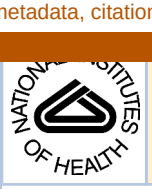


Published in final edited form as:

Cancer Detect Prev. 2008 ; 32(3): 200–208. doi:10.1016/j.cdp.2008.08.008.

Genetic variation in Transaldolase1 and risk of squamous cell carcinoma of the head and neck

Patricia V. Basta, Ph.D.[1,3], **Jeannette T. Bensen, Ph.D.**[1], **Chiu-Kit Tse, MSPH**[1], **Charles M. Perou, Ph.D.**[2], **Patrick F. Sullivan, Ph.D.**[2], and **Andrew F. Olshan, Ph.D.**[1]

1*Dept. of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7435, USA*

2*Dept. of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7435, USA*

Abstract

Background—The Pentose Phosphate Pathway (PPP) is involved in the body's protection against oxidative stress and resistance/susceptibility to apoptosis and thus has been implicated in tumor development and progression. Here we present data examining the association of genetic variation in one of the key enzymes of the PPP, Transaldolase 1 (*TALDO1*) with squamous cell carcinoma of the head and neck (SCCHN).

Methods—We performed sequencing analysis to identify common genetic variations in *TALDO1* and then investigated their association with SCCHN using samples from a population-based case/control study with both European American (EA) and African American (AA) former and current smokers.

Results—We identified three polymorphisms in *TALDO1* that were associated with SCCHN risk in our EA study population. Specifically the 5' upstream variant -490C>G or T (rs10794338), which we identified as tri-allelic, showed a reduced risk compared with any presence of the common allele, odds ratio (OR) [95% confidence interval (95% CI)]: 0.57 (0.38-0.86). Additionally two intronic high frequency polymorphisms demonstrated a positive association with disease, with the presence of the variant IVS1+1874T>A (rs3901233), 1.76 (1.19-2.61) and IVS4+2187A>C (rs4963163), 1.71 (1.16-2.53).

Conclusion—These results provide preliminary evidence that genetic polymorphisms in *TALDO1* are associated with SCCHN.

Author's key words

genetic polymorphism; head and neck cancer; Pentose Phosphate Pathway (PPP); single nucleotide polymorphism (SNP); squamous cell carcinoma

Introduction

Head and neck cancer typically includes cancers of the oral cavity (lip, tongue, salivary glands, gum, and mouth floor), pharynx (nasopharynx, hypopharynx, and pharynx), and the larynx. In

3Corresponding Author, Patricia V. Basta, Ph.D., Patricia_basta@unc.edu, Phone: (919)-843-3860 Fax: (919)-966-2089, Department of Epidemiology, Rm 2103A McGavran-Greenberg / CB# 7435, University of North Carolina, Chapel Hill, NC, 27599-7435.

Conflict of Interest: None declared

the United States (US) there are expected to be 28,900 new cases of head and neck cancer and 7,400 new deaths per year [1]. Head and neck cancers are the sixth most frequent cancers in the world [1]. Tobacco use is the most important established risk factor for head and neck cancer. In addition, in the US, tobacco and alcohol use account for approximately 75% of all oral and pharyngeal cancers [2]. Other factors that have been associated with the risk of head and neck cancer include diet, certain occupational exposures, and infection with human papilloma virus [3-7].

Based on histology, head and neck tumors are relatively homogeneous, and are primarily squamous cell carcinomas. In spite of the histological homogeneity, the clinical behaviors and outcomes of squamous cell carcinomas of the head and neck (SCCHN) are very heterogeneous. This heterogeneity poses difficulties for prevention and treatment of these tumors, in fact greater than 70% of the variation in survival among SCCHN can not be accounted for under current diagnostic staging schemes [8]. In an attempt to classify these tumors into reproducible outcome subtypes Chung et al., used genome-wide expression profiling and found that they could classify SCCHN into four distinct and reproducible subtypes that demonstrated significant differences in recurrence-free survival [9].

In the present study we expand on this work to explore the importance of germline genetic variation in *TALDO1*, a highly expressed gene in the subgroup of SCCHN tumors that showed the best recurrence-free survival rates [9]. This subgroup of tumors had high expression levels of several antioxidant enzymes involved in xenobiotic metabolism including Glutathione S-Transferase M3, Thioredoxin Reductase 1, Glutathione Peroxidase 2, Aldo-Keto Reductase 1, as well as two genes involved in the Pentose Phosphate Pathway (PPP), Phosphogluconate Dehydrogenase, and Transaldolase 1 (*TALDO1*). This same set of genes was also found to be more highly expressed in the lung epithelium of smokers versus non-smokers [10]. Genetic variation in the *TALDO1* gene has not previously been examined in SCCHN.

Transaldolase 1 is a key enzyme in the PPP. The primary functions of the PPP are to generate both ribose-5-phosphate (R5P) for the synthesis of nucleic acids and reducing equivalents, in the form of NADPH. NADPH is used in the body's biosynthetic reactions as well as in the maintenance of glutathione (GSH) in a reduced state for protection against oxidative stress [11]. The PPP includes an oxidative and a non-oxidative arm. The two arms of the PPP are functionally inter-related, and the balance between them can affect a cell's proliferation rate and its ability to respond to apoptosis induction. The body's protection against oxidative stress and resistance or susceptibility to apoptosis is relevant to both tumor development and progression [12-19].

The *TALDO1* gene is located on the short arm of human chromosome 11 in the region 11p15.5-p15.4 [20]. A 17,479 nucleotide genomic region encompasses all eight exons. The 11p15 region of chromosome 11 has been linked to a number of developmental disorders such as multiple sclerosis [21] and to numerous malignancies including bladder, brain, breast, esophageal, ovarian and testicular cancers [22-23]. *TALDO1* is transcribed as a single 1.3 kb mRNA and the start of transcription is located 48 bases upstream of the translation start site. The core promoter of *TALDO1* has been mapped to positions -49 to -1 and the basal promoter activity has been shown to be primarily mediated by the transcription factor ZNF143 [24]. Additional positive regulatory elements were shown experimentally to map to the sequence -152 to +53, and negative regulatory regions were shown to map between -903 and -387 [24].

The goal of the present study is to describe genetic variation in the *TALDO1* gene and to determine whether the variation is associated with the occurrence of SCCHN and whether the patterns differ by race, present or past smoking status and stage of the cancer. We first performed a re-sequencing analysis of the *TALDO1* gene to identify common SNPs (minor

allele frequency > 5%), in our study population. We then utilized the SNPs identified through re-sequencing of *TALDO1* and the SNPs identified using the HapMap data to evaluate the association of *TALDO1* polymorphisms with SCCHN.

Material and Methods

Study Population, Survey Development and Sample Collection

The study population consisted of a random sample of case and control smokers (former and current) from the Carolina Head and Neck Cancer Study ("CHANCE"; A. Olshan, PI). The goal of the CHANCE study is to comprehensively evaluate the role of genetic susceptibility in the etiology of squamous cell carcinoma of the head and neck. Cases consist of adults aged 20-80 that were newly diagnosed with a first primary invasive squamous cell carcinoma of the head and neck cancer (pharynx, larynx, oral cavity) between January 1, 2002 and February 28, 2006. To be eligible, cases had to be residents of a 46 county region in North Carolina. Subjects with tumors of other histologies or at other head and neck sites, or a history of recurrent or second primary tumors were not eligible. Cases were identified using a rapid case ascertainment system in conjunction with the North Carolina Central Cancer Registry. For cases, subsequent to obtaining consent, pathology reports and corresponding slides of tumor specimens from the patient's diagnostic surgery were obtained and histological confirmation and stage categorization was performed by a study pathologist. The overall cooperation rate of the cases was 81%. Controls from the same geographic region were identified using driver's license data. Controls were randomly sampled with stratification on age, race, and sex to obtain comparability to the expected case distribution on these factors. The overall cooperation rate of controls was 61%. Written informed consent for an interview, the collection of biologic specimens, and, for cases the collection of medical records was obtained prior to conducting the interview. In-person interviews were conducted to obtain lifetime histories of tobacco and alcohol use, dietary habits, medical history, and other factors.

The present analysis was performed as a pilot study on a random subset of smokers (ex and current) consisting of 328 SCCHN case and 354 controls from the eligible cases and controls that had been accrued at the start of the present analysis. Because smoking is the major risk factor for SCCHN we chose to analyze a random sample of cases and controls that were all either ex- or current smokers, to minimize heterogeneity and to increase the power for detecting gene-disease association [25]. This research protocol was approved by the University of North Carolina's Institutional Review Board.

DNA Extraction

Two 10ml yellow-top vacutainer tubes (containing ACD) that were used for DNA extraction from mononuclear cells were transported to the lab within 1-16 hours and processed immediately upon receipt. A high-salt DNA extraction method using either the Puregene chemistry from Gentra systems (Minneapolis, MN) or the BloodPrep® DNA Chemistry from Applied Biosystems (Foster City, CA) was used for all DNA extractions

Gene Re-sequencing for SNP Identification

DNA obtained from 96 cases (48 AA / 48 EA) from the CHANCE study was used for the *TALDO1* gene re-sequencing. PCR primers were designed to amplify all coding regions, and approximately 200 basepairs (bp) of the exon-intron boundaries, and 650 bp of the 5' and 3' genomic regions. These primers were then used to PCR amplify and directly sequence genomic DNA. All sequencing was performed by Polymorphic DNA Technologies Inc. (Alameda, Ca.) using high throughput capillary Sanger dideoxy DNA sequencing. Sequencing was performed in both the 5' and 3' direction for SNP validation.

Genotyping

SNPs identified via re-sequencing were chosen for the genotyping analysis if their Minor Allele Frequencies (MAF) were greater than 5% within all 96 SCCHN samples or within the AA or EA populations only. Twelve SNPs were chosen for genotyping. SNP genotyping assays were successfully developed for 9 of the 12 SNPs; 3 SNPs failed assay development and were excluded from further analyses. SNPs identified in our re-sequencing project were supplemented with two SNPs from the HapMap project Release 16c (phase1/build 34/assembly 124) and one SNP from Release 19 (phase2/build 34/assembly 124). These 3 SNPs were selected because 1) they were located within introns of *TALDO1* not re-sequenced, and 2) they demonstrated MAFs ≥5% among either the Utah residents with ancestry from Northern and Western Europe (CEU) or the Yoruba in Ibadan, Nigeria in West Africa (YRI) populations. Genotyping was performed on an Applied Biosystems (ABI) 7900 machine using ABI custom designed TaqMan MGB probes and primers and TaqMan Universal Master Mix (no UNG), and analyzed using ABI software version 2.2. Five percent of the samples from each genotyping assay were duplicated. For the bi-allelic SNPs 99.6% of the genotyping assayed were successfully assayed. For the tri-allelic SNP the success rate was 98.4%. Concordance between duplicates was 100%. The tri-allele SNP #1 (rs10794338), identified in our re-sequencing analysis, was genotyped using two sets of TaqMan assays to detect the third allele, a method that has been shown to be highly accurate for tri-allelic polymorphisms [26,27].

Statistical Analysis

Hardy-Weinberg equilibrium (HWE) was assessed among controls and within each race group using the exact test implemented in SAS Genetics (version 8.2) (SAS Institute, Cary NC). Pairwise linkage disequilibrium (LD) using r^2 was evaluated using LDSelect, since this allows for the inclusion of tri-alleles [28]. Genotypic association analyses with disease and disease stage (stages I & II grouped as early and stages III & IV grouped as late) were conducted using unconditional logistic regression to estimate odds ratios (ORs) adjusted for age and gender. For each SNP the most frequent homozygote genotype among EA (denoted major/major) was the referent for both race analyses allowing comparison of results across ethnic groups. In the case of the tri-allelic SNP, the referent genotype was C/C with the two other groups comprised of genotypes with either one C allele (C/G and C/T) or no C allele (G/G, G/T, T/T). Four genetic models, general, additive, dominant and recessive, were evaluated. For the general model two ORs [major/major (referent) vs major/minor (OR1) and major/major (referent) vs minor/minor (OR2)] and 4 p-values were calculated with the third p-value the chi-square or “global” p (reported in Table 3), and the fourth p-value the “trend” p (reported in Table 3), calculated by coding each genotype as an ordinal value. Since all SNPs were located in a single gene, and thus not entirely independent of one another, no corrections for multiple comparisons were made. The likelihood ratio test (LRT) was used to identify any interactions between SNP and smoking (ever vs. current, and pack years), SNP and drinking, as well as to investigate SNP-SNP interactions, using an alpha level of 0.20. Power calculations were carried out using the method of Dupont and Plummer [29].

Results

SNP Identification

Re-sequencing DNA from 96 head and neck cancer cases identified 27 SNPs within the coding and immediate adjacent regions of *TALDO1*. The sequencing contig coverage of the gene resulting from this analysis is illustrated in Figure 1. This re-sequencing effort discovered 20 SNPs that had not been previously identified, however only three of these had a MAF of >5% in either EA or AA. Of the three newly identified SNPs, one (#2) was located in the extended promoter region 5’ of the gene and the other two (#’s 11 and 12) were located downstream (3’) from the last exon. In addition, we identified SNP #1, as a tri-allelic SNP, which was a new

observation. This SNP is located in a putative repressor region in the extended promoter region, 5' of the *TALDO1* gene as assessed by promoter deletion analysis [24]. Our re-sequencing analysis also allowed us to provide allelic frequency data for SNPs rs11302 (#8) and rs2075545 (#10) for both AA and EA and for SNP rs3901233 (#3) for EA and for SNP rs7933290 (#9) for AA for which frequency data were not available in the literature (http://www.hapmap.org).

In summary, a total of 12 of the 27 SNPs (3 novel and 9 previously identified) from the re-sequencing project had a MAF of >5%, either in the entire population studied or within the EA or AA component of the population only. As described in the Material and Methods, 9 of these 12 SNPs, as well as three SNPs derived from the HapMap project were used in genotyping.

Genotyping

Population characteristics for the CHANCE subgroup used in this association analysis are described in Table 1. Briefly, this analysis was conducted on 328 SCCHN patients as well as 354 controls. To attempt to minimize heterogeneity and to increase the power for detecting gene-disease association in this pilot study, all subjects analyzed are either ex or current smokers. A total of 527 EA and 115 AA were included in the analysis. The position of each SNP within the *TALDO1* gene and their MAF in 293 EA control subjects or 61 AA control subjects is shown in Table 2. Among controls three SNPs (#'s 1, 3, and 6) failed HWE in AA and one (#5) failed HWE among EA (data not shown). These violations are not likely to be due to genotyping errors for several reasons. First, 100% concordance was observed between duplicate genotyping samples and second, samples for which SNPs were assayed by both TaqMan and sequencing were 100% concordant. In addition, the violations seen in the AA controls were not seen in the EA controls even though placement of samples on the genotyping plates was randomly distributed with regard to race. Therefore, the HWE violations are likely due to small sample size, particularly in AA. SNPs out of HWE (#5 in EA and #'s 1, 3, and 6 in AA) were not evaluated for association.

Weak LD was observed across *TALDO1* either among the SNPs evaluated in our study or among those in HapMap. The LD structure was evaluated using LDSelect [28] using an $r^2 \geq 0.8$ to creates bins. This analysis placed the 12 SNPs into 10 bins. Only two pair of SNPs were placed in common bins, SNPs 5 & 12 ($r^2=0.89$) and SNPs 3 & 6 ($r^2=0.87$). All other SNPs had pairwise LD r^2 values < 0.8. When a lower r^2 threshold (0.64) was used no additional bins were formed. Due to the little LD structure between these SNPs, haplotype analyses were not performed.

Disease Association

The association of individual *TALDO1* polymorphisms with SCCHN was examined in both AA (supplementary data) and EA. Effect estimates for AA were unstable as a result of small sample size, thus results are shown only for EA. Table 3 illustrates the results for the 7 SNPs out of the original 12 identified whose MAF were >5% in EA (8/12) and whose frequencies did not violate HWE (7/8). The ORs in Table 3 (adjusted for age and gender) demonstrate evidence of association between three SNPs (#'s 1, 3 & 6) and SCCHN in EA. Specifically, for SNP #1 (the tri-allele) an inverse association was seen with SCCHN status and the homozygous presence of the minor allele(s), which was consistent with a recessive model of inheritance. An increased association with disease was seen for SNPs #3 and #6 with any presence of the minor allele (A or C respectively), which is consistent with a dominant model of genetic inheritance. Additionally, it should be noted that the ORs in Table 3 did not change substantially after adjusting for smoking duration (pack years) or dose status (ex-smoker versus current).

The LRT was used to investigate both SNP-SNP and SNP-smoking interactions associated with SCCHN in both population groups (supplementary data: Tables S-1 and S-2). No clear patterns of interaction were found, but the sample sizes for some analyses were small.

We also examined the influence of alcohol using reported history of beer, wine, or hard liquor consumption. In our data, average amount of beer consumption per week was most strongly associated with the risk of SCCHN (>= 15 beers a week OR=3.11; CI=2.01, 4.80). We evaluated beer consumption as a confounder of the relationship between TALDO1 SNPs and the risk of SCCHN and found no material differences from the results reported in Table 3. For example the OR for the homozygous minor variant genotype for SNP # 3 (rs3901233) in EA for SCCHN was OR=1.66; CI=1.03 , 2.68 and for SNP #6 was OR=1.56; CI=0.97 , 2.51 without beer as a confounder and OR=1.53; CI=0.93, 2.50 and OR=1.48; CI=0.91 , 2.41 with beer as a confounder. Further, adjustment for beer use did not alter the smoking-SNP interaction results. Finally, beer consumption did not significantly interact with any TALDO1 SNP.

Finally, we investigated OR in EA for association of each SNP and later stage at diagnosis (Table 4). For SNPs that showed a significant positive association with disease (SNP #'s3 and 6) a trend toward later stage disease at diagnosis was observed, while a trend toward earlier stage disease was observed for the SNP negatively associated with disease (SNP# 1). These associations however, did not reach statistical significance due to small sample size among some of the stage groups (I+II vs III +IV). It should be noted that we do not have data on recurrence, and data on metastasis is incorporated into the staging and thus is a more useful overall outcome measure.

Discussion

The occurrence of SCCHN is due in large part to smoking, and possibly other exposures and interaction with genetic factors [2-7]. In this study we investigated for the first time whether genetic variations in *TALDO1* are associated with cancer of the head and neck. Although, *TALDO1* is a key enzyme of the PPP, which is involved in the production of R5P for synthesis of nucleic acids as well as in the generation of NADPH for biosynthetic reactions and response to oxidative stress, it has not traditionally been examined for association with cancer and smoking status. The PPP metabolic pathway enzymes however can contribute to tumor initiation and progression in a variety of ways. For instance an earlier expression study identifying *TALDO1* and another PPP enzyme as being up-regulated in a subgroup of SCCHN patients with increased survival indicated that this subset of tumors might have an altered metabolic profile [9]. In addition, Hackett, et al, have shown that 5/12 of the enzymes in the PPP are up-regulated in the airway epithelium of smokers [10], providing further evidence that smoking, one of the key risk factors for SCCHN, induces metabolic changes in normal tissues. Both the study of Chung et al., and Hackett et al., indicate that changes in transaldolase1 levels or activity may be important for understanding the initiation and progression of SCCHN tumors. The present study investigated whether germline genetic variation in *TALDO1* is associated with the development of SCCHN in a case-control population (of 328 SCCHN cases and 354 controls) of ex- or current smokers stratified by race. This analysis identified one polymorphism that provides suggestive evidence of a protective effect for SCCHN, as well as two polymorphisms that provide suggestive evidence for an increased risk of disease in EA.

We first conducted a re-sequencing analysis to identify SNPs in *TALDO1* in our study population. Our sequencing results identified 20 new SNPs (3 with MAF >5%), in the intronic and 5' and 3' regions of the gene. In addition, this re-sequencing effort allowed us to recognize a previously reported bi-allelic SNP, SNP #1 (rs10794338), in the extended promoter region of *TALDO1* as a tri-allelic SNP allowing development of a tri-allelic assay for genotyping.

The present analysis of seven SNPs in EA identified three (#'s 1, 3 and 6) that were associated with SCCHN. The fact that one of the associations (SNP #1) identified in EA showed an inverse correlation with disease was interesting particularly in light of the fact that over expression of transaldolase1 can result in an increased susceptibility to apoptosis [15]. This then might suggest that SNP#1 may be associated with increased levels of transaldolase1. This scenario is biologically plausible since this SNP is located in a putative repressor region in the *TALDO1* promoter [24], suggesting that perhaps this polymorphism might positively affect expression levels. The results with SNP #1 also correlate with the genome-wide expression profiling study of Chung et al., which showed that increased levels of *TALDO1* helped classify a subpopulation of SCCHN cases into a disease subtype that had the best recurrence-free survival [9]. Additionally, two SNPs #3 (rs3901233) and #6 (rs4963163) in strong LD, were most strongly associated with disease when a dominant model of inheritance was considered. The potential exists that either of these intronic SNPs may also effect expression levels of *TALDO1* or are in LD with an unknown SNP that does. In this case however, since these SNPs demonstrate a positive association with disease we might expect a down-regulation of *TALDO1* expression. In addition, the present results suggest that SNP #1 may be associated with early stage disease whereas SNP #'s 3 and 6 seem to be associated with late stage disease (Table IV).

Several strengths of our study should be noted; especially the population-based design and rapid case ascertainment of this unique group of SCCHN cases and controls that provides for a systematic identified and representative case group and the confirmation of case status by a comprehensive pathologic review. In addition, this study provides more complete identification and characterization of *TALDO1* genetic variants in both African and European Americans. Furthermore, scientific evidence supports the role of *TALDO1*, an under-studied candidate gene, in SCCHN tumor biology and the extensive case-control subject characterization provided a unique opportunity to determine the effects of specific risk factors such as smoking. However, like many candidate gene studies, our coverage of *TALDO1* was not based on complete sequencing data in our study population, in part due to the prohibitive cost of sequencing extensive intronic regions, however our use of multiple approaches to genotyping did not restrict our analysis to only validated assays under a single platform. In addition, due to the low minor allele frequency for some SNPs among EA, and low sample number of AA, there was limited power to evaluate the main effects of SNPs among AA and among SNPs with low minor allele frequencies or to test for interactions. Nonetheless, our exploratory study had good power (83% power to detect associations at an OR $\geq$1.5; p-value = 0.05 for genotypes of 20% frequency) to detect moderate effects of most of the *TALDO1* variants examined.

In summary, this exploratory study is the first to examine the association of genetic variations in the *TALDO1* gene with the risk of SCCHN. In this study we found suggestive evidence of a reduced risk for SCCHN with one polymorphism in the *TALDO1* gene in EA, a tri-allelic SNP located in the 5' extended promoter region and for an increased risk with two SNPs one located in the first and the other in the fourth intronic region. Further studies are needed to confirm these initial observations. In addition, the mechanism(s) responsible for these effects remains to be elucidated.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to thank Robert C. Millikan (University of North Carolina-Chapel Hill) for his critical manuscript review and for his invaluable suggestions, Joanna Smith and Kim Ludwig for facilitating CHANCE data analysis and Jason Luo, Director, of the University of North Carolina-Chapel Hill's Mammalian Genotyping Core for performing the TaqMan genotyping of the tri-allelic SNP, and to Chris R. Miller and Applied Biosystems (Foster City, CA) for generously developing and providing the primers for the tri-allele TaqMan assays.

Source of Support:

This research was supported in part by a grant from the National Cancer Institute (R01-CA90731) and funding from the National Institute of Environmental Health Sciences (P30ES10126).

References

1. Kademani D. Symposium on solid tumors: Oral Cancer. Mayo Clin Proc 2007 Jul;82(7):878–87. [PubMed: 17605971]
2. Hashibe M, Brennan P, Benhamou S, Castellsague X, Chen C, Curado MP, 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 May 16;99(10):777–89. [PubMed: 17505073]
3. Gillison ML. Current topics in the epidemiology of oral cavity and oropharyngeal cancers. Head Neck 2007 Aug;29(8):779–92. [PubMed: 17230556]
4. Ho T, Wei Q, Sturgis EM. Epidemiology of carcinogen metabolism genes and risk of squamous cell carcinoma of the head and neck. Head Neck 2007 Jul;29(7):682–99. [PubMed: 17274053]
5. Huang WY, Olshan AF, Schwartz SM, Berndt SI, Chen C, Llaca V, et al. Selected genetic polymorphisms in MGMT, XRCC1, XPD, and XRCC3 and risk of head and neck cancer: a pooled analysis. Cancer Epidemiol Biomarkers Prev 2005 Jul;14(7):1747–53. [PubMed: 16030112]
6. Schottenfeld, D.; Fraumeni, JF. Sadie Jenkins Harmon Collection. Cancer Epidemiology and Prevention. 3. Oxford ; New York: Oxford University Press; 2006. p. xviii-1392.
7. Sturgis EM, Wei Q, Spitz MR. Descriptive epidemiology and risk factors for head and neck cancer. Semin Oncol 2004 Dec;31(6):726–33. [PubMed: 15599850]
8. Choi P, Chen C. Genetic expression profiles and biologic pathway alterations in head and neck squamous cell carcinoma. Cancer 2005 Sep 15;104(6):1113–28. [PubMed: 16092115]
9. Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell 2004 May;5 (5):489–500. [PubMed: 15144956]
10. Hackett NR, Heguy A, Harvey BG, O'Connor TP, Luettich K, Flieder DB, et al. Variability of antioxidant-related gene expression in the airway epithelium of cigarette smokers. Am J Respir Cell Mol Biol 2003 Sep;29(3 Pt 1):331–43. [PubMed: 12702543]
11. Reitzer LJ, Wice BM, Kennell D. The pentose cycle. Control and essential function in HeLa cell nucleic acid synthesis. J Biol Chem 1980 Jun 25;255(12):5616–26. [PubMed: 6445904]
12. Banki K, Hutter E, Colombo E, Gonchoroff NJ, Perl A. Glutathione levels and sensitivity to apoptosis are regulated by changes in transaldolase expression. J Biol Chem 1996 Dec 20;271(51):32994–3001. [PubMed: 8955144]
13. Banki K, Hutter E, Gonchoroff NJ, Perl A. Elevation of mitochondrial transmembrane potential and reactive oxygen intermediate levels are early events and occur independently from activation of caspases in Fas signaling. J Immunol 1999 Feb 1;162(3):1466–79. [PubMed: 9973403]
14. Boros LG, Torday JS, Lim S, Bassilian S, Cascante M, Lee WN. Transforming growth factor beta2 promotes glucose carbon incorporation into nucleic acid ribose through the nonoxidative pentose cycle in lung epithelial carcinoma cells. Cancer Res 2000 Mar 1;60(5):1183–5. [PubMed: 10728670]
15. Juhnke H, Krems B, Kotter P, Entian KD. Mutants that show increased sensitivity to hydrogen peroxide reveal an important role for the pentose phosphate pathway in protection of yeast against oxidative stress. Mol Gen Genet 1996 Sep 25;252(4):456–64. [PubMed: 8879247]

16. Kuo PC, Abe K, Schroeder RA. Superoxide enhances interleukin 1beta-mediated transcription of the hepatocyte-inducible nitric oxide synthase gene. Gastroenterology 2000 Mar;118(3):608–18. [PubMed: 10702213]

17. Lachaise F, Martin G, Drougard C, Perl A, Vuillaume M, Wegnez M, et al. Relationship between posttranslational modification of transaldolase and catalase deficiency in UV-sensitive repair-deficient xeroderma pigmentosum fibroblasts and SV40-transformed human cells. Free Radic Biol Med 2001 Jun 15;30(12):1365–73. [PubMed: 11390181]

18. Poulsen HS, Frederiksen P. Glucose-6-phosphate dehydrogenase activity in human breast cancer. Lack of association with oestrogen receptor content. Acta Pathol Microbiol Scand 1981 Jul;89(4): 263–70.

19. Ramos-Montoya A, Lee WN, Bassilian S, Lim S, Trebukhina RV, Kazhyna MV, et al. Pentose phosphate cycle oxidative and nonoxidative balance: A new vulnerable target for overcoming drug resistance in cancer. Int J Cancer 2006 Dec 15;119(12):2733–41. [PubMed: 17019714]

20. Banki K, Eddy RL, Shows TB, Halladay DL, Bullrich F, Croce CM, et al. The human transaldolase gene (TALDO1) is located on chromosome 11 at p15.4-p15.5. Genomics 1997 Oct 1;45(1):233–8. [PubMed: 9339383]

21. Esposito M, Venkatesh V, Otvos L, Weng Z, Vajda S, Banki K, et al. Human transaldolase and cross-reactive viral epitopes identified by autoantibodies of multiple sclerosis patients. J Immunol 1999 Oct 1;163(7):4027–32. [PubMed: 10491006]

22. Lam CT, Tang CM, Lau KW, Lung ML. Loss of heterozygosity on chromosome 11 in esophageal squamous cell carcinomas. Cancer Lett 2002 Apr 8;178(1):75–81. [PubMed: 11849744]

23. Sonoda Y, Iizuka M, Yasuda J, Makino R, Ono T, Kayama T, et al. Loss of heterozygosity at 11p15 in malignant glioma. Cancer Res 1995 May 15;55(10):2166–8. [PubMed: 7743518]

24. Grossman CE, Qian Y, Banki K, Perl A. ZNF143 mediates basal and tissue-specific expression of human transaldolase. J Biol Chem 2004 Mar 26;279(13):12190–205. [PubMed: 14702349]

25. Dwyer T, Ponsonby AL, Stankovich J, Blizzard L, Easteal S. Measuring environmental factors can enhance the search for disease causing genes? J Epidemiol Community Health 2004 Jul;58(7):613–5. [PubMed: 15194726]

26. Huebner C, Petermann I, Browning BL, Shelling AN, Ferguson LR. Triallelic single nucleotide polymorphisms and genotyping error in genetic epidemiology studies: MDR1 (ABCB1) G2677/T/A as an example. Cancer Epidemiol Biomarkers Prev 2007 Jun;16(6):1185–92. [PubMed: 17548683]

27. Morita A, Nakayama T, Doba N, Hinohara S, Mizutani T, Soma M. Genotyping of triallelic SNPs using TaqMan PCR. Mol Cell Probes 2007 Jun;21(3):171–6. [PubMed: 17161935]

28. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Human Genet 2004 Jan;74(1):106–20. [PubMed: 14681826]

29. Dupont WD, Plummer WD Jr. Power and sample size calculations. A review and computer program. Control Clin Trials 1990 Apr;11(2):116–28. [PubMed: 2161310]

30. den Dunnen JT, Antonarakis SE. Nomenclature for the description of human sequence variations. Hum Genet 2001;109:121–4. [PubMed: 11479744]

Abbreviations

AA
: African Americans

CI
: 95% confidence interval

bP
: basepairs

EA
: European Americans

GSH
glutathione

HWE
Hardy-Weinberg Equilibrium

LD
linkage disequilibrium

LRT
likelihood ratio test

MAF
Minor Allele Frequencies

OR
odds ratio

PPP
pentose phosphate pathway

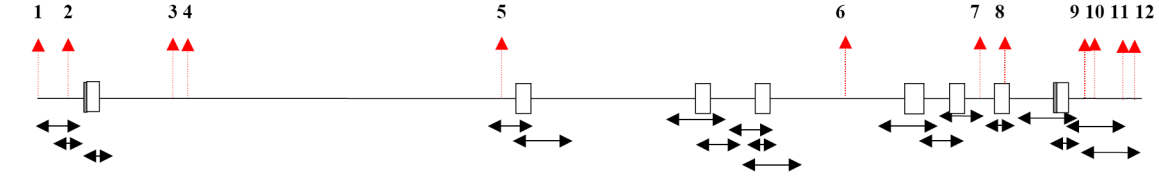
R5P
ribose-5-phosphate

SCCHN
squamous cell carcinoma of the head and neck

SNP
single nucleotide polymorphism

TALDO1
Transaldolase 1

a.



b.

SNP #	rs #
1	rs10794338
2	NA
3	rs3901233
4	rs11246300
5	rs3895063
6	rs4963163
7	rs3891028
8	rs11302
9	rs7933290
10	rs2075545
11	NA
12	NA

Figure 1.

a. Layout of the coverage of the 21 sequencing contigs (↔) that were used to identify SNPs in the *TALDO1* gene and intervening sequences in 96 head and neck cancer patients. Dotted arrows indicate location of SNPs that were used in the genotyping analyses. SNPs (#3, 4, and 6) outside of the 21 sequencing contigs were taken from HapMaP project NCBI phase1/build 34/assembly 124 genotyped SNPs. b. SNP designation by rs #.

Table 1

Characteristics of Study Population

		Cases N = 328	Controls N = 354
Age[a]		59.79 (10.24)	63.03 (11.11)
Gender[b]	**Male**	265 (81)	284 (80)
	Female	63 (19)	70 (20)
Race[b]	**EA**	274 (84)	293 (83)
	AA	54 (16)	61 (17)
Smoking Status[b]	**Former**	127 (39)	253 (71)
	Current	201 (61)	101 (29)
Pack Years[b,c]	**>0-20**	50 (15)	157 (45)
	>20-40	75 (23)	97 (27)
	>40	200 (62)	100 (28)
Cancer Stage[b]	**I**	82 (25)	NA
	II	63 (19)	NA
	III	64 (20)	NA
	IVA,B,C	119 (36)	NA

[a] Mean (SD)

[b] N (%)

[c] There was no data for this variable for 3 cases

NA = Not applicable

Table 2

SNP Description and Minor Allele Frequency in Controls

SNP #	Polymorphism[a]	Minor Allele	Minor Allele Frequency[b]	
			European Americans N=586 n (%)	African Americans N=122 n (%)
1	-490C>G or T (rs10794338)	C	272 (48)	49 (42)
		T	173 (30)	11 (9)
		G	127 (22)	58 (49)
		Missing	14	4
2	-394T>G (chanceaa2)	G	1 (0.2)	8 (7)
		Missing	4	2
3	IVS1+1874T>A (rs3901233)	A	268 (46)	34 (28)
		Missing	8	0
4	IVS1+2198C>T (rs11246300)	T	125 (21)	6 (5)
		Missing	0	0
5	IVS2-209G>A (rs3895063)	A	64 (11)	5 (4)
		Missing	0	0
6	IVS4+2187A>C (rs4963163)	C	278 (47)	35 (29)
		Missing	0	0
7	IVS6+54G>C (rs3891028)	C	143 (24)	55 (45)
		Missing	2	0
8	EX7+102T>C (rs11302)	C	1 (0.2)	6 (5)
		Missing	2	0
9	*193G>A (rs7933290)	A	0	6 (5)
		Missing	2	0
10	*222C>T (rs2075545)	T	4 (0.7)	12 (10)
		Missing	0	0
11	*444G>A (chancec7)	A	27 (5)	1 (0.8)
		Missing	2	0
12	*513G>C (chancec8)	C	66 (11)	6 (5)
		Missing	0	2

[a] Based on Nomenclature scheme of den Dunnen and Antonarakis [30] – indicates the number of the nucleotide 5' of the ATG translation initiation codon,

* indicates the number of the nucleotide 3' of the translation codon

[b] Calculated from controls only, based on number of chromosomes

Table 3

Genotype Frequencies and Odds Ratios in European Americans for SCCHN

SNP #	SNP Name	Genotype	Cases N=274 n (%)	Controls N=293 n (%)	Odds Ratio (95% CI)[a]	P Value[a] (trend)[d]
1	rs10794338	C/C	71 26)	71 (25)	referent	
		C/G	57 (21)	64 (22)	1.12 (0.74 , 1.68)	
		C/T	92 (34)	66 (23)		**0.022**[a]
		G/G	9 (3)	11 (4)		
		G/T	25 (9)	41 (14)	**0.62 (0.38 , 1.00)**	
		T/T	18 (7)	33 (12)		
		C/C vs C/minor allele + minor/minor[b]			0.93 (0.63 , 1.36)	0.694[b]
		C/C + C/minor vs minor/minor[c]			**0.57 (0.38 , 0.86)**	**0.007**[c]
						(0.057)[d]
		Missing	2	7		
3	rs3901233	T/T	54 (20)	91 (32)	referent	
		T/A	145 (54)	128 (44)	**1.81 (1.20 , 2.75)**	**0.017**[a]
		A/A	70 (26)	70 (24)	**1.66 (1.03 , 2.68)**	
		T/T vs T/A+ A/A[b]			**1.76 (1.19 , 2.61)**	**0.005**[b]
		T/T+T/A vs A/A[c]			1.12 (0.76 , 1.65)	0.567[c]
						(0.037)[d]
		Missing	5	4		
4	rs11246300	C/C	151 (55)	183 (63)	referent	
		C/T	110 (40)	95 (32)	1.41 (0.99 , 2.01)	0.1617[a]
		T/T	13 (5)	15 (5)	1.12 (0.51 , 2.44)	
		C/C vs C/T+T/T[b]			1.37 (0.98 , 1.93)	0.069[b]
		C/C+C/T vs T/T[c]			0.97 (0.45 , 2.1)	0.948[c]
						(0.135)[d]
		Missing	0	0		
6	rs4963163	A/A	55 (20)	90 (31)	referent	
		A/C	146 (54)	128 (44)	**1.79 (1.18 , 2.72)**	**0.022**[a]
		C/C	72 (26)	75 (25)	**1.56 (0.97 , 2.51)**	
		A/A vs A/C+C/C[b]			**1.71 (1.16 , 2.53)**	**0.007**[b]
		A/A+A/C vs C/C[c]			1.06 (0.72 , 1.55)	0.763[c]
						(0.069) [d]
		Missing	1	0		
7	rs3891028	G/G	171 (63)	163 (56)	referent	
		G/C	93 (34)	115 (39)	0.77 (0.54 , 1.10)	0.207[a]
		C/C	8 (3)	14 (5)	0.56 (0.23 , 1.37)	
		G/G vs G/C+C/C[b]			0.75 (0.53 , 1.06)	0.102[b]
		G/G+G/C vs C/C[c]			0.62 (0.25 , 1.5)	0.286[c]
						(0.076)[d]
		Missing	2	1		
11	chancec7	G/G	249 (91)	267 (91)	referent	
		A/G	25 (9)	23 (8)	1.19 (0.65 , 2.17)	0.852[a]
		A/A	0	2 (0.7)	NA	
		G/G vs A/G+A/A[b]			1.10 (0.61 , 1.97)	0.759[b]
						(0.976)[d]
		Missing	0	1		
12	chancec8	G/G	206 (75)	231 (79)	referent	
		C/G	67 (24)	58 (20)	1.31 (0.88 , 1.96)	0.196[a]
		C/C	1 (0.4)	4 (1)	0.26 (0.03 , 2.39)	

SNP #	SNP Name	Genotype	Cases N=274 n (%)	Controls N=293 n (%)	Odds Ratio (95% CI)[a]	P Value[a] (trend)[d]
		G/G vs G/C+C/C[b]			1.24 (0.83 , 1.84)	0.288[b]
		G/G+G/C vs C/C[c]			0.25 (0.03 , 2.25)	0.215[c]
						(0.473)[d]
		Missing	0	0		

[a] General model; Adjusted for age (continuous) and gender

[b] Dominant model (major/major = the referent)

[c] Recessive model (major/major + major/minor = the referent)

[d] Trend test; Adjusted for age (continuous) and gender

Table 4

OR for SNP and Later Stage at Diagnosis in European Americans

SNP #	SNP Name	Genotype	Late Stage (III+IV) N=146 n (%)	Early Stage (I+II) N=128 n (%)	Odds Ratio (95% CI)[a]	P Value[a]
1	rs10794338	2 C alleles	40 (28)	31 (24)	referent	
		1 C allele	83 (58)	66 (52)	0.97 (0.54 , 1.73)	0.215
		0 C allele	21 (14)	31 (24)	0.56 (0.27 , 1.18)	
		Missing	2	0		
3	rs3901233	T/T	22 (15)	32 (25)	referent	
		T/A	82 (58)	63 (50)	1.71 (0.90 , 3.26)	0.242
		A/A	39 (27)	31 (25)	1.68 (0.81 , 3.50)	
		Missing	3	2		
4	rs11246300	C/C	78 (53)	73 (57)	referent	
		C/T	62 (43)	48 (38)	1.23 (0.74 , 2.03)	0.606
		T/T	6 (4)	7 (5)	0.78 (0.24 , 2.46)	
		Missing	0	0		
6	rs4963163	A/A	22 (15)	33 (26)	referent	
		A/C	83 (57)	63 (49)	1.81 (0.95 , 3.43)	0.176
		C/C	40 (28)	32 (25)	1.76 (0.85 , 3.65)	
		Missing	1	0		
7	rs3891028	G/G	94 (65)	77 (60)	referent	
		G/C	49 (34)	44 (35)	0.88 (0.52 , 1.47)	0.375
		C/C	2 (1)	6 (5)	0.32 (0.06 , 1.66)	
		Missing	1	1		
11	chancec7	G/G	130 (89)	119 (93)	referent	
		A/G	16 (11)	9 (7)	1.48 (0.62 , 3.53)	0.378
		A/A	0	0	NA	
		Missing	0	0		
12	chancec8	G/G	113 (77)	93 (73)	referent	
		C/G	33 (23)	34 (27)	0.78 (0.44 , 1.37)	0.691
		C/C	0	1 (0.8)	NA	
		Missing	0	0		

[a] General Model; Adjusted for age (continuous) and gender