Published in final edited form as:

Cancer. 2012 January 1; 118(1): 205–215. doi:10.1002/cncr.26224.

Comprehensive pathway-based interrogation of genetic variations in the nucleotide excision DNA repair pathway and risk of bladder cancer

Jinliang Xing^{1,2}, Colin P. Dinney³, Sanjay Shete², Maosheng Huang², Michelle A. Hildebrandt², Zhinan Chen^{1,4}, and Jian Gu^{2,4,5}

¹ State Key Laboratory of Cancer Biology, Cell Engineering Research Centre & Department of Cell Biology, Fourth Military Medical University, Xi'an, 710032, P. R. China

² Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030

³ Department of Urology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030

Abstract

Background—Growing evidence suggests that single nucleotide polymorphisms (SNPs) in nucleotide excision repair (NER) pathway genes play an important role in bladder cancer etiology. However, only a limited number of genes and variations in this pathway have been evaluated to date.

Methods—In this study, we applied a comprehensive pathway-based approach to assess the effects of 207 tagging and potentially functional SNPs in 26 NER genes on bladder cancer risk using a large case-control study consisting of 803 bladder cancer cases and 803 controls.

Results—A total of 17 SNPs were significantly associated with altered bladder cancer risk at $P < 0.05$, of which 7 SNPs retained noteworthiness after assessed by a Bayesian approach for the probability of false discovery. The most noteworthy SNP was rs11132186 in *ING2* gene. Compared to the major allele-containing genotypes, the odds ratio (OR) was 0.52 (95% confidence interval [CI] 0.32–0.83, $P = 0.005$) for the homozygous variant genotype. Three additional *ING2* variants also exhibited significant associations with bladder cancer risk. Significant gene-smoking interactions were observed for three of the top 17 SNPs. Furthermore, through an exploratory classification and regression tree (CART) analysis, we identified potential gene-gene interactions.

Conclusions—We conducted a large association study of NER pathway with bladder cancer risk and identified several novel predisposition variants. We identified potential gene-gene and gene-environment interactions in modulating bladder cancer risk. Our results reinforce the importance of a comprehensive pathway-focused and tagging SNP-based candidate gene approach to identify low-penetrance cancer susceptibility loci.

Keywords

bladder cancer; genetic susceptibility; nucleotide excision repair; SNP; gene-smoking interaction

⁵To whom requests for reprints should be addressed, at Department of Epidemiology, Unit 1340, The University of Texas MD Anderson Cancer Center, 1155 Pressler Blvd., Houston, TX 77030. Phone: 713-792-8016; Fax: 713-792-4657; jjangu@mdanderson.org.

⁴These authors contributed equally to this work.

Financial disclosures: The authors have no financial disclosures related to the content of this article.

INTRODUCTION

Bladder cancer ranks ninth in worldwide cancer incidence. It is the seventh most common cancer in men and seventeenth in women¹. In the United States, it is the fifth most common cancer, fourth in men and eleventh in women². Environmental exposures account for the majority of bladder cancer cases. For example, around half of bladder cancer incidence in men and one third in women are due to tobacco smoking. In addition, occupational exposures to aromatic amines and polycyclic aromatic hydrocarbons, and polluted drinking water containing arsenic and chlorination by-product also significantly contribute to the development of bladder cancer³. Other environmental risk factors such as dietary factors, hair dye usage, artificial sweeteners, and phenacetin-containing analgesic drugs have also been reported, although the associations have not been consistent across different studies³.

Although the molecular mechanisms underlying these bladder cancer etiological factors are not fully understood, it is widely recognized that environmental carcinogens induce DNA damages that lead to genomic instability. Tobacco carcinogens mainly induce bulky DNA adducts and the nucleotide excision repair (NER) is the major cellular pathway to repair bulky DNA adducts. Other cellular DNA repair pathways, such as base excision repair (BER) and double-strand break repair (DSBR), also play important roles in the prevention of bladder carcinogenesis through repairing single strand and double strand DNA breaks caused by smoking, reactive oxygen species (ROS), ionizing radiation, and other DNA damaging agents⁴⁻⁷. In bladder cancer, NER pathway is one of the most commonly studied pathways that repairs bulky DNA lesions, such as pyrimidine dimers, photo-products, larger chemical adducts and cross-links. NER pathway is also critical to the maintenance of genomic stability. There are two types of NER processes in human cells: global genomic repair (GGR) and transcription-coupled repair (TCR). Both processes have four major steps: (1) recognition of DNA lesions by a complex of interactive proteins, including XPC-RAD23B, XPA and RPA in GGR, or ERCC6 and Cockayne syndrome type A (CSTA) in TCR; (2) unwinding of DNA strands within the region of lesions by transcription factor IIH (TFIIH) complex that includes XPD and XPB proteins; (3) elimination of damaged DNA fragments by a protein complex including ERCC1, XPF, and XPG; and (4) synthesis of new DNA strands by various DNA polymerases⁸. Defects in these critical genes have been frequently identified in many cancers including bladder cancer^{9, 10}.

Single nucleotide polymorphisms (SNPs) of NER pathway genes have been implicated in bladder cancer etiology⁹. For example, we and others have previously reported that SNPs in major NER genes such as *XPC*, *XPD*, *CCNH*, *RAD23B* genes were significantly associated with bladder cancer risk¹¹⁻¹⁵. However, these studies mostly took a limited candidate gene approach that evaluates a small number of genes and potential functional SNPs. In this study, we performed a comprehensive pathway-based study to assess the effects of 207 haplotype-tagging and potentially functional SNPs in 26 major genes in the NER pathway on bladder cancer risk in a total of 803 bladder cancer patients and 803 healthy controls. We then conducted a series of exploratory analyses to evaluate the cumulative effects and interaction of these variants on bladder cancer risk.

MATERIALS AND METHODS

Study population and epidemiology data

The study participants were derived from an ongoing hospital-based bladder cancer case-control study. The cases were newly diagnosed, histologically confirmed, and previously untreated bladder cancer patients recruited at The University of Texas MD Anderson Cancer Center and Baylor College of Medicine between 1999 and 2007. There were no age, gender,

ethnicity, and cancer-stage restrictions on the recruitment. The controls were healthy individuals with no prior history of cancer (except non-melanoma skin cancer) recruited from Kelsey Seybold Clinic, the largest multi-specialty, managed-care physician group in the Houston metropolitan area. Controls were matched to cases by age (± 5 years), gender and ethnicity. To control for confounding of population stratification, we restricted both cases and controls to self-reported non-Hispanic Caucasians for the current analysis. These cases and controls have been included in our recent genome-wide association study of bladder cancer and there was no evidence of population substructure among cases and controls¹⁶. The potential controls were first surveyed by a Kelsey-Seybold staff member during clinical registration using a short questionnaire to elicit their willingness to participate in the study and to provide preliminary demographic data for matching. They were contacted by telephone at a later date to schedule an interview appointment at a Kelsey-Seybold clinic convenient to the participant. The response rate for the ongoing study was 92% for cases and 76.7% for controls. All cases and controls in this study completed a 45-minute structured questionnaire administered by trained MD Anderson staff interviewers. The questionnaire collected information about demographics, smoking history, family history of cancer, and medical history. At the end of the interview, a 40-mL blood sample was drawn into a coded heparinized tube and sent to the laboratory for immediate DNA extraction and molecular analyses. The study was approved by all relevant institutional review boards, and the signed informed consent was obtained from each participant.

Selection of genes and polymorphisms

A comprehensive list of genes in the NER pathway was developed through the interrogation of the Gene Ontology (GO) database (<http://www.geneontology.org>) and PubMed-based literature review, as previously described¹⁷. Tagging SNPs were selected by the binning algorithm of LDSelect software (<http://droog.gs.washington.edu/LdSelect.pl>) with r^2 threshold of 0.8 and minor allele frequencies (MAF) > 0.05 within 10 kb upstream of the 5' untranslated region (UTR) and 10 kb downstream of the 3' UTR of each gene. We also included a few non-synonymous SNPs identified in the dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>). Since this is a gene-centered candidate region search, we achieved 100% coverage of our targeted genomic regions. The number of SNPs for each gene region was as follows: CENH, 3; CDK7, 2; DDB2, 9; ERCC1, 2; ERCC2, 8; ERCC3, 5; ERCC4, 8; ERCC5, 14; ERCC6, 12; ERCC8, 12; GTF2H1, 6; GTF2H3, 3; GTF2H4, 6; GTF2H5, 3; ING2, 6; LIG1, 9; MMS19L, 4; MNAT1, 9; RAD23A, 3; RAD23B, 17; RPA1, 15; RPA2, 3; RPA3, 28; XAB2, 5; XPA, 7; XPC, 8.

Genotyping

Genomic DNA was extracted from peripheral blood using the Qiagen Whole Blood DNA Extraction Kit (Qiagen, Valencia, CA). The genotyping was performed using the iSelect Infinium II platform according to Illumina's protocol, as described previously^{12, 17}. Briefly, 750 ng DNA of each sample was amplified 1,000–1,500 fold overnight. The amplified DNA was fragmented, precipitated, and resuspended before being hybridized to the iSelect Beadchip, which contains SNP-locus-specific oligonucleotide primers (50 bp long) covalently attached to the bead surface. After specific hybridization of genomic DNA to the bead array, each SNP locus-specific primer (attached to beads) was extended with a single hapten-labeled dideoxynucleotide in a single base extension reaction. Incorporated haptens were converted to fluorescent signal by multilevel immunohistochemistry staining and imaged using the BeadStation Scanner. The genotype for each SNP was auto-called using the BeadStudio software package and processed for further statistical analyses. The average call rate for SNPs is $> 99\%$. Individuals with $> 5\%$ missing genotypes, and SNPs with $> 5\%$ missing calls were excluded from downstream analyses. Randomly selected 2% of samples

were run in duplicates and the concordance of SNP genotype calls was >99.9% for duplicated samples.

Statistical analysis

Statistical analyses were performed using Intercooled STATA software, version 10.1 (STATA Corp., College Station, TX) and SAS/Genetics, version 9.0 (SAS Institute). χ^2 test was used to assess the differences between cases and controls with regard to categorical variables such as gender and smoking status. Student's *t* test was used to test for continuous variables, including age and pack-years. The Hardy-Weinberg equilibrium (HWE) was tested using a goodness-of-fit χ^2 test. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated using multivariable logistic regression to identify SNPs with significant associations with bladder cancer risk, adjusting for confounding factors including age, gender, and smoking status. For each SNP, genotypes containing the homozygous major allele were used as the reference group to calculate the ORs and 95% CIs for genotypes containing the variant allele. The definitions of smoking status were the same as previously described¹⁸. *P* value was calculated using the likelihood ratio test comparing the model with and without the variables of interest in multivariate logistic regression. Dominant, recessive and additive models were tested for each SNP and the reported *P* value was the smallest of the above three tests. We assessed the noteworthiness of an observed significant association using a Bayesian False Discovery Probability (BFDP) approach proposed by Wakefield¹⁹. The approach has been adopted and used in a number of epidemiologic studies to control for multiple testing^{20–25}. We used all *p*-values from different inheritance models (additive, recessive, and dominant) to perform the BFDP analyses. We adjusted all the 207 tests at the SNP level. We assumed a range of prior probability from 0.01 to 0.05. We set the prior probability that OR is bigger than 2.0 as 0.025. An association was declared as being significant if BFDP was below 0.8 for prior probability 0.05 (considered to be a moderate prior for pathway-based association study)¹⁹. The test for interaction between genotypes and smoking was done by including an interaction term in the logistic regression. We tested for interactions between SNPs and smoking status (never and ever) and between SNPs and packyears of smoking (as continuous variable). For the SNPs with significant main effects, we reported results of stratified analyses by smoking status and the interactions of these SNPs with smoking status. We also reported SNPs with nominally significant interactions with packyears of smoking ($P < 0.05$). We also performed exploratory classification and regression tree (CART) analysis to identify potential gene-gene interactions using the HelixTree software (version 4.1.0, Golden Helix). CART is a binary recursive partitioning method that produces a decision tree to identify subgroups of subjects at different risk level. Specifically, the recursive partitioning algorithm starts at the first node (with the entire data set) and determines the first locally optimal split and each subsequent split of the dataset with multiplicity adjusted *P* values to control tree growth. We used $P = 0.001$ to grow the tree and $q = 0.05$ to prune the over-grown tree and control the final tree size. All *P* values in this study were two-sided.

RESULTS

Characteristics of the study population

Our participants consisted of 803 Caucasian bladder cancer patients and 803 age (± 5 years) and gender frequency-matched Caucasian controls (Table 1). Cases had a significantly higher percentage of current smokers and recent-quitters than controls (23.3% vs. 8.3%, $P = 5.2 \times 10^{-21}$). Among ever smokers, cases reported significantly higher levels of cigarette consumption than controls (mean pack-years: 43.0 vs. 29.9, $P = 2.8 \times 10^{-12}$).

Main-effect analysis of Individual NER SNPs

Among the 207 NER SNPs, 17 (8.2%) were significantly associated with bladder cancer risk at the 5% level (Table 2), among which one SNP (rs4151330) had a statistically significant deviation from HWE ($P < 0.05$) in controls, consistent with that expected by chance. The most noteworthy finding of our study was related to *ING2* gene. Four of the six *ING2* SNPs conferred a significantly altered risk of bladder cancer; the most significant one was rs11132186, a SNP located in the 3' region of *ING2*. Under a recessive genetic model, the homozygous variant genotype of rs11132186 was associated with a reduced risk of bladder cancer (OR=0.52, 95% CI 0.32–0.83, $P=0.005$). Another *ING2* SNP in the 3' region, rs11735038, also conferred a reduced bladder cancer risk (OR=0.66, 95% CI 0.49–0.90, $P=0.008$) under a recessive genetic model. The other two significant *ING2* SNPs were rs6854224 in the 5' region (OR=0.70, 95% CI 0.53–0.93, $P=0.013$, recessive model) and rs11732255 in the 3' region (OR=0.84, 95% CI 0.72–0.98, $P=0.025$, additive model) (Table 2). There are two additional SNPs exhibiting a best-fitting P value less than 0.01 and conferred an increased bladder cancer risk: rs11039130 in the 5' region of *DDB2* (OR=1.64, 95% CI 1.14–2.35, $P=0.007$) and rs4150667 in the intron of *GTF2H1* (OR=1.55, 95% CI 1.12–2.15, $P=0.008$). In addition, we further assessed the noteworthiness of these significant associations using BFDP. We found that 7 out of the 17 significant SNPs had a BFDP value less than 0.8 (range 0.72–0.80), suggesting the potential noteworthiness of their associations with bladder cancer risk (Table 2). Table 3 shows the detailed distributions of different genotypes (homozygous major allele, heterozygote, and homozygous variant) of these 7 SNPs in cases and controls and the associations of each genotype with bladder cancer risk as well as the best fitting model for each SNP.

Stratified analyses by smoking status

We also performed stratified analysis by smoking status (data not shown). There were significant overlaps of top SNPs between analysis of overall population and smokers, for example, the top two SNPs in smokers, *DDB2* rs1685404 (OR=0.67, 95% CI, 0.52–0.86, $P=0.0021$ under dominant model) and *GTF2H1* rs4150667 (OR=1.90, 95% CI, 1.24–2.89, $P=0.0024$ under recessive model), were the 8th and 4th most significant SNPs in overall analysis. For never smokers, the most significant SNP, *XPA* rs10817938 (OR=0.26, 95% CI, 0.12–0.60, $P=0.0003$ under dominant model), was not significant in overall analysis ($P=0.67$). We then tested interactions between significant SNPs and smoking status in modulating bladder cancer risk. Many of the top SNPs exhibited similar effect on bladder cancer risk in never and ever smokers, but there were significant SNP-smoking interactions for a few SNPs (Table 2). *CCNH* rs10065575 was the 3rd most significant SNP in never smokers ($P=0.0012$), but was not significant in ever smokers, and the test of interaction showed a significant interaction of this SNP with smoking status ($P=0.034$, Table 2). Significant interactions with smoking status were also observed for *DDB2* rs1685404 ($P=0.038$) and *ERCC6* rs4453140 ($P=0.009$) (Table 2).

SNP-smoking (packyears) interactions

We then performed detailed analysis of the interactions between all the assayed SNPs and packyears of smoking (as continuous variable). Table 4 listed the 17 SNPs showing nominally significant interactions with packyears ($P < 0.05$). The most significant SNP was rs4151150 in *MNAT1* that exhibited a significant negative interaction with packyears ($\beta = -0.026$, $P=0.0018$). Among the 17 SNPs, only rs1685404 on *DDB2* showed significant main effect. The other two SNPs with significant main effect as well as significant interactions with smoking status (*CCNH* rs10065575 and *ERCC6* rs4453140 in Table 2) were not significant in this analysis, suggesting that quantitative measures of smoking provide more information than smoking status. Since we did not have sufficient power to detect SNP-

smoking interaction, this analysis was exploratory and no multiple testing correction was applied.

Haplotype analysis

We performed a haplotype analysis for *ING2* and *DDB2*, the two genes with several significant SNPs (Table 5). We calculated the pairwise D' to measure the linkage disequilibrium between SNPs in each gene and found no significant linkages. (D' range, 0.13–0.51 for *ING2* SNPs, 0.18–0.31 for *DDB2* SNPs, data not shown). A total of 12 and 9 haplotypes with a frequency of more than 1% were identified for *ING2* and *DDB2*, respectively. For *ING2*, compared to the most common haplotype H1, three minor haplotypes, H8, H10, and H11, were associated with decreased bladder cancer risks with ORs of 0.60 (95% CI 0.39–0.93), 0.43 (0.20–0.90), and 0.42 (0.20–0.89), respectively. For *DDB2*, compared to H1, haplotype H6 was associated with a decreased bladder cancer risk (OR=0.67, 95% CI 0.50–0.91, $P=0.009$) (Table 5).

CART analysis

CART analysis uses a binary recursive partitioning method to identify subgroups of high-risk subjects and detects higher-order interactions among a large number of variables. Figure 1A depicts the resulting tree structure generated by CART analysis. The initial split was rs1051315 in *RPA1* gene. In subjects with the homozygous major allele-containing genotype of rs1051315, the tree structure was further generated according to the genotype information of *CCNH* rs10065575, *ERCC2* rs1799787, *RPA1* rs3744467, *ING2* rs6854224, *MNAT1* rs1885094, and *MNAT1* rs4151330, resulting in different subgroups (terminal nodes), each with a distinct combination of genotypes and a different risk estimate. Figure 1B summarizes the risk estimates for individuals in each terminal node. Compared with individuals in terminal node 1, individuals in terminal node 3 exhibited a significantly increased bladder cancer risk (OR=2.58, 95% CI, 1.56–4.26, $P=2.1\times 10^{-4}$), whereas individuals in terminal node 6 had a significantly reduced risk (OR=0.42, 95% CI, 0.26–0.67, $P=4.0\times 10^{-4}$). We tested the interactions between SNPs identified from this CART analysis. There were significant interactions between *CCNH* rs10065575 and *ING2* rs6854224 ($P=0.014$), *ING2* rs6854224 and *MNAT1* rs4151330 ($P=0.017$), and between *CCNH* rs10065575 and *MNAT1* rs1885094 ($P=0.043$). These interactions resulted in terminal nodes 4 to 7. Due to the post-hoc data mining nature of CART analysis, these results are exploratory.

DISCUSSION

In this study, we assessed the effects of a comprehensive panel of 207 SNPs in 26 genes in the NER pathway on the risk of bladder cancer. *ING2* gene was the most noteworthy finding, with four of six evaluated SNPs exhibiting significant associations with bladder cancer risk. Furthermore, we found potential gene-smoking interaction and higher-order interactions among these NER SNPs in the modulation of bladder cancer susceptibility.

Our findings of *ING2* are biologically plausible. *ING2* is a member of the inhibitor of growth (ING) gene family and encodes a putative tumor suppressor protein involved in the regulation of DNA repair, cell cycle progression, apoptosis, and epigenetic functions in a p53-dependent manner. *ING2* gene was first cloned in 1998 as a homologue of the first ING family member, *ING1*²⁶. *ING2* gene is 6 kb in length, located in chromosome region 4q35.1. The implication of *ING2* in NER was first established by Wang et al., who found that overexpression of *ING2* gene significantly enhanced the repair of UV-induced DNA damage²⁷. This function of *ING2* is dependent on the normal functions of p53 protein since small interfering RNA (siRNA)-mediated degradation of either *ING2* or p53 abolished the

observed repair capacity²⁷. Furthermore, Want et al. showed that ING2 protein is not a component of the NER core protein complex; instead, ING2 enhance NER through recruiting XPA to the core complex²⁷. The function of ING2 in DNA repair is also mediated by the interaction with the trimethylated and dimethylated H3K4, which stabilizes the mSin3a-HDAC1 protein complex to enhance the transcriptional activity of many relevant genes²⁸. In addition to its involvement in NER, ING2 has also been implicated in G1 phase cell cycle arrest through increasing the transcriptional activation ability of p53²⁹. Furthermore, ING2 interacts with phosphoinositides to activate p53-dependent apoptosis pathway³⁰. In our study, all 4 significant *ING2* SNPs, which are not in strong linkage, were associated with reduced risks of bladder cancer. Among them, 3 SNPs are located in the 3' region of gene and one located in the 5' region of *ING2*, 5.5 kb upstream of the translation start codon. Since the promoter and enhancer sequences of *ING2* have not been well-characterized experimentally, it remains to be determined whether these SNPs have any functional significance. It is more likely that they are tagging SNPs but not the causal variants. Therefore, high-density mapping in combination with functional characterizations are warranted to further elucidate the molecular mechanisms underlying the association between *ING2* SNPs and bladder cancer risk observed in this study.

Another noteworthy gene we found in this study is *DDB2*. *DDB2* gene contains 8 exons and spans approximately 24 kb in chromosomal 11p12-p11³¹. As a response protein to DNA damage induced by genotoxic agents such as UV-irradiation, DDB2 interacts with CUL4A, RBX1, and COPS2 to form a protein complex that binds to chromatin and initiates the NER process³². DDB2 enhances the DNA binding activity of DDB1 and serves as an crucial component in p53-mediated DNA repair process^{33, 34}. Two SNPs of *DDB2* exhibited a significant association with bladder cancer risk. One SNP, rs11039130, located ~7 kb upstream of transcription start site, was associated with a 1.64-fold increased risk under a recessive model. The other SNP, rs1685404, is located in the 3' of the gene and conferred a reduced bladder cancer risk under a dominant model. Several studies have reported that the transcription of *DDB2* gene is tightly regulated by a wide array of transcription factors such as E2F family, BRCA1, SP1, Myc, and NF1³⁵⁻³⁷. However, whether rs1685404 is a direct causative locus or surrogate remains to be determined through further fine mapping and functional characterization.

We performed stratified analysis by smoking status and tested SNP-smoking (smoking status and packyears) interactions. Due to the limited power to detect SNP-smoking interaction, this analysis was exploratory and further validation with larger sample size is needed. Interestingly, the effects of *CCNH* and *ERCC6* SNPs were stronger in never smokers, whereas the effect of *DDB2* SNP was only evident in ever smokers and other SNPs showed similar effects in never and ever smokers. This observation is in line with a recent genome-wide association study (GWAS) of bladder cancer in which the *NAT2* slow acetylator genotype was associated with an increased bladder cancer risk in ever smokers but not never smokers, but the effect of *GSTM1* null genotype was the strongest in never smokers and grew progressively weaker in former and current smokers³⁸. Eight other GWAS-confirmed SNPs showed similar effect among never and ever smokers³⁸. It is intriguing that different genotypes in carcinogen metabolism and DNA repair exhibited differential effect on bladder cancer risk among never and ever smokers. Other exposures may explain such interactions. Occupational exposure is the second major environmental risk factor for bladder cancer. Our previous publication has shown that prolonged exposure to diesel fuel or fumes on a regular basis, exposure to tar/mineral oil, dry cleaning fluids, leather and tanning solutions, rubber products, glues, pesticides, insecticides, or herbicides, fertilizers, arsenic, zinc, radioactive materials, and aromatic amine, were all associated with an increased risk of bladder cancer³⁹. It would be interesting to assess the NER SNPs and bladder cancer risk in the context of these different DNA damaging exposures. Since only a

small percentage of our study populations were exposed to these different occupational exposures, the power to detect significant associations in exposed populations was limited. Future studies are warranted to address this question.

We also conducted exploratory CART analyses to assess potential higher-order gene-gene interactions within the NER pathway genes. The process of tumorigenesis in sporadic cancers is a multifactorial and multistep process that involves complicated interactions of various low-penetrance genetic and environmental components. The CART analysis identified subsets of individuals with different cancer risks based on different combinations of genotypes, with the ORs of individuals in each terminal node ranging from 0.42 to 2.58 (Figure 1). A paired interaction analysis supports some of the SNP-SNP interactions. These data suggest that gene-gene interactions play an important role in bladder cancer etiology. Nevertheless, given that CART analysis is a post-hoc data mining tool that was applied to the same dataset, the results are preliminary and should be interpreted with caution.

The strengths of our study include a large and homogenous study population, a comprehensive panel of genes in the relatively well-characterized NER pathway, and the use of an htSNP-based genotyping approach. The limitation of this study is that although we applied BFDP approach, one of several available statistical methods, to control for multiple testing, it is possible that some of our reported SNPs are false-positive findings. The main reason for our choice of BFDP is that the noteworthy threshold defined in BFDP approach accounts for the costs of false discovery and non-discovery. The other alternative approaches to correct for multiple testing, such as the Bonferroni correction for independent tests and the P(ACT) method to compute *P* values adjusted for correlated tests⁴⁰, do not consider the cost of non-discovery and are more conservative. Regardless of the method used for correcting multiple testing, the ultimate way to eliminate false positive findings is through independent validation. We reported both the significant SNPs after multiple testing correction by the BFDP method and the nominally significant SNPs in main effect and SNP-smoking interaction analyses. External validations in independent epidemiology studies with adequate sample sizes are warranted to confirm the results of our studies.

Acknowledgments

Funding sources: This study was supported by National Cancer Institute grants CA131335, CA74880, CA91846, and CA127615.

References

1. Ploeg M, Aben KK, Kiemeny LA. The present and future burden of urinary bladder cancer in the world. *World J Urol.* 2009; 27(3):289–93. [PubMed: 19219610]
2. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics. *CA Cancer J Clin.* 2010; 60(5):277–300. [PubMed: 20610543]
3. Wu X, Ros MM, Gu J, Kiemeny L. Epidemiology and genetic susceptibility to bladder cancer. *BJU Int.* 2008; 102(9 Pt B):1207–15. [PubMed: 19035883]
4. 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(2): 125–31. [PubMed: 11219769]
5. Joenje H. Genetic toxicology of oxygen. *Mutat Res.* 1989; 219(4):193–208. [PubMed: 2671707]
6. Bentley J, Diggle CP, Harnden P, Knowles MA, Kiltie AE. DNA double strand break repair in human bladder cancer is error prone and involves microhomology-associated end-joining. *Nucleic Acids Res.* 2004; 32(17):5249–59. [PubMed: 15466592]
7. Park SL, Bastani D, Goldstein BY, Chang SC, Cozen W, Cai L, et al. Associations between NBS1 polymorphisms, haplotypes and smoking-related cancers. *Carcinogenesis.* 31(7):1264–71. [PubMed: 20478923]

8. Bergoglio V, Magnaldo T. Nucleotide excision repair and related human diseases. *Genome Dyn.* 2006; 1:35–52. [PubMed: 18724052]
9. Kiltie AE. Molecular epidemiology of DNA repair genes in bladder cancer. *Methods Mol Biol.* 2009; 472:281–306. [PubMed: 19107438]
10. van Steeg H. The role of nucleotide excision repair and loss of p53 in mutagenesis and carcinogenesis. *Toxicol Lett.* 2001; 120(1–3):209–19. [PubMed: 11323179]
11. Gu J, Zhao H, Dinney CP, Zhu Y, Leibovici D, Bermejo CE, et al. Nucleotide excision repair gene polymorphisms and recurrence after treatment for superficial bladder cancer. *Clin Cancer Res.* 2005; 11(4):1408–15. [PubMed: 15746040]
12. Chen M, Kamat AM, Huang M, Grossman HB, Dinney CP, Lerner SP, et al. High-order interactions among genetic polymorphisms in nucleotide excision repair pathway genes and smoking in modulating bladder cancer risk. *Carcinogenesis.* 2007; 28(10):2160–5. [PubMed: 17728339]
13. Neumann AS, Sturgis EM, Wei Q. Nucleotide excision repair as a marker for susceptibility to tobacco-related cancers: a review of molecular epidemiological studies. *Mol Carcinog.* 2005; 42(2):65–92. [PubMed: 15682379]
14. Garcia-Closas M, Malats N, Real FX, Welch R, Kogevinas M, Chatterjee N, et al. Genetic variation in the nucleotide excision repair pathway and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2006; 15(3):536–42. [PubMed: 16537713]
15. Andrew AS, Karagas MR, Nelson HH, Guarrera S, Polidoro S, Gamberini S, et al. DNA repair polymorphisms modify bladder cancer risk: a multi-factor analytic strategy. *Hum Hered.* 2008; 65(2):105–18. [PubMed: 17898541]
16. Wu X, Ye Y, Kiemeny LA, Sulem P, Rafnar T, Matullo G, et al. Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. *Nat Genet.* 2009; 41(9):991–5. [PubMed: 19648920]
17. Wu X, Spitz MR, Lee JJ, Lippman SM, Ye Y, Yang H, et al. Novel susceptibility loci for second primary tumors/recurrence in head and neck cancer patients: large-scale evaluation of genetic variants. *Cancer Prev Res (Phila Pa).* 2009; 2(7):617–24.
18. Leibovici D, Grossman HB, Dinney CP, Millikan RE, Lerner S, Wang Y, et al. Polymorphisms in inflammation genes and bladder cancer: from initiation to recurrence, progression, and survival. *J Clin Oncol.* 2005; 23(24):5746–56. [PubMed: 16110031]
19. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet.* 2007; 81(2):208–27. [PubMed: 17668372]
20. Zienoldind S, Campa D, Lind H, Ryberg D, Skaug V, Stangeland LB, et al. A comprehensive analysis of phase I and phase II metabolism gene polymorphisms and risk of non-small cell lung cancer in smokers. *Carcinogenesis.* 2008; 29(6):1164–9. [PubMed: 18258609]
21. McKay JD, Hashibe M, Hung RJ, Wakefield J, Gaborieau V, Szeszenia-Dabrowska N, et al. Sequence variants of NAT1 and NAT2 and other xenometabolic genes and risk of lung and aerodigestive tract cancers in Central Europe. *Cancer Epidemiol Biomarkers Prev.* 2008; 17(1):141–7. [PubMed: 18199719]
22. Hoefl B, Linseisen J, Beckmann L, Muller-Decker K, Canzian F, Husing A, et al. Polymorphisms in fatty-acid-metabolism-related genes are associated with colorectal cancer risk. *Carcinogenesis.* 31(3):466–72. [PubMed: 20042636]
23. Need AC, Ge D, Weale ME, Maia J, Feng S, Heinzen EL, et al. A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet.* 2009; 5(2):e1000373. [PubMed: 19197363]
24. Resch K, Jockusch H, Schmitt-John T. Assignment of homologous genes, Peli1/PELI1 and Peli2/PELI2, for the Pelle adaptor protein Pellino to mouse chromosomes 11 and 14 and human chromosomes 2p13.3 and 14q21, respectively, by physical and radiation hybrid mapping. *Cytogenet Cell Genet.* 2001; 92(1–2):172–4. [PubMed: 11306823]
25. Liu Y, Shete S, Etzel CJ, Scheurer M, Alexiou G, Armstrong G, et al. Polymorphisms of LIG4, BTBD2, HMGA2, and RTEL1 genes involved in the double-strand break repair pathway predict glioblastoma survival. *J Clin Oncol.* 28(14):2467–74. [PubMed: 20368557]

26. Shimada Y, Saito A, Suzuki M, Takahashi E, Horie M. Cloning of a novel gene (ING1L) homologous to ING1, a candidate tumor suppressor. *Cytogenet Cell Genet.* 1998; 83(3–4):232–5. [PubMed: 10072587]
27. Wang J, Chin MY, Li G. The novel tumor suppressor p33ING2 enhances nucleotide excision repair via inducement of histone H4 acetylation and chromatin relaxation. *Cancer Res.* 2006; 66(4):1906–11. [PubMed: 16488987]
28. Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, et al. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature.* 2006; 442(7098):96–9. [PubMed: 16728974]
29. Nagashima M, Shiseki M, Miura K, Hagiwara K, Linke SP, Pedoux R, et al. DNA damage-inducible gene p33ING2 negatively regulates cell proliferation through acetylation of p53. *Proc Natl Acad Sci U S A.* 2001; 98(17):9671–6. [PubMed: 11481424]
30. Gozani O, Karuman P, Jones DR, Ivanov D, Cha J, Lugovskoy AA, et al. The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. *Cell.* 2003; 114(1):99–111. [PubMed: 12859901]
31. Dualan R, Brody T, Keeney S, Nichols AF, Admon A, Linn S. Chromosomal localization and cDNA cloning of the genes (DDB1 and DDB2) for the p127 and p48 subunits of a human damage-specific DNA binding protein. *Genomics.* 1995; 29(1):62–9. [PubMed: 8530102]
32. Groisman R, Polanowska J, Kuraoka I, Sawada J, Saijo M, Drapkin R, et al. The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell.* 2003; 113(3):357–67. [PubMed: 12732143]
33. Hwang BJ, Ford JM, Hanawalt PC, Chu G. Expression of the p48 xeroderma pigmentosum gene is p53-dependent and is involved in global genomic repair. *Proc Natl Acad Sci U S A.* 1999; 96(2):424–8. [PubMed: 9892649]
34. Hwang BJ, Toering S, Francke U, Chu G. p48 Activates a UV-damaged-DNA binding factor and is defective in xeroderma pigmentosum group E cells that lack binding activity. *Mol Cell Biol.* 1998; 18(7):4391–9. [PubMed: 9632823]
35. Takimoto R, MacLachlan TK, Dicker DT, Niitsu Y, Mori T, el-Deiry WS. BRCA1 transcriptionally regulates damaged DNA binding protein (DDB2) in the DNA repair response following UV-irradiation. *Cancer Biol Ther.* 2002; 1(2):177–86. [PubMed: 12170778]
36. Nichols AF, Itoh T, Zolezzi F, Hutsell S, Linn S. Basal transcriptional regulation of human damage-specific DNA-binding protein genes DDB1 and DDB2 by Sp1, E2F, N-myc and NF1 elements. *Nucleic Acids Res.* 2003; 31(2):562–9. [PubMed: 12527763]
37. Prost S, Lu P, Caldwell H, Harrison D. E2F regulates DDB2: consequences for DNA repair in Rb-deficient cells. *Oncogene.* 2007; 26(24):3572–81. [PubMed: 17173070]
38. Rothman N, Garcia-Closas M, Chatterjee N, Malats N, Wu X, Figueroa JD, et al. A multistage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat Genet.* 2007; 39(11):978–84. [PubMed: 17524388]
39. Wu X, Lin J, Grossman HB, Huang M, Gu J, Etzel CJ, et al. Projecting individualized probabilities of developing bladder cancer in white individuals. *J Clin Oncol.* 2007; 25(31):4974–81. [PubMed: 17971596]
40. Conneely KN, Boehnke M. So Many Correlated Tests, So Little Time! Rapid Adjustment of P Values for Multiple Correlated Tests. *Am J Hum Genet.* 2007; 81(6)

Table 1

Selected host characteristics of bladder cancer cases and normal controls

Variables	Cases (%) (N=803)	Controls (%) (N=803)	P value [*]
Age (Mean $\pm$ SD^{**,†})	64.7 $\pm$ 11.1	63.8 $\pm$ 10.9	0.10
Gender			
Male	640 (79.7)	639 (79.6)	0.95
Female	163 (20.3)	164 (20.4)	
Smoking Status			
Never	212 (26.4)	355 (44.2)	5.2×10^{-21}
Former	404 (50.3)	381 (47.4)	
Current & Recent quitter	187 (23.3)	67 (8.3)	
Pack-years (Mean $\pm$ SD)[†]	43.0 $\pm$ 30.7	29.9 $\pm$ 27.9	2.8×10^{-12}

^{*} P values were derived from the χ^2 test for categorical variables (gender and smoking status) and *t* test for continuous variables (age and pack-years)

^{**} SD, standard deviation

Table 2

Overall and Stratified Analysis by Smoking Status of Significant SNPs

SNP	Gene	Position in gene	Best Fit Model	Overall			Never Smokers		Ever Smokers		p ^{##} for interaction
				OR (95% CI) [*]	P	BFD ^{##}	HWE [#]	OR (95% CI) [*]	OR (95% CI) [*]		
rs11132186	ING2	3' of gene	Recessive	0.52 (0.32–0.83)	0.005	0.72	0.1412	0.37 (0.15–0.92)	0.59 (0.34–1.04)	0.327	
rs11039130	DDB2	5' of gene	Recessive	1.64 (1.14–2.35)	0.007	0.79	0.5940	1.69 (0.95–2.98)	1.60 (1.01–2.56)	0.806	
rs11735038	ING2	3' of gene	Recessive	0.66 (0.49–0.90)	0.008	0.73	0.2054	0.78 (0.60–1.01)	0.67 (0.47–0.97)	0.769	
rs4150667	GTF2H1	Intron	Recessive	1.55 (1.12–2.15)	0.008	0.73	0.0714	0.92 (0.65–1.29)	1.90 (1.24–2.89)	0.197	
rs10065575	CCNH	5' of gene	Dominant	0.76 (0.61–0.94)	0.010	0.78	0.7101	0.55 (0.38–0.79)	0.89 (0.69–1.16)	0.034	
rs1051315	RPA1	3' UTR	Dominant	0.65 (0.46–0.92)	0.013	0.80	0.4360	0.65 (0.37–1.12)	0.66 (0.43–1.01)	0.989	
rs6854224	ING2	5' of gene	Recessive	0.70 (0.53–0.93)	0.013	0.79	0.1650	0.70 (0.49–0.99)	0.67 (0.47–0.95)	0.755	
rs1685404	DDB2	3' of gene	Dominant	0.78 (0.63–0.95)	0.016	NS [§]	0.2868	1.01 (0.78–1.31)	0.67 (0.52–0.86)	0.038	
rs3744767	RPA1	3' UTR	Dominant	1.28 (1.04–1.58)	0.018	NS	0.1194	1.25 (0.88–1.77)	1.31 (1.01–1.69)	0.796	
rs4453140	ERCC6	5' of gene	Dominant	0.78 (0.63–0.96)	0.019	NS	0.5680	0.53 (0.37–0.77)	0.95 (0.73–1.23)	0.009	
rs10978792	RAD23B	Intron	Dominant	1.53 (1.06–2.20)	0.021	NS	0.4174	1.55 (0.89–2.68)	1.52 (0.94–2.46)	0.956	
rs11732255	ING2	3' of gene	Additive	0.84 (0.72–0.98)	0.025	NS	0.9866	0.82 (0.63–1.07)	0.85 (0.71–1.02)	0.720	
rs1799787	ERCC2	Intron	Additive	1.19 (1.02–1.39)	0.029	NS	0.9281	1.14 (0.64–2.05)	1.29 (1.06–1.57)	0.214	
rs3176639	XPA	Intron	Recessive	0.69 (0.49–0.97)	0.032	NS	0.2289	0.96 (0.74–1.25)	0.59 (0.39–0.90)	0.261	
rs1885094	MNAT1	5' of gene	Dominant	0.81 (0.66–0.99)	0.037	NS	0.1655	0.92 (0.65–1.29)	0.75 (0.58–0.97)	0.436	
rs4151330	MNAT1	Intron	Recessive	1.41 (1.01–1.98)	0.044	NS	0.0344	0.97 (0.68–1.36)	1.64 (1.08–2.50)	0.250	
rs1124303	XPC	Intron	Dominant	0.76 (0.58–1.00)	0.049	NS	0.3793	0.74 (0.47–1.15)	0.78 (0.56–1.10)	0.896	

* Adjusted by age, gender, and smoking status when appropriate

** Assuming a prior probability at 0.05 and significance level of 0.8

HWE, Hardy-Weinberg equilibrium for controls

P for interaction between SNPs and smoking status, bold font: p<0.05

§ NS, non-significant

Table 3

Risk Estimates of Genotypes of the Significant SNPs

SNP	Gene	Genotype*	Case/control	OR (95%CI)*
rs11132186	ING2	GG	498/471	1 (reference)
		GT	274/278	0.93 (0.75–1.15)
		TT	31/54	0.50 (0.31–0.81)
		TT vs. GG+GT		0.52 (0.32–0.83)
rs11039130	DDB2	CC	396/421	1 (reference)
		CT	321/325	1.05 (0.86–1.31)
		TT	85/57	1.68 (1.16–2.44)
		TT vs. CC+CT		1.64 (1.14–2.35)
rs11735038	ING2	TT	353/321	1 (reference)
		TA	358/360	0.94 (0.75–1.16)
		AA	91/122	0.64 (0.47–0.89)
		AA vs. TT+TA		0.66 (0.49–0.90)
rs4150667	GTF2H1	CC	360/356	1 (reference)
		CT	341/374	0.89 (0.72–1.10)
		TT	102/73	1.47 (1.04–2.07)
		TT vs. CC+CT		1.55 (1.12–2.15)
rs10065575	CCNH	GG	299/247	1 (reference)
		GA	361/392	0.77 (0.61–0.97)
		AA	143/164	0.73 (0.54–0.97)
		GA+AA vs. GG		0.76 (0.61–0.94)
rs1051315	RPA1	GG	738/702	1 (reference)
		GT	62/99	0.63 (0.45–0.89)
		TT	3/2	1.66 (0.27–10.14)
		GT+TT vs. GG		0.65 (0.46–0.92)
rs6854224	ING2	TT	317/292	1 (reference)
		TC	376/368	0.99 (0.79–1.24)
		CC	105/142	0.70 (0.51–0.95)
		CC vs. TT+TC		0.70 (0.53–0.93)

Table 4

SNP-Smoking (packyears) Interaction

SNP	Gene	Model	OR (95% CI)*	P value	β	P** for interaction
rs4151150	MNAT1	Dominant	0.80 (0.46–1.40)	0.4375	-0.026	0.0018
rs807812	XAB2	Dominant	1.03 (0.84–1.27)	0.7759	0.013	0.0029
rs10817938	XPA	Dominant	0.92 (0.63–1.34)	0.6666	0.028	0.0059
rs794078	XAB2	Additive	1.02 (0.86–1.21)	0.8525	0.010	0.0072
rs4253082	ERCC6	Dominant	0.95 (0.76–1.19)	0.6562	-0.011	0.0079
rs12926685	ERCC4	Recessive	0.77 (0.57–1.04)	0.0909	-0.013	0.0112
rs1131636	RPA1	Dominant	1.14 (0.92–1.41)	0.2323	-0.011	0.0133
rs4253211	ERCC6	Dominant	0.89 (0.69–1.16)	0.4042	-0.011	0.0168
rs794083	XAB2	Recessive	1.14 (0.86–1.52)	0.3547	-0.012	0.0197
rs7503021	RPA1	Dominant	1.17 (0.86–1.61)	0.3230	-0.013	0.0232
rs1685404	DDB2	Dominant	0.78 (0.63–0.95)	0.0161	-0.009	0.0249
rs2291120	DDB2	Dominant	0.87 (0.69–1.11)	0.2630	-0.010	0.0290
rs4253162	ERCC6	Dominant	0.85 (0.65–1.10)	0.2159	0.012	0.0342
rs158937	ERCC8	Dominant	1.11 (0.86–1.45)	0.4169	0.012	0.0414
rs794085	XAB2	Additive	1.05 (0.89–1.23)	0.5844	0.007	0.0433
rs7072383	ERCC6	Dominant	0.85 (0.66–1.11)	0.2357	0.011	0.0447
rs50871	ERCC2	Additive	0.94 (0.81–1.09)	0.3950	0.006	0.0474

* Adjust for age, gender, and smoking status

** P for interaction between SNPs and packyears (as continuous variable)

Table 5

Association of Haplotypes of NER Genes with Bladder Cancer Risk

Haplotype*	Haplotype sequence**	Haplotype frequency	Case/control	OR (95% CI)#	P value
ING2					
H1	M-M-M-M-M-M	31.0%	433/361	1	
H2	M-M-M-M-M-V	21.8%	290/268	0.90 (0.72–1.12)	0.34
H3	M-V-V-V-V-M	11.7%	151/149	0.85 (0.65–1.12)	0.25
H4	V-V-V-V-M-M	7.7%	92/104	0.79 (0.57–1.09)	0.15
H5	V-V-V-M-V-M	5.6%	70/73	0.83 (0.58–1.20)	0.33
H6	M-M-M-V-M-M	5.0%	66/61	0.93 (0.63–1.37)	0.71
H7	M-M-V-M-M-V	4.1%	49/55	0.78 (0.51–1.20)	0.26
H8	V-V-V-V-V-M	4.0%	45/57	0.60 (0.39–0.93)	0.02
H9	M-V-M-V-M-M	2.9%	36/39	0.78 (0.48–1.28)	0.33
H10	M-M-V-M-M-M	1.4%	11/24	0.43 (0.20–0.90)	0.02
H11	V-V-M-V-M-M	1.4%	12/23	0.42 (0.20–0.89)	0.02
H12	V-V-V-M-M-M	1.3%	15/18	0.68 (0.32–1.40)	0.29
OTHER (14)		2.2%	24/32	0.59 (0.33–1.03)	0.07
DDDB2					
H1	V-M-M-M-M-M-M-M	26.9%	419/376	1	
H1	M-V-M-M-M-M-M-M	19.4%	276/297	0.84 (0.67–1.05)	0.13
H3	M-M-V-M-M-V-V-V	13.3%	201/193	0.95 (0.74–1.22)	0.67
H4	M-V-M-V-M-M-M-M	11.5%	162/178	0.81 (0.62–1.05)	0.11
H5	M-M-V-M-M-V-M-M	10.2%	149/154	0.82 (0.62–1.08)	0.15
H6	M-M-V-M-M-M-M-M	8.2%	105/137	0.67 (0.50–0.91)	0.009
H7	M-M-M-M-V-M-V-M	3.4%	56/45	1.03 (0.67–1.59)	0.88
H8	V-M-M-M-M-M-M-M	1.6%	24/23	1.05 (0.58–1.92)	0.87
H9	M-M-M-M-M-M-M-M	1.4%	21/19	0.89 (0.46–1.71)	0.72
OTHER (26)		4.2%	61/62	0.93 (0.63–1.38)	0.72

* *ING2* haplotype includes 6 SNPs in the order of rs6854224, rs4862213, rs6830958, rs11132186, rs11735038, and rs11732255. *DDDB2* haplotype includes 8 SNPs in the order of rs11039130, rs2029298, rs2291120, rs10742797, rs1685404, rs2957873, rs3824866, rs901746, and rs1050244.

** M: major allele, V: variant allele

Adjusted for age, gender, and smoking status.