

available at www.sciencedirect.comjournal homepage: www.ejconline.com

Genetic polymorphisms in double-strand break DNA repair genes associated with risk of oral premalignant lesions

Hushan Yang^a, Scott M. Lippman^b, Maosheng Huang^a, J. Jack Lee^c, Wei Wang^a, Margaret R. Spitz^a, Xifeng Wu^{a,*}

^aDepartment of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Unit 1340, 1515 Holcombe Blvd., Houston, TX 77030, USA

^bDepartment of Thoracic/Head and Neck Medical Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA

^cDepartment of Biostatistics, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA

ARTICLE INFO

Article history:

Received 22 February 2008

Received in revised form 6 May 2008

Accepted 12 May 2008

Available online 23 June 2007

Keywords:

Double-strand break

Polymorphism

Haplotype

Oral premalignant lesion

ABSTRACT

Oral premalignant lesions (OPLs) are early genetic events en route to oral cancer. To identify individuals susceptible to OPL is critical to the prevention of oral cancer. In a case-control study consisting of 147 patients with histologically confirmed OPL and 147 matched controls, we evaluated the associations of 10 genetic variants in nine genes of the double-strand break (DSB) DNA repair pathway with OPL risk. The most notable finding was an intronic polymorphism (A17893G) of the XRCC3 gene. Compared with the homozygous wild-type AA genotype, the odds ratio (OR) (95% confidence interval [CI]) for the heterozygous AG and homozygous variant GG genotype was 0.85 (0.49–1.48) and 0.18 (0.07–0.47), respectively (P for trend = 0.002). In addition, compared with the most common A–C haplotypes of XRCC3 (in the order of A17893G–T241M), the G–C haplotypes were associated with a significantly decreased risk of OPL (OR = 0.40, 95% CI 0.23–0.68). Moreover, compared with individuals without the G–C haplotype, the OR was 1.04 (0.56–1.95) and 0.20 (0.08–0.51) for subjects with one copy and two copies of the G–C haplotypes, respectively (P for trend = 0.005). Classification and regression tree (CART) analysis further revealed potential high-order gene–gene and gene–environmental interactions and categorised subjects into different risk groups according to their specific polymorphic signatures. Overall, our study provides the first epidemiological evidence supporting a connection between DSB gene variants and OPL development. Our data also suggest that the effects of high-order interactions should be taken into consideration when evaluating OPL predisposition.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Cancers of the oral cavity accounted for 274,000 new cases in 2002 worldwide.¹ Western and southern Europe areas have the highest incidence of oral cancer in men.¹ There was a steady increase in the mortality rate of oral cancer in most

European countries and United States from 1950s to 1980s, although a trend of decline was observed after late 1980s, mostly due to the reduced tobacco and alcohol exposures.² However, this recent decline in mortality rate in Europe was not homogeneous, as evidenced by the persistent rising number of oral cancer-related deaths in some central and eastern

* Corresponding author. Tel.: +1 713 745 2485; fax: +1 713 792 4657.

E-mail address: xwu@mdanderson.org (X. Wu).

0959-8049/\$ - see front matter © 2008 Elsevier Ltd. All rights reserved.

doi:10.1016/j.ejca.2008.05.006

European countries such as Hungary, Slovakia and Russia.² Most oral cancers are preceded by the development of oral premalignant lesions (OPLs), which mainly include oral leukoplakia and erythroplakia.³ The rate at which OPLs transform into oral cancers varied worldwide, dependent upon the geographical location and length of follow-up.^{4,5} A prospective follow-up study of a hospital-based population of 166 patients with OPLs from the Netherlands reported a 2.9% annual malignant transformation rate.⁶ Both genetic and environmental factors contribute to OPL development. Epidemiological studies have identified environmental risk factors such as tobacco smoking, tobacco chewing, alcohol drinking and virus infection.^{3,7,8} However, the genetic components responsible for OPL formation are largely unknown.

DNA repair pathways are important in the defense against genetic insults, the maintenance of genome integrity and the prevention of tumorigenesis. Of the four major DNA repair pathways in mammals, the double-strand break (DSB) repair pathway is mainly involved in the repair of DSBs arising during many pivotal cellular processes such as DNA replication, meiosis, telomere metabolism and V(D)J recombination.⁹ Exogenous challenges, such as ionising radiation and certain chemotherapy drugs, can also induce DSBs. DSBs are sensed by the MRN (MRE11, RAD50 and NBS1) complex, after which the signals are transmitted to the ATM protein. One of two distinct mechanisms may next be induced, depending on the source of the DSB damage – the homologous recombination (HR) or the non-homologous end joining (NHEJ) pathway. The HR pathway repairs DSBs with high fidelity, as it uses one sister DNA strand as the repair template. The direct interaction of RAD51 and RAD52 proteins is an essential step in the HR pathway, and this interaction is facilitated by other important proteins such as BRCA1, BRCA2, XRCC2 and XRCC3.^{10,11} In the NHEJ pathway, the heterodimer of KU70 and KU80 binds to the end of DSBs and activates the DNA-dependent protein kinase catalytic subunit – DNAPKcs. LIG4 and XRCC4 proteins are then recruited to ligate and fill the gap in the broken DNA strands.¹² Deficiencies in these DSB repair genes have been extensively reported to be associated with high incidence of chromosome aberrations, elevated radiation sensitivity and tumorigenesis.^{13–16}

DNA polymorphisms of DSB pathway genes have also been found to be significantly associated with the aetiology of many malignancies including breast, lung, skin and epithelial ovarian cancers.^{17–22} We previously reported our finding of functional variants in the nucleotide excision repair genes that affect OPL risk individually and jointly.²³ We have also found that a single nucleotide polymorphism (SNP) influencing the splicing of CCND1, a key regulator of mammalian cell cycle progression, is associated with a significantly altered smoking-dependent risk of OPL.²⁴ However, to the best of our knowledge, no studies have been done that examine an association of DSB gene variants with the risk of OPL. In this study, we conducted a case-control analysis to assess the individual and joint effects on OPL predisposition of 10 potentially functional SNPs in nine DSB genes. We further evaluated the high-order gene–gene and gene–environment interactions between these factors using Classification and Regression Tree (CART) methodology.

2. Materials and methods

2.1. Study population

A total of 147 OPL patients were identified at The University of Texas M.D. Anderson Cancer Center for the years 1997–2006. All cases were histologically confirmed by the finding of leukoplakia and/or erythroplakia. Patients less than 18-years old and patients with a prior history of cancer (except non-melanoma skin cancer) that had been treated within the preceding two years were excluded. Patients with acute intercurrent illnesses or infections were also excluded. A self-administered questionnaire was used to gather epidemiological data from the patients. Healthy controls ($N = 147$) were identified from a control database used for ongoing case-control studies. The controls in our study were recruited in collaboration with Kelsey-Seybold Clinic, the largest multi-specialty medical organisation in Houston, Texas. The controls were identified by reviewing short survey forms distributed to individuals coming to Kelsey-Seybold Clinic for annual health check-ups or for addressing health concerns. A Kelsey-Seybold staff member provided the form to each potential control subject during clinical registration. They were subsequently contacted by telephone at a later date to confirm their willingness to participate and to schedule an interview appointment at a Kelsey-Seybold clinic convenient to the participant. On the day of the interview, the controls visited the clinic specifically for the purpose of participating in this study but not for any treatment purposes. Controls had no cancer history (except non-melanoma skin cancer) and were matched to cases with regard to age (± 5 yr), gender and ethnicity (Caucasian, African-American and Hispanic). This control selection strategy has been well described and proved to be feasible and effective for molecular epidemiological studies when population-based control selection poses a practical challenge.²⁵ Epidemiological questionnaire data were obtained through in-person interviews. After the interview, 40 ml of blood was collected from each participant into a heparinised tube and sent to the lab for immediate molecular analysis. Approval for the use of human subjects was obtained from the institutional review boards of both M.D. Anderson and Kelsey-Seybold. Informed consent was also obtained from each participant.

2.2. Selection of DSB polymorphisms and genotyping

A total of 10 potentially functional SNPs in nine DSB genes were genotyped in this study. All polymorphisms have been reported to affect either the expression or function of their host genes or to be associated with the risk of malignancy.^{26–34} These SNPs consisted of six non-synonymous SNPs (nsSNPs) with an amino acid substitution (ATM D1853N, NBS1 E185Q, BRCA2 N372H, XRCC3 T241M, RAG1 K820R and LIG4 T91I), one intronic SNP (XRCC3 A17893G), a splicing site SNP (XRCC4 IV7-1), one SNP in the 3' untranslated region (UTR) of KU80 gene and one SNP in the 3' region of XRCC2. Genotyping was performed using a 5' nuclease assay-based TaqMan assay. Probes and primers for the genotyping were either acquired from the SNP500Cancer database or designed using PrimerExpress 2.0 software (Applied Biosys-

tems, Foster, CA) and were available upon request. The probes were labelled fluorescently with either 6-FAM or VIC on the 5' end and a non-fluorescent minor groove binder quencher on the 3' end. The genotyping procedure was done exactly as described in a previous study.³⁵ Briefly, genomic DNA was extracted from peripheral blood lymphocytes using the Human Whole Blood Genomic DNA Extraction Kit (Qiagen, Valencia, CA). The PCR amplification mix (5 µl) included sample DNA (5 ng), 1× TaqMan buffer A, 200 µM deoxynucleotide triphosphates, 5 mM MgCl₂, 0.65 units of AmpliTaq Gold, 900 nM each primer and 200 nM each probe. The PCR conditions consisted of one cycle for 10 min at 95 °C and 40 cycles for 15 s at 95 °C and 1 min at 60 °C. PCR was performed using the ABI PRISM® 7900HT sequence detection system (Applied Biosystems). SDS 2.1 software (Applied Biosystems) was used to analyse the end-point genotyping data. Internal quality controls and negative controls were used to ensure genotyping accuracy. Laboratory personnel performing the genotyping were blinded to the case-control status.

2.3. Statistical analysis

Statistical analyses were performed using the Intercooled Stata 8.0 statistical software package (Stata Corp., College Station, TX). χ^2 and student's *t*-tests were used to assess patient characteristics. The Hardy-Weinberg equilibrium (HWE) was tested using a goodness-of-fit χ^2 test. The OPL risks were estimated as odds ratios (OR) and 95% confidence intervals (95% CI) using multivariate logistic regression adjusted by age, gender, ethnicity, smoking status and alcohol consumption, where appropriate. To account for the use of multiple comparisons, the Bonferroni correction was used to adjust *P*-values while setting the family wise significance level at 0.05. Definition of smoking status was as previously described.²⁴ Never drinkers were defined as those who never consumed alcohol or consumed less than or equal to one drink per month. One drink was defined as one bottle or can (12 ounce) of beer, one medium glass of wine (4 ounce), one mixed drink or one straight shot. Ever drinkers were those who consumed more than one drink per month. Haplotype frequencies were estimated using the Expectation-Maximisation algorithm implemented in the HelixTree program (Golden Helix Inc., Bozeman, MT). The adjusted OR and 95% CI for each haplotype were assessed using multivariate logistic regression. High-order gene-gene and gene-environment interactions were evaluated using CART analysis implemented in the HelixTree software. CART is a non-parametric decision tree-based data-mining approach that can identify specific combinations of contributing factors associated with disease risk. The recursive-partitioning algorithm uses a statistical hypothesis-testing method – formal inference-based recursive modelling – to determine the first locally optimal split and each subsequent split of the data set, with multiplicity-adjusted *P*-values generated 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 pre-specified minimum size. Subgroups of individuals with differential risk patterns were then identified in the different order of nodes of the tree structure, indicating the presence of gene-gene and gene-environment

interactions. Logistic regression was used to calculate the OR and 95% CI in each terminal node of the tree. *P* ≤ 0.05 was considered as the threshold of significance in this study. All statistical analyses were two-sided.

3. Results

3.1. Characteristics of the study population

Table 1 lists the selected characteristics of the study population. There were 147 OPL patients and 147 cancer-free controls adequately matched on age (cases versus controls [mean ± standard deviation]: 57.5 ± 13.6 years versus 59.1 ± 11.0 years, *P* = 0.26), gender (*P* = 1.00) and ethnicity (*P* = 1.00). As expected, significant differences in smoking status existed between cases and controls. The percentage of never and ever smokers was 31.3% and 68.7%, respectively, in the cases and 55.1% and 44.9%, respectively, in the controls (*P* < 0.001). Higher percentage of subjects in controls were ever drinkers compared to cases (Never drinkers versus ever drinkers: 46.3% versus 53.7% in cases; 32.0% versus 68.0% in controls; *P* = 0.01).

3.2. Individual DSB polymorphisms and OPL risk

Table 2 summarises the name, reference number, genetic position and genotype distribution of each SNP and its association with OPL risk. None of the 10 SNPs departed from the HWE amongst controls, indicating that the chance of selection bias or genotyping errors was small. The most notable finding was an intronic SNP in the XRCC3 gene (A17893G), which showed a protective effect on OPL risk. Compared with the wild-type genotype (AA), the heterozygous genotype (AG) was associated with a non-significantly reduced OPL risk (OR = 0.85, 95% CI 0.49–1.48, *P* = 0.57) while the homozygous variant genotype (GG) was associated with

Table 1 – Distribution of selected host characteristics by case-control status

Variable	Case (n = 147)	Control (n = 147)	<i>P</i> -value ^a
Age, mean (SD) ^b	57.5 (13.6)	59.1 (11.0)	0.26
Gender, n (%)			
Male	82 (55.8)	82 (55.8)	1.00
Female	65 (44.2)	65 (44.2)	
Ethnicity, n (%)			
Caucasian	129 (87.8)	129 (87.8)	1.00
Hispanic	11 (7.4)	11 (7.4)	
African-American	7 (4.8)	7 (4.8)	
Smoking status, n (%)			
Never	46 (31.3)	81 (55.1)	<0.001
Ever	101 (68.7)	66 (44.9)	
Alcohol consumption, n (%)			
Never	68 (46.3)	47 (32.0)	0.01
Ever	79 (53.7)	100 (68.0)	

^a *P*-values were derived from the χ^2 -test for categorical variables (gender, ethnicity and smoking status) and *t*-test for continuous variable (age).

^b SD, standard deviation.

Table 2 – Case-control distribution of genotypes of DSB polymorphisms

Polymorphism (rs number)	Position	Genotype	Control/case (N)	OR (95% CI) ^a	P-value	HWE ^b P-value
ATM D1853N rs1801516	nsSNP ^c	GG	114/102	Reference		0.41
		GA	28/30	1.40 (0.74–2.62)	0.30	
		AA	3/5	2.85 (0.62–13.11)	0.18	
		GA/AA	33/35	1.52 (0.84–2.78)	0.17	
		P trend = 0.11				
NBS1 E185Q rs1805794	nsSNP	CC	67/67	Reference		0.45
		CG	59/59	1.06 (0.62–1.81)	0.83	
		GG	18/11	0.54 (0.22–1.30)	0.17	
		CG/GG	77/70	0.93 (0.56–1.54)	0.77	
		P trend = 0.36				
BRCA2 N372H rs144848	nsSNP	TT	76/70	Reference		0.53
		TG	54/58	1.01 (0.59–1.72)	0.98	
		GG	13/8	0.64 (0.23–1.80)	0.40	
		TG/GG	67/66	0.94 (0.56–1.58)	0.82	
		P trend = 0.59				
XRCC2 C41657T rs718282	3' region	CC	131/119	Reference		1.00
		CT	11/16	1.59 (0.67–3.77)	0.29	
		TT	0/1	NA ^d	NA	
		CT/TT	11/17	1.01 (0.59–1.71)	0.98	
		P trend = 0.16				
XRCC3 T241M rs861539	nsSNP	CC	66/63	Reference		1.00
		CT	65/63	1.15 (0.68–1.95)	0.60	
		TT	16/21	1.41 (0.64–3.11)	0.40	
		CT/TT	81/84	1.21 (0.73–1.98)	0.46	
		P trend = 0.38				
XRCC3 A17893G rs1799796	Intron	AA	54/63	Reference		0.30
		AG	63/66	0.85 (0.49–1.48)	0.57	
		GG	27/7	0.18 (0.07–0.47)	0.0005 ^e	
		AG/GG	90/73	0.63 (0.38–1.07)	0.09	
		P trend = 0.002				
RAG1 K820R rs2227973	nsSNP	AA	116/107	Reference		0.09
		AG	25/28	1.13 (0.59–2.14)	0.72	
		GG	4/2	0.47 (0.08–2.82)	0.41	
		AG/GG	29/30	1.03 (0.56–1.91)	0.92	
		P trend = 0.85				
XRCC4 IV7-1G > A rs1805377	Splicing site	GG	109/98	Reference		0.09
		GA	31/32	1.13 (0.61–2.06)	0.70	
		AA	6/5	0.61 (0.14–2.64)	0.51	
		GA/AA	37/37	1.05 (0.59–1.88)	0.87	
		P trend = 0.92				
KU80 Exon21 + 466A > G rs1051685	3' UTR	AA	108/103	Reference		0.17
		AG	31/31	1.13 (0.62–2.09)	0.69	
		GG	5/3	0.87 (0.19–4.12)	0.86	
		AG/GG	36/34	1.10 (0.61–1.98)	0.74	
		P trend = 0.84				
LIG4 T91I rs1805388	nsSNP	CC	93/93	Reference		0.29
		CT	49/40	0.74 (0.43–1.29)	0.29	
		TT	3/3	1.34 (0.23–7.69)	0.74	
		CT/TT	52/43	0.77 (0.45–1.32)	0.34	
		P trend = 0.46				

a Adjusted by age, gender, ethnicity, smoking status and alcohol consumption.

b HWE: Hardy–Weinberg equilibrium.

c nsSNP: non-synonymous SNP.

d NA, not available.

e Remained significant after Bonferroni adjustment for multiple comparisons.

a significantly decreased risk (OR = 0.18, 95% CI 0.07–0.47) (remained significant after Bonferroni adjustment for multiple comparisons). The trend test is also significant (P for

trend = 0.002). The risk of the variant-containing genotypes (AG plus GG) was 0.63 (95% CI 0.38–1.07) compared with the risk associated with the homozygous wild-type geno-

type, which showed a borderline statistical significance ($P = 0.09$). In an exploratory analysis, we further stratified the risk associations of XRCC3 A17893G with various host characteristics. This showed that the reduction in risk was only significant in males (OR = 0.13, 95% CI 0.03–0.52, $P = 0.004$) (remained significant after Bonferroni adjustment for multiple comparisons) and in ever smokers (OR = 0.22, 95% CI 0.07–0.69, $P = 0.01$) (Table 3). The protective effect of the homozygous variant genotype was significant in both young subjects (OR = 0.18, 95% CI 0.04–0.73, $P = 0.02$) and old subjects (OR = 0.14, 95% CI 0.02–0.78, $P = 0.02$) and in both ever drinkers (OR = 0.12, 95% CI 0.02–0.63, $P = 0.01$) and never drinkers (OR = 0.06, 95% CI 0.11–0.33, $P = 0.001$) (remained significant after Bonferroni adjustment for multiple comparisons) (Table 3). No significant associations with OPL risk were identified for the other SNPs examined in this study (Table 2).

3.3. Associations of XRCC3 haplotypes with OPL risk

Table 4 shows the associations of OPL risk with XRCC3 haplotypes constructed in the order of A17893G-T241M. The A–C haplotype, which accounts for 38.9% and 46.3% of all haplotypes in cases and controls, respectively, were used as the referent. The G–C haplotype containing the wild-type allele of T241M and the variant allele of A17893G were associated with a significantly decreased OPL risk (OR = 0.40, 95% CI 0.23–0.68, $P = 0.001$) (remained significant after Bonferroni adjustment for multiple comparisons), which was consistent with the result from the single SNP analysis. The A–T haplotype were associated with a reduced OPL risk (OR = 0.67, 95% CI 0.40–1.11, $P = 0.12$) compared with the reference A–C haplotype. The global haplotype test indicated that the distribution pattern of the four haplotypes differed significantly between cases and controls ($\chi^2 = 12.64$, $P_{3df} = 0.006$).

Table 3 – Association of XRCC3 A17893G polymorphism with OPL risk stratified by host characteristics

Host characteristic	Control/case	OR (95% CI) ^a P-value	Control/case	OR (95% CI) P-value
Age ^b		Old		Young
AA	22/23	Reference	32/40	Reference
AG	29/24	0.96 (0.38–2.40) $P = 0.93$	34/42	0.93 (0.45–1.90) $P = 0.84$
GG	14/4	0.14 (0.02–0.78) $P = 0.02$	13/3	0.18 (0.04–0.73) $P = 0.02$
AG/GG	43/28	0.66 (0.28–1.58) $P = 0.35$	47/45	0.71 (0.36–1.41) $P = 0.33$
Gender		Female		Male
AA	29/29	Reference	25/34	Reference
AG	25/25	1.02 (0.45–2.35) $P = 0.96$	38/41	0.75 (0.34–1.64) $P = 0.47$
GG	10/3	0.28 (0.06–1.30) $P = 0.10$	17/4	0.13 (0.03–0.52) $P = 0.004^c$
AG/GG	35/28	0.82 (0.37–1.81) $P = 0.62$	55/45	0.56 (0.26–1.18) $P = 0.13$
Smoking status		Never		Ever
AA	30/21	Reference	24/42	Reference
AG	33/22	0.93 (0.40–2.13) $P = 0.86$	30/44	0.66 (0.30–1.42) $P = 0.28$
GG	15/0	NA ^d	12/7	0.22 (0.07–0.69) $P = 0.01$
AG/GG	48/22	0.64 (0.29–1.44) $P = 0.28$	42/51	0.52 (0.25–1.08) $P = 0.08$
Alcohol consumption		Never		Ever
AA	16/32	Reference	38/31	Reference
AG	18/26	0.39 (0.13–1.19) $P = 0.10$	45/40	1.17 (0.57–2.43) $P = 0.67$
GG	12/4	0.06 (0.11–0.33) $P = 0.001^c$	15/3	0.12 (0.02–0.63) $P = 0.01$
AG/GG	30/30	0.26 (0.09–0.75) $P = 0.01$	60/43	0.90 (0.45–1.81) $P = 0.76$

a Adjusted for age, gender, ethnicity, smoking status and alcohol consumption, where appropriate.

b Used median age in controls as cut-off point.

c Remained significant after Bonferroni adjustment for multiple comparisons.

d NA, not available.

Table 4 – Associations of XRCC3 haplotypes with OPL risk

Haplotype ^a	Case, N (%)	Control, N (%)	OR (95% CI) ^b	P-value
A–C	112 (38.9%)	126 (46.3%)	Reference	
G–C	81 (28.1%)	49 (18.0%)	0.40 (0.23–0.68)	0.001 ^b
A–T	59 (20.5%)	66 (24.3%)	0.67 (0.40–1.11)	0.12
G–T	36 (12.5%)	31 (11.4%)	0.45 (0.20–1.01)	0.05
Global test P-value ^c				$\chi^2 = 12.64$, $P_{3df} = 0.006$

a In the order of A17893G-T241M with nucleotides A and C as the common allele, respectively.

b Remained significant after Bonferroni adjustment for multiple comparisons.

c Adjusted for age, gender, ethnicity, smoking status and alcohol consumption.

3.4. Combined XRCC3 diplotypes and OPL risk

To further investigate the effect of the G–C haplotype of XRCC3 on OPL risk, we performed a trend analysis using the study subjects without this haplotype as the reference group. A statistically significant trend towards a reduction in OPL risk was noted for subjects with increasing number of the G–C haplotype. That is, the OR was 1.04 (0.56–1.95, $P = 0.90$) and 0.20 (0.08–0.51, $P = 0.001$) for subjects with one copy and two copies of the G–C haplotype, respectively (P for trend = 0.005).

3.5. CART analysis

Fig. 1A depicts the tree structure generated by the CART analysis, which included all investigated genetic variants and the smoking status variable. Smoking status was singled out in the first splitting node (well-matched variables such as age, gender and ethnicity were excluded from the CART analysis), confirming that smoking is the predominant risk factor in

OPL development. XRCC3 A17893G was important in the tree generation in both never and ever smokers. Never smokers with the XRCC3 A17893G GG genotype exhibited the lowest case ratio (no cases in the 15 subjects). For never smokers with the XRCC3 A17893 AA/AG genotypes, BRCA2 N372H and XRCC3 T241M were the determining factors for downstream tree structures. The case ratio for subjects with the BRCA2 N372H TG/GG-XRCC3 T241M CC, BRCA2 N372H TG/GG-XRCC3 T241M CT/TT and BRCA2 N372H TT genotype combinations was 18%, 46% and 54%, respectively. The key predictors downstream of XRCC3 A17893G SNP in ever smokers were NBS1 E185Q and XRCC2 C41657T. Individuals with the GG genotype of XRCC3 A17893G and the CG/GG genotypes of NBS1 E185Q had the lowest OPL risk, with a 10% case ratio. In comparison, subjects with the AA/AG genotypes of XRCC3 A17893G and the CT/TT genotypes of XRCC2 C41657T exhibited the highest OPL risk, with a 92% case ratio. Fig. 1B shows the OR estimates generated for the three different risk groups determined on the basis of the case ratio of each CART terminal node. Compared with the low-risk group combining terminal nodes with

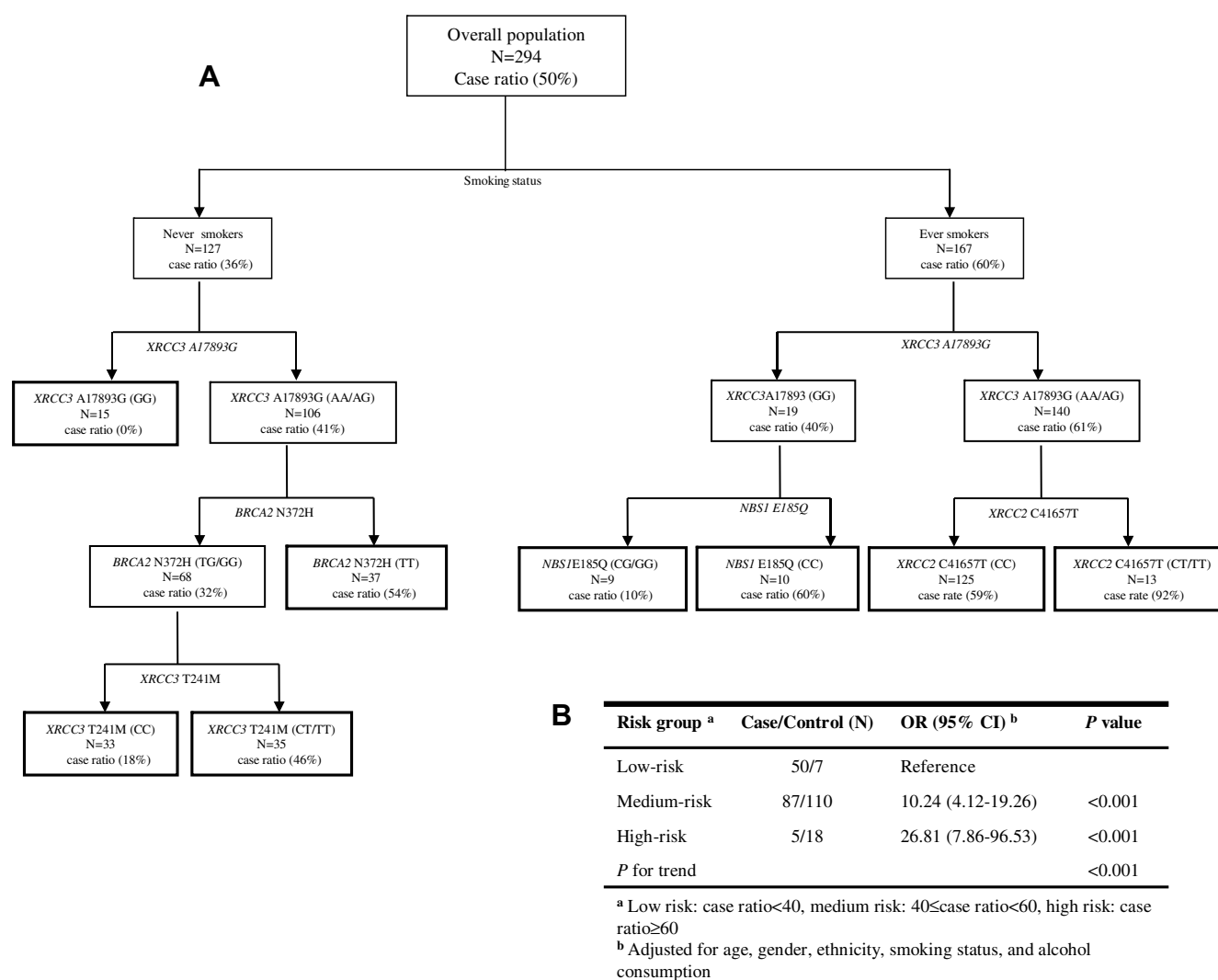


Fig. 1 – (A) CART analysis of genetic variants in the DSB repair pathway and smoking in modulating OPL risk. (B) Combined analysis of the effects of genetic variants in the DSB repair pathway on OPL risk based on the results of the CART analysis.

a case ratio less than 40%, the medium-risk (case ratio between 40% and 60%) and high-risk groups (case ratio more than 60%) were both associated with a significantly increased OPL risk, with ORs of 10.24 (95% CI 4.12–19.26) and 26.81 (95% CI 7.86–96.53), respectively (P for trend < 0.001).

4. Discussion

Given that patients with OPLs are at a dramatically elevated risk of oral cancer, identification of subjects susceptible to OPL development is critical for oral cancer prevention. In this study, we assessed the effects of 10 potentially functional polymorphisms in the DSB DNA repair pathway on OPL susceptibility. We found that the variant allele of the XRCC3 A17893G SNP exhibited a significant protective effect on OPL risk which remained robust after conservative Bonferroni adjustment for multiple comparisons. The effect was further confirmed in the haplotype and diplotype analyses of XRCC3. Remarkably, through the CART analysis we were also able to identify distinctive polymorphic signatures in the DSB repair pathway that differentiated the study participants into different risk subsets, indicating the potential presence of high-order gene–gene and gene–environment interactions.

XRCC3 is a key member of the RecA/Rad51-related protein family that is essential for HR DNA repair.⁹ The A17893G SNP is located in the intronic region, and its functional significance has not been evaluated. Nonetheless, multiple studies have shown an association between this SNP and cancer risk. For example, Kuschel and colleagues noted a significant protective effect of the variant allele of XRCC3 A17893G SNP against breast cancer in a UK population.³⁰ This was later confirmed by two population-based studies, one conducted in the United States and the other in Poland.³⁶ This variant allele has also been reported to confer a significantly decreased susceptibility to bladder,³⁵ invasive ovarian²¹ and upper aerodigestive³⁷ cancers. However, a few studies have shown no such association with cancer^{37–40}. Factors such as study design, sample size, cancer sites, population heterogeneity and the biological complexity of low-penetrance cancer susceptibility genes may explain the observed discrepancies. In keeping with our results showing a protective role for the variant allele of XRCC3 A17893G in OPL development, a meta-analysis of a total of 24,795 cancer patients and 34,209 controls reported a significantly reduced cancer risk associated with this allele under a dominant genetic model with an OR of 0.92 (95% CI 0.87–0.96, P = 0.0004).⁴¹ Whether this intronic SNP has any functional impact on the splicing of the XRCC3 transcript needs to be further evaluated through the assessment of XRCC3 mRNA levels in subjects with different genotypes. Until that is done, we cannot rule out the possibility that the decrease in OPL risk is due to other variants that are in linkage disequilibrium with this SNP.

The T241M SNP, which is located in exon 8 of the XRCC3 gene, has been investigated in many studies, with conflicting results. For example, some studies have shown the variant allele to be associated with an increased risk for bladder, breast and lung cancers,^{30,35,36,40} and these results were in agreement with the results from various functional assays showing that the variant allele was associated with an increased DNA

Table 5 – Combined XRCC3 diplotypes and OPL risk

Haplotype pair ^a	Case, N (%)	Control, N (%)	OR (95% CI) ^b	P-value
Other ^c :other	89 (61.8%)	94 (69.1%)	Reference	
Other:G–C	29 (20.1%)	35 (25.7%)	1.04 (0.56–1.95)	0.90
G–C:G–C	26 (18.1%)	7 (5.1%)	0.20 (0.08–0.51)	0.001
P for trend				0.005

a In the order of A17893G–T241M.

b Adjusted for age, gender, ethnicity, smoking status and alcohol consumption.

c A–C, A–T or G–T haplotype.

adduct level and compromised capacity to repair X-ray-induced chromosome aberrations.^{42,43} In contrast, other groups found a lack of association of the T241M SNP with a risk for ovarian, skin and endometrial cancers.^{20,21,39} The findings from these studies were supported by an *in vitro* study conducted by Araujo and colleagues, who showed that the XRCC3 T241M SNP did not affect intrinsic DSB repair capacity.⁴⁴ In further contrast, in our study, the variant T allele was associated with a non-significantly increased OPL risk in the main analysis. However, the haplotype analysis comparing the A–T haplotype with the A–C haplotype (Table 4) Table. 5 showed that, in conjunction with the A17893G polymorphism, the variant allele of T241M might be associated with a reduced OPL risk. This indicates the presence of a potential interaction between these two variants. However, a statistical interaction analysis did not yield any significant finding (data not shown). Further studies using tagged SNPs performed in a larger study are warranted to provide additional insight into these observations.

OPL is a multifactorial and multistep disease with complex interactions amongst these factors. We therefore performed an exploratory CART analysis to further elucidate the high-order gene–gene and gene–environment interactions of the DSB repair pathway in OPL development, and the results were informative. For example, the XRCC3 A17893G SNP was identified to be the best factor for discriminating between cases and controls amongst both never and ever smokers, which is consistent with the finding from the main effect analysis. We then identified distinctive polymorphic signatures that grouped the subjects according to different risk levels. This latter finding indicates that the development of OPL involves complex genetic interactions and follows different pathways depending on the specific genetic background and environmental exposures of the subjects. By combining risk groups with comparable case ratios, we found a statistically significant dose effect on OPL risk in groups with different case ratios, suggesting a good discriminating ability of the CART analysis. However, it should be noted that in this study we focused on those well-established DSB genes and SNPs with reported functional roles in tumorigenesis. A collection based on an unbiased selection of all DSB-related genes/SNPs would be necessary to provide more in-depth insights. Furthermore, our results should also be interpreted with a certain amount of caution because of the *post hoc* data-mining approach and the limited number of subjects in the stratified analyses and some CART terminal nodes. Independent validations done

in a larger study population may shed further light on the mechanisms underlying the complex high-order interactions.

We calculated the power of our study.⁴⁵ We found that with our current sample size, assuming a conservative 0.09% prevalence of OPL in the general population,⁴⁶ an additive genetic model and a two-sided significance level of 0.05, the power to detect an increased OR of 1.6 ranged from 60% to 80% given minor allele frequency from 0.15 to 0.45. The power to detect an increased OR of 1.8 was from 80% to 94% under the same model and the power to detect an increased OR of 2.0 was even higher, ranging from 91% to 98%. In addition, the major finding of our study on XRCC3 A17893G remained significant after the conservative Bonferroni correction for multiple comparisons. However, the possibility of chance finding cannot be completely ruled out and further larger epidemiologic and functional studies are needed to validate the results.

In summary, we used strict matching criteria to eliminate the potential confounding effects of age, gender and ethnicity. In addition, we have complete information on smoking status and alcohol consumption data for each subject. This study presents the first epidemiological evidence supporting the associations between the genotypes and haplotypes of genetic variants in the DSB DNA repair pathway and OPL risk. The XRCC3 gene was identified as a possible culprit in OPL development. Furthermore, we demonstrated the presence of complex interactions between DSB gene variations and smoking that influenced OPL susceptibility.

Conflict of interest statement

None declared.

Acknowledgement

Supported by National Cancer Institute Grants CA 106451 and CA 097007.

REFERENCES

- Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;**55**(2):74–108.
- La Vecchia C, Lucchini F, Negri E, Levi F. Trends in oral cancer mortality in Europe. *Oral Oncol* 2004;**40**(4):433–9.
- Neville BW, Day TA. Oral cancer and precancerous lesions. *CA Cancer J Clin* 2002;**52**(4):195–215.
- Silverman Jr S, Gorsky M, Lozada F. Oral leukoplakia and malignant transformation. A follow-up study of 257 patients. *Cancer* 1984;**53**(3):563–8.
- Lind PO. Malignant transformation in oral leukoplakia. *Scand J Dent Res* 1987;**95**(6):449–55.
- Schepman KP, van der Meij EH, Smelee LE, van der Waal I. Malignant transformation of oral leukoplakia: a follow-up study of a hospital-based population of 166 patients with oral leukoplakia from The Netherlands. *Oral Oncol* 1998;**34**(4):270–5.
- Hashibe M, Sankaranarayanan R, Thomas G, et al. Alcohol drinking, body mass index and the risk of oral leukoplakia in an Indian population. *Int J Cancer* 2000;**88**(1):129–34.
- Gillison ML. Current topics in the epidemiology of oral cavity and oropharyngeal cancers. *Head Neck* 2007;**29**(8):779–92.
- Scott SP, Pandita TK. The cellular control of DNA double-strand breaks. *J Cell Biochem* 2006;**99**(6):1463–75.
- Masson JY, Tarsounas MC, Stasiak AZ, et al. Identification and purification of two distinct complexes containing the five RAD51 paralogs. *Genes Dev* 2001;**15**(24):3296–307.
- Pellegrini L, Yu DS, Lo T, et al. Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature* 2002;**420**(6913):287–93.
- Karran P. DNA double strand break repair in mammalian cells. *Curr Opin Genet Dev* 2000;**10**(2):144–50.
- Pierce AJ, Johnson RD, Thompson LH, Jasin M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev* 1999;**13**(20):2633–8.
- Barnes DE, Stamp G, Rosewell I, Denzel A, Lindahl T. Targeted disruption of the gene encoding DNA ligase IV leads to lethality in embryonic mice. *Curr Biol* 1998;**8**(25):1395–8.
- Thompson LH, Schild D. Homologous recombinational repair of DNA ensures mammalian chromosome stability. *Mutat Res* 2001;**477**(1–2):131–53.
- Deans B, Griffin CS, O'Regan P, Jasin M, Thacker J. Homologous recombination deficiency leads to profound genetic instability in cells derived from Xrcc2-knockout mice. *Cancer Res* 2003;**63**(23):8181–817.
- Ralhan R, Kaur J, Kreienberg R, Wiesmuller L. Links between DNA double strand break repair and breast cancer: accumulating evidence from both familial and nonfamilial cases. *Cancer Lett* 2007;**248**(1):1–17.
- Zienoldind S, Campa D, Lind H, et al. Polymorphisms of DNA repair genes and risk of non-small cell lung cancer. *Carcinogenesis* 2006;**27**(3):560–7.
- Kadouri L, Kote-Jarai Z, Hubert A, et al. A single-nucleotide polymorphism in the RAD51 gene modifies breast cancer risk in BRCA2 carriers, but not in BRCA1 carriers or noncarriers. *Br J Cancer* 2004;**90**(10):2002–5.
- Han J, Colditz GA, Samson LD, Hunter DJ. Polymorphisms in DNA double-strand break repair genes and skin cancer risk. *Cancer Res* 2004;**64**(9):3009–13.
- Auranen A, Song H, Waterfall C, et al. Polymorphisms in DNA repair genes and epithelial ovarian cancer risk. *Int J Cancer* 2005;**117**(4):611–8.
- Fu YP, Yu JC, Cheng TC, et al. Breast cancer risk associated with genotypic polymorphism of the nonhomologous end-joining genes: a multigenic study on cancer susceptibility. *Cancer Res* 2003;**63**(10):2440–6.
- Wang Y, Spitz MR, Lee JJ, et al. Nucleotide excision repair pathway genes and oral premalignant lesions. *Clin Cancer Res* 2007;**13**(12):3753–8.
- Huang M, Spitz MR, Gu J, et al. Cyclin D1 gene polymorphism as a risk factor for oral premalignant lesions. *Carcinogenesis* 2006;**27**(10):2034–7.
- Hudmon KS, Honn SE, Jiang H, et al. Identifying and recruiting healthy control subjects from a managed care organization: a methodology for molecular epidemiological case-control studies of cancer. *Cancer Epidemiol Biomarkers Prev* 1997;**6**(8):565–71.
- Ryk C, Kumar R, Thirumaran RK, Hou SM. Polymorphisms in the DNA repair genes XRCC1, APEX1, XRCC3 and NBS1, and the risk for lung cancer in never- and ever-smokers. *Lung Cancer* 2006;**54**(3):285–92.
- Maillet P, Chappuis PO, Vaudan G, et al. A polymorphism in the ATM gene modulates the penetrance of hereditary non-polyposis colorectal cancer. *Int J Cancer* 2000;**88**(6):928–31.
- Wu X, Gu J, Grossman HB, et al. Bladder cancer predisposition: a multigenic approach to DNA-repair and cell-cycle-control genes. *Am J Hum Genet* 2006;**78**(3):464–79.

29. Roddam PL, Rollinson S, O'Driscoll M, et al. Genetic variants of NHEJ DNA ligase IV can affect the risk of developing multiple myeloma, a tumour characterised by aberrant class switch recombination. *J Med Genet* 2002;**39**(12):900–5.
30. Kuschel B, Auranen A, McBride S, et al. Variants in DNA double-strand break repair genes and breast cancer susceptibility. *Hum Mol Genet* 2002;**11**(12):1399–407.
31. Figueroa JD, Malats N, Rothman N, et al. Evaluation of genetic variation in the double-strand break repair pathway and bladder cancer risk. *Carcinogenesis* 2007;**28**(8):1788–93.
32. Hayden PJ, Tewari P, Morris DW, et al. Variation in DNA repair genes XRCC3, XRCC4, XRCC5 and susceptibility to myeloma. *Hum Mol Genet* 2007;**16**(24):3117–27.
33. Wang N, Dong XJ, Zhou RM, et al. An investigation on the polymorphisms of two DNA repair genes and susceptibility to ESCC and GCA of high-incidence region in northern China. *Mol Biol Rep* 2007.
34. Palli D, Falchetti M, Masala G, et al. Association between the BRCA2 N372H variant and male breast cancer risk: a population-based case-control study in Tuscany, Central Italy. *BMC Cancer* 2007;**7**:170.
35. Matullo G, Guarrera S, Sacerdote C, et al. Polymorphisms/haplotypes in DNA repair genes and smoking: a bladder cancer case-control study. *Cancer Epidemiol Biomarkers Prev* 2005;**14**(11 Pt 1):2569–78.
36. Garcia-Closas M, Egan KM, Newcomb PA, et al. Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. *Hum Genet* 2006;**119**(4):376–88.
37. Matullo G, Dunning AM, Guarrera S, et al. DNA repair polymorphisms and cancer risk in non-smokers in a cohort study. *Carcinogenesis* 2006;**27**(5):997–1007.
38. Han J, Hankinson SE, Ranu H, De Vivo I, Hunter DJ. Polymorphisms in DNA double-strand break repair genes and breast cancer risk in the Nurses' Health Study. *Carcinogenesis* 2004;**25**(2):189–95.
39. Han J, Hankinson SE, Hunter DJ, De Vivo I. Genetic variations in XRCC2 and XRCC3 are not associated with endometrial cancer risk. *Cancer Epidemiol Biomarkers Prev* 2004;**13**(2):330–1.
40. Jacobsen NR, Raaschou-Nielsen O, Nexø B, et al. XRCC3 polymorphisms and risk of lung cancer. *Cancer Lett* 2004;**213**(1):67–72.
41. Han S, Zhang HT, Wang Z, et al. DNA repair gene XRCC3 polymorphisms and cancer risk: a meta-analysis of 48 case-control studies. *Eur J Hum Genet* 2006;**14**(10):1136–44.
42. Matullo G, Palli D, Peluso M, et al. XRCC1, XRCC3, XPD gene polymorphisms, smoking and (32)P-DNA adducts in a sample of healthy subjects. *Carcinogenesis* 2001;**22**(9):1437–45.
43. Au WW, Salama SA, Sierra-Torres CH. Functional characterization of polymorphisms in DNA repair genes using cytogenetic challenge assays. *Environ Health Perspect* 2003;**111**(15):1843–50.
44. Araujo FD, Pierce AJ, Stark JM, Jasin M. Variant XRCC3 implicated in cancer is functional in homology-directed repair of double-strand breaks. *Oncogene* 2002;**21**(26):4176–80.
45. Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol* 2002;**155**(5):478–84.
46. Hashibe M, Mathew B, Kuruvilla B, et al. Chewing tobacco, alcohol, and the risk of erythroplakia. *Cancer Epidemiol Biomarkers Prev* 2000;**9**(7):639–45.