Carcinogenesis vol.30 no.5 pp.763–768, 2009
doi:10.1093/carcin/bgp046
Advance Access publication February 23, 2009

Genetic polymorphisms in 85 DNA repair genes and bladder cancer risk

Stefan Michiels[1], Agnès Laplanche[1], Thomas Boulet[1], Philippe Dessen[2], Bertrand Guillonneau[3], Arnaud Méjean[4], François Desgrandchamps[5], Mark Lathrop[6], Alain Sarasin[2] and Simone Benhamou[2,7,*]

[1]Department of Biostatistics and Epidemiology, Institut Gustave Roussy, Villejuif 94805, France, [2]CNRS, FRE2939, Université Paris-Sud, Institut Gustave Roussy, Villejuif 94805, France, [3]Department of Urology, Institut Mutualiste Montsouris, Paris 75014, France, [4]Department of Urology, Hôpital Necker, Paris 75015, France, [5]Department of Urology, Hôpital Saint-Louis, Paris 75010, France, [6]Centre National de Génotypage, Evry 91057, France and [7]INSERM, U794, Fondation Jean-Dausset – CEPH, 27 rue Juliette Dodu, Paris 75010, France

[*]To whom correspondence should be addressed. Tel: +33 1 42114139; Fax: +33 1 42115315;
Email: simone.benhamou@inserm.fr

Several defense mechanisms have been developed and maintained during the evolution to protect human cells against damage produced from exogenous or endogenous sources. We examined the associations between bladder cancer and a panel of 652 polymorphisms from 85 genes involved in maintenance of genetic stability [base excision repair, nucleotide excision repair, double-strand break repair (DSBR) and mismatch repair, as well as DNA synthesis and cell cycle regulation pathways] in 201 incident bladder cancer cases and 326 hospital controls. Score statistics were used to test differences in haplotype frequencies between cases and controls in an unconditional logistic regression model. To account for multiple testing, we associated to each *P*-value the expected proportion of false discoveries (*q*-value). Haplotype analysis revealed significant associations ($P < 0.01$) between bladder cancer and two genes (*POLB* and *FANCA*) with an associated *q*-value of 24%. A permutation test was also used to determine whether, in each pathway analyzed, there are more variants whose allelic frequencies are different between cases and controls as compared with what would be expected by chance. Differences were found for cell cycle regulation ($P = 0.02$) and to a lesser extent for DSBR ($P = 0.05$) pathways. These results hint to a few potential candidate genes; however, our study was limited by the small sample size and therefore low statistical power to detect associations. It is anticipated that genome-wide association studies will open new perspectives for interpretation of the results of extensive candidate gene studies such as ours.

Introduction

Bladder cancer is the fifth most common cancer in male populations of developed countries. Although cigarette smoking is the predominant risk factor, accounting for ~65% of male cases and 30% of female cases (1), only a fraction of exposed individuals actually develop bladder cancer, suggesting the involvement of predisposing genetic factors.

Several defense mechanisms have been developed and maintained during the evolution to protect human cells against damage produced from exogenous or endogenous sources. Four major repair pathways [base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR) and double-strand break repair (DSBR)] are responsible for repairing most DNA lesions according to their chemical structure (2). These pathways are involved in removing DNA damage produced directly or indirectly from tobacco carcinogens. The NER pathway mainly removes bulky DNA lesions typically generated from exposure to polycyclic aromatic hydrocarbons in tobacco smoke. BER is necessary to remove oxidized or chemically modified bases. MMR is a pathway that is necessary for corrections of the errors made during normal DNA replication by the replicative DNA polymerases. It is also involved in the recognition of certain misrepaired nucleotides opposite DNA lesions, such as those produced by cigarette smoking. The DSBR pathway is essential for all living organisms, as even one single unrepaired double-strand break (DSB) can be lethal for a cell (3). Consequently, eukaryotic cells have evolved specialized and redundant systems to detect and repair chromosomal DSB in various parts of the cell cycle. Eukaryotic cells possess two major subpathways to repair DSB: homologous recombination and non-homologous end joining.

In the absence of full repair of DNA lesions on genomic DNA, replication can occur, leading to DNA synthesis inhibition that may induce cell apoptosis or allow switching to tolerance mechanisms involving specific DNA polymerases with mutagenic translesion synthesis activities. Beside DNA repair, DSB and other DNA lesions can trigger a cellular response known as DNA damage response. Genes involved in this pathway are also partly implicated in DNA damage signaling, cell cycle regulation and apoptosis induction.

Hundreds of polymorphisms in genes involved in the maintenance of genome integrity have been identified and reported in public databases (4,5); however, for many of these polymorphisms, the impact on repair phenotype and cancer susceptibility remains unknown. A few functional studies in humans suggested that variant alleles in DNA repair genes are associated with variable phenotypic effects such as increased DNA levels (6–8) or decreased individual capacity of DNA repair (9,10) and might therefore influence the consequent risk of smoking-related cancers. However, other biomarker investigations did not provide consistent observations on genotype–phenotype correlations (reviewed in refs 11,12).

Most of the previously reported studies on associations between genetic variants and bladder cancer investigated only one or a few selected variants at a time (13–21). However, several cancers are thought to result from genetic variation in many genes, most of which having modest effects. Analysis of multiple sequence variants in a gene and multiple genes within an entire pathway might therefore provide refinement in predicting risk. Specific haplotypes in genes *XRCC4* (21), *XRCC1* and *XRCC3* (22), *XPD*/*ERCC2* (18,23) and *XPC* (24) have been suggested to be associated with bladder cancer risk. Altogether, all these results cannot be directly compared with each other due to the genotyping of different genes and of different SNPs within repair genes.

Multigenic approaches were recently used to examine global effects of genetic variation in whole biological pathways and to evaluate gene–gene and gene–smoking interactions on bladder cancer risk. In a large Spanish case–control study, significant overall associations between bladder cancer and two repair pathways were found (NER $P = 0.04$, 22 variants in seven genes; DSB $P = 0.01$, 29 variants in seven genes) (19,21); in contrast, no association was reported for the BER pathway (20). The use of classification and regression tree (CART) analyses yielded potential interactions between smoking and NER genotypes (19). This issue of interactions was also addressed by Wu *et al.* (25) using a similar classification tree approach on a panel of 44 polymorphisms in DNA repair (BER, NER and DSBR) and cell cycle control genes. A few possible higher risk subgroups for bladder cancer were put forward among smokers, categorized mainly by NER genotypes and to a lesser extent by BER and DSBR genotypes.

In the present study, we examined the associations between bladder cancer and a panel of 652 polymorphisms from 85 genes categorized as involved in maintenance of genetic stability in 201 bladder cancer cases and 326 controls. We analyzed genes in the four main repair

Abbreviations: BER, base excision repair; CART, classification and regression tree; CI, confidence interval; DSB, double-strand break; DSBR, double-strand break repair; MMR, mismatch repair; NER, nucleotide excision repair; OR, odd ratio; SNP, single nucleotide polymorphism.

pathways (BER, NER, MMR and DSBR), in 'DNA replication, translesion synthesis and transcription' called hereafter DNA synthesis pathway and in 'DNA damage signaling, cell cycle and apoptosis'' called hereafter cell cycle regulation pathway. We evaluated effects of individual SNPs, of haplotypes and of the global genetic variation in each pathway. Finally, we explored gene–gene and gene–smoking interactions on bladder cancer risk by a regression tree analysis.

Materials and methods

Study population

The study was conducted in France between 1997 and 2001 in three general hospitals located in Paris. Patients with newly diagnosed bladder cancer were eligible. All patients had histopathologically confirmed transitional cell carcinoma and none had received chemotherapy or radiotherapy before enrollment. Each time a new case was included in the study, we sought one urological control submitted to surgery for prostatic hyperplasia or urinary incontinence at the same urology department and one non-urological control treated at the medical departments for non-tobacco-related diseases. Indeed, one of the objectives of our case–control study was to assess the predictive value of smoking-induced DNA adducts in normal bladder tissue on bladder cancer risk (26). For practical and ethical reasons, the only possible conditions allowing resection of a sample of normal bladder tissue among controls were prostatic hyperplasia or urinary incontinence. Patients submitted to surgery for these conditions were therefore chosen as urological controls.

Controls were individually matched to the cases for sex and age (±3 years). For both case and control populations, non-Caucasian individuals and subjects with previous malignant disease were not eligible. There were a total of 527 subjects (201 bladder cancer cases and 326 controls—160 urological and 166 non-urological) recruited for this study. The 1:2 ratio was achieved for 129 cases; for the 72 remaining cases, fewer than two controls were included because of a lack of controls fulfilling the eligibility criteria during the data collection. The main medical diagnoses were prostatic hyperplasia (90%) among urological controls and arthropathia (70%) among non-urological controls.

Cases and controls underwent an identical in-person structured interview. Detailed information on sociodemographic characteristics, medical history prior to the date of hospitalization, lifetime use of tobacco products and occupational history was collected by the same trained physician throughout data collection. At the end of the interviews, 30 ml blood samples were drawn into coded heparinized tubes. The participation rates were 100% for the cases and 99% for the controls. All participants signed informed consent agreements and the study was approved by the Ethical Committee for the Protection of Human Subjects at Kremlin-Bicetre Hospital (France).

SNP genotyping

Genotyping of bladder cancer cases and controls was performed using several procedures (Golden Gate assay on Illumina BeadArray genotyping platform, Taqman, or direct sequencing) as described previously (27,28). The genotype data were subjected to various quality control procedures at the Center National de Génotypage (Evry, France). Quality control on Illumina genotype data was evaluated by including two Center d'Etude du Polymorphisme Humain control DNAs in duplicate in each DNA sample plate. These DNA sample plates were all genotyped against two Illumina genotyping panels that included all markers analyzed. From a total of 935 successfully genotyped variants, we removed (i) variants for which >50% of the genotype data among patients were missing (17 variants); or (ii) variants for which the minor allele frequency was <2.5% in the controls (217 variants); or (iii) variants for which there was a significant deviation from the Hardy–Weinberg proportions in the controls [$P < 0.05$ as assessed by Fisher's exact test (49 variants)]. Therefore, a total of 652 variants in 85 genes (72 SNPs in 10 BER genes, 99 SNPs in 12 NER genes, 58 SNPs in seven MMR genes, 139 SNPs in 16 DSBR genes, 231 SNPs in 34 DNA synthesis genes and 53 SNPs in six cell cycle genes) were retained for analysis at the end of these steps. The average number of retained SNPs by gene was eight (range 1–42).

Statistical analysis

Subjects who had smoked at least one cigarette, one cigar or one pipe a day for 6 months or longer were classified as ever smokers. Former smokers were defined as people who had stopped smoking at least 1 year prior to the diagnosis. The daily consumption of each type of tobacco smoked was expressed in grams per day (1 g for cigarette, 2 g for cigar and 3 g for pipe) (29). The average number of grams of tobacco smoked per day was calculated by dividing the cumulative lifetime tobacco consumption by the overall duration of smoking. Lifetime smoking exposure was also expressed in pack years of smoking (years smoked × the number of packs of cigarettes per day).

The main characteristics in urological and non-urological controls were, respectively, as follows: mean age 67 years (SD = 8) and 66 years (SD = 11) ($P = 0.27$), ever smokers 64 and 65% ($P = 0.98$) and mean duration of smoking 30 years (SD = 15) and 29 years (SD = 14) ($P = 0.83$). Because bias might arise if there is an association between admission diagnoses of controls and genetic variants, the allele frequencies for each of the 652 SNPs have been compared between the two control groups using a Fisher exact test (supplementary Table S1 is available at *Carcinogenesis* Online). A significant difference (raw P-value <0.05) was found for 23 of the 652 SNPs (3.5%). To control for multiple testing, we used the Benjamini and Hochberg (30) procedure to control for the false discovery rate or the expected proportion of false positives within a subset where the null hypothesis of no association is unlikely to be true. All false discovery rate-adjusted P-values were equal to 1, indicating that the observed significant differences between control groups can be expected to be false positives. Furthermore, the allele frequencies observed in our control groups were found to be similar to those reported in single nucleotide polymorphism database (dbSNP) for Caucasian populations (supplementary Table S1 is available at *Carcinogenesis* Online). Altogether, as no clear difference in single nucleotide polymorphism (SNP) allele frequencies was found, the two control groups were consequently combined to assess the associations between bladder cancer and polymorphisms in DNA repair genes.

For each SNP, we compared genotype frequencies between cases and controls using the Armitage test for trend and controlled for multiple testing (30). The false discovery rate-adjusted P-values are called q-values throughout the text.

For each gene, we tested for associations between haplotypes and cancer risk using the method implemented in the Haplo Stats library v1.2.2 for R (31,32). This method is based on a prospective likelihood that depends on haplotype frequencies estimated by an improved expectation-maximisation algorithm to test the statistical association between haplotype and phenotype when linkage phase is ambiguous. It is based on score statistics and provides both global test and haplotype-specific tests. No imputation procedures were applied in the case of missing data. For each gene, we retained the largest haplotype block for which all SNPs were in pairwise linkage disequilibrium ($|D'| > 0.85$) among controls. A global test was then performed to test the hypothesis that a difference in haplotype frequencies is seen between bladder cancer cases and controls. Further interpretation of tests for single haplotype effects was provided for genes for which the global test yielded a putative significant result ($P < 0.05$). Haplotype-specific odds ratios (ORs) were calculated assuming an additive model and using unconditional logistic regression controlling for sex, age and pack years of smoking. The most common haplotype among controls was used as the reference in the logistic regression, and rare haplotypes (frequency <5% in the pooled case and control set) were combined.

For each pathway analyzed in this study, we performed a permutation test to determine whether there are more SNPs in that pathway whose allelic frequencies are different between cases and controls as compared with what would be expected by chance. For each pathway, we calculated the sum of the individual SNP effects measured by the Armitage statistic. In order to avoid the assumption of independence of SNPs, we used a permutation method to determine the distribution of this sum statistic under the null hypothesis of no association (33). We randomly permuted 1000 times the labels of cases and controls and recalculated each time a sum statistic; the P-value is then equal to the proportion of times this relabeling resulted in a higher sum statistic.

Finally, we used a CART to develop a prediction rule for bladder cancer susceptibility. CART uses recursive partitioning based on binary splitting rules to stratify data into homogenous risk groups. One major advantage of this method is its ability to detect high-level gene–gene and gene–environment interactions. Smoking status (binary: never smokers versus former or current smokers) and SNPs were used to grow the tree. *A priori* class probabilities were equal to the observed frequencies of cases and controls. We used a 10-fold cross-validation scheme in order to estimate the error rate and to choose the optimal complexity parameter, corresponding to a particular tree size. Determination of the optimal tree size was performed using the one-standard error rule (34). The one standard error rule chooses the smallest tree whose estimated error rate is within one-standard error of the minimum error rate of all trees. We assumed for each SNP a dominant genetic model and redefined within each cross-validation step the homozygote wild type as the homozygote genotype with the highest frequency in controls. ORs were estimated among the smoker population on the final tree using a logistic regression model based on indicator variables of the terminal nodes while adjusting for age, sex and pack years of smoking.

Results

Characteristics of subjects

Selected characteristics of the 201 bladder cancer cases and the 326 controls are shown in Table I. The age distribution was comparable among cases and controls. About 85% of cases and controls were

Table I. Demographic and smoking data on bladder cancer cases and controls

Characteristics	Cases (n = 201)	Controls (n = 326)	P-value
Males	168 (84%)	278 (85%)	0.69
Age, mean (SD)	66 (11)	67 (10)	0.50
<50	14 (7%)	13 (4%)	
50–59	39 (19%)	57 (18%)	
60–69	67 (33%)	123 (38%)	
≥70	81 (40%)	133 (41%)	
Education, *n* (%)			
Less than primary school	24 (12%)	29 (9%)	0.52
Primary to high school	99 (49%)	163 (50%)	
High school or more	78 (39%)	134 (41%)	
Smoking status			
Never	33 (16%)	116 (36%)	<0.001
Former	98 (49%)	154 (47%)	
Current	70 (35%)	56 (17%)	
Age at smoking initiation, mean (SD)	19 (4)	21 (5)	<0.01
Pack years of smoking, mean (SD)	33 (25)	26 (24)	<0.01
≤20	48 (29%)	105 (50%)	
21–30	38 (23%)	42 (20%)	
>30	82 (49%)	63 (30%)	

males. As would be predicted, the cases had a significantly higher percentage of current (35%) and ever (84%) smokers than controls (17 and 64%, respectively, $P < 0.0001$). Among ever smokers, cases reported significantly higher levels of tobacco consumption than controls (mean pack years 33 versus 26, $P < 0.01$).

Individual SNP analysis

Genotype frequencies in bladder cancer cases were compared with those in controls for each of the 652 variants (supplementary Table S2 is available at *Carcinogenesis* Online). Unadjusted P-values <0.05 were found for 45 SNPs within 17 genes belonging to the six pathways analyzed [two genes in BER, one gene in NER and in MMR, four genes in DSB, seven genes in DNA synthesis and two genes in cell cycle control (Table II)]. Of those 45 SNPs, five (11%) were within the coding regions: four synonymous SNPs in the *LIG1* gene (*Ala*170*Ala*, rs20580), the *BRCA2* gene (*Lys*1132*Lys*, rs1801406) and the *FANCA* gene (*Thr*381*Thr*, rs1800331; Ser967Ser, ss69355534) and one non-synonymous SNP in the *FANCA* gene (*Met*717*Ile*, rs17232980). The associated proportion of false discoveries or q-value was equal to 0.72.

Haplotype-based analysis

No haplotype analysis could be performed for seven genes (*MLH3*, *MPG*, *POLD1*, *POLR2F*, *POLR2J*, *RECQL4* and *RFC5*) inasmuch as only one SNP was kept after filtration and for three additional genes (*APEX1*, *POLH* and *RFC2*) for which we found only one haplotype with an estimated frequency >5%. For the remaining 75 genes, we calculated a global score statistic to test the hypothesis of a difference in haplotype frequencies between cancer cases and controls for the largest haplotype block within that gene while adjusting for the covariates.

Four genes (*OGG1*, *POLG*, *POLB* and *FANCA*) showed a global P-value <0.05 with an associated q-value of 62%. When applying a global P-value cutoff of 0.01, only *POLB* and *FANCA* passed the bar, with an associated q-value of 24%. The haplotypes for these four genes are provided in Table III together with their estimated frequencies and haplotype-specific ORs, both unadjusted and adjusted for the covariates. We found a potentially increased risk for bladder cancer associated with one haplotype of the *OGG1* gene (haplotype 2, adjusted OR = 1.59, 95% confidence interval (CI): 1.05–2.41, $P = 0.03$), of the *POLB* gene (haplotype 2, OR = 4.22, 95% confidence interval (CI): 1.66–10.74, $P = 0.003$) and of the *FANCA* gene (haplotype 3, OR = 2.45, 95% CI: 1.38–4.35, $P = 0.002$) as compared with the most frequent haplotype of the corresponding genes.

Pathway-based analysis

Cell cycle control and to a lesser extent DSBR pathways showed a significantly higher number of SNPs with differential allele frequencies between bladder cancer cases and controls as compared with what would be expected by chance ($P = 0.02$ and $P = 0.05$, respectively). For the other pathways, the P-values were clearly non-significant (BER: $P = 0.10$, DNA synthesis: $P = 0.15$, NER: $P = 0.48$ and MMR: $P = 0.88$).

Regression tree analysis

Smoking status and genotype data for the 652 SNPs were incorporated in a CART analysis to explore gene–gene and gene–smoking interactions. The tree structure generated is shown in Figure 1. Not surprisingly, there was an initial split on smoking status. The tree structure then suggested distinct patterns for never smokers and ever smokers. In never smokers, no genetic variants were further selected, whereas in ever smokers, a potential two-order interaction between the two intronic *XRCC5* rs4674066 and *LIG1* rs2288878 polymorphisms was found. This suggests an increased adjusted risk for bladder cancer among smokers carrying both *XRCC5 CC* and *LIG1 CT* or *TT* genotypes (OR = 2.45; 95% CI: 1.52–3.95) as compared with those carrying the variant *XRCC5* allele. However, the estimated error rate by 10-fold cross-validation of this regression tree was rather high, 42% among cases and controls.

Discussion

In this study, we investigated the association between bladder cancer and 652 polymorphisms in 85 genes involved in the maintenance of genome integrity.

We applied both individual SNPs and haplotype analyses to determine whether sequence variations were associated with cancer. The univariate analysis yielded differences in genotype frequencies between cases and controls for 45 SNPs but came with a high expected proportion of false discoveries. Most of the previous genetic association studies on bladder cancer risk have focused on identifying effects of single sequence variants in DNA repair genes. Those that have been most extensively studied are *ERCC2 Asp*312*Asn* and *Lys*751*Gln*, *XRCC1 Arg*399*Gln*, and *XRCC3 Thr*241*Met*. These three genes belong to three different repair pathways (NER, BER and homologous recombination, respectively). We did not genotype these SNPs in our study. Moreover, none of these polymorphisms were clearly associated with bladder cancer occurrence in recent meta-analyses (12,21,35,36) or subsequent individual studies (18,19); the only significant finding was from the study by Wu *et al.* (25) that found an increased risk for carriers of the variant *ERCC2* 312*Asn* allele.

In our study, the largest overall differences in haplotype distribution between cases and controls were found for *POLB* and *FANCA* genes ($P = 0.006$ and $P = 0.004$, respectively) with an expected percentage of false discoveries of 24%. The *POLB* '*GATG*' haplotype was associated with a 4-fold elevated bladder cancer risk; however, none of the individual polymorphism was significantly related to risk and a causative genetic variant that could be attributed to the observed association for the *GATG* haplotype is unknown. An increased risk for bladder cancer was also associated with the *FANCA* '*TAA*' haplotype compared with the most common haplotype '*TTA*'; the single variant contrasting haplotypes (rs11644967) met also the $P < 0.05$ criterion for potential association. The *POLB* gene codes the polymerase β that we have classified inside the DNA replication but this enzyme is considered as the main polymerase used during the BER pathway. Interestingly, this polymerase functions in concert with DNA glycosylases initiating the BER process and particularly the glycosylase OGG1 in removing oxidized bases that can be produced in response

Table II. Associations ($P < 0.05$) between genotype frequencies and bladder cancer

Gene	SNP ID dbSNP build 126	Alleles *A1/A2*	Cases			Controls			Adjusted OR[a], *A1/A2* versus *A1/A1* (95% CI)	Adjusted OR[a], *A2/A2* versus *A1/A1* (95% CI)	*P*-value[b]	*q*-Value
			genotypes *A1/A1*	genotypes *A1/A2*	genotypes *A2/A2*	genotypes *A1/A1*	genotypes *A1/A2*	genotypes *A2/A2*				
BER												
LIG1	rs2288883	T/G	82	91	17	109	155	42	0.76 (0.51–1.15)	0.58 (0.30–1.13)	0.043	0.716
	rs274893	T/G	60	99	30	79	169	68	0.83 (0.54–1.29)	0.61 (0.34–1.08)	0.048	0.716
	rs2288878	C/T	41	105	43	92	168	56	1.40 (0.88–2.23)	1.64 (0.93–2.90)	0.045	0.716
	rs20580	T/G	58	102	30	77	169	70	0.87 (0.56–1.35)	0.59 (0.33–1.05)	0.044	0.716
	rs2386522	T/C	55	100	35	68	172	76	0.74 (0.47–1.16)	0.58 (0.33–1.01)	0.036	0.716
TDG	rs3829301	A/C	159	31	0	284	32	0	1.76 (1.00–3.10)	na	0.041	0.716
NER												
ERCC6	rs3750751	T/C	61	23	0	149	21	0	2.86 (1.38–5.93)	na	0.003	0.36
MMR												
MLH1	rs4647255	T/C	185	5	0	293	23	0	0.34 (0.12–0.95)	na	0.027	0.716
DSBR												
BRCA2	rs11571613	C/G	108	69	11	153	132	30	0.75 (0.50–1.12)	0.49 (0.23–1.05)	0.033	0.716
	rs3752451	A/T	95	77	17	126	148	41	0.67 (0.45–1.01)	0.53 (0.28–1.02)	0.02	0.716
	rs1801406	A/G	108	70	11	155	130	31	0.78 (0.53–1.17)	0.49 (0.23–1.04)	0.041	0.716
FANCA	rs11644967	T/A	152	33	4	285	30	1	2.15 (1.22–3.78)	11.88 (1.17–120.74)	0.001	0.335
	rs11648689	A/G	159	26	4	290	26	0	1.84 (1.00–3.39)	na	0.002	0.335
	rs11649162	A/C	159	26	4	290	25	0	1.91 (1.03–3.54)	na	0.002	0.335
	rs11639788	A/G	159	25	4	290	26	0	1.76 (0.95–3.25)	na	0.003	0.375
	rs2074903	T/C	160	26	4	290	26	0	1.83 (1.00–3.38)	na	0.002	0.335
	rs11641147	T/C	162	25	3	290	26	0	1.76 (0.95–3.25)	na	0.008	0.522
	rs3785281	G/C	160	26	4	290	26	0	1.83 (1.00–3.38)	na	0.002	0.335
	ss69355534	A/G	157	27	0	290	26	0	1.91 (1.04–3.5)	na	0.024	0.716
	ss69355533	A/G	160	26	4	290	26	0	1.83 (1.00–3.38)	na	0.002	0.335
	ss69355540	A/C	150	6	0	214	19	2	0.49 (0.18–1.29)	na	0.041	0.716
	rs1800339	A/C	161	25	1	290	26	0	1.73 (0.93–3.20)	na	0.032	0.716
	ss69355539	T/C	158	26	4	290	26	0	1.85 (1.00–3.40)	na	0.002	0.335
	rs1800331	A/C	161	25	3	290	26	0	1.73 (0.93–3.20)	na	0.008	0.522
	rs11648881	T/C	160	26	3	290	26	0	1.79 (0.97–3.29)	na	0.005	0.404
	rs2074963	C/T	158	27	3	289	27	0	1.86 (1.02–3.40)	na	0.004	0.391
FANCD2	rs2075310	T/C	107	69	14	192	100	11	1.25 (0.83–1.88)	2.10 (0.88–5.04)	0.048	0.716
XRCC5	ss69355507	A/G	160	27	1	290	25	1	1.83 (1.00–3.37)	2.37 (0.15–38.47)	0.022	0.716
DNA replication, translesion synthesis and transcription												
POLR2C	rs1114156	A/T	176	13	0	271	44	1	0.46 (0.24–0.91)	na	0.011	0.599
	rs601194	C/T	65	84	34	81	157	69	0.70 (0.45–1.10)	0.6 (0.34–1.05)	0.049	0.716
POLR2E	rs2238586	G/T	122	53	10	164	128	18	0.57 (0.38–0.87)	0.80 (0.34–1.87)	0.016	0.716
	rs1046911	C/T	123	54	11	165	128	22	0.58 (0.38–0.87)	0.70 (0.32–1.57)	0.013	0.644
	rs3787016	G/A	119	58	11	158	135	23	0.57 (0.38–0.85)	0.66 (0.3–1.46)	0.01	0.583
POLR2K	rs2453639	T/G	65	98	25	141	140	30	1.65 (1.09–2.5)	1.68 (0.89–3.19)	0.018	0.716
	rs2254883	G/A	67	100	23	147	140	29	1.65 (1.1–2.48)	1.71 (0.89–3.30)	0.018	0.716
	rs2453640	C/G	66	100	23	143	141	31	1.67 (1.11–2.52)	1.58 (0.83–3.03)	0.033	0.716
	rs2453641	T/G	66	96	25	142	141	32	1.60 (1.06–2.41)	1.61 (0.85–3.02)	0.033	0.716
	rs2453643	A/G	67	99	24	143	141	32	1.62 (1.08–2.44)	1.53 (0.81–2.90)	0.039	0.716
REV3L	rs240998	T/C	151	38	0	231	79	6	0.66 (0.41–1.06)	na	0.041	0.716
	rs240969	T/C	152	36	0	232	78	6	0.63 (0.39–1.01)	na	0.028	0.716
RFC2	ss68316964	T/C	175	15	0	305	11	0	1.97 (0.86–4.54)	na	0.029	0.716
RFC5	rs5745811	A/G	144	39	6	258	56	0	1.38 (0.85–2.24)	na	0.023	0.716
RPA1	rs2270412	A/G	121	62	6	172	129	15	0.80 (0.54–1.2)	0.66 (0.23–1.83)	0.035	0.716
DNA damage signaling, cell cycle and apoptosis												
ATM	rs189037	G/A	61	100	20	123	127	23	1.59 (1.04–2.44)	1.58 (0.77–3.21)	0.022	0.716
CDKN2A	rs3731238	T/C	173	17	0	306	10	0	3.25 (1.39–7.63)	na	0.005	0.404

[a]Adjusted for sex, age, pack years of smoking. na, not applicable.
[b]Armitage test.

to cigarette carcinogens. Interestingly, the *POLB* gene has already been associated with bladder cancer (37). The *POLB* locus is often lost in bladder cancers and numerous splice variants have been reported in tumor tissues. Fanconi A (FANCA) is a major protein in the FANC complex absolutely necessary for repairing DSB and allowing stalled DNA replication forks to recover normally from lesions such as DSB or cross-links. *FANCA* represents the major gene involved in the Fanconi anemia disease markedly associated with high cancer risk and has been implicated in several cancer predispositions (38). Its own molecular role in DNA repair is linked to the early step of homologous DNA repair of DSB (39). It is, moreover, necessary to shuttle DNA repair complexes between cytoplasm and nucleus to allow repair of genomic DNA.

Global pathway effects on bladder cancer risk were recently investigated in a large Spanish study (19–21). Specifically, 22 genetic variants in seven NER genes were analyzed and results suggested that NER pathway significantly contributes to cancer risk. Our data based on 99 SNPs in 12 NER genes do not support these findings. In contrast, we found evidence for an association between bladder cancer and DSBR and cell cycle regulation pathways. Cigarette smoking gives rise to numerous DNA lesions. Obviously, repair of these lesions by recombination associated with efficient regulation of the cell cycle

Table III. Significant variation ($P < 0.05$) in haplotypic frequencies between cases and controls

Gene	Haplotype	SNP[a]				Haplotype frequency[b]		Unadjusted OR (95% CI)	Adjusted OR[c] (95% CI)	*P*-value	*q*-Value
		1	2	3	4	Cases	Controls				
OGG1	1	A	A			0.70	0.78	1 (reference)[d]	1 (reference)[d]		
	2	G	A			0.24	0.18	1.42 (0.97–2.10)	1.59 (1.05–2.41)	0.03	
	3	A	G			0.06	0.04	1.72 (0.81–3.61)	2.05 (0.89–4.73)	0.09	
									Global test	0.032	0.62
POLG	1	A	A			0.58	0.53	1 (reference)[d]	1 (reference)[d]		
	2	G	A			0.27	0.30	0.81 (0.54–1.22)	0.72 (0.46–1.12)	0.14	
	3	G	G			0.13	0.17	0.67 (0.39–1.13)	0.58 (0.33–1.02)	0.06	
	4	A	G			0.02	0.004	5.53 (0.59–52.17)	6.50 (0.63–66.94)	0.11	
									Global test	0.024	0.60
POLB	1	A	G	C	A	0.83	0.87	1 (reference)[d]	1 (reference)[d]		
	2	G	A	T	G	0.10	0.03	3.23 (1.44–7.25)	4.22 (1.66–10.74)	0.003	
	RARE	*	*	*	*	0.07	0.01	0.85 (0.43–1.67)	0.84 (0.41–1.71)	0.62	
									Global test	0.006	0.24
FANCA	1	T	T	A		0.75	0.77	1 (reference)[d]	1 (reference)[d]		
	2	C	T	A		0.12	0.13	0.96 (0.62–1.50)	1.18 (0.73–1.92)	0.50	
	3	T	A	A		0.12	0.05	2.11 (1.23–3.62)	2.45 (1.38–4.35)	0.002	
	RARE	*	*	*		0.02	0.05	0.46 (0.19–1.12)	0.52 (0.21–1.32)	0.17	
									Global test	0.004	0.24

*All rare haplotypes have been grouped.
[a]OGG1: rs2304277 (SNP1), rs3218995 (SNP2); POLG: rs2072267 (SNP1), rs2074885 (SNP2); POLB: rs2272615 (SNP1), rs2953983 (SNP2), rs3136717 (SNP3), rs3136795 (SNP4); FANCA: rs17684004 (SNP1), rs11644967 (SNP2), ss69355540 (SNP3).
[b]Estimated haplotype frequencies were recalculated from the parameter estimates in the adjusted unconditional logistic regression model.
[c]Adjusted for sex, age and pack years of smoking.
[d]Reference category determined by the haplotype with the highest frequency (total of cases and controls).

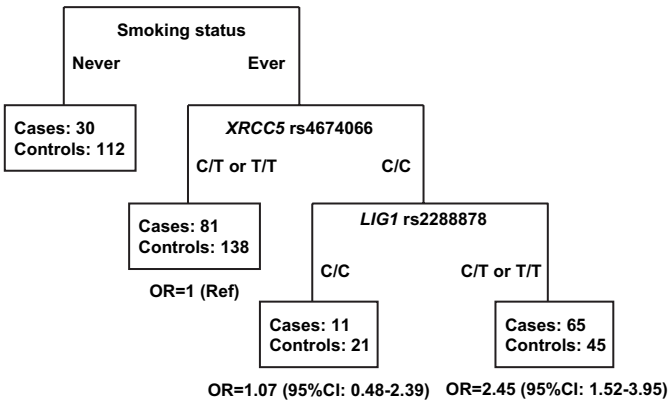

Fig. 1. Classification tree determined on the DNA repair genotype data (652 SNPs) and smoking status. ORs were estimated among the smoker population using a logistic regression model based on indicator variables of the terminal nodes while adjusting for age, sex and pack years of smoking.

represents a major pathway to protect targeted organ to develop cancer following cigarette smoking. Our results for DSB pathway are consistent with those previously reported by Figueroa *et al.* (21). To our knowledge, our study is the first looking for overall involvement of the cell cycle regulation pathway in bladder cancer susceptibility. Replication of our results in future studies is needed.

The CART technique was recently used to explore gene–gene and gene–smoking interactions on multiple SNPs in DNA repair pathways (19,25). Possible higher risk subgroups for bladder cancer were identified among smokers categorized by NER genotypes (19,25) and to a lesser extent by BER and DSBR genotypes (25). We used the same analytical method and did internal cross-validation to determine the optimal tree model and to estimate the error rate. In our population, a two-order interaction was suggested between the DSBR *XRCC5*-rs4674066 and the BER *LIG1*-rs2288878 polymorphisms among ever smokers. No gene polymorphism in the NER pathway was selected. This multigenic approach should nevertheless be treated with caution: the estimated error rate for classifying cases and controls using this regression tree was extremely high (42%), the number of possible interactions is large so that very large sample sizes are required to obtain reliable results and there is a strong possibility that the suggested increased risk for bladder cancer observed in one subgroup is very specific to this particular case–control study.

These results hint to a few potential candidate genes; however, our study was limited by the small sample size and therefore low statistical power to detect associations. It is anticipated that genome-wide association studies on bladder cancer, such as the recent study by Kiemeney *et al.* (40), will open new perspectives for interpretation of the results of extensive candidate gene studies such as ours.

Supplementary material

Supplementary Tables S1 and S2 can be found at http://carcin.oxfordjournals.org/

Funding

Ligue Contre le Cancer du Val-de-Marne; Fondation de France; Groupement d'Entreprises Françaises dans la Lutte contre le Cancer; Association pour la Recherche sur le Cancer, France.

Acknowledgements

Conflict of Interest Statement: None declared.

References

1. Stewart,B.W. *et al*. (2003) *World Cancer Report*. International Agency for Research on Cancer, IARC Press, Lyon.
2. Hoeijmakers,J.H. (2001) Genome maintenance mechanisms for preventing cancer. *Nature*, **411**, 366–374.
3. Bassing,C.H. *et al*. (2004) The cellular response to general and programmed DNA double strand breaks. *DNA Repair*, **3**, 781–796.
4. Packer,B.R. *et al*. (2006) SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucleic Acids Res.*, **34**, 617–621.

5. Thorisson,G.A. *et al*. (2005) The International HapMap Project Web site. *Genome Res.*, **15**, 1591–1593.
6. Palli,D. *et al*. (2001) DNA adduct levels and DNA repair polymorphisms in traffic-exposed workers and a general population sample. *Int. J. Cancer*, **94**, 121–127.
7. Hou,S.M. *et al*. (2002) The XPD variant alleles are associated with increased aromatic DNA adduct level and lung cancer risk. *Carcinogenesis*, **23**, 599–603.
8. Matullo,G. *et al*. (2003) Combination of DNA repair gene single nucleotide polymorphisms and increased levels of DNA adducts in a population-based study. *Cancer Epidemiol. Biomarkers Prev.*, **12**, 674–677.
9. Spitz,M.R. *et al*. (2001) Modulation of nucleotide excision repair capacity by XPD polymorphisms in lung cancer patients. *Cancer Res.*, **61**, 1354–1357.
10. Qiao,Y. *et al*. (2002) 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.*, **509**, 165–174.
11. Benhamou,S. *et al*. (2005) *ERCC2/XPD* gene polymorphisms and lung cancer: a HuGE review. *Am. J. Epidemiol.*, **161**, 1–14.
12. Manuguerra,M. *et al*. (2006) XRCC3 and XPD/ERCC2 single nucleotide polymorphisms and the risk of cancer: a HuGE review. *Am. J. Epidemiol.*, **164**, 297–302.
13. Stern,M.C. *et al*. (2002) XPD codon 751 polymorphism, metabolism genes, smoking, and bladder cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 1004–1011.
14. Shen,M. *et al*. (2003) 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.*, **12**, 1234–1240.
15. Kelsey,K.T. *et al*. (2004) A population-based case-control study of the XRCC1 Arg399Gln polymorphism and susceptibility to bladder cancer. *Cancer Epidemiol. Biomarkers Prev.*, **13**, 1337–1341.
16. Sanyal,S. *et al*. (2004) Polymorphisms in DNA repair and metabolic genes in bladder cancer. *Carcinogenesis*, **25**, 729–734.
17. Schabath,M.B. *et al*. (2005) Polymorphisms in XPD exons 10 and 23 and bladder cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 878–884.
18. Andrew,A.S. *et al*. (2006) Concordance of multiple analytical approaches demonstrates a complex relationship between DNA repair gene SNPs, smoking and bladder cancer susceptibility. *Carcinogenesis*, **27**, 1030–1037.
19. Garcia-Closas,M. *et al*. (2006) Genetic variation in the nucleotide excision repair pathway and bladder cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **15**, 536–542.
20. Figueroa,J.D. *et al*. (2007) Genetic variation in the base excision repair pathway and bladder cancer risk. *Hum. Genet.*, **121**, 233–242.
21. Figueroa,J.D. *et al*. (2007) Evaluation of genetic variation in the double-strand break repair pathway and bladder cancer risk. *Carcinogenesis*, **28**, 1788–1793.
22. Matullo,G. *et al*. (2005) Polymorphisms/haplotypes in DNA repair genes and smoking: a bladder cancer case-control study. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 2569–2578.
23. Matullo,G. *et al*. (2006) DNA repair polymorphisms and cancer risk in nonsmokers in a cohort study. *Carcinogenesis*, **27**, 997–1007.
24. Sak,S.C. *et al*. (2006) Comprehensive analysis of 22 XPC polymorphisms and bladder cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **15**, 2537–2541.
25. Wu,X. *et al*. (2006) Bladder cancer predisposition: a multigenic approach to DNA-repair and cell-cycle-control genes. *Am. J. Hum. Genet.*, **78**, 464–479.
26. Benhamou,S. *et al*. (2003) DNA adducts in normal bladder tissue and bladder cancer risk. *Mutagenesis*, **18**, 445–448.
27. Michiels,S. *et al*. (2007) Polymorphism discovery in 62 DNA repair genes and haplotype associations with risks for lung and head and neck cancers. *Carcinogenesis*, **28**, 1731–1739.
28. Danoy,P. *et al*. (2008) Variants in DNA double-strand break repair and DNA damage-response genes and susceptibility to lung and head and neck cancers. *Int. J. Cancer*, **123**, 457–463.
29. International Agency for Research on Cancer. (1986) *Tobacco Smoking. Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans*. IARC Scientific Publications, Lyon, Vol. 38, pp. 55.
30. Benjamini,Y. *et al*. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B*, **57**, 289–300.
31. Schaid,D.J. *et al*. (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am. J. Hum. Genet.*, **70**, 425–434.
32. Lake,S.L. *et al*. (2003) Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. *Hum. Hered.*, **55**, 56–65.
33. Ott,J. *et al*. (2003) Set association analysis of SNP case-control and microarray data. *J. Comput. Biol.*, **10**, 569–574.
34. Breiman,L. *et al*. (1984) *Classification and Regression Trees*. Chapman & Hall, New York.
35. Hung,R.J. *et al*. (2005) Genetic polymorphisms in the base excision repair pathway and cancer risk: a HuGE review. *Am. J. Epidemiol.*, **162**, 925–942.
36. Han,S. *et al*. (2006) DNA repair gene XRCC3 polymorphisms and cancer risk: a meta-analysis of 48 case-control studies. *Eur. J. Hum. Genet.*, **14**, 1136–1144.
37. Thompson,T.E. *et al*. (2002) Splice variants but not mutations of DNA polymerase beta are common in bladder cancer. *Cancer Res.*, **62**, 3251–3256.
38. D'Andrea,A.D. *et al*. (2003) The Fanconi anaemia/BRCA pathway. *Nat. Rev. Cancer*, **3**, 23–34.
39. Yang,Y.G. *et al*. (2005) The Fanconi anemia group A protein modulates homologous repair of DNA double-strand breaks in mammalian cells. *Carcinogenesis*, **26**, 1731–1740.
40. Kiemeney,L.A. *et al*. (2008) Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat. Genet.*, **40**, 1307–1312.

Received December 16, 2008; revised February 10, 2009; accepted February 18, 2009