Gene expression in uninvolved oral mucosa of OSCC patients facilitates identification of markers predictive of OSCC outcomes

Pawadee Lohavanichbutr

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2012

Committee:

Chu Chen

Stephen M. Schwartz

Program Authorized to Offer Degree:

Epidemiology

University of Washington

Abstract

Gene expression in uninvolved oral mucosa of OSCC patients facilitates identification of markers predictive of OSCC outcomes

Pawadee Lohavanichbutr

Chair of the Supervisory Committee:
Chu Chen, PhD
Epidemiology

Oral and oropharyngeal squamous cell carcinomas (OSCC) are among the most common cancers worldwide with approximately 60% 5-yr survival rate. To identify potential markers for disease progression, we used Affymetrix U133 plus 2.0 arrays to examine the gene expression profiles of 167 primary tumor samples from OSCC patients, 58 uninvolved oral mucosa from OSCC patients and 45 normal oral mucosa from patients without oral cancer, all of whom were enrolled at one of the three University of Washington-affiliated medical centers from 2003 to 2008. We found 2,596 probe sets differentially expressed between 167 tumor samples and 45 normal samples. Among 2,596 probe sets, 71 were significantly and consistently up- or down-regulated in the comparison between normal samples and uninvolved oral samples and between uninvolved oral samples and tumor samples. Cox regression analyses showed that 20 of the 71 probe sets were significantly associated with progression-free survival. The risk score for each patient was calculated from coefficients of a Cox model incorporating these 20 probe sets. The hazard ratio (HR) associated with each unit change in the risk score adjusting for age, gender, tumor stage, and high-risk HPV status was 2.7 (95% CI: 2.0-3.8, p = 8.8E-10). The risk scores in an independent dataset of 74 OSCC patients from the MD Anderson Cancer Center was also significantly associated with progression-free survival independent of age, gender, and tumor stage (HR 1.6, 95% CI: 1.1 – 2.2, p = 0.008). Gene Set Enrichment Analysis showed that the most prominent biological pathway represented by the 71 probe sets was integrin cell surface interactions pathway. In conclusion, we identified 71 probe sets in which dysregulation occurred in both uninvolved oral samples and cancer samples. Dysregulation of 20 of the 71 probe sets was associated with progression-free survival and was validated in an independent dataset.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Number Page

INTRODUCTION

Oral and oropharyngeal squamous cell carcinoma (OSCC) is among the most common cancers with an estimated nearly 400,000 new cases and approximately 200,000 deaths in 2008 worldwide (http://www-dep.iarc.fr/). Approximately 40,000 new cases and almost 8,000 deaths from OSCC are estimated to occur in the United States in 2012.[1] The overall 5-yr survival rate of OSCC patients is approximately 60%.[1] The prognosis of OSCC patients is adversely influenced by the development of recurrent cancer, which occurred between 5-50% of patients.[2-4] A need exists to predict which patients are most at risk for recurrence or disease progression. Several factors have been found to be predictive of the development of recurrent OSCC, including tumor stage, tumor depth, nodal status, lymphovascular or perineural invasion, positive surgical margins, and extracapsular spread.[5-8] However, further improvement in the prediction of risk for recurrence or disease progression could help physicians identify patients who need more aggressive treatment or frequent follow-up. Genes that play roles in the progression of normal tissue to cancer may serve as markers to predict recurrence or disease progression of OSCC patients. Based on the field cancerization concept proposed by Slaughter et al in 1953[9], the changes in the mucosa of the entire upper aerodigestive tract may be the result of long term exposure to carcinogens and may explain the occurrence of local recurrence or second primary disease. Identifying these changes may help advance our understanding of the disease progression process and lead to discovery of markers to predict disease progression. The purpose of the current study is to identify genes that are deregulated in uninvolved oral mucosa from OSCC patients compared with normal oral mucosa from non-cancerous patients, and show further dysregulation in cancer tissue. We believe that these genes may play an important role in the progression of OSCC and we tested our

hypothesis by determining whether the expressions of deregulated genes are associated with disease progression or OSCC-specific mortality.

METHODS

Study Population

Eligible cases are patients with first primary OSCC treated at one of the three University of Washington-affiliated medical centers in Seattle, WA from December 2003 to March 2010. Eligible controls are patients without OSCC who had oral surgery, such as tonsillectomy or uvulopalatopharyngoplasty, at the same institutions and during the same time period in which the OSCC cases were treated.

Data and Tissue Collection

Each patient was interviewed using a structured questionnaire regarding his or her demographic, medical, and lifestyle history, including tobacco and alcohol use. Data on tumor characteristics were obtained from medical records. The data on tumor recurrence were obtained from telephone interview and confirmed by medical record abstraction if patients reported having a tumor recurrence. If patients were not followed at one of the three University of Washington-affiliated medical centers, we attempted to obtain medical record from their physicians. Vital status was obtained from Social Security Death Index (SSDI) and Fred Hutchinson Cancer Research Center's Cancer Surveillance System (CSS), which is part of the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute. Death certificates and medical records, if available, were reviewed by otolaryngologists to determine the cause of death. The last search for vital status of all patients was in September 2011.

The tumor samples and uninvolved oral mucosa were obtained from OSCC patients at the time of resection prior to chemo/radiation therapy, if any. The uninvolved oral mucosa was collected either from the opposite side of the tumor or from the same side but far from the tumor margin. From controls we obtained normal mucosa from

buccal, uvula or anterior tonsillar pillar, the latter with effort to avoid surrounding lymphoid tissues. The tissue samples were soaked in RNAlater™ immediately after surgical removal and transferred to long term storage at – 80 ° C prior to use. After September 2008, we no longer treated samples with RNAlater™. Instead, the tissue samples were flash frozen in liquid nitrogen immediately after surgical removal.

From December 2003 to March 2010, we recruited 291 cases and 58 controls. Gene expression data from tumor samples of 167 cases and normal oral mucosa of 45 controls were generated in our previous study [10] using samples treated with RNAlater™. For comparability, the selection of uninvolved oral mucosa in the current study was limited to RNAlater™-treated samples. To maximize the variation of gene expression among the uninvolved oral samples, we included uninvolved oral samples from all patients who had recurrence/second primary OSCC as of March 2010 (n=29). We then used stratified sampling to select another 29 OSCC patients with a similar follow-up time distribution as the 29 recurrence/second primary cases. Forty-nine of the 58 selected OSCC patients also provided tumor samples that had already been processed in the previous studied (part of the 167 tumor samples). This research was conducted with written informed consent and institutional review board approval.

Laboratory Methods

The DNA and RNA from each specimen were simultaneously extracted using the TRIzol method (Invitrogen, Carlsbad, CA). RNA was further purified using RNeasy mini kit (Qiagen, Valencia, California) and then converted to double-stranded complementary DNA (cDNA) using a GeneChip Expression 3'-Amplification One-cycle DNA Synthesis Kit (Affymetrix). The cRNA was produced from cDNA and was hybridized to a U133 2.0 Plus GeneChip (Affymetrix) as previously described [10]. HPV DNA was tested using a nested

PCR based protocol and confirmed by LINEAR ARRAY HPV Genotyping Test (Roche, Indianapolis, IN) under a research use only agreement as described in Lohavanichbutr et al. [11].

Quality Control

For quality control, we re-extracted and processed two tumor samples, whose genome-wide gene expression had been assessed in our previous study, along with the uninvolved oral samples. We used Pearson correlation to determine whether the previous and new gene expression were comparable. We found a good correlation between samples previously processed and samples processed along with uninvolved oral samples. The Pearson's correlation coefficients for all probe sets of the two pair were 0.96 and 0.97.

The quality of the hybridized arrays was evaluated using the "affyQCReport" and "affyPLM" software in the Bioconductor package (http://bioconductor.org/). This included evaluation of RNA degradation and detection for possible outlier array. We examined 58 arrays of uninvolved oral samples separately and also together with 212 arrays (167 from tumor samples and 45 from normal oral samples) previously processed in order to detect a batch effect. All 58 arrays for uninvolved oral samples passed quality control and no batch effect was observed.

Statistical Analyses

Assessment of Differential Gene Expression

All 260 CEL files were normalized using the RMA algorithm in Partek® Genomics Suite™ software. We first identified "OSCC-related genes" by comparing gene expression profiles of normal oral samples from 45 controls to tumor samples from 167 OSCC cases using ANOVA implemented in Partek® Genomics Suite™ software, adjusting for age

(continuous variable), sex (male vs. female), cigarette-smoking (current smoker vs. never/former smoker), alcohol use (current vs. never/former alcohol use) and HPV status (high risk vs. negative/low risk). We set the false discovery rate at 0.05 and required at least a 2-fold difference in gene expression as criteria for differential expression. The purpose of this first step is to reduce the number of genes for further comparison. The next step is to identify genes, among the "OSCC-related genes", that show dysregulation in a field of carcinogenic exposure (uninvolved oral mucosa) and increased level of dysregulation in cancer stage. To identify these genes, we used linear regression to compare the gene expression level between 45 normal oral samples and 58 uninvolved oral samples, and compare the gene expression level of 58 uninvolved oral samples to that of 167 tumor samples. We used three criteria to select the gene list: 1) the Bonferroni adjusted p-value must be less than 0.05 in both comparisons; 2) the magnitude of the difference in expression level must be greater than one standard deviation of the expression in the uninvolved oral samples; 3) the direction of the coefficients of each gene must be the same in both comparisons, i.e. the coefficients must be positive in both comparisons for up-regulated genes, and must be negative for both comparisons for down-regulated genes. The analyses were performed using STATA 11.1 (StataCorp, College Station, TX).

Evaluation of Gene Expression Profile in relation to Disease Outcome

To determine whether the selected genes are associated with disease progression or death due to OSCC, we performed a Cox regression with robust standard error on each selected gene adjusting for age, gender, high-risk HPV status, and tumor stage (stage I/II vs. stage III/IV). In this study, the disease progression is defined as a persistence or recurrence of squamous cell carcinoma in oral cavity, oropharynx, or in head and neck

area. For patients who were alive as of September 2011 or patients who died with other causes, they were censored at time of last known disease status either at the last follow-up interview or at the last clinic visit. We used a Bonferroni adjusted p-value of 0.05 as a criterion to select genes that are associated with disease progression/OSCC-related death. We then built a Cox regression model with the genes associated with disease progression/OSCC-related death and used coefficients from this model to calculate a risk score for each patient.

Validation using External Dataset

An independent dataset of 74 frozen tumor samples from OSCC patients treated at the MD Anderson Cancer Center was used for validation. The 74 tumor samples were hybridized at the MD Anderson Cancer Center to the same type of Affymetrix array as used in our study. We normalized the CEL files using RMA algorithm in Partek® Genomics Suite™ software. A risk score for each patient was calculated using coefficients from a Cox regression model from our study. We then investigated the association between risk score and disease progression/OSCC-related death using a Cox regression analysis adjusting for age (continuous variable), gender (male vs. female), and tumor stage (I/II vs, III/IV). We compared the model with tumor stage alone and tumor stage plus risk score using a log likelihood ratio test. The patients were divided into three equal size groups based on the risk scores (low, medium, and high). We then used a Kaplan-Meier method to compare progression free survival for patients in each group.

Pathway analyses

We used Gene Set Enrichment Analysis (GSEA) [12] to investigate pathways of the genes deregulated in uninvolved oral samples and tumor samples. GSEA compute overlaps between genes of interest and the gene sets in the Molecular Signatures

Database (MSigDB). The gene sets of the pathways in the MSigDB are derived from three pathway database: the Biocarta pathway database (www.biocarta.com), KEGG pathway database (www.genome.jp/keg), and Reactome pathway database (www.reactome.org). We also used GSEA to compare genes that we found in this study to the 131 genes that we previously reported to be associated with survival of OSCC patients. [13]

RESULTS

Selected characteristics of the study participants are showed in Table 1. Compared to controls, OSCC patients tended to be older, more likely to be white and current smoker. Approximately two-thirds of the cases had an advanced stage tumor.

Gene Selection

In the ANOVA analysis to identify “OSCC-related genes”, we found 2,596 probe sets differentially expressed between 167 tumor samples and 45 normal oral samples form controls, using the criteria of a FDR of 0.05 and at least a two-fold difference in the expression level. The result of linear regression comparing gene expression level of 2,596 probe sets between 45 normal oral samples from controls and 58 uninvolved oral samples from OSCC cases, and between 58 uninvolved oral samples and 167 tumor samples (both from OSCC cases) showed that 60 probe sets were significantly and consistently up-regulated and 11 probe sets were significantly and consistently down-regulated in both comparisons, using the three criteria described in the Methods section. The list of the 71 probe sets is presented in Table 2.

Survival analyses

We excluded nine of 167 cases who died within 30 days of surgery (due to complication of surgery) or had been followed for less than 30 days. Among 158 patients included in the survival analyses, 70 had disease progression/OSCC-related death. The follow-up time for patients without progression/OSCC-related death ranged from 3.6 to 83.9 months with a median follow-up time of 43.3 months. The result of Cox regression analyses of each of the 71 probe sets adjusting for age, sex, tumor stage, and high-risk HPV status showed 20 of 71 probe sets significantly associated with disease progression/OSCC-related death with a p-value < 0.0007 (Table 3). We then built a

prediction model based on the Cox regression model incorporating the 20 probe sets. A coefficient of each probe set (Table 3) was multiplied with the expression of that probe set and summed up to be a risk score for each patient. The risk score ranged from 11.0 to 18.6 (mean 14.9, standard deviation 1.2). In our study, one unit higher in risk score was associated with 2.7 time higher in the risk of disease progression/OSCC-related death after adjusting for age, gender, tumor stage, and high-risk HPV status (95% CI: 2.0 – 3.8, p-value < 0.001).

Analyses of 20 probe sets in an independent dataset

Among 74 OSCC patients from MD Anderson Cancer Center, five patients had follow-up time less than 30 days and were excluded from the survival analyses. The age range of the patients was 22 to 84 years with an average age of 58.2 years. The majority of patients had stage III or IV disease (73.9%). Twenty-five of 69 patients had disease progression/OSCC-related death. The follow-up time for 69 patients ranged from one month to 92.7 months. The median follow-up time for patients without events was 22.7 months (range 1.6 to 92.7 months). A risk score for each patient was calculated using coefficients of the prediction model from the University of Washington data and the expression values from each MD Anderson Cancer Center patients as described above. The risk score ranged from 14.8 to 20.4 (mean 17.8, standard deviation 1.4). The crude hazard ratio (HR) for each unit increase in the risk score was 1.63 (95% CI: 1.16 – 2.29, p-value 0.004). The hazard ratio associated with a risk score after adjusting for age, gender, and tumor stage was 1.59 (95% CI: 1.13 – 2.23, p-value 0.008). The hazard ratio for each variable in the model is shown in Table 4. Data on HPV status were not available for the MD Anderson dataset, thus was not adjusted for. Higher tumor stage (stage III/IV) was associated with higher risk of disease progression/OSCC-specific mortality; however, it did

not reach statistically significance in the MD Anderson dataset (crude HR 2.3, 95% CI: 0.79 – 6.75, p-value 0.13, HR adjusted for age, gender, and risk score 1.65, 95%CI: 0.53 – 5.19, p-value 0.39). The prediction model incorporating the risk score and tumor stage provided a better fit to the data than the model with tumor stage alone (log likelihood ratio test p-value = 0.006); however, it was not better than the model with risk score alone (log likelihood ratio test p-value = 0.3). The HR adjusted for age, gender, and tumor stage of patients with medium risk score and high risk score compare to patients with low risk score was 1.79 (95% CI: 0.54 – 5.91, p-value 0.34) and 3.67 (95% CI:1.2 – 11.2, p-value 0.02), respectively. Kaplan-Meier curves provided additional illustration for a progression-free survival of patients in each group (Figure 1). Patients with high risk score had poorer progression free survival than patients with medium and low risk score with a Log-rank p-value of 0.019.

Pathway analyses

Results of GSEA of the 71 probe sets show that the most prominent biologic pathway belongs to the integrin cell surface interactions pathway. The complete list of the pathways is presented in Table 5. When compared the 71 probe sets to the 131 probe sets that we previously reported to be associated with survival of OSCC patients,[13] we found eight genes (*KLF7, OSMR, PDPN, PADI1, CLEC3B, COL7A1, COL27A1*, and *NETO2*) overlapped between the two gene lists. GSEA showed 10 pathways common to both gene lists (Table 6).

DISCUSSION

OSCC has a high mortality rate, with a 5-year survival rate of 30-50% for late stage cancer (www.cancer.org). Fortunately, the 5-year survival rate exceeds 80% in early stage cancer. Thus early detection or prevention of disease progression may help improve survival of OSCC patients. Identifying the key genes that play an important role in the progression of the carcinogenesis process may have potential clinical implications, e.g. as targets for OSCC prevention or treatment, or as biomarkers for early detection or prediction of disease progression. Our study is unique in that we collected not only normal oral mucosa from controls and tumor samples from OSCC, but also collected uninvolved oral samples from OSCC patients. This design provided an opportunity to study the effect of field cancerization by comparing gene expression between normal oral mucosa from controls and uninvolved oral mucosa from OSCC patients which may help identify genes that play a role in a very early stage of carcinogenesis. In addition, by comparing gene expression of uninvolved oral samples to that of tumor samples, we can further select genes that not only play a role in an early stage but also play a role in the later stage. We believe that the genes found through both of these comparisons are important in the progression of normal mucosa to cancer. A limitation of this study is that the uninvolved oral samples were not processed at the same time as normal oral samples and tumor samples thus a batch effect is a potential issue to consider. We attempted to investigate and minimize the batch effect by: 1) Re-processing some tumor samples along with uninvolved oral samples and comparing the gene expression to that of the previously processed tumor samples from the same patients. The results showed a good correlation with correlation coefficients for all probe sets of 0.96-0.97; 2) Examining the quality of all arrays together to detect batch effects, and then normalizing all arrays together; and 3)

Performing a first step analysis to compare gene expression between tumor samples and normal oral samples from controls. The benefit from this first step is to minimize the penalty from multiple comparisons by reducing the number of genes from more than 50,000 genes to approximately 2,500 “OSCC-related” genes for further comparison. Since tumor samples and normal oral samples were processed at the same period of time with each batch containing both tumor sample and normal oral sample, these 2,596 probe sets were unlikely to be affected by batch effect.

We identified 71 probe sets that showed dysregulation in the uninvolved oral samples and showed even higher level of dyregulation in tumor samples. As mentioned earlier, one potential clinical implication of these genes is to predict disease progression. Thus we further tested whether some of the 71 probe sets were associated with disease progression or death due to OSCC and built a prediction model based on these genes. The result that 20 genes were associated with disease progression/OSCC-related death and can be used to predict disease progression independent of age, sex, tumor staging, and high-risk HPV status support our hypothesis.

We validated our results using data from 74 OSCC patients recruited at the MD Anderson Cancer Center. We found a significant association between risk score calculated from the prediction model based on the Cox model incorporating 20 probe sets and disease progression/OSCC-related death in the MD Anderson dataset. This association was independent of age, gender, and tumor stage. Moreover, the prediction model with risk score plus stage was better than the model with stage alone suggesting that adding risk score to tumor stage improve the prediction of disease progression or OSCC-related death. To the best of our knowledge, this is the first gene signature using microarray data to predict disease progression/OSCC-related death that has been

validated in an independent dataset from a different institution. One limitation in the MD Anderson dataset is the lack of information on HPV status. Patients with HPV-positive tumors are more likely to have better survival, and HPV-positive tumors are more commonly found in the oropharynx[14,15]. Among the 69 MD Anderson Cancer Center tumor samples, only three tumors were from the oropharynx. Thus, it is unlikely that HPV status would confound the association between risk scores and disease progression/OSCC-related death in the MD Anderson dataset. The fact that the tumor samples from the MD Anderson Cancer Center were frozen samples suggested that the use of this risk score is not limited to the RNAlater™-treated samples only.

Another potential use of the 20 or the 71 probe sets is to predict which premalignant lesions are likely to progress to cancer. Future study is needed to address this potential. In addition to prediction of disease progression, some of the 71 probe sets may serve as targets for detection or treatment of OSCC. For instance, SART2 (squamous cell carcinoma antigen recognized by T-cells 2) protein was found to be over expressed in several types of cancer but not in normal cells.[16] One potential study would be to investigate the level of SART2 protein in saliva or oral rinse to determine whether it could help improve detection of OSCC, especially those that are located in the areas that are difficult to visualize. Since SART2 is a tumor antigen recognized by cytotoxic T cell, it could potentially serve as a target for cancer immunotherapy as well. SART2-derived peptide has been shown to be immunogenic in hepatocellular carcinoma patients.[17] Further investigation is needed to explore the potential use of SART2 as a tumor marker or as a target for cancer immunotherapy for OSCC patients. Another gene that has been investigated as a potential anti-cancer target is Integrin α3β1.[18] Genes in the Integrin family have functional roles in migration/invasion of tumor cells [19-21]. Among the 71 probe

sets, four were genes in the Integrin family (*ITGA3, ITGAV, ITGB6*, and *ITGA6*). The most prominent pathway for the 71 probe sets is integrin cell surface interaction pathway. Our results lend support to the important role of integrins in cancer.

Previously, we reported that a 131 gene expression signature provided high discrimination between OSCC samples and normal oral mucosa from controls, and was associated with OSCC-specific mortality.[13] Most of the 131 probe sets were in the list of 2,596 probe sets differentially expressed between tumor samples and normal oral samples in the current analyses. However, the difference in selection criteria provided different gene lists. The 131 probe sets were selected based on the most significant difference in gene expression between tumor and normal samples but the 71 probe sets were selected by emphasizing their potential involvement in both early and late stage of carcinogenesis. Although not many genes overlapped between 131 and 71 gene lists, there were ten pathways common to both gene lists suggesting that both signatures were correlated in the molecular pathway level.

CONCLUSION

We found dysregulation of gene expression of 71 probe sets, corresponding to 61 known genes, occurring early in uninvolved oral mucosa from OSCC patients, and the level of dysregulation was even higher in tumor samples. Dysregulation in the expression of 20 of the 71 probe sets was associated with disease progression or death due to OSCC in OSCC patients. The result was validated in an independent dataset from the MD Anderson Cancer Center. If further confirmed in future studies, the expression of these genes has the potential to be developed into a clinical test. Such a test could help physicians to identify patients who need more aggressive treatment or frequent follow-up.

REFERENCES

1. Siegel R, Naishadham D and Jemal A. Cancer statistics, 2012. CA Cancer J Clin, 2012;62:10-29.

2. Gonzalez-Garcia R, Naval-Gias L, Roman-Romero L, Sastre-Perez J and Rodriguez-Campo FJ. Local recurrences and second primary tumors from squamous cell carcinoma of the oral cavity: a retrospective analytic study of 500 patients. Head Neck, 2009;31:1168-80.

3. Hockel M and Dornhofer N. The hydra phenomenon of cancer: why tumors recur locally after microscopically complete resection. Cancer Res, 2005;65:2997-3002.

4. Mucke T, Wagenpfeil S, Kesting MR, Holzle F and Wolff K. Recurrence interval affects survival after local relapse of oral cancer. Oral Oncol, 2009;45:687-91.

5. Mishra RC, Parida G, Mishra TK and Mohanty S. Tumour thickness and relationship to locoregional failure in cancer of the buccal mucosa. Eur J Surg Oncol, 1999;25:186-9.

6. Larsen SR, Johansen J, Sorensen JA and Krogdahl A. The prognostic significance of histological features in oral squamous cell carcinoma. J Oral Pathol Med, 2009;38:657-62.

7. Woolgar JA, Rogers S, West CR, Errington RD, Brown JS and Vaughan ED. Survival and patterns of recurrence in 200 oral cancer patients treated by radical surgery and neck dissection. Oral Oncol, 1999;35:257-65.

8. Brandwein-Gensler M, Teixeira MS, Lewis CM, Lee B, Rolnitzky L, Hille JJ, Genden E, Urken ML and Wang BY. Oral squamous cell carcinoma: histologic risk assessment, but not margin status, is strongly predictive of local disease-free and overall survival. Am J Surg Pathol, 2005;29:167-78.

9. Slaughter DP, Southwick HW and Smejkal W. "Field Cancerization" in Oral Stratified Squamous Epithelium: Clinical Implications of Multicentric Origin. Cancer, 1953;6:963-8.

10. Chen C, Mendez E, Houck J, Fan W, Lohavanichbutr P, Doody D, Yueh B, Futran ND, Upton M, Farwell DG, Schwartz SM and Zhao LP. Gene expression profiling identifies genes predictive of oral squamous cell carcinoma. Cancer Epidemiol Biomarkers Prev, 2008;17:2152-62.

11. Lohavanichbutr P, Houck J, Fan W, Yueh B, Mendez E, Futran N, Doody DR, Upton MP, Farwell DG, Schwartz SM, Zhao LP and Chen C. Genomewide gene expression profiles of HPV-positive and HPV-negative oropharyngeal cancer: potential implications for treatment choices. Arch Otolaryngol Head Neck Surg, 2009;135:180-8.

12. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA,

Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 2005;102:15545-50.

13. Mendez E, Houck JR, Doody DR, Fan W, Lohavanichbutr P, Rue TC, Yueh B, Futran ND, Upton MP, Farwell DG, Heagerty PJ, Zhao LP, Schwartz SM and Chen C. A genetic expression profile associated with oral cancer identifies a group of patients at high risk of poor survival. Clin Cancer Res, 2009;15:1353-61.

14. Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, Zahurak ML, Daniel RW, Viglione M, Symer DE, Shah KV and Sidransky D. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer Inst, 2000;92:709-20.

15. Schwartz SR, Yueh B, McDougall JK, Daling JR and Schwartz SM. Human papillomavirus infection and survival in oral squamous cell carcinoma: a population based study. Otolaryngol Head Neck Surg, 2001;125:1-9.

16. Nakao M, Shichijo S, Imaizumi T, Inoue Y, Matsunaga K, Yamada A, Kikuchi M, Tsuda N, Ohta K, Takamori S, Yamana H, Fujita H and Itoh K. Identification of a gene coding for a new squamous cell carcinoma antigen recognized by the CTL. J Immunol, 2000;164:2565-74.

17. Mizukoshi E, Nakamoto Y, Arai K, Yamashita T, Sakai A, Sakai Y, Kagaya T, Yamashita T, Honda M and Kaneko S. Comparative analysis of various tumor-associated antigen-specific t-cell responses in patients with hepatocellular carcinoma. Hepatology, 2011;53:1206-16.

18. Subbaram S and Dipersio CM. Integrin alpha3beta1 as a breast cancer target. Expert Opinion on Therapeutic Targets, 2011;15:1197-210.

19. Lee CY, Huang CY, Chen MY, Lin CY, Hsu HC and Tang CH. IL-8 increases integrin expression and cell motility in human chondrosarcoma cells. J Cell Biochem, 2011;112:2549-57.

20. Wang Y, Shenouda S, Baranwal S, Rathinam R, Jain P, Bao L, Hazari S, Dash S and Alahari SK. Integrin subunits alpha5 and alpha6 regulate cell cycle by modulating the chk1 and Rb/E2F pathways to affect breast cancer metastasis. Molecular Cancer, 2011;10:84.

21. Morgan MR, Jazayeri M, Ramsay AG, Thomas GJ, Boulanger MJ, Hart IR and Marshall JF. Psoriasin (S100A7) associates with integrin beta6 subunit and is required for alphavbeta6-dependent carcinoma cell invasion. Oncogene, 2011;30:1422-35.

Figure 1. Kaplan-Meier curves comparing progression free survival of 69 OSCC patients in the MD Anderson dataset with low, medium, and high risk score.

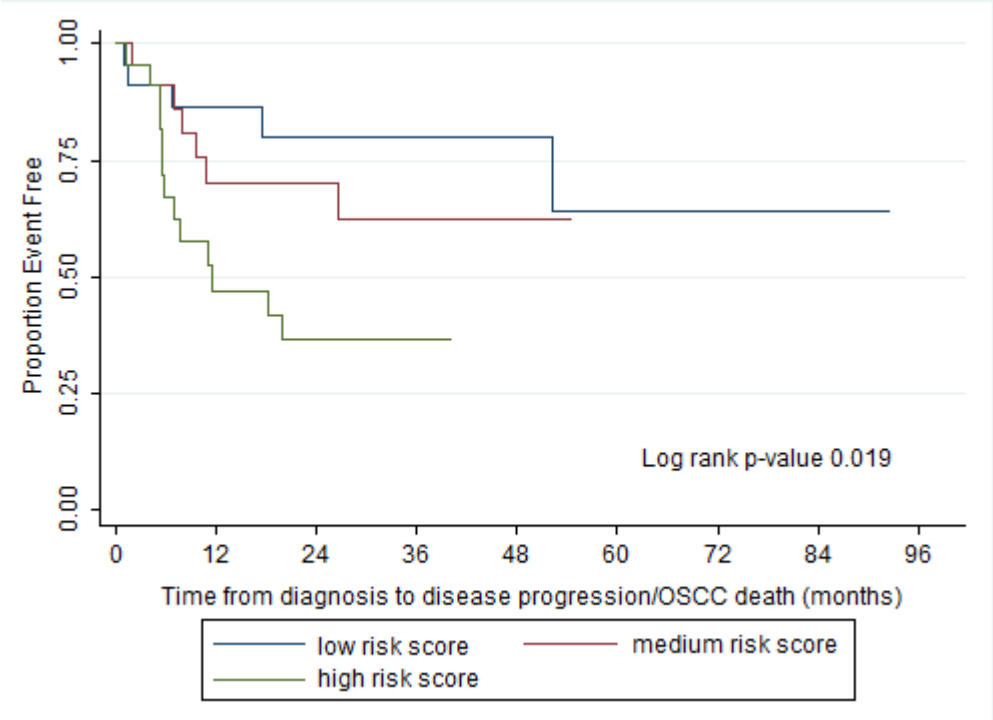

Table 1. Selected characteristics of OSCC patients by sample type and controls, University of Washington Affiliated Medical Centers, 2003-2010

	OSCC				Control	
	Tumor sample (n=167*)		Uninvolved oral sample (n=58)		Normal oral sample (n=45)	
Characteristic	n	%	n	%	n	%
Age						
19-39	7	4.2	4	6.9	17	37.8
40-49	26	15.6	10	17.3	14	31.1
50-59	57	34.1	18	31.0	5	11.1
60-90	77	46.1	26	44.8	9	20.0
Gender						
Male	120	71.9	41	70.7	32	71.1
Female	47	28.1	17	29.3	13	28.9
Race						
White	152	91.0	54	93.1	31	68.9
Non-white	15	9.0	4	6.9	14	31.1
Smoking history						
Never/Former	86	51.5	37	63.8	33	73.3
Current	81	48.5	21	36.2	12	26.7
Alcohol use						
Never/Former	55	33.5	23	39.7	11	25.0
Current	109	66.5	35	60.3	33	75.0
Unknown	3				1	
AJCC staging						
I	39	23.3	15	26.3		
II	16	9.6	9	15.8		
III	22	13.2	8	14.0		
IV	90	53.9	25	43.9		
Unknown			1			

*49 of 167 OSCC patients provided both tumor tissue and uninvolved oral tissue

Table 2. Seventy-one probe sets dysregulated in uninvolved oral samples and tumor samples of OSCC patients comparing to normal oral mucosa from non-cancerous patients, University of Washington Affiliated Medical Centers, 2003-2010

Probe set ID	Gene Symbol	Gene Title	Fold difference (tumor vs. normal)
231867_at	ODZ2	odz, odd Oz/ten-m homolog 2 (Drosophila)	7.2
221898_at	PDPN	podoplanin	6.3
213110_s_at	COL4A5	collagen, type IV, alpha 5	5.9
217820_s_at	ENAH	enabled homolog (Drosophila)	5.1
225105_at	OCC1	overexpressed in colon carcinoma-1	5.1
225288_at	COL27A1	collagen, type XXVII, alpha 1	4.9
226535_at	ITGB6	integrin, beta 6	4.7
218717_s_at	LEPREL1	leprecan-like 1	4.4
226448_at	FAM89A	family with sequence similarity 89, member A	4.4
201505_at	LAMB1	laminin, beta 1	4.3
235683_at	SESN3	sestrin 3	4.1
201250_s_at	SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1	4.0
213139_at	SNAI2	snail homolog 2 (Drosophila)	4.0
211651_s_at	LAMB1	laminin, beta 1	3.7
218888_s_at	NETO2	neuropilin (NRP) and tolloid (TLL)-like 2	3.7
205122_at	TMEFF1	transmembrane protein with EGF-like and two follistatin-like domains 1	3.7
222774_s_at	NETO2	neuropilin (NRP) and tolloid (TLL)-like 2	3.6
1554018_at	GPNMB	glycoprotein (transmembrane) nmb	3.5
217312_s_at	COL7A1	collagen, type VII, alpha 1	3.5
201474_s_at	ITGA3	integrin, alpha 3 (antigen CD49C, alpha 3 subunit of VLA-3 receptor)	3.4
201656_at	ITGA6	integrin, alpha 6	3.4
225258_at	FBLIM1	filamin binding LIM protein 1	3.3
218847_at	IGF2BP2	insulin-like growth factor 2 mRNA binding protein 2	3.3
206581_at	BNC1	basonuclin 1	3.2
1552277_a_at	MSANTD3	Myb/SANT-like DNA-binding domain containing 3	3.2
204136_at	COL7A1	collagen, type VII, alpha 1	3.1
202351_at	ITGAV	integrin, alpha V (vitronectin receptor, alpha polypeptide, antigen CD51)	3.1
201976_s_at	MYO10	myosin X	3.1
221538_s_at	PLXNA1	plexin A1	2.9

Table 2 continued

Probe set ID	Gene Symbol	Gene Title	Fold difference (tumor vs. normal)
209935_at	ATP2C1	ATPase, Ca++ transporting, type 2C, member 1	2.9
205796_at	TCP11L1	t-complex 11 (mouse)-like 1	2.9
208636_at	ACTN1	actinin, alpha 1	2.9
1558152_at	LOC100131262	hypothetical LOC100131262	2.8
202599_s_at	NRIP1	nuclear receptor interacting protein 1	2.8
204334_at	KLF7	Kruppel-like factor 7 (ubiquitous)	2.8
201249_at	SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1	2.7
235492_at	RNF217	ring finger protein 217	2.5
225150_s_at	RTKN	rhotekin	2.5
212285_s_at	AGRN	agrin	2.4
202872_at	ATP6V1C1	ATPase, H+ transporting, lysosomal 42kDa, V1 subunit C1	2.3
1554008_at	OSMR	oncostatin M receptor	2.3
204068_at	STK3	serine/threonine kinase 3	2.3
218854_at	SART2	squamous cell carcinoma antigen recognized by T cells 2	2.3
202066_at	PPFIA1	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 1	2.2
209011_at	TRIO	triple functional domain (PTPRF interacting)	2.2
238933_at	IRS1	insulin receptor substrate 1	2.2
203935_at	ACVR1	activin A receptor, type I	2.2
1554795_a_at	FBLIM1	filamin binding LIM protein 1	2.2
1554016_a_at	C16orf57	chromosome 16 open reading frame 57	2.2
209934_s_at	ATP2C1	ATPase, Ca++ transporting, type 2C, member 1	2.2
202027_at	TMEM184B	transmembrane protein 184B	2.2
214853_s_at	SHC1	SHC (Src homology 2 domain containing) transforming protein 1	2.2
217788_s_at	GALNT2	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2 (GalNAc-T2)	2.2
212589_at	RRAS2	related RAS viral (r-ras) oncogene homolog 2	2.1
224747_at	UBE2Q2	ubiquitin-conjugating enzyme E2Q family member 2	2.0
228914_at	MSANTD3-TMEFF1	MSANTD3-TMEFF1 readthrough	2.0
202896_s_at	SIRPA	signal-regulatory protein alpha	2.0
224791_at	ASAP1	ArfGAP with SH3 domain, ankyrin repeat and PH domain 1	2.0
209081_s_at	COL18A1	collagen, type XVIII, alpha 1	2.0

Table 2 continued

Probe set ID	Gene Symbol	Gene Title	Fold difference (tumor vs. normal)
206335_at	GALNS	galactosamine (N-acetyl)-6-sulfate sulfatase	2.0
225795_at	C22orf32	chromosome 22 open reading frame 32	-2.1
222368_at	no gene symbol	EST384442 MAGE resequences, MAGL Homosapiens cDNA	-2.2
215913_s_at	GULP1	GULP, engulfment adaptor PTB domain containing 1	-2.2
234233_s_at	no gene symbol	cDNA FLJ20924 fis, clone ADSE00928	-2.5
220962_s_at	PADI1	peptidyl arginine deiminase, type I	-2.7
207057_at	SLC16A7	solute carrier family 16, member 7 (monocarboxylic acid transporter 2)	-3.1
206453_s_at	NDRG2	NDRG family member 2	-3.8
228335_at	CLDN11	claudin 11	-5.5
1569608_x_at	no gene symbol	clone IMAGE: 4720764	-5.8
210085_s_at	ANXA9	annexin A9	-7.1
205200_at	CLEC3B	C-type lectin domain family 3, member B	-7.3

Table 3. Gene expression of 20 probe sets associated with disease progression or death due to OSCC, University of Washington Affiliated Medical Centers, 2003-2010

Probe ID	Gene Symbol	HR*	95% CI		p-value	Coefficient in the Cox model**
225105_at	OCC1	1.9	1.5	2.4	4.00E-07	0.46727
218854_at	SART2	3.0	1.9	4.6	5.80E-07	0.22577
208636_at	ACTN1	2.3	1.6	3.3	1.90E-06	0.47392
212589_at	RRAS2	2.5	1.7	3.8	6.80E-06	0.61303
201474_s_at	ITGA3	2.1	1.5	2.9	6.90E-06	-0.10248
201976_s_at	MYO10	1.9	1.4	2.6	7.70E-06	0.16939
204334_at	KLF7	2.1	1.5	2.9	2.30E-05	0.59466
214853_s_at	SHC1	2.7	1.7	4.3	6.00E-05	0.59933
1552277_a_at	MSANTD3	2.2	1.5	3.3	6.10E-05	0.63538
202896_s_at	SIRPA	2.3	1.5	3.5	6.10E-05	-0.48287
225795_at	C22orf32	0.3	0.2	0.6	7.70E-05	-0.60779
202872_at	ATP6V1C1	2.6	1.6	4.2	9.00E-05	0.44730
205122_at	TMEFF1	1.5	1.2	1.8	1.00E-04	-0.16921
213139_at	SNAI2	2.1	1.4	3.0	1.00E-04	-0.59571
228914_at	MSANTD3-TMEFF1	2.2	1.4	3.3	2.30E-04	-0.29491
235492_at	RNF217	1.9	1.3	2.7	2.80E-04	-0.23651
221898_at	PDPN	1.5	1.2	1.9	2.90E-04	-0.19016
202599_s_at	NRIP1	1.9	1.3	2.7	3.00E-04	0.16372
206581_at	BNC1	1.7	1.3	2.3	4.80E-04	-0.25078
1558152_at	LOC100131262	1.5	1.2	2.0	6.10E-04	-0.15511

* Hazard ratio of each gene from Cox regression analysis adjusting for age, sex, tumor stage, and high-risk HPV status.
** Cox model incorporating 20 probe sets, used for calculating a risk score.

Table 4. Association of a risk score calculated from a 20 probe sets prediction model with disease progression or death due to OSCC among 69 OSCC patients in the MD Anderson dataset

Variable	HR*	SE	p-value	95% CI	
Risk score	1.59	0.28	0.008	1.13	2.23
Age	0.98	0.01	0.294	0.96	1.01
Gender (male vs female)	0.93	0.49	0.894	0.33	2.60
Stage (III/IV vs. I/II)	1.65	0.97	0.388	0.53	5.19

*hazard ratio from a multivariable Cox regression model including risk score, age, gender, and tumor stage

Table 5. Gene set enrichment analysis identified biologic pathways of the 71 probe set dysregulated in both uninvolved oral samples and cancer samples from OSCC patients, University of Washington Affiliated Medical Centers, 2003-2010.

Description of gene set pathway	K*	k**	p value^
Genes involved in Integrin cell surface interactions	81	5	6.39E-04
Genes involved in Cell surface interactions at the vascular wall	94	5	1.26E-03
Vitamin C in the Brain	11	2	3.87E-03
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	76	4	4.11E-03
Integrin Signaling Pathway	78	4	4.51E-03
ECM-receptor interaction	84	4	5.88E-03
Small cell lung cancer	84	4	5.88E-03
Genes involved in Cell junction organization	84	4	5.88E-03
Focal adhesion	201	6	7.56E-03
Genes involved in Cell-extracellular matrix interactions	16	2	8.20E-03
Genes involved in Basigin interactions	25	2	1.95E-02
Genes involved in Hemostasis	274	6	3.05E-02
Hypertrophic cardiomyopathy (HCM)	85	3	3.71E-02
Regulation of actin cytoskeleton	216	5	3.85E-02
Integrin Signaling Pathway	38	2	4.26E-02
Dilated cardiomyopathy	92	3	4.52E-02
Genes involved in Signaling in Immune system	366	6	9.58E-02
Cell adhesion molecules (CAMs)	134	3	1.10E-01
Adherens junction	75	2	1.38E-01
Pathways in cancer	328	5	1.54E-01
Genes involved in Glucose and other sugar SLC transporters	82	2	1.59E-01
Genes involved in Axon guidance	161	3	1.63E-01
Hematopoietic cell lineage	88	2	1.77E-01
Leukocyte transendothelial migration	118	2	2.73E-01
Tight junction	134	2	3.24E-01
Genes involved in SLC-mediated transmembrane transport	169	2	4.33E-01
Genes involved in Signalling by NGF	215	2	5.61E-01
Genes involved in Transmembrane transport of small molecules	218	2	5.69E-01
Cytokine-cytokine receptor interaction	267	2	6.80E-01

* A total number of genes in each gene set
** number of genes overlap with genes in 71 gene list.
^ p-value calculated based on the hypergeometirc distribution (identical to the corresponding one-tailed version of Fisher's exact test).

Table 6. Gene set enrichment analysis showed ten pathways common to both gene lists of the 71 probe sets and the 131 probe sets, identified by comparing gene expression among oral samples from OSCC patients and non-cancer patients, University of Washington Affiliated Medical Centers, 2003-2010.

Description of gene set pathway	Genes in the 71 gene list*	Genes in the 131 gene list**
Genes involved in Integrin cell surface interactions	ITGA6, ITGA3, ITGAV, COL4A5, LAMB1	COL4A1, COL1A1, COL1A2, THBS1,
Genes involved in Cell surface interactions at the vascular wall	ITGA6, ITGA3, ITGAV, SHC1, SIRPA	COL1A1, COL1A2, SLC16A1
ECM-receptor interaction	ITGA6, ITGA3, ITGAV, LAMB1	THBS1, COL4A1, LAMC2, COL1A1, COL1A2, COL5A1, COL5A2, TNXB
Small cell lung cancer	ITGA6, ITGA3, ITGAV, LAMB2	COL4A1, LAMC2
Genes involved in Cell junction organization	ITGA6, ACTN1, FBLIM1, CLDN11	LAMC2, CDH3
Focal adhesion	ITGA6, ITGA3, ITGAV, LAMB1, SHC1, ACTN1	THBS1, COL4A1, LAMC2, COL1A1, COL1A2, COL5A1, COL5A2, TNXB
Genes involved in Hemostasis	ITGA6, ITGA3, ITGAV, SHC1, SIRPA, ACTN1	COL1A1, COL1A2, THBS1, PLAUR SERPINE, TGFB1, SLC16A1,
Genes involved in Axon guidance	COL4A5, ENAH, PLXNA1	COL4A1, COL1A1, COL1A2, COL5A1, COL5A2, MYH11, TREM2
Leukocyte transendothelial migration	ACTN1, CLDN11	NCF2, THY1
Cytokine-cytokine receptor interaction	OSMR, ACVR1	IL1B, TGFB1, CCL4, CXCL9, CXCL2, CXCL3, OSMR,

***71 probe sets were deregulated in both uninvolved oral mucosa and cancer tissues of OSCC patients
**131 probe sets were previously reported to be differentially expressed between OSCC samples and normal oral mucosa

Acknowledgements

I wish to express sincere appreciation to all of those that I have worked with in the Chen laboratory, those that have contributed to the Oralchip study, and those from the MD Anderson Cancer Center for their support. The Oralchip study was supported by grants from the National Institutes of Health, National Cancer Institute (NIH NCI 095419), and by institutional funds from the Fred Hutchinson Cancer Research Center. The study at the MD Anderson Cancer Center was supported by Cancer Center Support Core Grant CA16672 (Affymetrix Microarray Core Facility; the Bioinformatics Core), Specialized Program of Research Excellence in Head and Neck Cancer Grant P50 CA97007 from the National Cancer Institute, “Clinician Investigator Program in Translational Research” K12 CA88084, NIH Loan Repayment Program, Clinical Research Program 2 L30 CA117652-02A1, and THANC Foundation Young Investigator Award in Head and Neck Cancer.