

ORIGINAL ARTICLE

Reduced mRNA expression of nucleotide excision repair genes in lymphocytes and risk of squamous cell carcinoma of the head and neck

Peng Han^{1,2,3}, Fengqin Gao^{1,2}, Hongliang Liu^{1,2}, Zhensheng Liu^{1,2}, Qiong Shi^{1,2}, Jesse D. Troy^{1,4}, Kouros Owzar^{1,5}, Walter Lee^{1,6}, Jose P. Zevallos^{7,8}, Erich M. Sturgis^{9,10} and Qingyi Wei^{1,2,*}

¹Duke Cancer Institute, Duke University Medical Center and ²Department of Medicine, Duke University School of Medicine, Durham, NC, 27710, USA, ³Department of Otorhinolaryngology-Head and Neck Surgery, Second Affiliated Hospital, Xi'an Jiaotong University College of Medicine, Xi'an, Shaanxi, 710004, PR China, ⁴Division of Blood and Marrow Transplantation and ⁵Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC, 27710, USA, ⁶Division of Head and Neck Surgery Communications Science, Department of Surgery, Duke University Medical Center, Durham, NC, 27710, USA, ⁷Department of Otolaryngology-Head and Neck Surgery and ⁸Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA and ⁹Departments of Head and Neck Surgery and ¹⁰Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, 77030, USA

*To whom correspondence should be addressed. Department of Medicine, Duke Cancer Institute, Mail box 103858, 10 Bryan Searle Drive, Durham, NC 27710, USA. Tel: +1 919 660 0562; Fax: +1 919 681 7385; E-mail: qingyi.wei@duke.edu

Abstract

Nucleotide excision repair (NER) plays a critical role in the development of smoking-related cancers. We hypothesize that mRNA expression levels of NER genes are associated with risk of the squamous cell carcinoma of head and neck (SCCHN). To test this hypothesis, we conducted a case-control study of 260 SCCHN patients and 246 cancer-free controls by measuring the mRNA expression levels of eight core NER genes in cultured peripheral lymphocytes. Compared with the controls, cases had statistically significantly lower expression levels of *DDB1* and *ERCC3* ($P = 0.015$ and 0.041 , respectively). Because *DDB1* and *ERCC3* expression levels were highly correlated, we used *DDB1* for further multivariate analyses and modeling. After dividing the subjects by controls' quartiles of expression levels, we found an association between an increased risk of SCCHN and low *DDB1* expression levels [adjusted ORs and 95% CIs: 1.92 and 1.11–3.32, 1.48 and 0.85–2.59, 2.00 and 1.15–3.45 for the 2nd–4th quartiles, respectively, compared with the 1st quartile; $P_{\text{trend}} = 0.026$]. We also identified a multiplicative interaction between sex and *DDB1* expression levels ($P = 0.007$). Finally, the expanded model with gene expression levels, in addition to demographic and clinical variables, on SCCHN risk was significantly improved, especially among men. In conclusion, reduced *DDB1* expression levels were associated with an increased risk of SCCHN. However, these results need to be validated in larger studies.

Introduction

Squamous cell carcinoma of head and neck (SCCHN) is among the most common malignancies worldwide, which includes cancers of the oral cavity, larynx and pharynx (1,2). In the United States, the estimated number of new SCCHN cases has been

increasing from 48 010 in 2009 to 61 760 in 2016, according to the American Cancer Society (3–5). Although it is well established that cigarette smoking and alcohol use are major risk factors for SCCHN in general as well as prior HPV type 16 infection for

Abbreviations

BPDE	benzo (a) pyrene diol epoxide
CI	confidence interval
DRC	DNA repair capacity
HCR	host-cell reactivation
NER	nucleotide excision repair
OR	odds ratio
ROC	receiver operating characteristic
SCCHN	squamous cell carcinoma of head and neck
XP	xeroderma pigmentosum

oropharyngeal cancer, only a small fraction of those who have the history of smoking, alcohol use or HPV infection develop one of these cancers in their lifetime, suggesting that there may be heterogeneity in SCCHN susceptibility (6,7).

Numerous chemicals in tobacco smoke can cause damage to cellular DNA (8,9). For example, one of these chemicals is benzo (a) pyrene diol epoxide (BPDE) that is a bio-activated form of benzo (a) pyrene, which is a classic tobacco carcinogen that can induce irreversible damage to DNA through the formation of covalent binding or oxidation by forming DNA adducts (10–12). Several DNA repair pathways have evolved to repair these adducts (13–15), among which nucleotide excision repair (NER) is a major and well-studied but complex and versatile mechanism (16). NER essentially uses eight core genes to restore the damaged DNA to normal one in eukaryotes. The basic steps in NER include: damage recognition (DDB1, XPA and XPC), unwinding the duplex and kinetic proofreading (ERCC2/XPD and ERCC3/XPB), dual incision (ERCC1, ERCC4/XPF and ERCC5/XPG), repair synthesis and ligation (17). Functional mutations in any of these genes may lead to abnormal NER and subsequently increase susceptibility to cancer. In humans, for example, several rare disorders are associated with NER deficiency (18), including xeroderma pigmentosum (XP) that is an autosomal recessive disease with a defective NER (19). XP patients have mutations in at least one of these eight NER genes and consequently have a remarkably increased risk of developing sun-light induced skin cancers (20). In addition to NER, double strand break (DSB) repair may also play an important role in HPV-positive oropharyngeal cancer, and we recently showed that reduced DSB repair was associated with increased risk of SCCHN as well (21).

In an early study of 55 SCCHN patients and 61 healthy controls by the host-cell reactivation (HCR) assay that measures DNA repair capacity (DRC), it was reported that the cases had a significantly lower DRC than that of controls, and subjects with a reduced DRC had a greater than 2-fold increased risk of developing head and neck cancer, compared with those with a higher DRC (22). These results were confirmed later by a much larger study of 744 SCCHN patients and 753 controls (23). To further understand the underlying molecular mechanism, we conducted a pilot study of 57 SCCHN patients and 105 cancer-free controls and reported that an increased risk of SCCHN was associated with reduced mRNA expression levels of NER genes in lymphocytes (24). However, previous association studies of NER mRNA expression and risk of SCCHN had relatively limited numbers of cases and controls, and also we were not able to measure the expression of DDB1, ERCC2, ERCC4 or XPA, because the sequences of these genes were unknown at that time. Thus, studies of a larger sample size are required to further evaluate the previously reported association between NER gene expression and SCCHN risk. Therefore, we conducted a larger case-control study to test associations between mRNA expression levels of eight core NER genes and risk of SCCHN. In the present study, because

the precise molecular mechanism underlying HPV-positive oropharyngeal cancer is still not well documented, we specifically tested expression levels of genes in the NER pathway.

Materials and Methods**Study subjects**

The SCCHN study cases and controls were recruited from The University of Texas M.D. Anderson Cancer Center between 2002 and 2012 for a molecular epidemiology study, and a subset of 506 subjects (260 cases and 246 controls) were selected for the present study, because they had additional blood samples for culture of T-lymphocytes that were sufficient for total RNA extraction. In order to evaluate selection bias, we compared the case-control samples included in the study with the samples excluded. The SCCHN patients were recruited, if they meet the following criteria: newly diagnosed, 18 years and older, untreated, no other previous cancers but diagnosed with histologically confirmed SCCHN. The cancer-free controls were recruited among visitors accompanying patients to clinics other than the Head and Neck clinic where cases were recruited; they were frequency-matched with cases on age and sex and had no blood transfusion in the last six months as well as no history of any previous cancers. In order to avoid genetic heterogeneity of different races, the subjects included in the analysis were all non-Hispanic white. A written informed consent was obtained from eligible patients and controls. Subjects who had smoked more than 100 cigarettes during their lifetime were defined as smokers, of which those who had stopped smoking at least 12 months were defined as former smokers and remaining was considered current smokers; others were considered never smokers. Subjects who consumed alcoholic beverages at least weekly for one year were defined as drinkers, of which those who had stopped drinking more than 12 months were defined as former and the remaining was considered current drinkers; others were considered never drinkers. Ever smokers and ever drinkers were also used to describe former and current smokers or drinkers, respectively. Each subject donated a 30 ml blood sample drawn into a heparinized tube before receiving treatment for cases and when consent was given for controls. The study protocol was approved by M.D. Anderson Institutional Review Board.

Quantitative real-time PCR

The mRNA expression levels of eight core NER genes were examined by quantitative real-time PCR using the total RNA that was isolated from each sample of lymphocytes from the 506 subjects with the TRIzol reagent (Invitrogen™, Carlsbad, CA). The PCR was performed using the ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA) by TaqMan gene expression assays with the master mix reagent (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions which was also described previously (25,26). Each amplification reaction was performed in a final volume of 5 μ l containing 5 ng of the complementary DNA, 0.25 μ l primers and 2.5 μ l Master mix. The PCR reaction was carried out in 384-well optical plates with a final volume of 5 μ l reaction mixtures. Thermal cycling conditions were as follows: 95°C for 5 min, followed by 40 cycles of denaturation at 95°C for 15 s, and annealing/extension at 60°C for 1 min. The 18S expression was assessed as an internal control to account for differences of complementary DNA. Each sample was analyzed in duplicate, and if the coefficient of variation of all reactions was <5%, the mean values of the duplicates were used. When coefficient of variation was >5%, the assay was repeated and the samples with new coefficient of variation >5% were removed. The threshold cycle or Ct value was the PCR cycle at which a significant increase in fluorescent signal was first detected. Ct values form the basis for quantitative comparison of genes in the real-time PCR. The expression levels of eight NER genes relative to that of 18S were calculated by delta Ct (Δ Ct). The Δ Ct value was the Ct value of the target gene subtracted its Ct value of 18S, Δ Ct = Ct (Gene) – Ct (18S). Therefore, the smaller the Δ Ct values, the higher expression levels of the target mRNA.

Statistical analysis

The Chi-square test was used to evaluate differences in selected demographic variables, smoking and alcohol consumption between cases and

controls. Student's *t*-test, Wilcoxon rank-sum test or ANOVA were used to compare differences in the relative mRNA expression levels analyzed as a continuous variable among groups. In addition, we performed a stepwise logistic regression analysis to explore the best model to predict SCCHN risk. We used the quartiles of ΔCt values in the control group as cutoff values for calculating crude odds ratio (OR) and their 95% confidence intervals (CI). The associations of NER genes mRNA expression levels with SCCHN risk were estimated by computing ORs and CIs from logistic regression models. These analyses were performed with or without adjustment for age (in years), sex (female versus male), smoking (never, former or current) and alcohol use (never, former or current).

Further stratification analyses were performed to identify effect modification of related NER genes and selected variables. A multiplicative interaction was suggested when $\text{OR}_{11} > \text{OR}_{10} \times \text{OR}_{01}$, in which OR_{11} was the OR when both factors were present, OR_{01} was the OR when only factor 1 was present, and OR_{10} was the OR when only factor 2 was present. We evaluated the interactions by looking for a multiplicative scale using standard unconditional logistic regression models.

To assess the improvement of SCCHN risk models by gene expression levels, we compared the area under the receiver operating characteristic (ROC) curve (AUC) among three risk models: the baseline model including only epidemiologic variables [e.g. age, sex, smoking and drinking status], the NER model including the epidemiologic variables and NER gene expression (continuous variable), and the interaction model including the epidemiologic variables, NER gene expression and interaction term. We further calculated the AUC upon stratification by related covariates. All tests were two-sided and $P < 0.05$ was considered significant. All statistical analyses were performed with SAS software (version 9.4; SAS Institute, Inc., Cary, NC).

Results

Characteristics of the study population

The summary of the distributions of selected characteristics of cases and controls is presented in Table 1. There were no significant differences in the distributions of age and sex between cases and controls. Furthermore, no significant differences were observed in the distributions of age, sex, smoking and drinking status between the included and excluded subjects, except for

Table 1. Distributions of demographic variables and tumor characteristics between 260 SCCHN cases and 246 cancer-free controls

Variable	Case (n = 260)	Control (n = 246)	P*
Age			0.258
Median (range)	58 (22–86)	56 (25–86)	
≤56	118 (45.4)	124 (50.4)	
>56	142 (54.6)	122 (49.6)	
Sex			0.314
Female	67 (25.8)	54 (21.9)	
Male	193 (74.2)	192 (78.1)	
Smoking			<0.001
Never	92 (35.4)	125 (50.8)	
Former	88 (33.8)	93 (37.8)	
Current	80 (30.8)	28 (11.4)	
Drinking			0.014
Never	70 (26.9)	94 (38.2)	
Former	63 (24.2)	59 (24.0)	
Current	127 (48.9)	93 (37.8)	
Tumor site			
Oral cavity	80 (30.8)		
Oropharynx	142 (54.6)		
Larynx/Hypopharynx	38 (14.6)		

SCCHN, squamous cell carcinoma of the head and neck.

*Chi-square tests for the distributions comparison of the demographic variables between cases and controls.

the distribution of drinkers in controls. Even though the difference in the distribution of drinkers between the included and excluded controls was statistically significant, the two control groups both had more never drinkers than current drinkers and more current drinkers than former drinkers (Supplementary Table 1). These results suggest there was no selection bias in the included samples. The average age was 58.4 years for the case subjects (median, 58; range, 22–86) and 55.7 years for the control subjects (median, 56; range, 25–86). Of all the subjects, 74.2% of cases and 78.1% of controls were male, 64.6% of cases and 49.2% of controls were ever smokers and 73.1% of cases and 61.8% of controls were ever drinkers. There were more current smokers (30.8%) and current drinkers (48.9%) in cases than in controls (11.4% and 37.8%, respectively). The primary SCCHN of 260 patients included the oral cavity (80, 30.8%), oropharynx (142, 54.6%), hypopharynx (6, 2.3%) and larynx (32, 12.3%).

Differences in NER mRNA expression levels between cases and controls

The cases showed lower relative mean expression levels in seven of the eight core NER genes analyzed than did controls, except for XPA (Table 2). In Student's *t*-test or Wilcoxon rank-sum test for differences in NER mRNA expression levels between cases and controls, only DDB1 and ERCC3 levels were statistically significantly lower in cases than in controls ($P = 0.015$ and $P = 0.041$, respectively; Figure 1A). Because the expression levels of the eight NER genes were measured at the same time, they were likely to be correlated with each other. As shown in Supplementary Table 2, mRNA expression levels of all eight NER genes were statistically significantly correlated with each other. Subsequently, when we included the expression levels of both DDB1 and ERCC3 in the stepwise logistic regression analysis after adjustment for age, sex, smoking and alcohol use, only DDB1 remained in the final model ($P = 0.017$).

Stratification analyses of mRNA expression levels of DDB1 and ERCC3 by selected variables

Stratification analyses of DDB1 and ERCC3 expression levels revealed that patients in subgroups of the age ≤ 56 , male, former smokers and current drinkers exhibited significantly lower mean expression levels of DDB1 than did controls ($P = 0.030$, $P < 0.001$, $P = 0.021$, and $P = 0.012$, respectively, Table 3). Likewise, patients in subgroups of the age ≤ 56 , male, and never smokers exhibited significantly lower mean expression of ERCC3 than did controls ($P = 0.013$, $P = 0.015$, and $P = 0.042$, respectively, Table 3). In both cases and controls, women had lower expression levels of DDB1 and ERCC3 than did men, but the sex differences in the expression levels were only significant in control groups ($P < 0.001$ and $P = 0.004$, respectively, Table 3). Moreover, there were more ever smokers and ever drinkers in males than in females, but the difference was only statistically significant in alcohol drinkers ($P < 0.001$), and not statistically significant in smokers ($P = 0.387$, Supplementary Table 3). There were no significant differences in the expression levels of DDB1 and ERCC3 by tumor sites, suggesting that expression levels of DDB1 and ERCC3 may not be different among tumors of SCCHN (Supplementary Table 4).

Associations between NER mRNA expression levels and risk of SCCHN

To estimate SCCHN risk, the relative expression levels were grouped into quartile values of the controls (Table 4 and Supplementary Table 5). The crude ORs for SCCHN risk associated with relative expression levels of DDB1 in the second quartile,

Table 2. Comparison of the mRNA expression levels of eight core nucleotide excision repair genes between the cases and controls

Gene	Mean ± SD		Median/IQR		Difference	
	Case (n = 260)	Control (n = 246)	Case (n = 260)	Control (n = 246)	Case - Control	P
DDB1	18.44 ± 2.20	17.95 ± 2.32	18.29/3.32	17.93/3.75	0.49	0.015*
ERCC1	19.29 ± 2.17	18.95 ± 2.19	19.32/3.55	18.76/3.38	0.34	0.125
ERCC2	21.54 ± 2.24	21.30 ± 2.05	21.65/3.66	21.50/3.53	0.24	0.243
ERCC3	19.52 ± 2.05	19.09 ± 2.08	19.54/3.21	19.25/3.09	0.43	0.041
ERCC4	21.90 ± 2.15	21.73 ± 2.33	22.10/3.57	21.67/3.75	0.17	0.568
ERCC5	20.07 ± 1.96	19.99 ± 1.88	20.16/3.06	20.07/3.02	0.08	0.670
XPA	20.45 ± 2.06	20.47 ± 1.96	20.33/3.50	20.41/2.88	-0.01	0.813
XPC	20.58 ± 2.06	20.48 ± 2.34	20.46/3.32	20.66/3.21	0.10	0.855

The expression levels of eight nucleotide excision repair genes were calculated by Δ Ct values, the smaller the Δ Ct represents the higher expression levels of the target mRNA. P value in Wilcoxon rank-sum tests. IQR, interquartile range; SD, standard deviation.

*P value in two-sided t-tests.

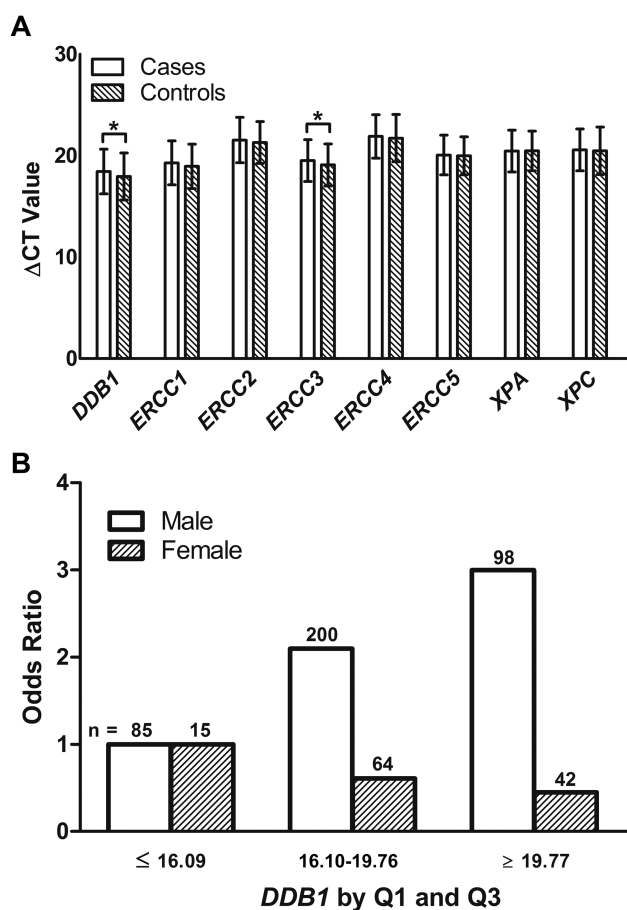


Figure 1. (A) Relative mRNA expression levels of eight NER genes between SCCHN patients and healthy controls. Quantitative real-time PCR was used to measure the relative expressions of eight NER genes. Housekeeping gene 18S was used as the reference. The threshold cycle or Ct value is the PCR cycle at which a significant increase in fluorescent signal is first detected. The expression levels of eight NER genes relative to that of 18S were calculated by Δ Ct. Δ Ct = Ct (Gene) - Ct (18S). The smaller the Δ Ct value, the higher the level of the measured target mRNA (*P < 0.05). (B) Modification effects of DDB1 by sex.

third quartile and fourth quartile were 2.00 (95% CI, 1.19–3.37), 1.64 (95% CI, 0.96–2.80) and 2.03 (95% CI, 1.20–3.42), while the corresponding crude ORs for ERCC3 were 1.64 (95% CI, 0.98–2.74), 1.34 (95% CI, 0.79–2.27) and 1.93 (95% CI, 1.16–3.21), respectively, compared with the first quartile (the highest expression levels of

the gene). After adjusting for age, sex, smoking status and alcohol consumption in multivariate logistic regression analysis, the ORs of DDB1 and ERCC3 remained essentially unchanged. When continuous Δ Ct values were used in the logistic regression model with adjustment for all covariates, there was also a dose-response relationship between the reduced expression levels and the increased SCCHN risk for both DDB1 ($P_{\text{trend}} = 0.026$) and ERCC3 ($P_{\text{trend}} = 0.036$).

Interactions between DDB1 expression levels and selected variables

We further assessed possible interactions on a multiplicative scale between expression levels of DDB1 and selected variables listed in Table 1. The multiplicative interaction was tested when we included the interaction term (i.e., relative expression levels of DDB1 × each of the risk factors) in a multivariate regression model that also included the main effects of NER gene expression levels and other covariates. We found that sex (male versus female) had a significantly multiplicative interaction with relative expression levels of DDB1 ($P = 0.007$, Table 3), but did not find any for ERCC3, in association with SCCHN risk. To further unravel these multiplicative interactions, we stratified the adjusted ORs by sex. It was apparent that ORs for the relative expression levels of DDB1 by Q1 and Q3 in groups of males were greater than those of females (Figure 1B).

We further assessed the prediction performance of models integrating demographic variables and gene expression levels on SCCHN using the ROC curves that measure the effect of DDB1 expression levels and the interaction with sex in three dimensions. The AUC was significantly improved in the model that included the combined effect of DDB1 expression levels and the interaction with sex, compared with the model that did not (Figure 2A, $P = 0.033$). Furthermore, the AUC was insignificantly improved in male subjects that included the combined effect of DDB1 expression levels and the interaction with sex, compared with the model that did not (Figure 2B, $P = 0.083$).

Discussion

In this larger case-control study, we further confirmed our previous studies that reduced NER gene expression was associated with an increased risk of SCCHN. Our results showed that the reduced relative mRNA expression levels of the two NER genes (e.g. DDB1 and ERCC3) were associated with an increased risk of SCCHN. We further assessed interactions between DDB1 expression levels and selected variables and found that sex had a significant multiplicative interaction with DDB1 expression

Table 3. Stratification analyses of mRNA expression levels of *DDB1* and *ERCC3* between SCCHN cases and controls

Variable	DDB1 (Mean ± SD)		P*	P^	ERCC3 (Mean ± SD)		P*	P^
	Case (n = 260)	Control (n = 246)			Case (n = 260)	Control (n = 246)		
Age				0.890				0.920
≤56	18.49 ± 2.19	17.88 ± 1.26	0.030		19.65 ± 2.01	19.05 ± 1.81	0.013**	
>56	18.40 ± 2.22	18.02 ± 2.49	0.191		19.41 ± 2.07	19.13 ± 2.33	0.626**	
P*	0.739	0.641			0.341	0.481**		
Sex				0.007				0.138
Female	18.59 ± 2.32	19.08 ± 2.09	0.233		19.76 ± 2.21	19.81 ± 1.77	0.877	
Male	18.39 ± 2.16	17.63 ± 2.29	0.001		19.44 ± 1.99	18.89 ± 2.13	0.015**	
P*	0.511	<0.001			0.293**	0.004		
Smoking				0.493				0.243
Never	18.37 ± 2.30	17.82 ± 2.50	0.098		19.54 ± 2.08	18.93 ± 2.22	0.042	
Former	18.71 ± 2.20	17.97 ± 2.11	0.021		19.64 ± 2.00	19.16 ± 1.87	0.160**	
Current	18.22 ± 2.08	18.46 ± 2.21	0.597		19.36 ± 2.07	19.59 ± 2.10	0.623	
P***	0.322	0.412			0.675	0.403**		
Drinking				0.255				0.626
Never	18.42 ± 2.55	17.93 ± 2.49	0.219		19.50 ± 2.31	18.95 ± 2.20	0.224**	
Former	18.52 ± 1.87	18.41 ± 2.33	0.781		19.58 ± 1.86	19.45 ± 2.11	0.720	
Current	18.41 ± 2.16	17.66 ± 2.12	0.012		19.50 ± 2.00	19.00 ± 1.89	0.063	
P***	0.944	0.153			0.904**	0.306		

The expression levels of eight nucleotide excision repair genes were calculated by ΔC_t , the smaller the ΔC_t values represents the higher expression levels of the target mRNA. SCCHN, squamous cell carcinoma of the head and neck.

*P value in two-sided t-tests.

**P value in Wilcoxon rank-sum tests.

***P value in one-way ANOVA.

^P value in multiplicative interaction analysis between selected variables and genes in relation to SCCHN risk.

Table 4. Logistic regression analysis of mRNA expression levels of eight core NER genes in SCCHN cases and controls

NER Genes	Quantile mRNA Levels***	Case n (%)	Control n (%)	Crude OR (95% CI)	Adjusted OR* (95% CI)
<i>DDB1</i>	≤16.09	39 (15.0)	61 (25.0)	1.00 (Ref)	1.00 (Ref)
	16.10–17.93	78 (30.0)	61 (25.0)	2.00 (1.19–3.37)	1.92 (1.11–3.32)
	17.94–19.76	64 (24.6)	61 (25.0)	1.64 (0.96–2.80)	1.48 (0.85–2.59)
	≥19.77	79 (30.4)	61 (25.0)	2.03 (1.20–3.42)	2.00 (1.15–3.45)
	P**			0.016	0.026
<i>ERCC3</i>	≤17.50	44 (16.9)	61 (25.0)	1.00 (Ref)	1.00 (Ref)
	17.51–19.25	72 (27.7)	61 (25.0)	1.64 (0.98–2.74)	1.66 (0.97–2.85)
	19.26–20.59	59 (22.7)	61 (25.0)	1.34 (0.79–2.27)	1.32 (0.76–2.29)
	≥20.60	85 (32.7)	61 (25.0)	1.93 (1.16–3.21)	1.90 (1.12–3.24)
	P**			0.022	0.036

CI, confidence interval; NER, nucleotide excision repair; OR, odds ratio; SCCHN, squamous cell carcinoma of the head and neck.

*Adjusted for age, sex, smoking and alcohol status.

**P value in trend test by continuous mRNA expression levels.

***Expression levels by quartile based on the quartile values of control subjects. The expression levels of eight NER genes were calculated by ΔC_t , the smaller the ΔC_t values represent the higher expression levels of the target mRNA.

on SCCHN risk. Moreover, the AUC model suggested that the combined effect of *DDB1* and its interaction with sex further improved the risk prediction.

In previous studies, we measured the DRC of SCCHN cases and healthy controls in their peripheral lymphocytes by using the HCR assay; the results suggested that suboptimal DRC was associated with risk of developing SCCHN (22,23). Later, another study demonstrated a correlation between NER mRNA expression levels and suboptimal DRC phenotype in the lymphocytes (27). Subsequently, we developed a multiplex RT-PCR to measure the association between mRNA expressions of five NER genes (*ERCC1*, *ERCC3*, *ERCC5*, *CSB* and *XPC*) and risk of SCCHN (28). The relative mRNA expression levels of *ERCC1*, *ERCC3*, *ERCC5* and *CSB* were significantly lower in cases than in controls, and the risk of SCCHN associated with low expressions of these genes was

higher by 2- to 6-fold (24). However, we were not able to measure the expressions of *DDB1*, *XPA*, *ERCC2* and *ERCC4*, because at that time the sequences of these genes were unknown. Moreover, the sample size of that study was relatively small, and most positive results in studies with small samples were not substantiated in later larger studies (29,30). In the current study, we confirmed that reduced expression of *ERCC3* was associated with SCCHN risk, with additional finding for *DDB1*. These results further support the notion that altered mRNA levels, which may have an effect on the NER capacity, may contribute to risk of SCCHN. Moreover, as the HPV infection is the main risk factor for oropharyngeal cancer, we performed a case-case analysis of 109 patients who had both the available HPV data and mRNA expression levels, of which 64 patients with oropharyngeal cancer. There were no significant differences in the expression

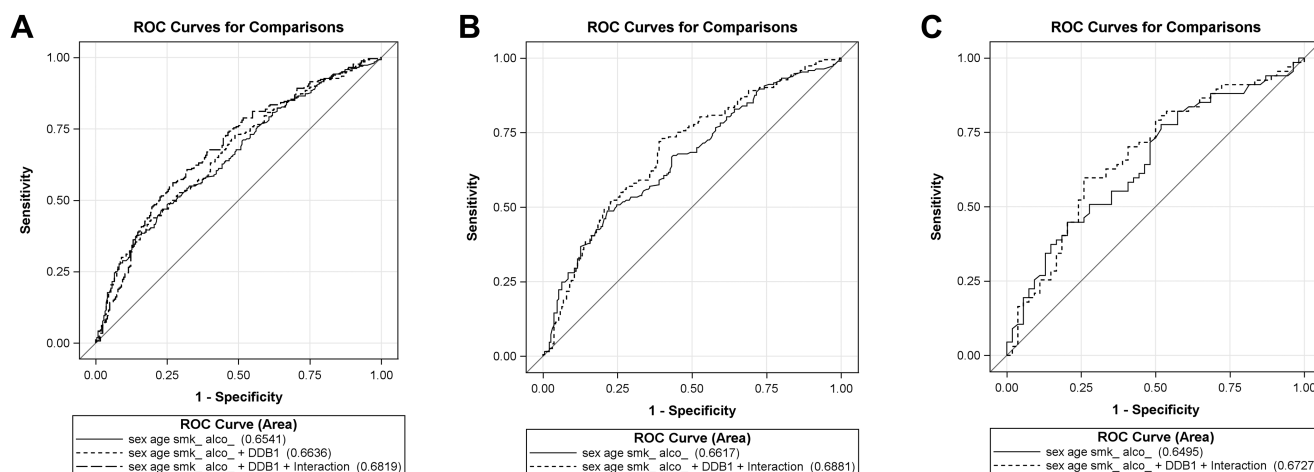


Figure 2. Overall and stratified receiver operating characteristic (ROC) curves by sex calculated in multivariate logistic models. (A) The area under the curve (AUC) was significantly improved in the model that included the combined effect of *DDB1* expression levels and the interaction with sex, compared with the model that did not ($P = 0.033$). (B) The AUC was insignificantly improved in male subjects that included the combined effect of *DDB1* expression levels ($P = 0.083$). (C) The AUC was insignificantly improved in female subjects ($P = 0.255$).

levels of all eight NER genes by the HPV status for oropharyngeal cancer patients nor for all the patients except for *ERCC4*, suggesting that expression levels of eight NER genes were not correlated with the HPV status (Supplementary Table 6).

Previous studies have suggested that women have a higher risk of developing smoking-related cancers than men do, even at similar levels of exposure (23,31,32). In the present study, we showed that a significantly lower mRNA expression levels of *DDB1* and *ERCC3* in female controls than in male controls, but not in the cases (Table 3), which is also consistent with our previous results on DRC (23,33). Furthermore, we found a modified effect of sex on *DDB1*, suggesting that the association between the reduced expression levels of *DDB1* and risk of SCCHN differed between sexes. Subsequently, we estimated the ORs of *DDB1* for both sexes and found that the adjusted ORs for *DDB1* in males were greater than in females, indicating that males might have a higher risk of developing SCCHN with a lower *DDB1* mRNA expression (Figure 1B).

Recently more data showed phenotypic and genotypic markers have been tentatively added to the AUC models (34–36). Previously, we reported the improvement on prediction of lung cancer risk by adding two markers of DRC in the AUC model, and the results showed the expanded models were statistically significantly better than the baseline models (37). Later, we assessed the performance of DRC in SCCHN risk prediction model and found that the addition of DRC improved the AUC, which was more evident in never smokers than in ever smokers (23). In the present study, we found further improvement in AUC by adding the combined effect of expression of *DDB1* and its interaction with sex, compared with baseline model. Taken together, these results suggest that reduced *DDB1* mRNA expression may play a more important role in SCCHN risk, especially in male subjects. Future mechanistic studies are needed for the role of expression of these genes in the etiology of SCCHN.

The protein encoded by *DDB1* recognizes some UV-damaged DNA lesions and initiates the NER process, but some studies suggested that *DDB1* might be involved in a general cellular response to DNA damage rather than a specific repair pathway (38–40). In the present study, we found that the addition of *DDB1* improved the sensitivity of the expanded model in male subjects. The possible reasons behind this are: on one hand, there might be genetic differences relating the gene expression of

DDB1 between men and women; on the other hand, the control group has more ever smokers (49.2%) and ever drinkers (61.8%) than the general population, in which male subjects tend to have more smokers and alcohol users than female subjects. The *ERCC3* gene encodes a DNA-dependent ATPase-helicase subunit of the transcription factor IIIH, which is involved in transcription and NER (41,42). Our data suggested an increased risk of SCCHN associated with reduced expression levels of *ERCC3*, and the current results were consistent with those of previously published SCCHN studies as well as a lung cancer study in which some tag SNPs of *ERCC3* were shown to serve as biomarkers of susceptibility to lung cancer (24,32,41).

There are both strengths and limitations in the current study. First, because the quantitative real-time PCR assay provides a more precise quantitative measurement of mRNA levels than conventional multiplex reverse transcriptase PCR, the current study provided more accurate measurements than the previous pilot SCCHN study. In addition, this method requires a much small amount of total RNA, and the peripheral blood lymphocytes are easier to obtain from population-based studies. Therefore, the quantitative real-time PCR assay using peripheral blood lymphocytes is an optimal assay for future epidemiologic studies. Second, like all other hospital-based studies, the control group may not be representative of the general population. In the current study, the control group had 49.2% ever smokers and 61.8% ever drinkers, much greater than those of the general population. To overcome possible selection bias from a hospital-based case-control studies, future studies may need a much larger sample size and recruit the controls from the community-based population. Third, our results showed sex, but not smoking and drinking, had a statistically significant interaction with SCCHN risk. Although smoking and drinking are the two major biological plausible factors for SCCHN, different hormone levels by sex may have played a role in SCCHN, which should be further investigated in future studies. Lastly, although we found that reduced expression levels of *DDB1* and *ERCC3* were associated with SCCHN risk, the exact mechanisms by which *DDB1* and *ERCC3* influence SCCHN risk are still unknown and to be investigated.

Supplementary material

Supplementary data are available at *Carcinogenesis* online.

Funding

This study was supported in part by National Institutes of Health (R01 ES 11740 and R01 CA 131274 to Q.W.); and a start-up funds to (Q.W.) from Duke Cancer Institute. QW was also supported from the Duke University Medical Center and the Duke Cancer Institute as part of the P30 Cancer Center Support of National Institutes of Health (CA014236). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Acknowledgments

We thank Li-E Wang, Margaret Lung and Jessica Fiske for their assistance in recruiting the subjects and gathering the questionnaire information.

Conflict of Interest Statement: None declared.

Reference

- Simard, E.P. et al. (2014) International trends in head and neck cancer incidence rates: differences by country, sex and anatomic site. *Oral Oncol.*, 50, 387–403.
- Wissinger, E. et al. (2014) The economic burden of head and neck cancer: a systematic literature review. *Pharmacoeconomics*, 32, 865–882.
- Jemal, A. et al. (2009) Cancer statistics, 2009. *CA. Cancer J. Clin.*, 59, 225–249.
- Siegel, R.L. et al. (2016) Cancer statistics, 2016. *CA. Cancer J. Clin.*, 66, 7–30.
- Torre, L.A. et al. (2015) Global cancer statistics, 2012. *CA. Cancer J. Clin.*, 65, 87–108.
- Hashibe, M. et al. (2007) Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *J. Natl. Cancer Inst.*, 99, 777–789.
- Hashibe, M. et al. (2009) Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiol. Biomarkers Prev.*, 18, 541–550.
- Jackson, S.P. et al. (2009) The DNA-damage response in human biology and disease. *Nature*, 461, 1071–1078.
- Wang, M. et al. (2013) Molecular epidemiology of DNA repair gene polymorphisms and head and neck cancer. *J. Biomed. Res.*, 27, 179–192.
- Li, D. et al. (2007) *In vitro* benzo[a]pyrene diol epoxide-induced DNA adducts and risk of squamous cell carcinoma of head and neck. *Cancer Res.*, 67, 5628–5634.
- Phillips, D.H. (1983) Fifty years of benzo(a)pyrene. *Nature*, 303, 468–472.
- Xiong, P. et al. (2007) *In vitro* benzo[a]pyrene diol epoxide-induced DNA damage and chromosomal aberrations in primary lymphocytes, smoking, and risk of squamous cell carcinoma of the head and neck. *Int. J. Cancer*, 121, 2735–2740.
- Wei, Q. et al. (1998) Reduced expression of hMLH1 and hGTBP/hMSH6: a risk factor for head and neck cancer. *Cancer Epidemiol. Biomarkers Prev.*, 7, 309–314.
- Wood, R.D. et al. (2005) Human DNA repair genes, 2005. *Mutat. Res.*, 577, 275–283.
- Wood, R.D. et al. (2001) Human DNA repair genes. *Science*, 291, 1284–1289.
- Li, C. et al. (2009) DNA repair phenotype and cancer susceptibility—a mini review. *Int. J. Cancer*, 124, 999–1007.
- Spivak, G. (2015) Nucleotide excision repair in humans. *DNA Repair (Amst.)*, 36, 13–18.
- de Boer, J. et al. (2000) Nucleotide excision repair and human syndromes. *Carcinogenesis*, 21, 453–460.
- Norgauer, J. et al. (2003) Xeroderma pigmentosum. *Eur. J. Dermatol.*, 13, 4–9.
- Cleaver, J.E. (2005) Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nat. Rev. Cancer*, 5, 564–573.
- Liu, Z. et al. (2016) Reduced DNA double-strand break repair capacity and risk of squamous cell carcinoma of the head and neck—A case-control study. *DNA Repair (Amst.)*, 40, 18–26.
- Cheng, L. et al. (1998) Reduced DNA repair capacity in head and neck cancer patients. *Cancer Epidemiol. Biomarkers Prev.*, 7, 465–468.
- Wang, L.E. et al. (2010) Reduced DNA repair capacity for removing tobacco carcinogen-induced DNA adducts contributes to risk of head and neck cancer but not tumor characteristics. *Clin. Cancer Res.*, 16, 764–774.
- Cheng, L. et al. (2002) Expression of nucleotide excision repair genes and the risk for squamous cell carcinoma of the head and neck. *Cancer*, 94, 393–397.
- Liu, Z. et al. (2003) Overexpression of hMTH1 in peripheral lymphocytes and risk of prostate cancer: a case-control analysis. *Mol. Carcinog.*, 36, 123–129.
- Liu, Z. et al. (2005) Methylation and messenger RNA expression of p15INK4b but not p16INK4a are independent risk factors for ovarian cancer. *Clin. Cancer Res.*, 11, 4968–4976.
- Vogel, U. et al. (2000) DNA repair capacity: inconsistency between effect of over-expression of five NER genes and the correlation to mRNA levels in primary lymphocytes. *Mutat. Res.*, 461, 197–210.
- Cheng, L. et al. (1999) Expression in normal human tissues of five nucleotide excision repair genes measured simultaneously by multiplex reverse transcription-polymerase chain reaction. *Cancer Epidemiol. Biomarkers Prev.*, 8, 801–807.
- Kumar, A. et al. (2012) Reduced expression of DNA repair genes (XRCC1, XPD, and OGG1) in squamous cell carcinoma of head and neck in North India. *Tumour Biol.*, 33, 111–119.
- Ma, H. et al. (2012) Polymorphisms of XPG/ERCC5 and risk of squamous cell carcinoma of the head and neck. *Pharmacogenet. Genomics*, 22, 50–57.
- Mollerup, S. et al. (2006) Sex differences in risk of lung cancer: Expression of genes in the PAH bioactivation pathway in relation to smoking and bulky DNA adducts. *Int. J. Cancer*, 119, 741–744.
- Muscat, J.E. et al. (1996) Gender differences in smoking and risk for oral cancer. *Cancer Res.*, 56, 5192–5197.
- Wei, Q. et al. (2000) Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: a molecular epidemiologic study. *J. Natl. Cancer Inst.*, 92, 1764–1772.
- Paynter, N.P. et al. (2009) Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3. *Ann. Intern. Med.*, 150, 65–72.
- Vachon, C.M. et al. (2007) Mammographic density, breast cancer risk and risk prediction. *Breast Cancer Res.*, 9, 217.
- van der Net, J.B. et al. (2009) Usefulness of genetic polymorphisms and conventional risk factors to predict coronary heart disease in patients with familial hypercholesterolemia. *Am. J. Cardiol.*, 103, 375–380.
- Spitz, M.R. et al. (2008) An expanded risk prediction model for lung cancer. *Cancer Prev. Res. (Phila.)*, 1, 250–254.
- Cang, Y. et al. (2007) DDB1 is essential for genomic stability in developing epidermis. *Proc. Natl. Acad. Sci. U. S. A.*, 104, 2733–2737.
- Scrima, A. et al. (2008) Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell*, 135, 1213–1223.
- Tang, J.Y. et al. (2000) Xeroderma pigmentosum p48 gene enhances global genomic repair and suppresses UV-induced mutagenesis. *Mol. Cell*, 5, 737–744.
- Hu, Z. et al. (2006) Polymorphisms in the two helicases ERCC2/XPD and ERCC3/XPB of the transcription factor IIH complex and risk of lung cancer: a case-control analysis in a Chinese population. *Cancer Epidemiol. Biomarkers Prev.*, 15, 1336–1340.
- Oh, K.S. et al. (2006) Phenotypic heterogeneity in the XPB DNA helicase gene (ERCC3): xeroderma pigmentosum without and with Cockayne syndrome. *Hum. Mutat.*, 27, 1092–1103.