

Salivary markers and risk factor data: A multivariate modeling approach for head and neck squamous cell carcinoma detection

Lutécia H. Mateus Pereira^a, Islamiyat Nancy Adebisi^a, Aymee Perez^b, Michael Wiebel^c,
Isildinha Reis^{d,e}, Robert Duncan^d, W. Jarrard Goodwin^{a,c}, Jennifer J. Hu^{a,d},
Vinata B. Lokeshwar^{b,f} and Elizabeth J. Franzmann^{a,c,*}

^aSylvester Comprehensive Cancer Center, University of Miami Leonard Miller School of Medicine, Miami, FL, USA

^bDepartment of Cell Biology and Anatomy, University of Miami Leonard Miller School of Medicine, Miami, FL, USA

^cDepartment of Otolaryngology, University of Miami Leonard Miller School of Medicine, Miami, FL, USA

^dDivision of Biostatistics, Department of Epidemiology and Public Health, University of Miami Leonard Miller School of Medicine, Miami, FL, USA

^eSylvester Comprehensive Cancer Center, Division of Biostatistics, University of Miami Leonard Miller School of Medicine, Miami, FL, USA

^fDepartment of Urology, University of Miami Leonard Miller School of Medicine, Miami, FL, USA

Abstract. *Background:* Head and neck squamous cell carcinoma (HNSCC) is a debilitating and deadly disease largely due to late stage diagnosis. Prior work indicates that soluble CD44 (solCD44) and total protein may be useful diagnostic markers for HNSCC. In this study we combine the markers solCD44, IL-8, HA, and total protein with demographic and risk factor data to derive a multivariate logistic model that improves HNSCC detection as compared to our previous data using biomarkers alone.

Methods: We performed the solCD44, IL-8, HA, and total protein assays on oral rinses from 40 HNSCC patients and 39 controls using ELISA assays. Controls had benign diseases of the upper aerodigestive tract and a history of tobacco or alcohol use. All subjects completed a questionnaire including demographic and risk factor data.

Results: Depending on cancer subsite, differences between cases and controls were found for all markers. A multivariate logistic model including solCD44, total protein and variables related to smoking, oral health and education offered a significant improvement over the univariate models with an AUC of 0.853. Sensitivity ranged from 75–82.5% and specificity from 69.2–82.1% depending on predictive probability cut points.

Conclusion: A multivariate model, including simple and inexpensive molecular tests in combination with risk factors, results in a promising tool for distinguishing HNSCC patients from controls.

Impact: In this case-control study, the resulting observations led to an unprecedented multivariate model that distinguished HNSCC cases from controls with better accuracy than the current gold standard which includes oral examination followed by tissue biopsy. Since the components are simple, noninvasive, and inexpensive to obtain, this model combining biomarkers, risk factor and demographic data serves as a promising prototype for future cancer detection tests.

Keywords: CD44, hyaluronic acid, IL-8, protein, head and neck cancer

1. Introduction

Each year 50 000 individuals in the United States [1] and more than 600,000 worldwide [2] are diagnosed with head and neck squamous cell carcinoma (HNSCC). The main risk factors, smoking, alcohol use and human papillomavirus (HPV) infection [3] are very

*Corresponding author: Elizabeth J. Franzmann, Department of Otolaryngology, Head and Neck Surgery Division, University of Miami, Leonard Miller School of Medicine, and Jackson Memorial Hospital, 1120 NW 14th Street, Clinical Research Bldg, Room 1513, Miami, FL 33136, USA. Tel.: +1 305 243 5955; Fax: +1 305 243 3383; E-mail: efranzman@med.miami.edu.

common. Thus millions are at risk for the disease. Yet screening is suboptimal. Consequently, most patients are diagnosed in late stage when 5-year relative survival rates reach only 30 percent.

There is currently no acceptable early detection test for HNSCC [4]. Adjunctive techniques for oral cancer detection are available, but whether they improve early detection rates remains unclear. Conventional oral examination, followed by tissue biopsy is the gold-standard for oral cancer screening, but lacks sensitivity and specificity and is costly [5]. Therefore, several studies have tested saliva for RNA expression profiles [6], microRNA discovery [7] and proteomic analysis [8]. While such molecular investigations have yielded promising results, none has been validated in large trials.

Our work focuses on cancer-related markers that can be measured simply and inexpensively [9]. CD44 is overexpressed in normal versus dysplastic and malignant upper aerodigestive tract (UADT) mucosa [10] and is involved in tumorigenesis [11]. Its soluble form (solCD44) is a major ligand for hyaluronic acid (HA). Interleukin-8 (IL-8) and total protein levels are all increased in saliva and/or oral rinses from HNSCC patients compared to controls [12,13]. Each of these markers is fairly stable and easily measured with ELISA and ELISA-like assays, detection platforms that are highly validated and widely used [14]. Our prior work shows that the combination of solCD44 and total protein results in a test that distinguishes cancer patients from controls with high accuracy. To improve the test further we assessed additional markers IL-8 and HA as well as demographic and risk factor data, since these can affect biomarker levels [15]. Then, using multivariate analysis, we derived a model combining the most significant biomarkers, demographic and risk factor data that distinguished cancer from benign disease with higher accuracy than any combination of the biomarkers alone.

2. Materials and methods

Subjects were recruited and consented for the study according to the Institutional Review Board approved protocol from the University of Miami Sylvester Comprehensive Cancer Center, University of Miami Hospital and Clinics (UM/SCCC), and Jackson Memorial Hospital (JMH) otolaryngology clinics. All experiments were undertaken with the understanding and written consent of each subject and conform to The

Code of Ethics of the World Medical Association (Declaration of Helsinki). Subjects completed a questionnaire that included demographic data as well as patient's behavioral risk factors and socioeconomic status (SES). HNSCC patients had biopsy-proven squamous cell carcinoma of the oral cavity (OC), oropharynx (OP), larynx (L), or hypopharynx (H). Controls included 39 consecutively enrolled individuals (16 from a prior study) [16] with a history of tobacco and/or alcohol use who were treated for benign diseases of the UADT. Cases included 40 consecutively enrolled, newly diagnosed HNSCC patients. Squamous cell carcinomas of the nasopharynx, paranasal sinuses, esophagus or salivary glands were also excluded, as were pregnant women and subjects with HIV infection.

Oral rinses were collected using previously published procedures [16]. Levels of solCD44 (normal and variant isoforms) were measured using a sandwich ELISA assay (Bender MedSystems), with previously published modifications [16]. The HA test uses a competitive binding principle where the HA coated on a microtiter plate competes with HA present in a subject's saliva (10). We modified the test to correct for matrix effects as previously described for the solCD44 assay [16]. An ELISA kit was used for IL-8 (R&D Systems, Inc.) according to the manufacturer's protocol. We performed the protein assay (Bio-Rad Laboratories) according to the manufacturer's protocol using saliva samples prepared as previously published [16]. Each sample was tested in triplicate, concentrations were averaged and the absorbances read in a microplate reader (Bio-Rad Laboratories). Concentrations were determined by a standard curve.

Statistical analyses were performed with the SAS software, version 9.2 (SAS Institute, Inc.). We compared patient groups with respect to the distribution of potentially important covariates using the chi-square test for categorical data. If indicated by small expected values, Fisher's exact test was used instead of chi-square test. Data on the four markers were log base-2 transformed to stabilize estimates of variance and improve the fit to the normal distribution. Continuous data were analyzed using Analysis of variance (ANOVA) followed by Fisher's least-significant-difference test for pairwise mean comparison, and tests of pre-specified contrasts. Univariate and multivariate logistic regression analyses were used to assess the effect of markers on risk for HNSCC. We report odds ratio estimates with corresponding 95% confidence interval (95% CI) and area under the curve (AUC) of the receiver operating characteristic (ROC) plots for fitted models. Also,

we report estimates of sensitivity, specificity, and accuracy derived from a fitted multivariate model which included significant interactions between markers and covariates.

3. Results

Demographics of the study participants are summarized in Table S1. The groups did not differ statistically in regards to gender, race, ethnicity, health insurance, smoking status, alcohol intake and number of teeth removed. The groups did differ significantly in age (cases were older, $P = 0.001$), education (cases were less educated, $P = 0.007$), and ability to gargle (cases had poorer gargle, $P = 0.004$). The data on income showed cases with significantly lower income. Since 25% of cases and 10% of controls had missing data on this variable, it was not used in subsequent analyses. Table S2 also shows cancer-specific characteristics among the cases.

Comparisons of the marker means of the log base-2 transformed data by group are shown in Table 1. OC and OP cancer cases and controls were significantly different with respect to all four markers with the exception of HA for OP cancer; however, no differences were found when comparing L/H cases and controls with respect to these markers. The comparison of all cases versus controls reached statistical significance at the 5% level for $\log_2\text{CD44}$ and $\log_2\text{total protein}$, but was borderline significant for IL-8 ($P = 0.058$) and not significant for HA ($P = 0.107$).

Mean marker level comparisons between group (cases versus controls) when key variables are considered are reported in Table 1. When smoking history was considered, $\log_2\text{solCD44}$ was significantly higher in all subgroups of cases compared to controls for never smokers ($P = 0.0002$) but not ever smokers ($P = 0.546$). Similarly, when oral health was considered, mean $\log_2\text{solCD44}$ and $\log_2\text{total protein}$ were significantly higher in OC and OP cases than in controls in subjects with five or fewer teeth removed (CD44: $P = 0.0008$, total protein: $P = 0.001$) but was not higher when subjects had 6 or more teeth removed (CD44: $P = 0.790$, total protein: $P = 0.838$). There were no differences between groups based on the following variables: ethnicity, alcohol use, education, gargle, and health care coverage. Significant within group differences were seen in cancer patients (Table 1), for $\log_2\text{solCD44}$ by disease site (lower for L/H than OP), for $\log_2\text{HA}$ by disease site (lower for L/H than OC and

lower for OP than OC) and for $\log_2\text{IL-8}$ by disease site (lower for L/H than OC and OP), and node status (lower for N0/Nx L/H compared to OC and lower for N0/Nx than N1-3 within OP cancer). Among controls (Table 1, column 5), differences in means were significant for $\log_2\text{solCD44}$ by smoking history (lower in never than in ever smoke) and by number of teeth removed (lower in 5 or less compared to 6 or more), and for $\log_2\text{total protein}$ by number of teeth removed (lower in 5 or less compared to 6 or more).

Univariate odds ratios and the predictive ability of the logistic models are shown in Table 2. Univariate, all markers seemed to have some predictive ability. A complete logistic model including all markers, smoking, interaction of $\log_2\text{solCD44}$ and smoking, teeth removed, interaction of $\log_2\text{total protein}$ and teeth removed, gender, ethnicity, education, health coverage, alcohol use and gargle ability was computed. The only significant terms were $\log_2\text{solCD44}$, smoking, $\log_2\text{solCD44}$ and smoking interaction, $\log_2\text{total protein}$, $\log_2\text{total protein}$ and teeth removed interaction, and education.

We also specifically tested whether there was an effect modification of age and/or gender. The ages were stratified as ≤ 55 and > 55 and a three-way ANOVA of group, coded age, and gender with main effects and all two-way interactions were performed for solCD44 , $\log_2\text{solCD44}$, and protein. None of the interactions was significant, indicating that there was no effect modification by age or gender in this data set.

To help clarify the interactions, we report in Table 2A odds ratios for $\log_2\text{solCD44}$ effect within smoking categories and odds ratios for $\log_2\text{total protein}$ within teeth removed categories. As expected, the effect of $\log_2\text{solCD44}$ is significant only among never smokers, and the effect of $\log_2\text{total protein}$ is significant only among those having had five or fewer teeth removed. The area under the ROC curve for the multivariate model is 0.853, which is a significant improvement over the univariate models. Two choices for cutpoints were chosen that resulted in the best estimates for determining sensitivity and specificity (Table 2B), providing up to 82.1% specificity and 75% sensitivity (cut point of 0.547) or sensitivity as high as 82.5%, with corresponding specificity of 69.2% (cut point of 0.509).

Bivariate cross tabulations of gargle, number of teeth removed, smoking and alcohol by patient group (3 sites and controls) were made. The only statistically significant association was between gargle and patient group. However, mean marker levels differences by gargle categories (good versus poor/fair) within patient group

Table 1
Molecular markers in oral rinses from 40 HNSCC patients and 39 controls: mean \pm standard deviation and (n)

Molecular marker/ variable category ^a	Larynx/ Hypopharynx can cer	Oral Cavity cancer	Oropharynx cancer	Controls	Cases vs. controls P value ^b
Log ₂ solCD44	1.17 ¹ \pm 0.90	1.73 ² \pm 0.77	2.02 ^{1,3} \pm 1.01	1.11 ^{2,3} \pm 0.86	0.010
Log ₂ total protein	-0.39 \pm 0.54	-0.15 ¹ \pm 0.48	-0.05 ² \pm 0.63	-0.62 ^{1,2} \pm 0.80	0.009
Log ₂ HA	6.95 ¹ \pm 1.72	9.59 ^{1,2,3} \pm 2.24	7.65 ² \pm 2.21	7.19 ³ \pm 2.64	0.107
Log ₂ IL-8	6.75 ^{1,2} \pm 0.96 (13)	8.41 ^{1,3} \pm 0.72 (12)	8.33 ^{2,4} \pm 1.45 (15)	7.26 ^{3,4} \pm 1.49 (39)	0.058
Log ₂ solCD44 Never smoke ^c	1.60 ¹ \pm 0.49 (4)	1.55 ² \pm 0.79 (7)	1.99 ³ \pm 0.96 (6)	0.58 ^{1,2,3,4} \pm 0.58 (17)	0.0002
Ever smoke	0.98 ^{5,6} \pm 1.00 (9)	1.98 ⁵ \pm 0.76 (5)	2.04 ⁶ \pm 1.10 (9)	1.51 ⁴ \pm 0.83 (22)	0.546
None/5 or less teeth removed	0.92 ¹ \pm 0.98 (7)	1.69 ³ \pm 0.70 (7)	2.50 ^{1,2,4} \pm 1.11 (8)	0.87 ^{2,3,5} \pm 0.76 (27)	0.0008
6 or more/all teeth removed	1.45 \pm 0.79 (6)	1.78 \pm 0.95 (5)	1.47 ⁴ \pm 0.52 (7)	1.65 ⁵ \pm 0.85 (12)	0.790
Log ₂ total protein None/5 or less teeth removed	-0.44 \pm 0.54 (7)	-0.21 ¹ \pm 0.35 (7)	0.17 ² \pm 0.65 (8)	-0.81 ^{1,2,3} \pm 0.68 (27)	0.001
6 or more/all teeth removed	-0.33 \pm 0.60 (6)	-0.08 \pm 0.66 (5)	-0.30 \pm 0.55 (7)	-0.19 ³ \pm 0.90 (12)	0.838
log ₂ IL-8 N0, Nx	6.43 ¹ \pm 0.92 (9)	8.24 ¹ \pm 0.44 (8)	7.20 ² \pm 2.05 (3)	NA NA	
N1, N2, N3	7.48 \pm 0.64 (4)	8.77 \pm 1.10 (4)	8.62 ² \pm 1.21 (12)	NA	

^aCD44 in ng/ml, total protein in mg/ml, HA in ng/ml, and IL-8 in pg/ml.

^bP-value for contrast comparing all cases v. controls. Same number identifies pairwise mean comparisons significant at $P < 0.05$ by Fisher's least-significant-difference test. For example, for log₂solCD44 within *never smoke* (second panel) the mean for controls was significantly lower than the means for each disease site (1, 2, 3), and there were no significant differences among disease sites. Mean comparisons between patient groups by the following variables – ethnicity, alcohol use, education, gargle, and health care coverage – were not statistically significant at the 5% level.

^cCategory 'Never' defined as having smoked < 100 cigarettes in lifetime or quit ≥ 10 years ago.

were not significant. Since impaired gargle is more likely caused by the tumor rather than a risk factor for the tumor, gargle ability was not included in the model.

We also performed multivariate logistic regression analyses by cancer site and stage as shown in Table S3. It is clear that the relative size of the Odds Ratios within subgroups and the associated significance levels are comparable to the overall multivariate model.

4. Discussion

HNSCC is a multifactorial disease marked by racial, gender and SES disparities [17]. Biomarker levels may differ based on certain demographic or risk variables [15]. Therefore, in order to better discriminate cancer patients from controls, we created a preliminary model that combines these important variables with promising biomarkers. In our model, we found better discrimination power (75–82.5% sensitivity and 69.2–82.1% specificity, AUC = 0.853) than the current gold standard of physical exam followed by biop-

sy (sensitivity 64%, specificity 74%) [18]. This study is unique and important because we show that specific demographics and risk factors can be incorporated into the combined marker model (CD44 and total protein) to improve accuracy of the test.

Although the study size is relatively small there are several mitigating factors in support of the findings. First, the sample sizes of cases and controls are virtually equal (40 and 39) which leads to maximum power among the statistical comparisons for this total sample size of 79 patients. Second, all the subgroup comparisons, including the ones performed by cancer site and stage (Table S3) are consistent among themselves and with the published literature. Finally, the areas under the curve for the various predictions models are uniformly high ($\sim 85\%$) and sensitivity ranges from 75%–83% and specificity from 69% to 82%. Thus, we feel that the findings reported in this manuscript are both internally and externally consistent in spite of relative small total sample size.

While this work is preliminary, we envision that a similar type of test combining solCD44 and total pro-

Table 2
Molecular predictors of HNSCC cases versus controls

(A) Logistic regression models ^a	Odds ratio (95%CI)	<i>P</i> value	AUC ^b	
Univariate				
log ₂ solCD44	1.96 (1.15–3.33)	0.013	0.670	
log ₂ total protein	2.51 (1.24–5.11)	0.011	0.667	
log ₂ HA	1.15 (0.95–1.39)	0.150	0.597	
log ₂ IL-8	1.36 (0.97–1.91)	0.076	0.643	
Multivariate^c				
log ₂ CD44, in ever smoke	0.88 (0.36–2.16)	0.772		
log ₂ CD44, in never smoke	18.10 (1.98–165.40)	0.010		
log ₂ total protein, in 6+/all teeth removed	0.40 (0.09–1.72)	0.219		
log ₂ total protein, in none/5 or less teeth removed	6.56 (1.27–33.84)	0.024		
Education ^d (low vs. high)	6.17 (1.58–24.11)	0.009		
(B) Examples of classification based on multivariate model				
	Predicted probability cut point	Sensitivity (%)	Specificity (%)	Accuracy (%)
Better sensitivity	0.509	82.5	69.2	75.9
Better specificity	0.547	75.0	82.1	78.5

^aBased on 40 cancer and 39 controls. ^bAUC: area under ROC (receiver operating curve). ^cMultivariate model including log₂CD44, log₂total protein, smoking status, teeth removed, education, interaction log₂CD44 by smoking status, and interaction log₂total protein by teeth removed. The effect of log₂solCD44 is reported by smoking categories because the corresponding interaction was significant ($P = 0.010$); likewise, the effect of log₂total protein is reported by number of teeth removed because the corresponding interaction was also significant ($P = 0.008$). The inclusion of log₂ HA, log₂ IL-8, and age in the model were tested and none reach statistical significance at the 5% level, therefore they were dropped from the model. ^dLow = grades 1–12 or GED or HSG, High = college, completed or not.

Note: Using log₂-transformed data, the odds ratio estimates represent the change in risk corresponding to a two-fold increase in the original variable.

tein with easily determined demographic and risk factor data could be performed in a clinical setting. Combined results above a predetermined cutoff point, would be referred to a specialist for examination, biopsy, and treatment. Those without identifiable lesions would be followed up with regular oral rinse tests and physical exam. Furthermore, they would be encouraged to stop smoking and optimize their oral health, since it is possible for premalignant lesions to regress [19]. ELISAs such for solCD44 are quite inexpensive and total protein can be measured for mere pennies a sample. Since economically disadvantaged groups suffer disproportionately from this disease, an inexpensive, easily accessible test is crucial.

Despite the need for a multi-institutional validation to test whether the significant findings related to tobacco and tooth loss (a surrogate for oral health) in this small cohort hold true, these findings are consistent with our prior published work [20]. We analyzed 102 HNSCC patients and 84 control subjects using the same tests. The latter study data on solCD44 for controls suggested a trend towards association between higher solCD44 levels and current tobacco use ($P = 0.07$) that was lost when cancer patients were concerned ($P = 0.46$). In addition, there are some studies suggesting

that smoking may affect circulating levels of CD44 isoforms, and that those levels decline rapidly even within 4 weeks of smoking cessation ($P < 0.001$) [21,22]. In order to investigate the true meaning of these findings, we have an ongoing study that will determine how changes in oral rinse solCD44 and protein levels correlate with smoking cessation status.

Tobacco exposure and poor oral health are both known potential risk factors for HNSCC [23,24]. Since solCD44 and total protein levels were elevated in controls with histories of heavy smoking or poor oral health, they may be indicators of very early tumorigenesis. Premalignancy and early malignancy of the UADT are often occult [25] and CD44 is a tumor initiation marker [26]; thus our test may actually have higher accuracy than reported here. In fact, we have shown in prior work that the test is effective at detecting early disease [16]. To better understand the relationship between these markers and early disease, we are currently following subjects with leukoplakia over time.

The addition of HA and IL-8 did not improve the model, though these markers have shown promise in other studies [6,13,27]. Our results do suggest that IL-8 levels are associated with more aggressive disease (lymph node metastases). This corroborates work by

Gokhale et al. [28] that found serum levels of IL-8 were elevated in patients with metastatic or recurrent HNSCC ($P = 0.03$ and $P = 0.001$, respectively) compared to healthy controls. Further work investigating IL-8 and HA with larger numbers of subjects may show a benefit.

Since HPV is an important risk and prognostic factor for oropharyngeal HNSCC we analyzed results based on tumor site (Table 1). All marker levels, with the exception of total protein, were correlated with site of disease, with higher levels found in either OC or OP compared to L and H. However, in this small sample, accuracy of the test did not differ based on tumor site (OC/OP versus L/H). The incidence of HNSCC involving the tongue and tonsils has increased compared to HNSCC at other sites which have actually remained constant or decreased [29]. HPV is such a major contributor to current HNSCC etiology that we have decided to focus primarily on oral cavity and oropharynx cancers in our ongoing study.

A modeling approach using simple and inexpensive tumor markers, solCD44 and total protein, along with risk factor and demographic data shows great promise in this pilot dataset. Investigations are underway to optimize the modeling approach described here in larger clinic-based cohorts.

Acknowledgements

This study was supported in part by NIH/National Cancer Institute grants CA107828 and CA118584, Flight Attendant Medical Research Institute grant 12536, Sylvester Comprehensive Cancer Center and Woman's Cancer Association (E.J.F).

Conflict of interest

The University of Miami and Drs. Elizabeth Franzmann and Vinata Lokeshwar hold a patent focused on CD44 and have the potential for financial benefit from its future commercialization.

All other authors: None declared.

References

- [1] American Cancer Society. Cancer Facts & Figures 2011. Atlanta: American Cancer Society; 2011.
- [2] D.M. Parkin, F. Bray, J. Ferlay, P. Pisani, Global cancer statistics, 2002, *CA Cancer J Clin* 55 (2005), 74-108.
- [3] M.L. Gillison, G. D'Souza, W. Westra, E. Sugar, W. Xiao, S. Begum et al., Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers, *J Natl Cancer Inst* 100 (2008), 407-420.
- [4] H.S. Lin, H.S. Talwar, A.L. Tarca, A. Ionan, M. Chatterjee, B. Yeet et al., Autoantibody approach for serum-based detection of head and neck cancer, *Cancer Epidemiol Biomarkers Prev* 16 (2007), 2396-2405.
- [5] M.W. Lingen, J.R. Kalmar, T. Karrison, P.M. Speight, Critical evaluation of diagnostic aids for the detection of oral cancer, *Oral Oncol* 44 (2008), 10-22.
- [6] Y. Li, M.A. St John, X. Zhou, Y. Kim, U. Sinha, R.C. Jordan et al., Salivary transcriptome diagnostics for oral cancer detection, *Clin Cancer Res* 10 (2004), 8442-8450.
- [7] N.J. Park, H. Zhou, D. Elashoff, B.S. Henson, D.A. Kastratovic, E. Abemayore et al., Salivary microRNA: discovery, characterization, and clinical utility for oral cancer detection, *Clin Cancer Res* 15 (2009), 5473-5477.
- [8] S. Hu, M. Arellano, P. Boonthueung, J. Wang, H. Zhou, J. Jianget al., Salivary proteomics for oral cancer biomarker discovery, *Clin Cancer Res* 14 (2008), 6246-6252.
- [9] E.P. Reategui, A.A. de Mayolo, P.M. Das, F.C. Astor, R. Singal, K.L. Hamilton et al., Characterization of CD44v3-containing isoforms in head and neck cancer, *Cancer Biol Ther* 5 (2006), 1163-1168.
- [10] E. Ioachim, D. Assimakopoulos, A.C. Goussia, D. Peschos, A. Skevas, N.J. Agnantis, Glycoprotein CD44 expression in benign, premalignant and malignant epithelial lesions of the larynx: an immunohistochemical study including correlation with Rb, p53, Ki-67 and PCNA, *Histol Histopathol* 14 (1999), 1113-1118.
- [11] L.Y. Bourguignon, E. Gilad, A. Brightman, F. Diedrich, P. Singleton, Hyaluronan-CD44 interaction with leukemia-associated RhoGEF and epidermal growth factor receptor promotes Rho/Ras co-activation, phospholipase C epsilon-Ca2+ signaling, and cytoskeleton modification in head and neck squamous cell carcinoma cells, *J Biol Chem* 281 (2006), 14026-14040.
- [12] E.J. Franzmann, G.L. Schroeder, W.J. Goodwin, D.T. Weed, P. Fisher, V.B. Lokeshwar, Expression of tumor markers hyaluronic acid and hyaluronidase (HYAL1) in head and neck tumors, *Int J Cancer* 106 (2003), 438-445.
- [13] M.A. St John, Y. Li, X. Zhou, P. Denny, C.M. Ho, C. Montemagno et al., Interleukin 6 and interleukin 8 as potential biomarkers for oral cavity and oropharyngeal squamous cell carcinoma, *Arch Otolaryngol Head Neck Surg* 130 (2004), 929-935.
- [14] R.R.-S. Lucille Beaudet, Marie-Hélène Venne, Mireille Caron, Julie Bédard, Véronique Brechler, Stéphane Parent, et al., AlphaLISA immunoassays: the no-wash alternative to ELISAs for research and drug discovery, *Nature Methods* 5 (2008).
- [15] L. Savage, Is an improved PSA screening test in sight? *J Natl Cancer Inst* 99 (2007), 1503-1504.
- [16] E.J. Franzmann, E.P. Reategui, F. Pedroso, F.G. Pernas, B.M. Karakullukcu, K.L. Carraway et al., Soluble CD44 is a potential marker for the early detection of head and neck cancer, *Cancer Epidemiol Biomarkers Prev* 16 (2007), 1348-1355.
- [17] W.J. Goodwin, G.R. Thomas, D.F. Parker, D. Joseph, S. Levis, E. Franzmann et al., Unequal burden of head and neck cancer in the United States, *Head Neck* 30 (2008), 358-371.
- [18] P. Brocklehurst, O. Kujan, A.M. Glenny, R. Oliver, P. Sloan, G. Ogdnet et al., Screening programmes for the early detection

- and prevention of oral cancer, *Cochrane Database Syst Rev* 11 (2010), CD004150.
- [19] A. Larsson, T. Axell, G. Andersson, Reversibility of snuff dippers' lesion in Swedish moist snuff users: a clinical and histologic follow-up study, *J Oral Pathol Med* 20 (1991), 258-264.
- [20] R.E. Franzmann EJ, Mateus Pereira LH, Pedrosa F, Joseph D, Allen GO, et al., Salivary protein and solCD44 levels as a potential screening tool for early detection of head and neck squamous cell carcinoma, *Head And Neck* Jul 11. doi: 10.1002/hed.21810. [Epub ahead of print] (2011).
- [21] D.A. Scott, J.A. Stapleton, R.M. Palmer, R.F. Wilson, G. Sutherland, P.Y. Coward et al., Plasma concentrations of reputed tumor-associated soluble CD44 isoforms (v5 and v6) in smokers are dose related and decline on smoking cessation, *Cancer Epidemiol Biomarkers Prev* 9 (2000), 1211-1214.
- [22] R.M. Palmer, J.A. Stapleton, G. Sutherland, P.Y. Coward, R.F. Wilson, D.A. Scott, Effect of nicotine replacement and quitting smoking on circulating adhesion molecule profiles (sICAM-1, sCD44v5, sCD44v6), *Eur J Clin Invest* 32 (2002), 852-857.
- [23] N. Guha, P. Boffetta, V. Wunsch Filho, J. Eluf Neto, O. Shangina, D. Zaridze et al., Oral health and risk of squamous cell carcinoma of the head and neck and esophagus: results of two multicentric case-control studies, *Am J Epidemiol* 166 (2007), 1159-1173.
- [24] M. Hashibe, P. Brennan, S. Benhamou, X. Castellsague, C. Chen, M.P. Curado 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* 99 (2007), 777-789.
- [25] C.F. Poh, S.P. Ng, P.M. Williams, L. Zhang, D.M. Laronde, P. Lane et al., Direct fluorescence visualization of clinically occult high-risk oral premalignant disease using a simple handheld device, *Head Neck* 29 (2007), 71-76.
- [26] M.E. Prince, R. Sivanandan, A. Kaczorowski, G.T. Wolf, M.J. Kaplan, P. Dalerba et al., Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma, *Proc Natl Acad Sci U S A* 104 (2007), 973-978.
- [27] F. Linkov, A. Lisovich, Z. Yurkovetsky, A. Marrangoni, L. Velikokhatnaya, B. Nolen et al., Early detection of head and neck cancer: development of a novel screening tool using multiplexed immunobead-based biomarker profiling, *Cancer Epidemiol Biomarkers Prev* 16 (2007), 102-107.
- [28] A.S. Gokhale, R.I. Haddad, L.A. Cavacini, L. Wirth, L. Weeks, M. Hallare et al., Serum concentrations of interleukin-8, vascular endothelial growth factor, and epidermal growth factor receptor in patients with squamous cell cancer of the head and neck, *Oral Oncol* 41 (2005), 70-76.
- [29] C.H. Shiboski, B.L. Schmidt, R.C. Jordan, Tongue and tonsil carcinoma: increasing trends in the U.S. population ages 20-44 years, *Cancer* 103 (2005), 1843-1849.

Supplemental material

Table S1. Demographics and other characteristics of study subjects

Variable		Cancer (n = 40)		Control (n = 39)		P value ^a
		N	% ¹	N	%	
Age, years	≤ 55	16	40.0	26	66.7	0.024
	> 55	24	60.0	13	33.3	
	Median (minimum – maximum)	62.5 (27–84)		50.0 (29–69)		
	Mean (standard deviation)	60.9 (13.4)		49.3 (11.1)		< 0.001
Gender	Female	11	27.5	15	38.5	0.300
	Male	29	72.5	24	61.5	
Race ^b	White	37	92.5	31	86.1	0.465
	Black	2	5.0	4	11.1	
	Asian	1	2.5	1	2.8	
	Missing (%)	–	–	3	(7.7)	
Ethnicity	Hispanic	18	45.0	20	51.3	0.576
	Non-Hispanic	22	55.0	19	48.7	
Education ^c	Grades 1–12 or GED or HSG	21	52.5	9	23.1	0.007
	College, complete or not	19	47.5	30	76.9	
Income	Less than \$25,000	17	56.7	9	25.7	0.011
	\$25,000 or more	13	43.3	26	74.3	
	Missing	10	(25.0)	4	(10.3)	
Health care	Yes	31	79.5	34	87.2	0.362
	No	8	20.5	5	12.8	
	Missing (%)	1	(2.5)	–	–	
Smoking status ^d	Ever	23	57.5	22	56.4	0.922
	Never	17	42.5	17	43.6	
Alcohol	Yes	22	57.9	24	61.5	0.746
	No	16	42.1	15	38.5	
	Missing (%)	2	(5.0)	–	–	
	Teeth removed	None	8	20.0	13	
	1 to 5	14	35.0	14	35.9	0.193
	6 or more but not all	9	22.5	11	28.2	
	All	9	22.5	1	2.6	
	None/1 to 5	22	55.0	27	69.2	
	6 or more but not all, or all	18	45.0	12	30.8	
Gargle ^e	Poor	8	20.5	1	2.6	0.004
	Fair	10	25.6	5	13.2	
	Good	21	53.9	32	84.2	
	Missing (%)	1	(2.5)	1	(2.6)	

^aMissing are not included in denominator for category percentage nor in group comparison using Fisher's exact test, chi-square test, or Student's t-test. ^bWhite versus Black/Asian. ^cGED = General Educational Development, HSG = High School Graduate. ^dCategory 'Never' defined as having smoked < 100 cigarettes in lifetime or quit ≥ 10 years ago. ^eGood versus Poor/Fair.

Table S2. Disease characteristics of the 40 cancer patients

Variable	Categories	N	%	Variable	Categories	N	%
Disease site	Laryngeal cancer	9	22.5	N stage	No, Nx (1 case)	20	50.0
	Hypopharyngeal cancer	4	10.0		N1	7	17.5
	Lip and oral cavity cancer	12	30.0		N2	10	25.0
	Oropharyngeal cancer	15	37.5		N3	3	7.5
Disease stage	I	6	15.0	M stage	M0	33	82.5
	II	5	12.5		M1	1	2.5
	III	12	30.0		Mx	6	15.0
	IVA	14	35.0				
	IVB	3	7.5				
T stage	T1, Tx (1 case)	8	20.0				
	T2	17	42.5				
	T3	8	20.0				
	T4 or T4a	7	17.5				

Table S3. Fit of multivariate logistic regression model in Table 2 to subset of cases vs. all controls

Multivariate final model ^a	Odds Ratio (95%CI)	P value	AUC ^b
All pts, 40 cancer cases vs. 39 controls:			0.853
log ₂ CD44, in ever smoke	0.88 (0.36–2.16)	0.772	
log ₂ CD44, in never smoke	18.10 (1.98–165.40)	0.010	
log ₂ protein, in 6+/all teeth removed	0.40 (0.09–1.72)	0.219	
log ₂ protein, in none/5 or less teeth removed	6.56 (1.27–33.84)	0.024	
Education (low vs. high)	6.17 (1.58–24.11)	0.009	
Subset of 27 OC/Oropharynx cancer cases vs. 39 controls:			0.852
log ₂ CD44, in ever smoke	1.49 (0.44–5.00)	0.518	
log ₂ CD44, in never smoke	14.91 (1.55–143.24)	0.019	
log ₂ protein, in 6+/all teeth removed	0.32 (0.06–1.65)	0.173	
log ₂ protein, in none/5 or less teeth removed	8.21 (1.09–62.14)	0.042	
Education (low vs. high)	5.53 (1.10–27.87)	0.038	
Subset of 13 Larynx/Hypopharynx cancer cases vs. 39 controls:			0.894
log ₂ CD44, in ever smoke	0.47 (0.12–1.91)	0.296	
log ₂ CD44, in never smoke	349.02 (0.79–>999)	0.059	
log ₂ protein, in 6+/all teeth removed	0.54 (0.09–3.30)	0.503	
log ₂ protein, in none/5 or less teeth removed	3.58 (0.33–39.05)	0.296	
Education (low vs. high)	15.29 (1.79–131.00)	0.013	
Subset of 11 stage I/II cancer cases vs. 39 controls:			0.846
log ₂ CD44, in ever smoke	0.31 (0.07–1.32)	0.112	
log ₂ CD44, in never smoke	11.01 (0.52–235.9)	0.125	
log ₂ protein, in 6+/all teeth removed	0.58 (0.07–4.80)	0.614	
log ₂ protein, in none/5 or less teeth removed	11.12 (0.64–192.0)	0.097	
Education (low vs. high)	2.38 (0.23–24.71)	0.013	
Subset of 29 stage III/IV cancer cases vs. 39 controls:			0.874
log ₂ CD44, in ever smoke	1.78 (0.53–6.00)	0.354	
log ₂ CD44, in never smoke	47.73 (2.17–1037.2)	0.014	
log ₂ protein, in 6+/all teeth removed	0.36 (0.07–1.73)	0.200	
log ₂ protein, in none/5 or less teeth removed	3.35 (0.51–21.92)	0.207	
Education (low vs. high)	9.32 (2.02–43.08)	0.004	

^aFitted model: $\text{Log}(p_{\text{case}}/p_{\text{control}}) = \text{intercept} + \log_2 \text{CD44} + \log_2 \text{protein} + \text{smoke} + \text{teeth removed} + \log_2 \text{CD44} \times \text{smoke} + \log_2 \text{protein} \times \text{teeth removed}$.

^bAUC: area under ROC (receiver operating curve).