-mL blood sample was drawn into coded heparinized tubes. Human subject approval was obtained from the M. D. Anderson Cancer Center, Baylor College of Medicine, and the Kelsey-Seybold institutional review boards. An individual who had smoked at least 100 cigarettes in his or her lifetime was defined as an ever smoker. Ever smokers include former smokers, current smokers, and recent quitters (those who had quit within the previous year).

DNA Isolation. Genomic DNA was isolated from peripheral blood using QIAamp DNA blood maxi kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. The working aliquots of the genomic DNA were stored at -20°C until use.

Genotype Assays. Each single nucleotide polymorphism (SNP) genotyping was done using the Taqman method with a 7900 HT sequence detector system (Applied Biosynthesis, Foster City, CA). The primer and probe sequences for each SNP are available on request. Typical amplification mixes (5 μL) contained sample DNA (5 ng), $1\times$ Taqman buffer A, 200 $\mu\text{mol/L}$ deoxynucleotide triphosphates, 5 mmol/L MgCl_2 , 0.65 units of AmpliTaq Gold, 900 nm of each primer, and 200 nmol/L of each probe. The reactions were carried out in the Dual 384-Well GeneAmp PCR System 9700. The thermal conditions were 95°C for 10 min followed by 50 cycles of 92°C for 30 sec and 60°C for 1 min. Following the amplification reaction, the reacted plates were read using the ABI Prism 7900HT Sequence Detection System. The analyzed fluorescence results were then automatically called into genotypes using the built-in software of the system. Water control, amplification internal controls, and previously genotyped samples were included in each plate to ensure accuracy of the genotyping, and 5% of the samples were randomly selected and run in duplicates with 100% concordance.

Statistical Analysis. Using the Intercooled Stata 8.0 statistical software package (Stata Co., College Station, TX), The Pearson χ^2 test was used to test for differences between the cases and the control subjects for the categorical variables of gender, smoking status, and each SNP genotype. The Student's *t* test was used to test for differences between the case and control subjects for the continuous variables of age and pack-

year. Hardy-Weinberg equilibrium for the genotypes was tested by a goodness-of-fit χ^2 test. Odds ratios (OR) and 95% confidence intervals (95% CI) were calculated as an estimate of relative risk. Unconditional multivariate logistic regression was used to control for possible confounding by age, gender, and smoking status, when appropriate as well as when examining interactions between SNPs and smoking. Interaction was tested using a multiplicative interaction term included in the multivariate model. Joint effects were analyzed using never smokers with the wild-type (WT) genotype as the reference group. Statistical significances of the interactions were assessed using likelihood ratio tests comparing the models with and without interaction terms.

Classification and Regression Tree Approach. For higher-order gene-gene interactions, classification and regression tree (CART) analysis was done using the HelixTree Genetics Analysis software (version 4.1.0; Golden Helix, Bozeman, MT). CART is a binary recursive partitioning method that produces a decision tree to identify subgroups of subjects at higher risk (14). Specifically, the recursive partitioning algorithm in HelixTree starts at the first node (with the entire data set) and uses a statistical hypothesis testing method, formal inference-based recursive modeling, to determine the first locally optimal split and each subsequent split of the data set, with multiplicity-adjusted *P* values to control tree growth ($P < 0.05$). This process continues until the terminal nodes have no subsequent statistically significant splits or the terminal nodes reach a prespecified minimum size (at least 10 subjects for each terminal node in our analysis).

Multifactor Dimensionality Reduction Approach. The nonparametric multifactor dimensionality reduction (MDR) approach was selected to complement logistic regression for the analysis of gene-gene and gene-environment interactions. The MDR method was first described by Moore et al. (15-18). Here, we briefly describe MDR method. MDR is a nonparametric and genetic model-free alternative to logistic regression for detecting and characterizing nonlinear interactions among discrete genetic and environmental attributes. The MDR method combines attribute selection, attribute construction, and classification with cross-validation and permutation testing to provide a comprehensive and powerful data mining approach to detecting nonlinear interactions. The method involved several steps. In step one, the data were divided into a training set (9 of 10 of the data) and an independent testing set (the remaining 1 of 10 of the data) as part of cross-validation. In step two, a set of *n* factors (in this case, factors) were selected, where $n = 1$ to 5. In steps 3 and 4, the *n* factors and their possible multifactor classes were represented in *n* dimensional space. The ratio for the number of cases to the number of controls was calculated within each multifactor class. Each multifactor class in *n* dimensional space was then labeled as "high risk" if the case to control ratio met or exceeded a threshold (for example, 1.1065) or as "low risk" if that threshold was not exceeded, thus reducing the *n* dimensional space to one dimension with two levels (low risk and high risk). In step five, the model that gave the lowest misclassification error was selected for each set of *n* factors. In step six, a prediction error was estimated for each model selected in step five, as a cross-validation procedure. Steps one to six were repeated 10 times using a random seed number. We did this entire 100-fold cross-validation procedure 10 times, using different random seed numbers, to reduce the chance of observing spurious results due to chance divisions of the data. In addition to prediction error, we also estimated a cross-validation consistency, defined as a percentage of the same combination of factors selected as the best model among different cross-validation data sets, for each set of *n* factors. A testing accuracy of 0.5 was expected under the null hypothesis. Statistical significance was determined using permutation

Table 1. Overall main effects of genotypes on bladder cancer risk and effects stratified by smoking status

Genes (SNP ID)	Genotypes	Controls, <i>n</i> (%)	Cases, <i>n</i> (%)	Adjusted OR (95% CI)*	Smoking status	
					Never	Ever
MBD4	GG	591 (57.2)	597 (55.2)	Ref.	Ref.	Ref.
E346K (rs140693)	GA & AA	5 (42.8)	8 (44.8)	1.74 (0.54-5.63)	1.29 (0.21-7.83)	2.05 (0.41-10.2)
MUTYH	GG	340 (57.2)	334 (55.2)	Ref.	Ref.	Ref.
Q335H (rs3219489)	GC & CC	254 (42.8)	271 (44.8)	1.08 (0.85-1.37)	1.11 (0.75-1.66)	1.31 (0.86-1.99)
OGG1	CC	348 (57.2)	375 (62.2)	Ref.	Ref.	Ref.
S326C (rs1052133)	CG & GG	260 (42.8)	228 (37.8)	0.82 (0.65-1.05)	0.93 (0.62-1.39)	0.74 (0.56-0.99)
APEX1	TT	166 (28.1)	176 (29.5)	Ref.	Ref.	Ref.
D148E (rs3136820)	TC & CC	424 (71.9)	420 (70.5)	0.91 (0.70-1.18)	0.86 (0.56-1.32)	0.94 (0.68-1.30)
XRCC1	CC	524 (87.3)	539 (87.8)	Ref.	Ref.	Ref.
R194W (rs1799782)	CT & TT	76 (12.7)	75 (12.2)	0.94 (0.66-1.34)	0.91 (0.50-1.68)	0.96 (0.63-1.47)
XRCC1	GG	267 (44.8)	266 (43.4)	Ref.	Ref.	Ref.
R399Q (rs25487)	GA & AA	329 (55.2)	347 (56.6)	1.05 (0.83-1.33)	0.96 (0.65-1.43)	1.06 (0.79-1.41)
POLD1	GG	495 (86.7)	501 (86.7)	Ref.	Ref.	Ref.
R119H (rs1726801)	GA & AA	76 (13.3)	77 (13.3)	0.97 (0.68-1.38)	1.09 (0.60-1.98)	0.94 (0.61-1.43)
ADPRT	TT	416 (69.9)	437 (72.1)	Ref.	Ref.	Ref.
V762A (rs1136410)	TC & CC	179 (30.1)	169 (27.9)	0.89 (0.68-1.15)	0.58 (0.37-0.91)	1.12 (0.81-1.54)

*Adjusted by gender, age, and smoking status.

testing. Here, the case-control labels were randomized *n* times, and the entire MDR model fitting procedure was repeated on each randomized data set to determine the expected distribution of testing accuracies under the null hypothesis. In this study, we used 100-fold cross-validation and 1,000-fold permutation testing. MDR results were considered statistically significant at the 0.05 levels. To better visualize interactions, we built an interaction dendrogram that places strongly interacting variables close together at the leaves of the tree. This method is included in the MDR software and was described by Moore et al. (19).

Results

Characteristics of Subjects. Due to the small numbers of minorities, our analyses were restricted to Caucasians, including 696 patients with bladder cancer and 629 healthy controls. The case group had more males than the control group (78.45% versus 72.66%; *P* = 0.014). There was no significant difference between cases (63.94 ± 11.17 years) and healthy controls (62.77 ± 10.50 years) on age (*P* = 0.06). In smoking status, there were more ever smokers among the cases than among the controls (73.55% versus 53.74%; *P* <

Table 2. Joint effects between genotypes and smoking status in the BER pathway

Gene	Smoking status	Cases, <i>n</i>	Controls, <i>n</i>	Joint effect, adjusted OR (95% CI)*	<i>P</i> _{interaction}
MUTYH Q335H					
GG	No	84	155	Ref.	
GG	Yes	250	185	2.40 (1.72-3.34)	
GC & CC	No	72	120	1.11 (0.75-1.66)	
GC & CC	Yes	199	134	2.60 (1.83-3.69)	0.915
OGG1 S326C					
CC	No	96	165	Ref.	
CC	Yes	279	183	2.48 (1.80-3.42)	
CG & GG	No	62	113	0.93 (0.62-1.38)	
CG & GG	Yes	166	147	1.84 (1.31-2.59)	0.378
APEX1 D148E					
TT	No	50	79	Ref.	
TT	Yes	126	87	2.21 (1.41-3.46)	
TC & CC	No	105	195	0.86 (0.56-1.32)	
TC & CC	Yes	315	229	2.08 (1.40-3.10)	0.733
XRCC1R194W					
CC	No	140	244	Ref.	
CC	Yes	399	280	2.37 (1.82-3.09)	
CT & TT	No	18	34	0.91 (0.50-1.67)	
CT & TT	Yes	57	42	2.28 (1.45-3.58)	0.884
XRCC1R399Q					
GG	No	76	129	Ref.	
GG	Yes	190	138	2.25 (1.57-3.24)	
GA & AA	No	82	148	0.96 (0.65-1.42)	
GA & AA	Yes	265	181	2.39 (1.69-3.38)	0.685
ADPRTV762A					
TT	No	122	184	Ref.	
TT	Yes	315	232	1.95 (1.46-2.61)	
TC & CC	No	35	91	0.58 (0.37-0.91)	
TC & CC	Yes	134	88	2.20 (1.53-3.14)	0.019
POLD R119H					
GG	No	129	227	Ref.	
GG	Yes	372	268	2.35 (1.79-3.09)	
GA & AA	No	20	32	1.08 (0.59-1.97)	
GA & AA	Yes	57	44	2.21 (1.41-3.49)	0.710

*Adjusted by gender, age, and smoking status.

0.001); the case group also had significantly higher pack-years than the control group (43.0 versus 28.3 pack-years; $P < 0.001$) among ever smokers.

Risk Associated with Individual SNPs Stratified by Smoking Status. The distributions of all selected SNP in the control subjects were in agreement with Hardy-Weinberg equilibrium ($P > 0.05$). By evaluating the independent effects of each SNP on bladder cancer susceptibility using unconditional multivariate logistic regression, we did not observe that the main effects of the BER polymorphisms at each SNP were related to bladder cancer risk. Among ever smokers, however, OGG1 S326C variant genotype was associated with a significantly reduced risk of bladder cancer (OR, 0.74; 95%CI, 0.56-0.99). In the never smoking group, ADPRT V762A variant genotypes conferred a significantly reduced risk (OR, 0.58; 95% CI, 0.37-0.91; Table 1).

Gene-Smoking Joint Effect and Interaction. Using a multiple logistic regression approach, we examined the joint effect and interaction between a SNP and the smoking status. We observed that there was a two-way interaction between ADPRT V762A and the smoking status (gene-smoking interaction $P = 0.019$) on a multiplicative scale. Using never smokers with the WT (T/T) genotypes as the reference group, the ORs (95% CIs) for never smokers with the variant genotypes (T/C + C/C), smokers with the T/T genotype, and smokers with the T/C + C/C genotypes were 0.58 (0.37-0.91), 1.95 (1.46-2.61), and 2.20 (1.53-3.14), respectively (Table 2).

CART Analysis. To further explore gene-gene and gene-environment interactions, we did CART analysis incorporating both the genetic and the smoking status variables. Figure 1 depicted the resulting tree structure generated. There was an initial split on smoking status, confirming that smoking was the most important risk factor for bladder cancer among the factors considered. Further inspection of the CART structure suggested distinct patterns for ever smokers and never

smokers. As documented in previous analyses, the ADPRT V762A gene polymorphism was relevant only in never smokers. In ever smokers, we identified gene-gene interactions for three SNPs (OGG1 S326C, XRCC1 R194W, and MUTYH H335Q). Figure 1 summarized the risk estimates of all the terminal subgroups compared with the subgroups with the least percentage of cases (Ref.). The subgroups with the highest bladder cancer risk were those smokers with WT genotypes of OGG1 S326C, variant genotypes of XRCC1 R194W, and MUTYH H335Q SNPs (OR, 31.86; 95% CI, 4.01-253.1).

MDR Analysis. In the MDR analysis (Table 3), smoking status was the best one-factor model with the highest cross-validation consistency (100%) and the lowest prediction error (39.42%) among all nine factors. Subjects who were ever smokers had a high risk for bladder cancer. The prediction error was statistically significant ($P < 0.001$). The rate of ever smoking was significantly higher in cases than in controls (73.71% and 53.72%; $P < 0.001$, respectively) and was the most significant among these nine factors. When factors were considered two at a time, the combination of ever smoking and MUTYH Q335H was the best two-factor model with the highest cross-validation consistency (75.9%) and the lowest prediction error (39.42%) among all the possible combinations of two factors ($P < 0.001$). The two high-risk groups existed in the subjects who were ever smokers; therefore, the lowest prediction error was equal to the lowest prediction error of the one-factor model (39.42%). When factors were considered three at a time, the factors from smoking status, XRCC1 R194W, and OGG1 S326C had the highest cross-validation consistency (85%) and the lowest prediction error (41.6%) among all the possible combinations of three factors. Compared with the two-factor model, the three-factor model decreased the ability to predict bladder cancer risk. Again, the prediction error was statistically significant, with an empirical $P < 0.001$ based on 1,000 permutations. The four factors from smoking status, OGG1 S326C, APEX1 D148E, and

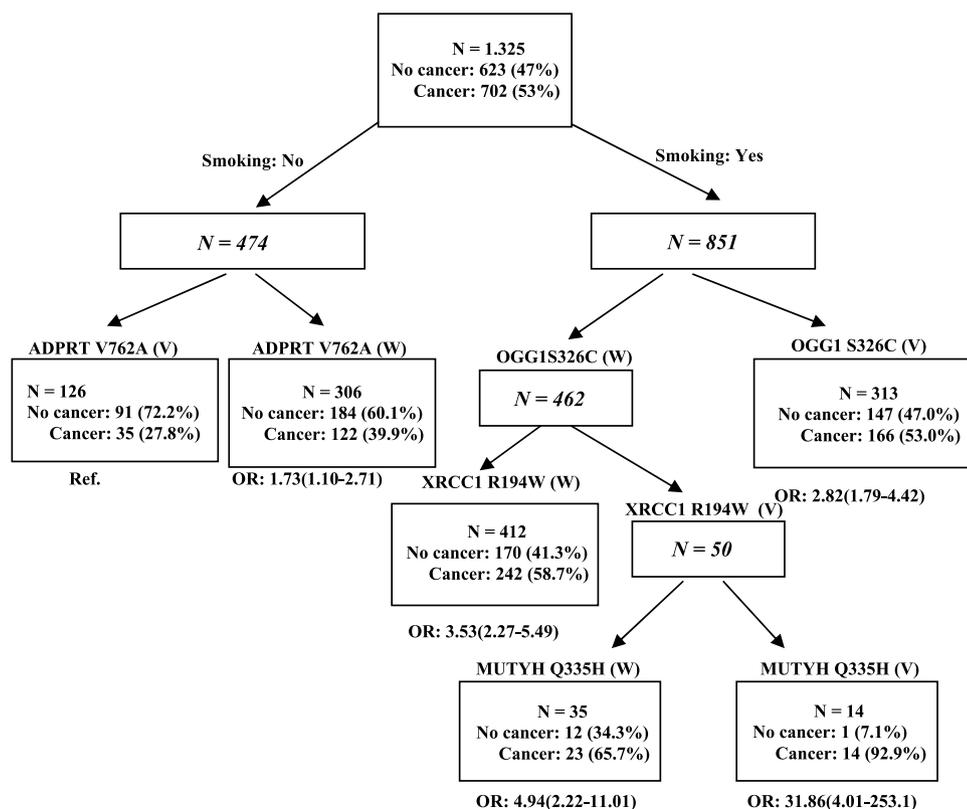


Figure 1. CART analysis of BER pathway genetic polymorphisms and smoking status.

Note: "W" represents Wild type genotypes; "V" represents Variant genotypes.

Table 3. Summary of results for bladder cancer risk prediction from MDR analysis

Best model	Low risk	High risk	Cross-validation consistency	Average prediction error (%)	<i>P</i> , permutation test
One factor					
Smoking status	Never Smoker	Ever smoker	100/100	39.42	<0.001
Two factors					
Smoking status MUTYH Q335H	Never Smoker and any type MUTYH Q335H	Ever smoker and any type MUTYH Q335H	75.9/100	39.42	<0.001
Three factors					
Smoking status XRCC1 R194W OGG1 S326C	Never smoker	Ever smoker except for OGG1 S326C (V) and XRCC1R194W (V)	85/100	41.60	<0.001
Four factors					
Smoking status OGG1 S326C APEX1 D148E ADPRT V762A	Never smoker except for OGG1 S326C (W), APEX1 D148E (W), and ADPRT V762A (W)	Ever smoker except for OGG1 S326C (V), APEX1 D148E (V), and ADPRT V762A (W)	100/100	37.02	<0.001
Five factors					
Smoking status OGG1 S326C APEX1 D148E ADPRT V762A XRCC1 R194W	Never smoker except for OGG1 S326C (W), APEX1 D148E (W), ADPRT V762A (W), and XRCC1 R194W (V)	Ever smoker except for (1) OGG1 S326C (W), APEX1 D148E (V), XRCC1 R194W (W), and ADPRT V762A (V); (2) OGG1 S326C (V), APEX1 D148E (W), XRCC1 R194W (W), and ADPRT V762A (W); (3) OGG1 S326C (V), APEX1 D148E (V), XRCC1 R194W (V), and ADPRT V762A (W)	100/100	41.62	0.002

NOTE: The multilocus model with maximum cross-validation consistency and maximum prediction accuracy is indicated in boldface. Abbreviations: W, WT genotypes; V, variant genotypes.

ADPRT V762A were the best four-factor model with the highest cross-validation consistency (100%) and the lowest prediction error (37.02%) among all the possible combinations of four factors. Compared with the three-factor model, the four-factor model had improved the capability of predicting bladder cancer risk. The prediction error was statistically significant, with an empirical $P < 0.001$ based on 1,000 permutations. Although this prediction error was far from a perfect 0%, it was an important improvement from the a priori 50% chance in predicting bladder cancer status. Eight combinations of these four factors had a high risk for bladder cancer (Fig. 2). These eight combinations did not follow simple dominant, recessive, or additive models. When XRCC1 R194W was added to the four-factor model, the five-factor model had the highest cross-validation consistency (100%) and the lowest prediction error (41.62%) among all the possible combinations of five factors; however, the capability to predict bladder cancer risk was decreased. Fourteen combinations of these five factors had a high risk for bladder cancer (data not shown). These 14 combinations did not follow simple dominant, recessive, or additive models. Figure 3 shows the interaction dendrogram for these four SNPs and smoking. The hierarchical cluster analysis clearly shows that smoking status, OGG1 S326C, APEX1 D148E, and ADPRT V762A exist on the same branch, indicating that there are strong interactions among these five factors and the strongest interaction exists between the smoking status and APEX1 D148E.

Discussion

In the study, we first investigated the association between risk for development of bladder cancer and individual genetic polymorphisms in seven BER genes by using traditional logistic regression. We did not find an association between any single BER gene SNP and bladder cancer risk. However, OGG1 S326C variant genotype was found to be associated with a significantly reduced risk of bladder cancer in ever smokers, and ADPRT V762A variant genotypes conferred a significantly reduced risk in never smokers. We further assessed the association between genotype and genotype combinations

with smoking status and bladder cancer risk using a multifaceted analytic approach (traditional logistic regression, the nonparametric CART, and MDR approach). In CART analysis, smoking is the most important risk factor for bladder cancer. The ADPRT V762A gene polymorphism was relevant only in never smokers. In ever smokers, we identified gene-gene interactions for three SNPs (OGG1 S326C, XRCC1 R194W, and MUTYH H335Q). In MDR analysis, the four-factor model, including smoking status, OGG1 S326C (rs1052133), APEX1 D148E (rs3136820), and ADPRT V762A (rs1136410), had the best ability to predict bladder cancer risk.

Several studies have in fact found associations between single genetic polymorphisms in some BER genes, such as OGG1 S326C, APEX1 D148E, MUTYH H335Q, APEX1 D148E, ADPRT V762A, and XRCC1 R194W, and risk of certain cancers, including human breast (20), colorectal (21), gastric (22), and endometrial cancer (23). Association of the common S326C polymorphism of OGG1 with an increased risk for cancer was observed in several case-control studies (24). However, no previous studies have found associations between OGG1S326C polymorphisms and bladder cancer risk, and only one study showed the OGG1S326C variant genotypes with a significantly reduced risk for superficial bladder cancer recurrence (25). This result was consistent with our finding that OGG1 S326C variant genotypes was associated with a significantly reduced risk of bladder cancer in ever smokers.

Two epidemiologic studies examined the effect of APEX1 D148E polymorphisms on cancer. One study found a significant positive association between APEX1 Glu/Glu genotype and lung cancer (26) in a Japanese population, whereas another study is consistent with our finding, in which there was no association between APEX1 genotype and bladder cancer risk (27). Six epidemiologic studies examined the effect of the XRCC1 polymorphisms on bladder cancer. A reduced risk of XRCC1 R399Q homozygous variant Q genotypes compared with those with one or two WTs was observed by Kelsey et al. (28). This association was particularly apparent among heavy smokers in a study by Shen et al. (29). Two other studies by Sanyal et al. (30) and Matullo et al. (31) suggested that XRCC1 R399Q had no effect on the risk of bladder cancer. Stern et al. (32, 33) reported that the XRCC1 R194W

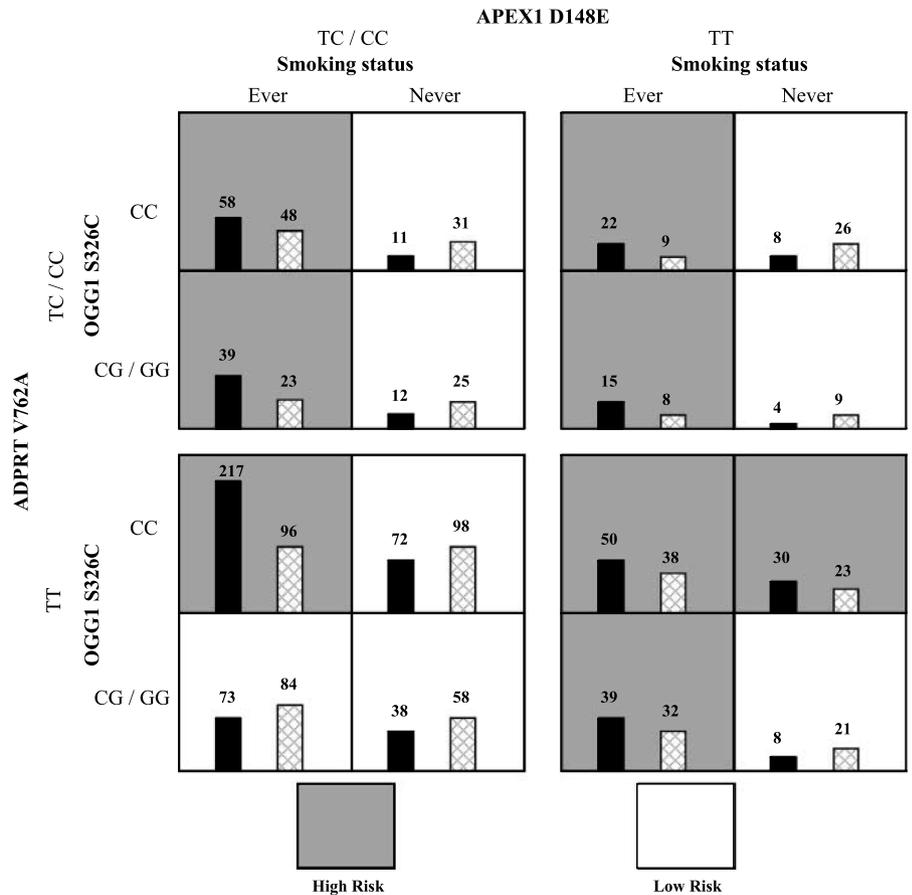


Figure 2. Summary of the four-factor (smoking status, OGG1 S326C, APEX1 D148E, and ADPRT V762A) combinations associated with high risk or low risk, along with the corresponding distribution of cases and controls, for each multilocus-genotype combination. Note that the patterns of high-risk and low-risk cells differ across each of the different multilocus dimensions. *Left, black column, cases (for each cell); right, crossed column, controls (for each cell).*

homozygous variant *W* genotypes have a protective effect on bladder cancer. The ADPRT *Val*¹⁷⁶²*Ala* polymorphism plays an important role in the development of gastric cancer, and the XRCC1 *Arg*³⁹⁹*Gln* polymorphism may serve as a risk modifier (34). Differences in ethnicity and sample size of the study populations and differences in the etiology of different cancer sites might account for some of the discrepancies among previous studies and our data.

The BER pathway involves a serial of critical actions from the genes we investigated in this study. MBD4, MUTYH, and OGG1 are three base-specific glycosylases that have active roles in releasing the modified base and creating a basic site. The APEX endonuclease then incises the DNA strand at the abasic site. XRCC1 functions as scaffold protein in BER by bringing DNA polymerase and ligase together at the site of repair. ADPRT is another important enzymes that can temporarily bind to and protect DNA single-strand interruptions and recruit other repair proteins. The proofreading domain of DNA polymerase δ (encoded by *POLD1* gene)

has a critical role in faithful DNA synthesis in this DNA repair process (35). Because the BER pathway is a group of proteins functioning cooperatively to repair base damages from environmental and exogenous insults, studies designed to analyze an individual gene has the obvious limitations to elucidate the effect of the entire BER pathway. Our results supported our hypothesis that multiple genes and smoking are involved in the predisposition to bladder cancer. The relationship between DNA BER polymorphisms, smoking, and cancer risk may be particularly complex because the effects of genetic variation in the repair process may depend on the presence of a DNA lesion (e.g., gene-environment interaction) or the presence or absence of polymorphisms in other genes in the same or a different pathway (Fig. 2). Thus, we suspect that some of the conflicts between the results of previous studies might also be due to uncharacterized gene-gene or gene-environment interactions.

For studies attempting to examine possible interactions among two or more genetic polymorphisms, traditional

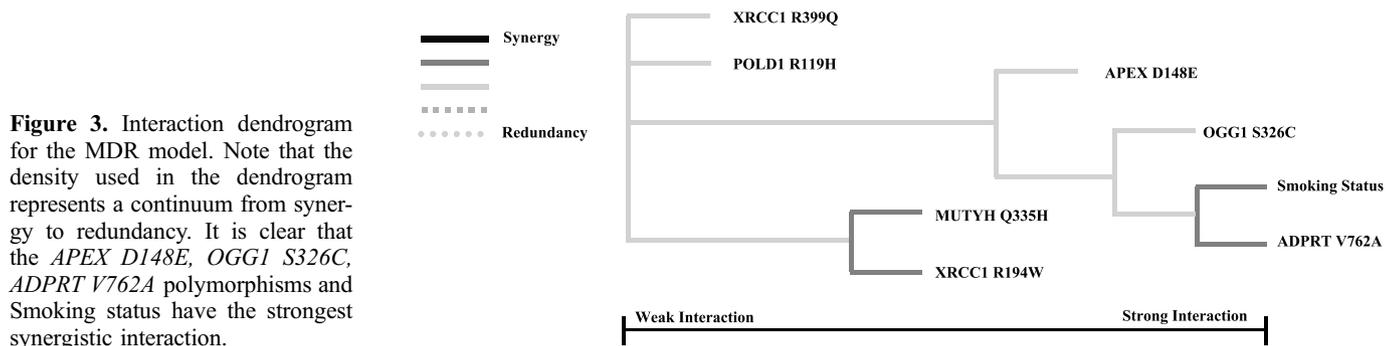


Figure 3. Interaction dendrogram for the MDR model. Note that the density used in the dendrogram represents a continuum from synergy to redundancy. It is clear that the *APEX D148E*, *OGG1 S326C*, *ADPRT V762A* polymorphisms and Smoking status have the strongest synergistic interaction.

methods, such as unconditional logistic regression, may either prove infeasible due to combinations of factors with no observations or have limited power to detect clinically relevant interactions due to a low number of events per variable in the model. The CART and MDR method was proposed as a possible solution in such settings (14-19). Andrew et al. (36) has recently applied MDR to analyze the gene-gene and gene-environment interactions and identified some interesting interactions among DNA repair gene polymorphisms and smoking in a bladder cancer case-control study. These approaches improve statistical power to efficiently identify potential gene-gene and gene-environment interactions. The results of these novel algorithms were consistent with our logistic regression analysis for the two-way interaction models. Using the logistic regression approach, we identified a positive interaction between smoking and ADPRT V762A ($P = 0.019$; Table 2). This is consistent with what we found in Fig. 1 using CART, where we identified that in nonsmokers, ADPRT V762A was the most important factor that influences bladder cancer risk. These findings also agreed with the effect of this SNP on bladder cancer risk stratified by smoking status showed in Table 1, in which ADPRT V762A variant genotypes had a protective role only in never smokers. Similarly, using CART, we also identified that OGG1 S326C was the most important factor in smokers for bladder cancer risk (Fig. 1). Although the interaction between OGG1 S326C and smoking was not significant in logistic regression analysis in Table 2, we did find that OGG1 S326C variant genotypes had a significant association with decreased risk for bladder cancer in ever smokers when relative risk was calculated (Table 1). These different analysis approaches have validated each other and have emphasized the reproducibility of our findings. When never smokers with ADPRT V762A variant genotypes were set as the reference group, using CART, we found that the smokers carrying WT genotype of OGG1 S326C, variant genotypes of XRCC1 R194W, and variant genotypes of MUTYH H335Q had a 31.86-fold (95% CI, 4.01-253.1) increased risk for bladder cancer. These data indicated the significant joint effects between smoking and genetic polymorphisms in the BER pathway.

We attempted to test four-way interactions to replicate our findings from the MDR analysis in logistic regression; however, the model failed to converge due to the small number of individuals in some cells. Thus, our experience highlights the need for alternative, more powerful methods. Of the entire possible two-factor combinations tested, MDR analysis selected smoking and MUTYH-335 as the best two predictors of bladder cancer risk. However, comparing with the one-factor model with smoking status as the only risk factor, this model did not improve on the testing accuracy and had a decreased cross-validation consistency. These data suggested that the two-factor model was not a good choice for bladder cancer risk prediction. Similarly, the three-factor model was worse in both cross-validation consistency and average testing accuracy when compared with the one-factor model. The five-factor model had a similar 100% cross-validation consistency as the one-factor model but had a decreased average testing accuracy. Only the four-factor model, including smoking, APEX1 D148E, ADPRT V762A, and OGG1S326C, was the strongest model overall because it had the highest level of testing accuracy and showed good cross-validation consistency. The MDR four-factor model indicated that smoking, APEX1 D148E, ADPRT V762A, and OGG1 S326C were a high-risk combination of factors but did not specify whether there was a synergistic relationship. Figure 2 helped us interpret the nature of the interactions in these multifactor models. In Fig. 2, we observed that although smoking was an established risk factor for bladder cancer, in some cases, depends on the genotypes the studied individuals were carrying, ever smokers (harboring TT genotype of ADPRT V762A and CG or GG genotype of OGG1

S326C and also TC or CC genotype of APEX D148E) could have low bladder cancer risk and never smokers (harboring TT genotype of ADPRT V762A and CC genotype of OGG1 S326C and TT genotype of APEX D148E) could have higher bladder cancer risk.

In summary, we used the multifaceted analytic approach (CART and MDR) to explore the complex interaction effect between multiple genes and smoking on bladder cancer susceptibility in large case-control populations in Texas. In this study, we have revealed that the interaction relationship among the SNPs, smoking, and bladder cancer risk. These results support the hypothesis that common polymorphisms in DNA repair genes modify bladder cancer risk and emphasize DNA repair is a complex process involving the cooperation of multiple enzymes in DNA BER pathways.

Acknowledgments

We thank Dr. Jason H. Moore (Dartmouth Medical School, Lebanon, NH) for introducing the MDR analysis to us.

References

- Mucci LA, Wedren S, Tamimi RM, Trichopoulos D, Adami HO. The role of gene-environment interaction in the aetiology of human cancer: examples from cancers of the large bowel, lung, and breast. *J Intern Med* 2001;249:477-93.
- Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. *Nature* 2001;411:366-74.
- Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA repair genes. *Science* 2001;291:1284-9.
- Barrett JC. Mechanisms of action of known human carcinogens [review]. *IARC Sci Publ* 1992;116:115-34.
- Hoover RN. Cancer-nature, nurture, or both. *N Engl J Med* 2000;343:135-6.
- Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer-analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343:78-85.
- MacMahon B. Gene-environment interactions in human disease. *J Psychiatr Res* 1968;6:393-402.
- Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2006. *CA Cancer J Clin* 2006;56:106-30.
- Doll R. Cancers related to smoking. *Proceedings Second World Conference on Smoking and Health*. London: Pitman Medical; 1971. p. 10-23.
- Petrovich Z, Baert L, Boyd SD, et al. Management of carcinoma of the bladder [review]. *Am J Clin Oncol* 1998;21:217-22.
- Ward EM, Sabbioni G, DeBord DG, et al. Monitoring of aromatic amine exposures in workers at a chemical plant with a known bladder cancer excess. *J Natl Cancer Inst* 1996;88:1046-52.
- Wilson DM III, Thompson LH. Life without DNA repair. *Proc Natl Acad Sci U S A* 1997;94:12754-7.
- Friedberg EC, McDaniel LD, Schultz RA. The role of endogenous and exogenous DNA damage and mutagenesis [review]. *Curr Opin Genet Dev* 2004;14:5-10.
- Zhang H, Singer B. *Recursive partitioning in the health sciences*. New York: Springer; 1999.
- Ritchie MD, Hahn LW, Roodi N, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* 2001;69:138-47.
- Hahn LW, Ritchie MD, Moore JH. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics* 2003;19:376-82.
- Ritchie MD, Hahn LW, Moore JH. Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. *Genet Epidemiol* 2003;24:150-7.
- Moore JH. Computational analysis of gene-gene interactions using multifactor dimensionality reduction. *Expert Rev Mol Diagn* 2004;4:795-803.
- Moore JH, Gilbert JC, Tsai CT, et al. A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *J Theor Biol* 2006;241:252-61.
- Rossner P, Jr., Terry MB, Gammon MD, et al. OGG1S326C polymorphisms and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2006;15:811-5.
- Hansen R, Saebo M, Skjelbred CF, et al. GPX Pro¹⁹⁸Leu and OGG1S326C Ser³²⁶Cys polymorphisms and risk of development of colorectal adenomas and colorectal cancer. *Cancer Lett* 2005;229:85-91.
- Takezaki T, Gao CM, Wu JZ, et al. hOGG1S326C Ser(326)Cys polymorphism and modification by environmental factors of stomach cancer risk in Chinese. *Int J Cancer* 2002;99:624-7.
- Arcand SL, Provencher D, Mes-Masson AM, Tonin PN. OGG1S326C Cys³²⁶

- variant, allelic imbalance of chromosome band 3p25.3, and TP53 mutations in ovarian cancer. *Int J Oncol* 2005;27:1315–20.
24. Chevillard S, Radicella JP, Levalois C, et al. Mutations in OGG1S326C, a gene involved in the repair of oxidative DNA damage, are found in human lung and kidney tumours. *Oncogene* 1998;16:3083–6.
 25. Kim EJ, Jeong P, Quan C, et al. Genotypes of TNF- α , VEGF, hOGG1S326C, GSTM1, and GSTT1: useful determinants for clinical outcome of bladder cancer. *Urology* 2005;65:70–5.
 26. Ito H, Matsuo K, Hamajima N, et al. Gene-environment interactions between the smoking habit and polymorphisms in the DNA repair genes, APE1 Asp¹⁴⁸Glu, and XRCC1 Arg³⁹⁹Gln, in Japanese lung cancer risk. *Carcinogenesis* 2004;25:1395–401.
 27. Terry PD, Umbach DM, Taylor JA. APE1 genotype and risk of bladder cancer: evidence for effect modification by smoking. *Int J Cancer* 2006;118:3170–3.
 28. Kelsey KT, Park S, Nelson HH, Karagas MR. A population-based case-control study of the XRCC1 Arg³⁹⁹Gln polymorphism and susceptibility to bladder cancer. *Cancer Epidemiol Biomarkers Prev* 2004;13:1337–41.
 29. Shen M, Hung RJ, Brennan P, et al. Polymorphisms of the DNA repair genes XRCC1, XRCC3, XPD, interaction with environmental exposures, and bladder cancer risk in a case-control study in northern Italy. *Cancer Epidemiol Biomarkers Prev* 2003;12:1234–40.
 30. Sanyal S, Festa F, Sakano S, et al. Polymorphisms in DNA repair and metabolic genes in bladder cancer. *Carcinogenesis* 2004;25:729–34.
 31. Matullo G, Guarrera S, Carturan S, et al. DNA repair gene polymorphisms, bulky DNA adducts in white blood cells and bladder cancer in a case-control study. *Int J Cancer* 2001;92:562–7.
 32. Stern MC, Umbach DM, van Gils CH, Lunn RM, Taylor JA. DNA repair gene XRCC1 polymorphisms, smoking, and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev* 2001;10:125–31.
 33. Stern MC, Umbach DM, Lunn RM, Taylor JA. DNA repair gene XRCC3 codon 241 polymorphism, its interaction with smoking and XRCC1 polymorphisms, and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev* 2002;11:939–43.
 34. Zhang Z, Miao XP, Tan W, Guo YL, Zhang XM, Lin DX. Correlation of genetic polymorphisms in DNA repair genes ADPRT and XRCC1 to risk of gastric cancer. *Ai Zheng* 2006;25:7–10.
 35. Matsumoto Y. Molecular mechanism of PCNA-dependent base excision repair. *Prog Nucleic Acid Res Mol Biol* 2001;68:129–38.
 36. Andrew AS, Nelson HH, Kelsey KT, et al. Concordance of multiple analytical approaches demonstrates a complex relationship between DNA repair gene SNPs, smoking, and bladder cancer susceptibility. *Carcinogenesis* 2006;27:1030–7.

BLOOD CANCER DISCOVERY

High-Order Interactions among Genetic Variants in DNA Base Excision Repair Pathway Genes and Smoking in Bladder Cancer Susceptibility

Maosheng Huang, Colin P. Dinney, Xin Lin, et al.

Cancer Epidemiol Biomarkers Prev 2007;16:84-91.

Updated version Access the most recent version of this article at:
<http://cebp.aacrjournals.org/content/16/1/84>

Cited articles This article cites 33 articles, 7 of which you can access for free at:
<http://cebp.aacrjournals.org/content/16/1/84.full#ref-list-1>

Citing articles This article has been cited by 2 HighWire-hosted articles. Access the articles at:
<http://cebp.aacrjournals.org/content/16/1/84.full#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

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

Permissions To request permission to re-use all or part of this article, use this link
<http://cebp.aacrjournals.org/content/16/1/84>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.