Published in final edited form as:

Mol Carcinog. 2018 June ; 57(6): 784–793. doi:10.1002/mc.22801.

Associations between Expression Levels of Nucleotide Excision Repair Proteins in Lymphoblastoid Cells and Risk of Squamous Cell Carcinoma of the Head and Neck

Peng Han^{1,2,3,*}, Hongliang Liu^{1,2,*}, Qiong Shi^{1,2,4,*}, Zhensheng Liu^{1,2}, Jesse D. Troy^{1,5}, Walter T. Lee^{1,6}, Jose P. Zavallos^{7,8}, Guojun Li^{9,10}, Erich M. Sturgis^{9,10}, and Qingyi Wei^{1,2,11,**}

¹Duke Cancer Institute, Duke University Medical Center, Durham, NC, 27710, USA

²Department of Medicine, Duke University School of Medicine, Durham, NC, 27710, USA

³Department of Otorhinolaryngology-Head and Neck Surgery, First Affiliated Hospital, Xi'an Jiaotong University College of Medicine, Xi'an, Shaanxi, 710004, PR China

⁴Department of Dermatology, Xijing Hospital, Fourth Military Medical University, Xi'an, Shaanxi, 710032, PR China

⁵Division of Blood and Marrow Transplantation, Duke University Medical Center, Durham, NC, 27710, USA

⁶Division of Head and Neck Surgery & Communication Sciences, Department of Surgery, Duke University Medical Center, Durham, NC, 27710, USA

⁷Department of Otolaryngology-Head and Neck Surgery, Washington University School of Medicine, St. Louis, MO, 63110, USA

⁸Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

⁹Departments of Head and Neck Surgery, The University of Texas M.D. Anderson Cancer Center, Houston, TX, 77030, USA

¹⁰Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, 77030, USA

¹¹Department of Population Health Sciences, Duke University School of Medicine, Durham, NC, 27710, USA

Abstract

Squamous cell carcinoma of head and neck (SCCHN) is one of the most common malignancies worldwide, and nucleotide excision repair (NER) is involved in SCCHN susceptibility. In this analysis of 349 newly diagnosed SCCHN patients and 295 cancer-free controls, we investigated

**Corresponding author: Qingyi Wei, M.D., Ph.D., Duke Cancer Institute, Duke University, Medical Center and Department of Medicine, Duke School of Medicine, 905 S LaSalle Street, Durham, NC 27710, USA, Tel.: (919) 660-0562, qingyi.wei@duke.edu.

*Contributed equally to this work.

Conflict of Interest: The authors declare no potential conflicts of interest.

whether expression levels of eight core NER proteins were associated with risk of SCCHN. We quantified NER protein expression levels in cultured peripheral lymphocytes using a reverse-phase protein microarray. Compared with the controls, SCCHN patients had statistically significantly lower expression levels of ERCC3 and XPA ($P = 0.001$ and 0.001 , respectively). After dividing the subjects by controls' median values of expression levels, we found a dose-dependent association between an increased risk of SCCHN and low expression levels of ERCC3 (adjusted OR, 1.75 and 95% CI: 1.26–2.42; $P_{\text{trend}} = 0.008$) and XPA (adjusted OR, 1.88; 95% CI, 1.35–2.60; $P_{\text{trend}} = 0.001$). We also identified a significant multiplicative interaction between smoking status and ERCC3 expression levels ($P = 0.014$). Finally, after integrating demographic and clinical variables, we found the addition of ERCC3 and XPA expression levels to the model significantly improved the sensitivity of the expanded model on SCCHN risk. In conclusion, reduced protein expression levels of ERCC3 and XPA were associated with an increased risk of SCCHN. However, these results need to be confirmed in additional large studies.

Keywords

biomarker; NER pathway; head and neck cancer; DNA repair; protein

Introduction

Squamous cell carcinoma of head and neck (SCCHN), which comprises malignancies of the oral cavity, pharynx and larynx, is the sixth most common cancer and seventh leading cause of cancer-related deaths worldwide [1,2]. It has been estimated that new SCCHN cases have been increasing from 45,660 in 2007 to 63,030 in 2017 in the United States [3–5]. Several key factors may contribute to SCCHN, such as cigarette smoking and alcohol use for non-HPV SCCHN as well as prior HPV infection for HPV-positive oropharyngeal cancer, but only a small fraction of the population exposed develop SCCHN in their lifetime [6,7]. Although the underlying mechanisms are still not fully studied, the heterogeneity in DNA repair mechanism may be responsible for the susceptibility to SCCHN.

Benzo(a)pyrene diol epoxide (BPDE) is a classic tobacco carcinogen that can cause irreversible DNA damage by forming DNA adducts [8–10]. DNA repair is a response to protect against DNA damage through several pathways, of which the nucleotide excision repair (NER) pathway is a major one [11–13]. Eight core proteins (DDB1, ERCC1–5, XPA and XPC) are involved in the NER process that includes damage recognition (DDB1, XPA and XPC), dual incision and oligonucleotide fragment removal (ERCC2/XPD and ERCC3/XPB), gap-filling synthesis and restore DNA duplex (ERCC1, ERCC4/XPF and ERCC5/XPG), and ligation [14–16]. Variation in these proteins' functions may lead to an aberrant NER process and subsequently increase the susceptibility to cancer [17]. Humans with defective NER proteins may have NER-associated diseases, such as xeroderma pigmentosum (XP), Cockayne's syndrome and trichothiodystrophy [18,19]. In addition, double strand break (DSB) repair may also play an important role in HPV-related oropharyngeal cancer, and our team recently reported an association between the reduced DSB repair capacity and increased risk of SCCHN as well [20,21].

In early phenotype studies, we first reported that reduced DNA repair capacity as measured by a host cell reactivation (HCR) assay was associated with an increased risk of SCCHN [22], and later we validated these results in a larger study [23]. In order to determine the molecular mechanism underlying the NER pathway, we conducted two studies measuring the transcript levels of the core NER genes and found an increased risk of SCCHN was associated with reduced mRNA expression levels in lymphoblastoid cells [24,25]. However, the transcript levels may not exactly represent the translational levels of NER proteins, because the latter more directly participate in the repair process. Thereafter, we had developed a high-throughput reverse-phase protein lysate microarray (RPPA) assay to measure the NER proteins expression levels in a pilot study of 120 subjects and reported that an increased risk of SCCHN was associated with reduced expression levels of ERCC2, ERCC4, XPA and XPC [26]. Nevertheless, the previous association study of NER proteins had a relatively small size, and due to the lack of appropriate antibodies in the past, we were not able to measure the expression levels of DDB1 and ERCC3, which may also play a crucial role in NER. Consequently, a larger study was warranted to validate the previous study of associations between NER proteins and risk of SCCHN. Thus, we conducted a larger case-control study with 644 subjects to evaluate associations between expression levels of eight core NER proteins and risk of SCCHN. As there were no significant differences in the transcript levels of eight NER genes by HPV status as we showed before [25], we tested expression levels of NER proteins in all SCCHN subjects.

Materials and Methods

Study subjects

We recruited SCCHN cases and cancer-free controls from The University of Texas M.D. Anderson Cancer Center during the period between 2002 and 2012, and a subset of 644 subjects (349 cases and 295 controls) were selected from a part of an ongoing large molecular epidemiology study, because they had additional blood sample for culturing the needed T-lymphocytes that were sufficient for transfection and protein extraction. Additionally, we evaluated the selection bias in the present study by comparing the case-control samples included in the current analysis with those excluded samples. Furthermore, 96 SCCHN cases had both the available HPV data and protein expression levels, of which 58 cases with oropharyngeal cancer. HPV infection status was determined by PCR assay. The cases were selected based on the following criteria: 18 years and older, newly diagnosed, histologically confirmed SCCHN but with no other previous cancers. The controls were recruited among visitors accompanying patients to M.D. Anderson; they were biologically unrelated to the cases, frequency-matched with cases by age (± 5 y) and sex, and have no history of prior cancers. To avoid genetic heterogeneity due to ethnic differences, the subjects included in currently study were all non-Hispanic whites. A written informed consent was obtained from eligible cases and controls. As the method described in most of epidemiology studies [27,28], subjects who smoked more than 100 cigarettes during their lifetime were defined as ever smokers, of which those who had quit smoking at least one year were defined as former smokers and remaining were considered current smokers; others were considered never smokers. Subjects who drank alcoholic beverages at least weekly for one year were considered as ever drinkers, of which those who had quit drinking more than

one year were considered as former and the remaining were defined current drinkers; others were defined never drinkers. The study protocol was reviewed and approved by M.D. Anderson Institutional Review Board.

Sample procurement, cell culture and protein preparation

The current methods have been described previously [26]. Briefly, each subject donated a 30-ml blood sample drawn into a heparinized tube before receiving treatment for cases and when consent was obtained for controls. We isolated T-lymphocytes from whole peripheral blood by Ficoll gradient centrifugation, and cryopreserved in a -80°C freezer within 24 h after the blood was drawn. To ensure a robust cellular viability ($>80\%$), which was confirmed by the exclusion assay using 0.4% trypan blue (Sigma Chemical, St. Louis, MO), the cryopreserved cells were quickly thawed and mixed with thawing medium (50% fetal bovine serum, 40% RPMI 1640, and 10% dextrose; purchased from Sigma Chemical). After being washed with the thawing medium, the cells were incubated in RPMI 1640 (Life Technologies, Grand Island, NY) supplemented with 20% fetal bovine serum (Life Technologies) and later stimulated with phytohemagglutinin at 37°C for 72 h.

These cells were harvested and transfected using the DEAE-dextran (Pharmacia Biotech, Piscataway, NJ) method with BPDE damaged plasmids. After being treated with the damaged plasmids for 40 h, which served as the substrate for the NER enzymes, the cells were ready for protein extraction. The cell suspension from each case sample was mixed with $4 \times$ SDS sample buffer (Roche Applied Science, Indianapolis, IN), and the cell lysate was then boiled for 5 min and stored at -80°C .

Cellular proteins were extracted from the cells and were serially diluted 1:1 with PBS (pH 7.5) to achieve final total protein concentration to approximately $1.5 \mu\text{g}/\mu\text{l}$. The minimal total protein concentration for detection was $0.75 \mu\text{g}/\mu\text{l}$. The extracted proteins were denatured by $4 \times 1\%$ SDS with Beta-mercaptoethanol and adjusted to a final concentration to be approximately $1 \mu\text{g}/\mu\text{l}$. Thereafter, cellular proteins were diluted in five 2-fold serial dilutions in dilution lysis buffer and a minimal $40 \mu\text{l}$ of protein lysate is required for the RPPA analysis.

Reverse-phase protein lysate microarrays

Serial diluted lysates applied to nitrocellulose-coated slides (Schleicher & Schuell BioScience, Inc., Keene, NH) by Aushon 2470 Arrayer (Aushon BioSystems, Burlington, MA). Each sample containing the antigens (the NER proteins) to be detected was spotted in duplicate with additional positive and negative controls prepared from mixed cell lysates or dilution buffer, respectively. Each slide was probed with a validated primary antibody plus a biotin-conjugated secondary antibody. Mouse anti-human monoclonal or anti-goat or anti-rabbit polyclonal antibodies were used against DDB1 (Invitrogen Carlsbad, CA); ERCC1, ERCC2, ERCC3, XPA and XPC (Santa Cruz Biotechnology, Santa Cruz, CA); and ERCC4 (Abcam, Cambridge, MA); ERCC5 (Protein tech, Tucson, AZ). The arrays were incubated with individual antibodies for 1 h at room temperature. The secondary antibodies were labeled with biotin, and added to the slides, which were then incubated at room temperature for 30 min.

Signals were amplified using a Dako Cytomation-Catalyzed system according to the manufacturer's protocol as previously described [29]. We then incubated the slides with a secondary conjugated streptavidin for 30 min and visualized the signals by DAB colorimetric reaction. The signals on the microarrays were scanned and analyzed using the Array-Pro Analyzer software (MediaCybernetics) to determine spot intensity, which later were processed by a logistic model from the R package ("Supercurve Fitting" developed by the Department of Bioinformatics and Computational Biology in MD Anderson Cancer Center, "<http://bioinformatics.mdanderson.org/OOMPA>"). A fitted curve ("supercurve") was plotted with the signal intensities on the Y-axis and the relative log₂ concentration of each protein on the X-axis using the non-parametric, monotone increasing B-spline model [30]. Protein concentrations were determined from the supercurve for each lysate by the curve-fitting. The relative concentrations of each protein were normalized for protein loading and transformed to linear values [29,31]. The RPPA_CF is the correction factor in RPPA. Samples were considered as an outlier, if the correction factor was above 2.5 or below 0.25.

Statistical Analysis

The distribution of selected demographic variables, tobacco smoking and alcoholic beverage consumption were evaluated between patients and controls by the χ^2 test. The differences in the relative expression levels of NER proteins as continuous variables were compared by Wilcoxon rank-sum test or ANOVA among groups. In addition, the correlations between the expression levels of the eight NER proteins were analyzed by the Spearman's correlation coefficient. In addition, we performed a stepwise logistic regression analysis to explore the best model to predict SCCHN risk.

The median and quartiles of protein expression values were used in the controls as the cutoff values for calculating crude odds ratio (OR) and their 95% confidence intervals (CI). The associations between NER protein expression levels and SCCHN risk were estimated by computing ORs and CIs from multivariate logistic regression models with adjustment for age (in years), sex (male vs. female), smoking status (never, former or current) and alcohol consumption (never, former or current).

Further stratification analyses were used to evaluate effect modification of related NER proteins and selected variables. A multiplicative interaction was defined as when $OR_{11} > OR_{10} \times OR_{01}$, 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 [23]. We analyzed the interactions by looking for a multiplicative scale using standard unconditional logistic regression models.

To evaluate the effects of protein expression levels on SCCHN risk prediction in the multivariate model, two risk models were constructed to examine the area under the receiver operating characteristic (ROC) curve (AUC): the baseline model including only epidemiologic variables [e.g., age, sex, smoking and drinking status], and the protein model including the protein expression (continuous variable) in addition to these epidemiologic variables. Furthermore, the AUC was calculated upon stratification by smoking status (never, former and current). All tests were two-sided, and $P < 0.05$ was considered significant. All

statistical analyses were performed using SAS software (version 9.4; SAS Institute, Inc., Cary, NC).

Results

Characteristics of the Subject Population

The summary of the distributions of demographic variables and tumor characteristics between cases and controls was presented in Table 1. There were no significant differences in the distributions of age and sex between SCCHN patients and cancer-free controls. Moreover, no significant differences were observed in the distributions of age, sex, smoking and drinking status between the included and excluded subjects (Supplementary Table 1), suggesting that no selection bias was introduced in the current study. The average age was 56.9 years for the case subjects (median, 56; range, 22–86) and 55.9 years for the control subjects (median, 57; range, 21–84). Of all the subjects, 74.2% of cases and 78.6% of controls were male, 62.5% of cases and 49.2% of controls were ever smokers, and 73.9% of cases and 62.7% of controls were ever drinkers. There were more current smokers (32.7%) and current drinkers (47.0%) in cases than in controls (12.2% and 42.0%, respectively). The primary sites of 349 SCCHN included the oral cavity (101, 28.9%), oropharynx (196, 56.2%), and hypopharynx/larynx (52, 14.9%).

Differences in relative expression levels of NER proteins between cases and controls

We evaluated the differences in NER protein expression levels between cases and controls by Wilcoxon rank-sum test (Table 2). Among all eight NER proteins, only ERCC3 and XPA expression levels were statistically significantly lower in cases than in controls ($P = 0.001$ and $P = 0.001$, respectively; Figure 1A). Because the expression levels of the eight NER proteins were measured at the same time, they were likely to be correlated with each other. As shown in Table 3, expression levels of ERCC3 were statistically significantly correlated with all other six NER proteins, except for XPA; while expression levels of XPA were statistically significantly correlated with ERCC1, ERCC2, ERCC4 and ERCC5. Subsequently, when we included the expression levels of ERCC3 and XPA in the stepwise logistic regression analysis after adjustment for age, sex, smoking and alcohol use, both ERCC3 and XPA remained in the final model ($P = 0.007$ and $P = 0.002$, respectively), suggesting that the effects of these two NER proteins were independent.

Stratification analyses of expression levels of ERCC3 and XPA

Stratification analyses of ERCC3 and XPA expression levels showed that cases in subgroups of the age > 56 , male, never and former smokers, and current drinkers exhibited significantly lower mean expression levels of ERCC3 than did controls ($P = 0.001$, $P = 0.002$, $P = 0.010$, $P = 0.002$, and $P = 0.007$, respectively, Table 4). Likewise, cases in subgroups of the age > 56 and age > 56 , male and female, current smokers, and current drinkers exhibited significantly lower mean expression of XPA than did controls ($P = 0.023$, $P = 0.018$, $P = 0.008$, $P = 0.017$, $P = 0.032$, and $P = 0.002$, respectively). In both cases and controls, the age and sex differences in the expression levels of ERCC3 and XPA were all statistically insignificant ($P = 0.065$, $P = 0.997$, $P = 0.774$, $P = 0.967$, $P = 0.861$, $P = 0.518$, $P = 0.663$, and $P = 0.245$, respectively). In controls, current smokers had lower expression levels of

ERCC3 than that of former and never smokers, and the differences of ERCC3 expression levels were statistically significant ($P=0.022$), but such a difference was not observed in cases ($P=0.154$, Table 4). The differences in the expression levels of ERCC3 and XPA were not statistically significant by sex in both cases and controls, nor by tumor sites (Supplementary Table 2). Furthermore, we conducted a case-case analysis of 96 SCCHN cases and 58 oropharyngeal cases who had both the available HPV status and protein expression data, and we found that expression levels of ERCC3 and XPA were not differed by the HPV status (Supplementary Table 3).

Associations between NER protein expression levels and risk of SCCHN

To estimate SCCHN risk, the relative NER protein expression levels were dichotomized by the median values of the controls (Table 5 and Supplementary Table 4). The crude ORs for SCCHN risk associated with lower relative expression levels of ERCC3 were 1.78 (95% CI, 1.30–2.44), while the corresponding crude ORs for XPA were 1.85 (95% CI, 1.35–2.54), compared with the high expression levels of each protein. Furthermore, we also used the quartile values of the controls to calculate the ORs, the ORs of ERCC3 and XPA were statistically significant by the first and second quartiles, respectively, compared with the fourth quartile. After adjustment for age, sex, smoking status and alcohol consumption in multivariate logistic regression analyses, the ORs of ERCC3 and XPA remained essentially unchanged. When continuous protein values were used in the logistic regression model with adjustment for all covariates, there was also a dose-response relationship between the reduced protein expression levels and the increased SCCHN risk for both ERCC3 ($P_{\text{trend}}=0.008$) and XPA ($P_{\text{trend}}=0.001$, Table 5).

Interactions between ERCC3 or XPA expression levels and selected variables

We further assessed possible interactions on a multiplicative scale between expression levels of ERCC3 or XPA and selected variables listed in Table 1. The multiplicative interaction was analyzed when we included the interaction term (i.e., relative expression levels of ERCC3 (or XPA) $\times$ each of the risk factors) in a multivariate regression model that also included the main effects of NER protein expression levels and other covariates. We found that smoking status had a significantly multiplicative interaction with the relative expression levels of ERCC3 ($P=0.014$, Table 4), but not of XPA, in association with SCCHN risk. To further unravel these multiplicative interactions, we stratified the adjusted ORs by smoking status. It was apparent that ORs for the relative expression levels of ERCC3 by quartiles in groups of former smokers were greater than those of never smokers (Figure 1B). However, the associations between the low expression levels of ERCC3 and risk of SCCHN in current smokers were not statistically significant.

We further assessed the prediction performance of models integrating demographic variables and protein expression levels on SCCHN using the ROC curves that measure the effect of ERCC3 and XPA expression levels in two dimensions. The AUC was significantly improved in the model that included the combined effect of ERCC3 and XPA expression levels, compared with the model that did not (Figure 2A, $P=0.023$). Furthermore, the AUC was significantly improved in never and former smokers that included the combined effect of

expression levels, but insignificantly improved in current smokers, compared with the model that did not (Figure 2B, 2C, and 2D, $P = 0.039$, $P = 0.015$, and $P = 0.158$, respectively).

Discussion

In the current study, we further confirmed the results in the previous pilot study that reduced NER protein expression levels as measured by the RPPA assay were associated with an increased risk of SCCHN. Specifically, our results demonstrated that the reduced relative protein expression levels of XPA and ERCC3 were associated with an increased risk of SCCHN. Furthermore, we found that smoking status had a significant multiplicative interaction with ERCC3 expression levels on SCCHN risk. The AUC model suggested that the combined effect of ERCC3 and XPA further improved the efficacy of risk prediction.

In previous studies, we developed the HCR assay to measure the association between the SCCHN susceptibility and impaired DRC. The results suggested that suboptimal DRC was associated with an increased SCCHN risk, and we further confirmed that DRC was an independent biomarker for SCCHN susceptibility rather than a tumor marker [22,23]. Later, another study indicated a correlation between suboptimal DRC phenotype and NER mRNA expression levels in the lymphocytes [32]. Subsequently, we measured the NER mRNA expression levels of SCCHN cases and healthy controls in their peripheral lymphocytes by using PCR assay. The results suggested that the relative mRNA expression levels of *DDB1* and *ERCC3* were significantly lower in cases than in controls, and the risk of SCCHN associated with low expression levels of the two genes was about two-fold [24,25].

As the mRNA is eventually translated into a protein, one might assume that there might be a correlation between the level of mRNA and that of protein. However, the association between NER mRNA expression and SCCHN risk may be spurious, as the transcript levels and translational levels may not be significantly correlated [33]. It was reported that transcription and translation are far from having a linear and simple relationship. The possible reason behind this is that different mechanisms involving *cis*-acting and *trans*-acting mechanisms generate a repertoire of systems that enhance or repress the synthesis of proteins from a certain copy number of mRNA molecules [34]. To address this underlying discrepancy, we used the RPPA assay to measure the NER protein expression levels in a preliminary study and found that there was an association between an increased risk of SCCHN and reduced expression levels of ERCC2, ERCC4, XPA and XPC [26], when appropriate antibodies for DDB1 and ERCC3 were not available. In the current larger study with more available antibodies for essential proteins, we confirmed that reduced XPA expression was associated with an increased SCCHN risk but found a new association with ERCC3 expression levels, which is also in accordance with our previous transcriptional studies [24,25]. Collectively, these results suggested that altered protein expression levels, which have a more direct effect on the NER capacity than that of transcript levels, may contribute to the risk of SCCHN.

As the etiologic mechanism for HPV-positive oropharyngeal cancer is different from other non-HPV SCCHN [35,36], we compared the expression levels of ERCC3 and XPA between HPV-positive cases and HPV-negative cases, the results confirmed that the expression levels

of ERCC3 and XPA were not correlated with the HPV status. Moreover, we also found that there are no differences in these two proteins between different tumor sites, suggesting that expression levels of ERCC3 and XPA may be a general biomarker for all SCCHN tumors.

Our previous studies found that significantly lower mRNA expression levels of NER genes in female controls than in male controls, and the reduced *DDB1* mRNA expression levels may play a more important role in the risk of SCCHN in males than in females [25]. However, we did not observe any significant differences in the protein expression levels of ERCC3 and XPA between male and female subjects, nor did find any interaction with sex. Interestingly, we have observed a modification effect of smoking status on ERCC3, which is in accordance with our DRC study [23], indicating that an association between the reduced expression levels of ERCC3 and increased risk of SCCHN may differ by smoking status. Subsequently, we stratified the ORs by smoking status and found that the adjusted ORs for ERCC3 in former smokers were greater than that in never smokers, indicating that former smokers have a higher risk of developing SCCHN with a lower ERCC3 expression. Even though the results in the protein study is different from that of the mRNA study, protein levels are likely more directly involved in the NER repair process.

The ERCC3 was a subunit of the transcription factor IIH functioning as a DNA-dependent ATPase-helicase, which is actively involved in NER [37,38]. The finding of reduced expression levels of ERCC3 associated with an increased risk of SCCHN in the present study is consistent with that of our previous study of NER transcript levels. These suggest that levels of both *ERCC3* transcript and protein are predictors of SCCHN risk.

The XPA is a zinc-finger DNA-binding protein that participates in NER to detect DNA damage [39]. We further confirmed the results of our previous translational analysis that reduced XPA expression levels were associated with an increased risk of SCCHN. These suggest that levels of XPA protein expression but not its transcripts may be a biomarker for SCCHN risk. Interestingly, of 461 SNPs reported to date for XPA, a common 23A>G SNP (rs1800975) in the 5'-UTR has been widely studied [40–42]. In a meta-analysis study, a significant association of XPA A23G variant with head and neck cancer risk was observed, providing further genotype evidence for the role of XPA in SCCHN risk [43].

Recently more studies showed the positive performance of phenotypic and genotypic markers in risk prediction models by calculating the AUC under the ROC curve [44,45]. Previously, we assessed the performance of two biomarkers of DRC in the AUC model and found that the addition of DRC significantly improved the prediction of lung cancer risk [46]. Later, we reported the improvement on prediction of SCCHN risk by adding DRC phenotype, and the results showed the improvement was more evident in never smokers than in ever smokers [23]. Furthermore, we found an improvement in AUC by adding the combined effect of NER mRNA expression and its interaction effects, compared with the baseline model [25]. In the current study, we found that the AUC model was significantly improved by including the combined effect of ERCC3 and XPA expression, compared with the model that did not, especially in former smokers. Collectively, these results suggested that reduced expression levels of NER proteins may play an important role in SCCHN risk, especially in former smokers.

To date, we have developed several methods to measure biomarkers at mRNA, protein or DRC levels that may predict the risk of tobacco-induced SCCHN [23–26]. The RPPA assay is a rapid, cost-effective and most importantly an efficient method to measure the DRC. First, compared with the PCR assay, the RPPA assay is more sensitive and reliable, as the expression levels measured by PCR may fluctuate in different cell stage. Second, compared with Western blot assay, the RPPA assay only requires a much small amount of protein lysate samples, and many samples can also be tested simultaneously under the same condition. Third, compared with the HCR assay, the PRRA assay demands a relatively fewer viable lymphocytes for protein extraction, and the peripheral blood lymphocytes are easier to obtain from population-based studies. Therefore, the RPPA assay using peripheral blood lymphocytes is an optimal assay for future epidemiologic studies.

Although the present study is an extension of our previous work for NER proteins with more antibodies and larger sample size, there are still several limitations needed to be resolved. Like our previous hospital-based studies, the control group may not be representative of the general population, and future studies may need a much larger sample size and recruit the controls from the community-based population. Future mechanistic studies are also needed for the role of NER proteins in the etiology of SCCHN.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Margaret Lung and Jessica Fiske for their assistance in recruiting the subjects and gathering the questionnaire information.

Funding

This study was supported in part by National Institutes of Health (R01 ES011740 and R01 CA131274 to Q.W.); and a start-up funds (to Q.W.) from Duke Cancer Institute. QW was also supported by 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.

Abbreviation

SCCHN	squamous cell carcinoma of head and neck
NER	nucleotide excision repair
RPPA	reverse-phase protein lysate microarray
BPDE	benzo (a) pyrene diol epoxide
OR	odds ratio
CI	confidence interval
XP	xeroderma pigmentosum

DRC	DNA repair capacity
HCR	host-cell reactivation
ROC	receiver operating characteristic

References

1. Simard EP, Torre LA, Jemal A. International trends in head and neck cancer incidence rates: differences by country, sex and anatomic site. *Oral oncology*. 2014; 50(5):387–403. [PubMed: 24530208]
2. Wissinger E, Griebisch I, Lungershausen J, Foster T, Pashos CL. The economic burden of head and neck cancer: a systematic literature review. *Pharmacoeconomics*. 2014; 32(9):865–882. [PubMed: 24842794]
3. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. *CA Cancer J Clin*. 2007; 57(1):43–66. [PubMed: 17237035]
4. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA: a cancer journal for clinicians*. 2017; 67(1):7–30. [PubMed: 28055103]
5. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA: a cancer journal for clinicians*. 2015; 65(2):87–108. [PubMed: 25651787]
6. Hashibe M, Brennan P, Benhamou S, et al. 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*. 2007; 99(10):777–789. [PubMed: 17505073]
7. Hashibe M, Brennan P, Chuang SC, et al. 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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2009; 18(2):541–550.
8. Li D, Wang LE, Chang P, El-Naggar AK, Sturgis EM, Wei Q. In vitro benzo[a]pyrene diol epoxide-induced DNA adducts and risk of squamous cell carcinoma of head and neck. *Cancer research*. 2007; 67(12):5628–5634. [PubMed: 17575128]
9. Phillips DH. Fifty years of benzo(a)pyrene. *Nature*. 1983; 303(5917):468–472. [PubMed: 6304528]
10. Xiong P, Hu Z, Li C, et al. 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*. 2007; 121(12):2735–2740. [PubMed: 17724733]
11. Wood RD, Mitchell M, Lindahl T. Human DNA repair genes, 2005. *Mutation research*. 2005; 577(1–2):275–283. [PubMed: 15922366]
12. Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA repair genes. *Science*. 2001; 291(5507):1284–1289. [PubMed: 11181991]
13. Li C, Wang LE, Wei Q. DNA repair phenotype and cancer susceptibility--a mini review. *International journal of cancer*. 2009; 124(5):999–1007. [PubMed: 19065660]
14. Spivak G. Nucleotide excision repair in humans. *DNA Repair (Amst)*. 2015; 36:13–18. [PubMed: 26388429]
15. Sancar A. Mechanisms of DNA Repair by Photolyase and Excision Nuclease (Nobel Lecture). *Angewandte Chemie*. 2016; 55(30):8502–8527. [PubMed: 27337655]
16. Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annual review of biochemistry*. 2004; 73:39–85.
17. de Boer J, Hoeijmakers JH. Nucleotide excision repair and human syndromes. *Carcinogenesis*. 2000; 21(3):453–460. [PubMed: 10688865]
18. Cleaver JE. Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nat Rev Cancer*. 2005; 5(7):564–573. [PubMed: 16069818]

19. Cleaver JE, Lam ET, Revet I. Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. *Nat Rev Genet.* 2009; 10(11):756–768. [PubMed: 19809470]
20. Weaver AN, Cooper TS, Rodriguez M, et al. DNA double strand break repair defect and sensitivity to poly ADP-ribose polymerase (PARP) inhibition in human papillomavirus 16-positive head and neck squamous cell carcinoma. *Oncotarget.* 2015; 6(29):26995–27007. [PubMed: 26336991]
21. Liu Z, Liu H, Gao F, Dahlstrom KR, Sturgis EM, Wei Q. Reduced DNA double-strand break repair capacity and risk of squamous cell carcinoma of the head and neck--A case-control study. *DNA repair.* 2016; 40:18–26. [PubMed: 26963119]
22. Cheng L, Eicher SA, Guo Z, Hong WK, Spitz MR, Wei Q. Reduced DNA repair capacity in head and neck cancer patients. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* 1998; 7(6):465–468.
23. Wang LE, Hu Z, Sturgis EM, et al. Reduced DNA repair capacity for removing tobacco carcinogen-induced DNA adducts contributes to risk of head and neck cancer but not tumor characteristics. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2010; 16(2):764–774. [PubMed: 20068090]
24. Cheng L, Sturgis EM, Eicher SA, Spitz MR, Wei Q. Expression of nucleotide excision repair genes and the risk for squamous cell carcinoma of the head and neck. *Cancer.* 2002; 94(2):393–397. [PubMed: 11900225]
25. Han P, Gao F, Liu H, et al. Reduced mRNA expression of nucleotide excision repair genes in lymphocytes and risk of squamous cell carcinoma of the head and neck. *Carcinogenesis.* 2017; 38(5):504–510. [PubMed: 28379545]
26. Wei Q, Wang LE, Sturgis EM, Mao L. Expression of nucleotide excision repair proteins in lymphocytes as a marker of susceptibility to squamous cell carcinomas of the head and neck. *Cancer Epidemiol Biomarkers Prev.* 2005; 14(8):1961–1966. [PubMed: 16103444]
27. Guo H, Liu H, Wei J, et al. Functional single nucleotide polymorphisms of the RASSF3 gene and susceptibility to squamous cell carcinoma of the head and neck. *Eur J Cancer.* 2014; 50(3):582–592. [PubMed: 24295637]
28. Liu Z, Wei S, Ma H, et al. A functional variant at the miR-184 binding site in TNFAIP2 and risk of squamous cell carcinoma of the head and neck. *Carcinogenesis.* 2011; 32(11):1668–1674. [PubMed: 21934093]
29. Gonzalez-Angulo AM, Hennessy BT, Meric-Bernstam F, et al. Functional proteomics can define prognosis and predict pathologic complete response in patients with breast cancer. *Clin Proteomics.* 2011; 8(1):11. [PubMed: 21906370]
30. Hu J, He X, Baggerly KA, Coombes KR, Hennessy BT, Mills GB. Non-parametric quantification of protein lysate arrays. *Bioinformatics.* 2007; 23(15):1986–1994. [PubMed: 17599930]
31. Tibes R, Qiu Y, Lu Y, et al. Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. *Mol Cancer Ther.* 2006; 5(10):2512–2521. [PubMed: 17041095]
32. Vogel U, Dybdahl M, Frentz G, Nexø BA. DNA repair capacity: inconsistency between effect of over-expression of five NER genes and the correlation to mRNA levels in primary lymphocytes. *Mutation research.* 2000; 461(3):197–210. [PubMed: 11056291]
33. Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome biology.* 2003; 4(9):117. [PubMed: 12952525]
34. Maier T, Guell M, Serrano L. Correlation of mRNA and protein in complex biological samples. *FEBS letters.* 2009; 583(24):3966–3973. [PubMed: 19850042]
35. Mannarini L, Kratochvil V, Calabrese L, et al. Human Papilloma Virus (HPV) in head and neck region: review of literature. *Acta Otorhinolaryngol Ital.* 2009; 29(3):119–126. [PubMed: 20140157]
36. Pytynia KB, Dahlstrom KR, Sturgis EM. Epidemiology of HPV-associated oropharyngeal cancer. *Oral oncology.* 2014; 50(5):380–386. [PubMed: 24461628]
37. Hu Z, Xu L, Shao M, et al. 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 epidemiology, biomarkers & prevention : a publication of the American*

- Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2006; 15(7):1336–1340.
38. Oh KS, Khan SG, Jaspers NG, et al. Phenotypic heterogeneity in the XPB DNA helicase gene (ERCC3): xeroderma pigmentosum without and with Cockayne syndrome. *Human mutation*. 2006; 27(11):1092–1103. [PubMed: 16947863]
39. Wang M, Chu H, Zhang Z, Wei Q. Molecular epidemiology of DNA repair gene polymorphisms and head and neck cancer. *J Biomed Res*. 2013; 27(3):179–192. [PubMed: 23720673]
40. Jelonek K, Gdowicz-Klosok A, Pietrowska M, et al. Association between single-nucleotide polymorphisms of selected genes involved in the response to DNA damage and risk of colon, head and neck, and breast cancers in a Polish population. *Journal of applied genetics*. 2010; 51(3):343–352. [PubMed: 20720310]
41. Park JY, Park SH, Choi JE, et al. Polymorphisms of the DNA repair gene xeroderma pigmentosum group A and risk of primary lung cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2002; 11(10 Pt 1):993–997.
42. Wu X, Zhao H, Wei Q, et al. XPA polymorphism associated with reduced lung cancer risk and a modulating effect on nucleotide excision repair capacity. *Carcinogenesis*. 2003; 24(3):505–509. [PubMed: 12663511]
43. Liu J, Zhang Z, Cao XL, et al. XPA A23G polymorphism and susceptibility to cancer: a meta-analysis. *Molecular biology reports*. 2012; 39(6):6791–6799. [PubMed: 22314912]
44. Paynter NP, Chasman DI, Buring JE, Shiffman D, Cook NR, Ridker PM. Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3. *Annals of internal medicine*. 2009; 150(2):65–72. [PubMed: 19153409]
45. Vachon CM, van Gils CH, Sellers TA, et al. Mammographic density, breast cancer risk and risk prediction. *Breast cancer research : BCR*. 2007; 9(6):217. [PubMed: 18190724]
46. Spitz MR, Etzel CJ, Dong Q, et al. An expanded risk prediction model for lung cancer. *Cancer prevention research*. 2008; 1(4):250–254. [PubMed: 19138968]

Figure 1.

(A) Relative expression levels of eight NER proteins between SCCHN patients and healthy controls. Reverse-phase protein lysate microarray was used to measure the relative expression levels of eight NER proteins. The relative concentrations of each protein were normalized for protein loading and transformed to linear values (* P value < 0.05); (B) Modification effects of ERCC3 by smoking status. ORs for the relative expression levels of ERCC3 by quartiles in groups of former smokers were greater than those of never smokers. However, the associations between the low expression levels of ERCC3 and risk of SCCHN in current smokers were not statistically significant.

Figure 2.

Overall and stratified ROC curves by smoking status calculated in multivariate logistic models. (A) The AUC was significantly improved in the model that included the combined effect of ERCC3 and XPA expression levels, compared with the model that did not ($P=0.023$); (B) The AUC was significantly improved in never smokers that included the combined effect of ERCC3 and XPA expression levels ($P=0.039$); (C) The AUC was significantly improved in former smokers that included the combined effect of ERCC3 and

XPA expression levels ($P = 0.015$); **(D)** The AUC was insignificantly improved in current smokers ($P = 0.158$)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Distributions of demographic variables and tumor characteristics between 349 SCCHN patients and 295 controls

Variable	Case (N = 349)	Control (N = 295)	P*
Age			0.823
Median (range)	56 (22–86)	57 (21–84)	
Mean $\pm$ SD	56.9 $\pm$ 10.3	55.9 $\pm$ 10.0	
56	177 (50.7)	147 (49.8)	
> 56	172 (49.3)	148 (50.2)	
Sex			0.188
Female	90 (25.8)	63 (21.4)	
Male	259 (74.2)	232 (78.6)	
Smoking			< 0.001
Never	131 (37.5)	150 (50.8)	
Former	104 (29.8)	109 (37.0)	
Current	114 (32.7)	36 (12.2)	
Drinking			0.007
Never	91 (26.1)	110 (37.3)	
Former	94 (26.9)	61 (20.7)	
Current	164 (47.0)	124 (42.0)	
Tumor site			
Oral cavity	101 (28.9)		
Oropharynx	196 (56.2)		
Hypopharynx/Larynx	52 (14.9)		

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

Table 2

Comparison of the expression levels of eight NER proteins between cases and controls

NER Protein	Mean $\pm$ SD		Difference	<i>P</i>
	Case (N =349)	Control (N = 295)	Case - Control	
DDB1	0.013 $\pm$ 0.010	0.014 $\pm$ 0.010	-0.001	0.291
ERCC1	0.039 $\pm$ 0.006	0.039 $\pm$ 0.006	0.000	0.659
ERCC2	0.534 $\pm$ 0.115	0.528 $\pm$ 0.114	0.006	0.611
ERCC3	0.714 $\pm$ 0.073	0.731 $\pm$ 0.075	-0.017	0.001
ERCC4	0.209 $\pm$ 0.024	0.207 $\pm$ 0.024	0.002	0.358
ERCC5	0.400 $\pm$ 0.040	0.399 $\pm$ 0.046	0.001	0.573
XPA	0.180 $\pm$ 0.019	0.186 $\pm$ 0.026	-0.006	0.001
XPC	0.021 $\pm$ 0.006	0.019 $\pm$ 0.005	0.002	0.071

*P*value in Wilcoxon rank-sum tests

The relative concentrations of each protein were normalized for protein loading and transformed to linear values

Table 3

Correlation between expression levels of eight NER proteins

Proteins	Spearman correlation coefficients/ <i>P</i> values						
	ERCC1	ERCC2	ERCC3	ERCC4	ERCC5	XPA	XPC
DDB1	0.199	-0.151	-0.120	-0.046	-0.227	0.072	-0.070
	<0.001	0.001	0.002	0.247	<0.001	0.068	0.077
ERCC1		-0.134	-0.127	0.343	-0.020	0.194	-0.166
		<0.001	0.001	<0.001	0.609	<0.001	<0.001
ERCC2			-0.484	-0.113	0.173	-0.239	0.177
			<0.001	0.004	<0.001	<0.001	<0.001
ERCC3				0.173	0.403	0.013	0.194
				<0.001	<0.001	0.750	<0.001
ERCC4					0.413	0.153	-0.029
					<0.001	<0.001	0.466
ERCC5						0.101	0.027
						0.010	0.494
XPA							0.003
							0.931

Table 4

Stratification analysis of expression levels of ERCC3 and XPA between cases and controls

Variable	ERCC3 (Mean ± SD)		P	P [^]	XPA (Mean ± SD)		P	P [^]
	Case (N=349)	Control (N=295)			Case (N=349)	Control (N=295)		
Age				0.382				0.922
≤ 56	0.720 ± 0.073	0.731 ± 0.079	0.144		0.180 ± 0.020	0.187 ± 0.031	0.023	
> 56	0.708 ± 0.074	0.731 ± 0.072	0.001		0.180 ± 0.019	0.185 ± 0.019	0.018	
P	0.065	0.997			0.774	0.967		
Sex				0.343				0.398
Female	0.719 ± 0.071	0.725 ± 0.076	0.225		0.179 ± 0.021	0.192 ± 0.038	0.008	
Male	0.713 ± 0.074	0.733 ± 0.075	0.002		0.180 ± 0.019	0.184 ± 0.021	0.017	
P	0.861	0.518			0.663	0.245		
Smoking				0.014				0.265
Never	0.712 ± 0.085	0.739 ± 0.071	0.010		0.180 ± 0.021	0.186 ± 0.031	0.190	
Former	0.705 ± 0.061	0.729 ± 0.082	0.002		0.183 ± 0.019	0.186 ± 0.019	0.058	
Current	0.726 ± 0.068	0.702 ± 0.067	0.186		0.177 ± 0.018	0.185 ± 0.021	0.032	
P	0.154	0.022			0.165	0.513		
Drinking				0.265				0.369
Never	0.720 ± 0.075	0.727 ± 0.078	0.381		0.181 ± 0.020	0.188 ± 0.032	0.077	
Former	0.708 ± 0.069	0.725 ± 0.091	0.064		0.180 ± 0.018	0.180 ± 0.020	0.868	
Current	0.715 ± 0.076	0.738 ± 0.064	0.007		0.179 ± 0.020	0.186 ± 0.021	0.002	
P	0.334	0.562			0.753	0.130		

P value of non-parametric tests (Wilcoxon rank-sum tests for comparison of two groups or Kruskal-Wallis tests for comparison of three groups)

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

The relative concentrations of each protein were normalized for protein loading and transformed to linear values

Table 5

Associations between dichotomized and quartered expression levels of two NER proteins and SCCN risk

NER Protein	Protein levels	Case No. (%)	Control No. (%)	Crude OR (95% CI)	Adjusted OR [†] (95% CI)
ERCC3 *	0.736	125 (35.8)	147 (49.8)	1.00 (Ref)	1.00 (Ref)
	< 0.736	224 (64.2)	148 (50.2)	1.78 (1.30–2.44)	1.75 (1.26–2.42)
ERCC3 **	0.781	53 (15.2)	73 (24.8)	1.00 (Ref)	1.00 (Ref)
	0.736–0.780	72 (20.6)	75 (25.4)	1.32 (0.82–2.14)	1.23 (0.75–2.01)
	0.687–0.735	117 (33.5)	73 (24.7)	2.21 (1.40–3.49)	2.15 (1.34–3.44)
	< 0.687	107 (30.7)	74 (25.1)	1.99 (1.26–3.16)	1.81 (1.12–2.91)
<i>P</i> ^{&}				0.005	0.008
XPA *	0.185	123 (35.2)	148 (50.2)	1.00 (Ref)	1.00 (Ref)
	< 0.185	226 (64.8)	147 (49.8)	1.85 (1.35–2.54)	1.88 (1.35–2.60)
XPA **	0.199	55 (15.7)	73 (24.8)	1.00 (Ref)	1.00 (Ref)
	0.185–0.198	68 (19.5)	75 (25.4)	1.20 (0.74–1.94)	1.17 (0.72–1.92)
	0.171–0.184	112 (32.1)	74 (25.1)	2.01 (1.27–3.17)	2.02 (1.27–3.23)
	< 0.171	114 (32.7)	73 (24.7)	2.07 (1.31–3.27)	2.06 (1.28–3.30)
<i>P</i> ^{&}				<0.001	0.001

* Based on the median values of control subjects

** Based on the quartile values of control subjects

[†] Adjustment for age, sex, smoking and alcohol status, ORs for association between protein levels and risk of SCCN[&] *P* value in trend test