January 2015

The Changing Landscape Of Oral Cavity Cancer: Analysis Of Epidemiological And Genomic Data

Ryan Aronberg
Yale School of Medicine, ryan.aronberg@yale.edu

The Changing Landscape of Oral Cavity Cancer: Analysis of Epidemiological and Genomic Data

A Thesis Submitted to:

Yale University School of Medicine

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Medicine

Ryan Aronberg

2015

Abstract

Background: Oral cavity squamous cell carcinoma (OCSCC) has been reported to have stagnant survival rates over the last generation. This report represents the first population-based study with a rigorous subsite analysis of OCSCC. Recently, The Cancer Genome Atlas (TCGA) released a broad molecular characterization of Head and Neck Squamous Cell Carcinoma (HNSCC) – however comparative genomics has not yet been performed on individual oral cavity subsites.

Methods: The Surveillance, Epidemiology, and End Results (SEER) database (1988–2010) was used to examine 16,298 adult cases of OCSCC. Trends in tumor subsite, staging, patient demographics, treatment characteristics, and survival over time were examined. Subsequently, data from TCGA were used to evaluate mutation, copy number, and expression profiles of clinical subgroups of interest identified by epidemiological data.

Results: The overall incidence of OCSCC decreased between 1988 and 2007, but there was a marked increase in the incidence oral tongue squamous cell carcinoma (OTSCC). There were also trends towards oral tongue (OT) cancers being diagnosed in younger individuals and at earlier stages. Five-year overall survival of OCSCC increased between 1988 and 2007 (39.9% to 50.4%, $p < .01$), independent of changes in patient and tumor characteristics. Much of this survival increase was specifically attributable to increases in survival of OT cancers. Multivariate analysis revealed that age, stage, and grade were important covariates with survival, but oral subsite was not. Genomic analyses aimed at characterizing OT tumors higher rates of mutation in p53 and CDKN2A, and lower rates of mutation in most other genes. CASP8 mutations were found almost exclusively in non-tongue oral subsites. OT and oral cavity (OC) cancers, even in non-smokers, did not show the characteristic molecular changes associated with Human Papillomavirus (HPV) -related cancers, but instead closely resembled traditional smoking-related tumors. Clustering analysis revealed that OT tumors possess a distinct expression signature.

Conclusions: Survival for OCSCC has improved significantly over the past 20 years. Additionally, OTSCC now has a superior 5-year survival compared to other OC subsites; this can be attributed to trends towards earlier staging and younger population. The molecular profile of OTSCC, including tumors occurring in young, non-smokers, resemble that of traditional head and neck cancers (related to environmental carcinogens) – and is very different from HPV-related oropharyngeal cancers.

Acknowledgements

First, I would like to thank my friends and family for their support. To my father, without whom I would not be where I am, and who every day shows me the meaning of dedication, loyalty, and sacrifice. To my mother, step-father, and brothers, who always keep me smiling. And last but not least, to my wonderful, soon-to-be-wife Kate Davis for her tireless support and endless capacity to make me happy.

I would like to especially thank Dr. Wendell Yarbrough, MD, MMHC for his support, inspiration, and mentorship over the past two years. Not only did you help me to learn and eventually contribute in my first experience in cancer research, but you showed me the epitome of a physician scientist and leader – one who values each patient, colleague and student as highly as groundbreaking discoveries in the lab. I would also like to thank Natalia Isaeva, Sergey Ivanov, Tissa Manavalan, Eriko Kanaya, Alex Panaccione, Jake Wamsley, Gary Bellinger, Michael Otremba and Benjamin Judson for their and guidance, teaching, support and great memories over the past two years.

Thank you to Dr. John Forrest, Donna Carranzo, and Mae Geter from the Yale School of Medicine Office of Student Research, for your support of my research proposal and resulting conference presentations.

Research Funding Support Provided By: Richard K. Gershon, M.D. Student Research Fellowship, Yale University School of Medicine Medical Research Fellowship, National Institutes of Health-NHLBI Medical Student Research Fellowship

Table of Contents

Common Abbreviations

Term	Abbreviation
Oral Cavity	OC
Oral Tongue	OT
Oropharynx	OP
Larynx/Hypopharynx	LH
Human Papilloma Virus	HPV
Copy Number Alteration	CNA
Floor of mouth	FOM
Squamous Cell Carcinoma	SCC
Oral cavity squamous cell carcinoma	OCSCC
Oral tongue squamous cell carcinoma	OTSCC
Head and neck s squamous cell carcinoma	HNSCC
The Cancer Genome Atlas	TCGA
Surveillance, Epidemiology, and End Results Database	SEER
American Joint Committee on Cancer	AJCC

Introduction

Epidemiology of Head and Neck Cancer

Head and neck cancer is the 6th most common malignancy worldwide, a statistic commonly quoted in the literature [1]. The vast majority (greater than 90%) of cancer in the head and neck is squamous cell carcinoma (SCC), which is the sole focus of this investigation. Squamous cell carcinomas of the head and neck (HNSCC's), though commonly and conveniently classified as a single entity, are actually a heterogeneous group of diseases. Cancers affecting the sinonasal region, for example, are quite different in many respects from those affecting the larynx. However, they are often lumped together due to their anatomical proximity, occurrence on mucosal surfaces, and similar (but not identical) risk factors. Nevertheless, even adjacent areas of the aero-digestive tract can be affected in drastically different ways by cancer. Such is the case with cancers of the oral cavity and oropharynx.

Oral Cavity Cancer

The oral cavity is the most commonly affected site within HNSCC. In 2012, oral cavity cancers accounted for 26,000 new cases and 6,000 deaths in the US [2]. Oral cavity squamous cell carcinoma (OCSCC) has been associated with poor and stagnant survival, particularly for advanced stage disease [3, 4]. Traditionally, it has been closely associated with tobacco and alcohol exposure, and has predominantly affected an older population, representing a typical dose-response effect [5-7]. These environmental toxins cause direct damage to mucosal surfaces, and possess dozens of carcinogenic compounds, such as polycyclic aromatic hydrocarbons (PAH's), oxidating substances,

and free radicals. Following activation by endogenous metabolic enzymes (cytochrome p450 enzymes), these can form DNA adducts or induce epigenetic changes. If enough of these genetic defects accumulate and affect critical genes over time, carcinogenesis is the result.

Oropharyngeal Cancer and HPV

In stark contrast to most cancers of the head and neck, cancers of the oropharynx (OPSCC) are on a meteoric rise in incidence [8], and affect an entirely different population. This is because greater than 80% of OPSCCs are thought to be caused by the human papilloma virus (HPV) [9].Papillomaviruses are sexually transmitted, double stranded DNA viruses which infect proliferating keratinocytes in the basal layer of epithelial surfaces. First discovered in cervical cancers, HPV possesses the ability to integrate its DNA into the host genome, utilizing host machinery to produce proteins which are inherently oncogenic (especially HPV E6 and E7 proteins) due to their inhibition of powerful tumor suppressors, Rb and p53, respectively. In the normal viral lifecycle, intracellular viral replication and assembly increases as keratinocytes differentiate and mature, ultimately resulting in shedding of infectious particles as epithelial cells are sloughed off the outer surfaces of the epithelium. As would be expected, the known risk factors for HPV-related cancers are primarily sexual (number of sexual contacts, early initiation of sexual contact, and oral sex) [10, 11]. In addition, the cancers that arise in this region tend to affect a younger and predominantly male demographic with little exposure to classic risk factors such as tobacco. They also have distinct clinical characteristics, such as their predilection for the oropharynx, tendency for

neck metastasis, better response to chemo-radiotherapy, and overall better prognosis ([12]).

The Anatomy of the Oral Cavity and Oropharynx

The oral cavity and oropharynx are located immediately adjacent to one another, working together for the purpose of swallowing, phonation, and other functions. The oral cavity begins anteriorly at the skin-vermilion junction of the lips, and is bound laterally by the cheeks. It extends posteriorly to the boundary of the oropharynx, which is the hard-soft palate junction superiorly, and the terminal sulcus (line of the circumvallate papillae) on the tongue surface. Importantly, the adjacent oropharyngeal area includes the soft palate, the base of the tongue (the 1/3 of the tongue posterior to the terminal sulcus), and the palatine tonsils (Appendix A). The oral cavity proper can be broken into well-defined anatomical subsites, including the mucosal lip anteriorly, the buccal surfaces laterally, the hard palate superiorly, the floor of mouth inferiorly, the gingival surfaces, the retromolar trigone laterally, and the "oral" tongue (the anterior 2/3 of the tongue) (Appendix A). The clinical relevance of these different subsites is not clear – and they are not component of staging or management. Instead, OCSCC has a single staging protocol. Previous studies examining the prognostic significance of oral subsites, especially the oral tongue, have demonstrated conflicting results [13-18]. A significant difference between subsites would indicate that subsite should be considered during the staging or treating of cancers. A goal of this study was to clarify whether various oral subsites differed in their clinical characteristics or prognosis.

Oral Tongue Cancer: Evidence of a Changing Disease

The oral tongue is the subsite within the oral cavity responsible for the highest number of cancers. Interestingly, it appears the clinical characteristics of oral tongue squamous cell carcinoma (OTSCC) have changed over the last few decades [19, 20]. For example, the overall rates of HNSCC have steadily declined, due to decreases in tobacco and alcohol usage [21]. Over the same period, the incidence of OTSCC is on the rise. OTSCCs are also afflicting a greater proportion of non-smokers, women, and a younger population [22-25]. Although this shift in the epidemiology of OCSCC raises many possible hypotheses, none have been validated. One of the goals of our investigation was to further clarify the characteristics of this trend.

Based on the distinct differences in etiology and clinical characteristics, a clear demarcation between the oral cavity and oropharynx would seem essential. However, many previous database studies examining OCSCC have failed to appropriately distinguish it from the oropharynx. In the SEER database, the "Tongue" site code is misleadingly composed mostly (>50%) of cancers which are actually oropharyngeal in origin (e.g. the base of tongue and lingual tonsil). Several previous studies using the SEER database have utilized this "Tongue" category to represent oral cavity cancers, inadvertently misclassifying a significant proportion of oropharyngeal cancers into their data for the oral cavity [26, 27].

Management of OCSCC has evolved over the past two decades, with efforts aimed at early detection, advances in surgical and reconstructive treatments, an increased role for adjuvant chemotherapy with radiation, and advances in radiotherapy techniques. Additionally, there has been increased standardization of therapy, as well as a trend toward more patients being treated in experienced, high-volume academic centers. Given

these changes, it is important to evaluate our progress in the battle against OCSCC. Though some head and neck cancers have traditionally been associated with stagnant survival rates[28, 29], recent studies have shown that survival in oral cavity cancers may be improving [26, 27, 30]. However, given that many previous studies have inadvertently grouped oral cavity cancers with the oropharyngeal cancers, which are known to have higher survival rates, it is important to clarify this analysis using appropriate anatomical distinctions. Another aim of this investigation is to determine the survival trends related to OCSCC as well as its subsites, and to develop a multivariate model to determine the clinical and treatment characteristics which correlate with survival.

One hypothesis addressing the increased incidence and potential improved survival of OCSCC is that HPV has become responsible for a proportion of oral tongue cancers [26]. Common sense might justify this assertion, given that HPV is known to affect mucosa only millimeters away (in the oropharynx), and given that HPV related tumors affect a younger population and carry improved prognosis. However, this has apparently not been the case, as HPV has not been detected in a significant number of oral cavity cancers [31-33]. Additionally, recent studies examining the genomics of these oral tongue tumors have found that they do not carry the characteristic molecular trademarks of cancers caused by HPV (such as p16 expression and lack of p53 mutations) ([34] [9].).

Specific Aims

- ***Aim 1 – Determine if oral cavity subsites vary in their clinical characteristics or prognosis and identify important clinical factors correlating with survival.*** An improved understanding of the typical characteristics and outcomes of oral subsites will assist clinicians in treatment decision making and prognostication.

- ***Aim 2 – Identify changes in survival over time for cancers of the oral cavity/oral tongue.*** Understanding whether/how this disease is changing and correlating changes with changes in diagnostic characteristics and therapeutic approaches will provide data to support further studies to accelerate positive correlations.

- ***Aim 3 – Identify characteristics of the genomic landscape for OTSCC, and determine if it is significantly different from other cancers in the oral cavity/head and neck.*** Identifying molecular characteristics of OTSCC and comparing them to those of other oral subsites and HNSCC may provide insight into distinctions that correlate with origin, pathogenesis, or identify new targets for therapy.

Materials and Methods

Data Collection

Data were obtained from the Surveillance, Epidemiology, and End Results (SEER - 18 Registry dataset, November 2012) database of the National Cancer Institute. The SEER database contains epidemiological information on the incidence and survival of cancer in the United States, along with routinely collected clinical information for each patient. The SEER program collects and publishes this data from state-run cancer registries covering roughly 28% of the US population, which have been chosen to represent the ethnic makeup of the US as a whole. SEER coverage includes 26% of African Americans, 38% of Hispanics, 44% of American Indians and Alaska Natives, 50% of Asians, and 67% of Hawaiian/Pacific Islanders.

We carefully selected and categorized oral cavity subsites for analysis by topography codes: lip (C00.3-C00.5), tongue (C02.0-C02.3), gum (C03.0-C03.1,C03.9), floor of the mouth (C04.0-C04.1,C04.8-C04.9) Hard palate (C05.0), buccal areas (C06.0-C06.1) and retromolar area (C06.2). Subsites representing the oropharynx were carefully excluded in order to avoid inclusion of oropharyngeal cancers, which are known to have different etiology and prognosis. Subsites excluded were the base of the tongue, uvula, soft palate, and lesions of the tongue and mouth that overlap with each of those regions. Malignant SCC diagnoses were selected using the histology codes 8050-8089. We included patients 1) with only one primary 2) Diagnosed from 1988-2007 3) age 18 or over 4) actively followed.

Statistical Analysis

Many clinical variables were examined, including patient demographics, tumor characteristics, treatment modality, and follow-up data. All raw data were downloaded during a single session, and many of the variables were then categorized into groups using custom programs in MATLAB (The MathWorks, Inc., Natick, Massachusetts, United States, version 7.8.0). Demographic variables included gender, year of diagnosis (stratified into groups of 1988-92, 1993-97, 1998-02, and 2003-07), age (stratified into 18-40, 41-55, 56-70, 71-85, 85+), race (recoded into 'White', 'Black', and 'Other: American Indian/AK Native, Asian/Pacific Islander'). Tumor characteristic variables included stage, grade, primary subsite. Disease stage was classified according to American Joint Commission of Cancer (AJCC) 7th edition staging guidelines. T classification was derived from measurements of the tumor size and degree of extension into adjacent organs. N classification prior to 2004 was determined from a single variable encoding lymph node involvement, while N classification from 2004-2010 was derived from two variables encoding lymph node size and lymph node involvement. Though the coding schemes changed between these periods, we carefully translated them into N0, N1, N2, and N3 classes. For M classification, metastatic tumors included metastases to distant sites. All T, N, and M classification was performed according to the *SEER Program Coding and Staging Manual 2010*. Treatment-related variables include treatment modality (surgery, radiation, or both), and whether or not neck dissection was performed. Radiation therapy included one or more of the following: external beam radiation, radiation implants, radiation NOS, or radioisotopes. The variable 'Scope of Regional Lymph Node Surgery' indicated whether and what type of neck dissection was performed. However, this was only available from 1998 and on. No variable explicitly

indicated the performance of a neck dissection before this, so we used the number of lymph nodes which were pathologically examined as a surrogate. While some lymph nodes are processed with a primary tumor specimen, it is quite rare for there to be more than a few lymph nodes in the resection of a primary site tumor. If more than 15 lymph nodes were pathologically examined, we used this as a surrogate to indicate that a lymph node dissection as performed. We chose this based on personal surgical experience, as well as previous published reports [35]. Upon retrospective analysis, this variable was found to have a very high rate of concordance with the "Scope of Regional Lymph Node Surgery", suggesting it was a reliable indicator of whether a dissection of the neck was performed.

The year of diagnosis, year of death, and current vital status were used to determine two and five year survivals. Five-Year survival was evaluated by using a Kaplan-Meier analyses, and survival of different groups was compared by log rank tests. Disease specific survival includes only death attributable to the oral cavity cancer as the event of interest. This helps account for changes in the population's baseline survival and approximates relative survival. Relative survival calculates the survival of the study group in comparison to the baseline survival of the general population matched for age, race and sex. This was calculated using the Ederer II method [36] and found to be extremely similar to disease specific survival, and thus not included. To examine changes in the above characteristics over time, the proportions of each attribute were tabulated for each time period. A chi squared test was used to test the significance of changes in these characteristics over time. To further explore the characteristics or oral

tongue cancer, we also performed the analyses described above for each individual subsite within the oral cavity.

A multivariate analysis was conducted using a Cox proportional hazards ratio. Using a Cox proportional hazards regression, a univariate analysis was first performed for each individual demographic, clinical, and treatment variable in order to determine whether each was potentially associated with survival. The final multivariate "adjusted model" included all the covariates deemed to be significantly associated with survival when evaluated individually. Using the Cox proportional hazards regression to create a model with all the relevant variable, hazard ratios were determined, and are displayed as compared to the reference value (1.00) within each category. All tests were 2-sided, and statistical significance was determined at the $p<.05$ level.

A case listing with clinical information for each subject was obtained from the SEER database and all statistics were computed using MATLAB Statistics Toolbox.

Genomic Analysis

We utilized data available from The Cancer Genome Atlas (TCGA) HNSCC [9] project to examine the mutational, copy number variation (CNV), and RNA expression profiles of head and neck cancer specimens. The TCGA is a collaborative, multi-institution effort supervised and funded by the National Cancer Institute (NCI) and the National Human Genome Research Institute. Begun in 2005, the goal was to catalogue genetic alterations responsible for different types of cancer. Raw data is publicly available, and can be downloaded through a portal TCGA website. Clinical data were downloaded via the TCGA data matrix and samples were identified a segregated according to anatomic subsite, patient age, and smoking status/pack year history. This

clinical data were used to further divide the sample into subgroups. The "Oral Tongue" group consisted of tumors originating in the oral tongue. The "Oral Cavity" group consisted of tumors originating elsewhere in the oral cavity (FOM, hard palate, buccal mucosa, alveolar ridge, lip, other). The "Oropharynx" (OP) group originated in the oropharynx (base of tongue, tonsil, or oropharynx). The "Larynx/Hypopharynx" (LH) group originated at those respective sites. Within the OT category, we further segregated samples based on traditional vs. non-traditional demographics. The "Young Tongue Non-Smoker" (YTN) subgroup was defined as being located in the oral tongue, age <= 55, and a confirmed history of 'Lifelong Non-Smoker". The more traditional, "Old Tongue Smoker" (OTS) subgroup, was defined as located in the oral tongue, age >55, and a confirmed history of smoking (minimum 5 pack-years). After segregating into anatomical sub-groups, we analyzed and compared the mutational, copy-number, and RNA expression profiles of each group.

For somatic mutation data, sequencing data were generated by the TCGA on a IlluminaGA system. The mutation calls were generated at Broad Institute Genome Sequencing Center using the Mutect method. These calls were then downloaded through the UCSC Cancer Genomics Browser[37, 38], and mutational frequencies were calculated using MATLAB. To determine significantly mutated genes for each clinical subgroup, level 2 mutational data were downloaded from the TCGA data portal and processed using the MutSigCV module within the Broad Institute's GenePattern portal [39, 40]. Significant calls in MutSigCV (based on background mutation rate, expression, and mutation types) were recorded for each subgroup and compared to one another.

For copy number analysis, copy number profile was measured experimentally using whole genome microarray at a TCGA genome characterization center. Subsequently, GISTIC2 method was applied using the TCGA FIREHOSE pipeline to produce gene-level copy number estimates. GISTIC2 estimates are thresholded to estimate values of -2,-2,0,1,2 representing homozygous deletion, single copy deletion, diploid normal copy, low-level copy number amplification, or high-level copy number amplification. Genes were then mapped onto human genome coordinates and visualized using UCSC software and tools available through the Cancer Genome Browser [37, 38]. To illustrate the full spectrum of CNVs, visual comparison of amplifications/deletions across subgroups was performed on the UCSC Cancer Genome Browser, along with student's t-tests comparing the mean value of each copy number locus between the two comparison groups.

Since CNVs and somatic mutations can have a similar impact on cancer-related genes, we performed a combined mutation+CNV analysis to gain an integrated perspective on the genes most commonly altered in HNSCC. This data were obtained/visualized and printed through the cbioportal website maintained by Memorial Sloan Kettering Cancer Center [41]. Calls for this mutation and copy number data were performed according to the TCGA HNSCC project, and genes cited as most commonly indicated were selected for analysis [9]. Samples were sorted according to clinical subgroup and genes were sorted according to pathway.

RNA sequencing and expression quantification were performed as previously described by TCGA using IlluminaHiSeq2000 [9]. This data, after being normalized across the TCGA HNSCC cohort and log-transformed by the UCSC Cancer Browser Team, was downloaded for further analysis. Expression data were segregated according to clinical

subgroup and analyzed using the ComparativeMarkerSelection through the GenePattern portal [39, 42]. This compared expression of all genes between the relevant clinical groups to select the genes most significantly upregulated or downregulated in a specific cohort. The most significantly up- or down-regulated genes for each cohort was recorded, and visualized using the HeatMapViewer through the GenePattern portal [39]. The list of genes which were up and down-regulated for each subgroup were analyzed for enrichment of certain pathways, functions, or components using Gene Ontology and Gorilla Enrichment analyses [43, 44]. Since most of the genes found to be upregulated in Oral Tongue Cancers vs. others were muscle-related, we also repeated this analysis using only known cancer-related genes according to the COSMIC database [45]. Lastly, to determine if there were separate subgroups of samples defined by significantly different gene expression, unsupervised clustering using Pearson correlation for both samples and genes and pairwise average linkage was performed and visualized through the HierarchicalClustering tool in the GenePattern portal [39, 46]. We selected those genes which were most variably expressed across all samples. This was performed for 1) the full set of HNSCC and 2) Oral Tongue tumors. The sample clustering was analyzed for the degree of clinical subgroup clustering, and significance of the clusters was tested using SigClust, though not shown here. We also performed a separate clustering using an 'intrinsic gene list' utilized by previously published HNSCC clustering studies, which is not shown here[9, 47, 48].

Results

Characteristics of OCSCC

From 1988 – 2007, SEER reported a total of 16,298 cases of OCSCC. Median age at diagnosis was 63 years. The male:female ratio was 1.5:1, and the white:black:other ratio was 12.5:1.3:1

Demographic trends in OCSCC

When broken into 5 year intervals, patient demographics underwent a small but gradual shift over the study period (table 1). The mean age at diagnosis decreased modestly from 64.3 to 63.5 years, and gradually included more middle aged patients (41-55) and fewer patients aged 56+ ($p < .001$). There was also an increase in the proportion of female subjects, comprising 35.7% of diagnoses in the first time period, and 39.2% in the final time period ($p <.001$). The proportional breakdown of affected races remained fairly stable over time, with 84.4% of cases affecting whites, 8.1% blacks, and 7.5% 'other' in the final time period. This represented a slight decline in the proportion of black patients from the first time period (9.5% to 8.1%), and an increase in 'other' (5.0% to 7.5%).

Incidence Trends of OCSCC and Affected Subsites

While incidence rates of all other oral cavity subsites have remained stagnant or decreased, incidence of oral tongue SCC rose 17% (figure 1). Oral tongue cancers comprised only 30% of all oral cavity cancers in 1988-92, but increased to 38% in 2003-07 (table 1). Notably, floor of mouth cancers, initially comprising 34.2% of OCSCC diagnoses, dropped to 23.3% between the initial and final time periods examined. Other oral cavity subsites remained relatively stable in incidence.

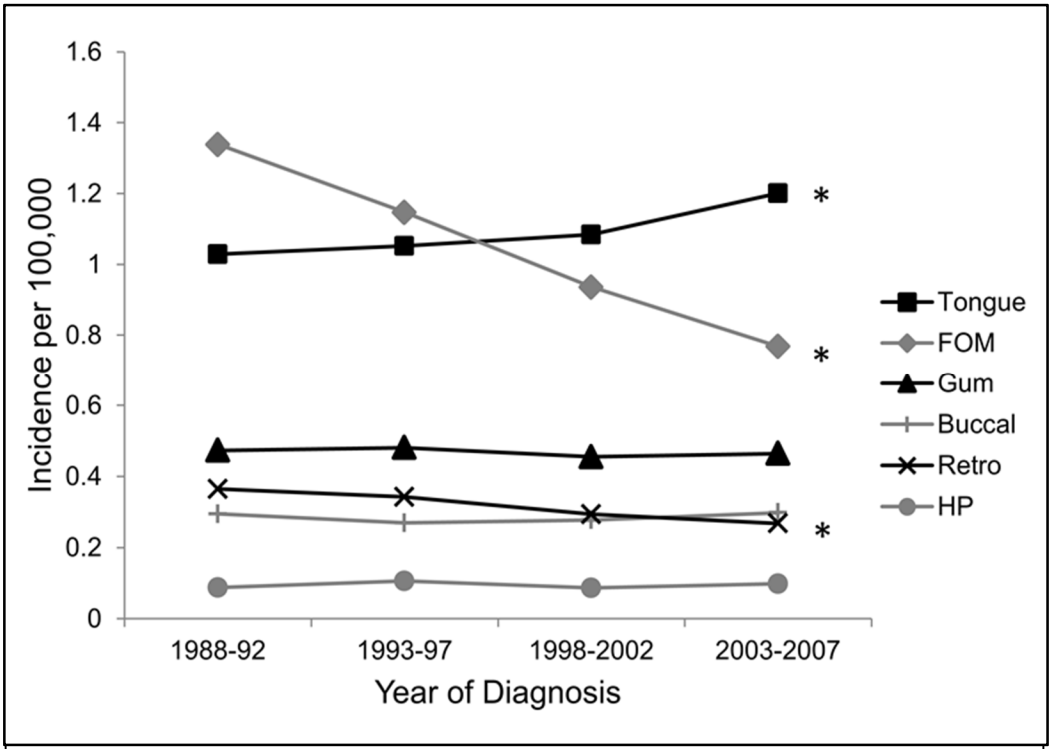


Figure 1 - OCC Subsite Incidence. Annual incidence rates for squamous cell carcinoma of each oral subsite averaged over five-year intervals. The overall incidence of all subsites combined was decreased (not shown, p<.05). * indicates a significant change from first to last interval (p<.05). FOM = floor of mouth, HP = hard palate, Retro = retromolar

Stage and Grade Trends of OCSCC

OCSCC showed a steady trend toward earlier-staged disease (table 1), with an increase in AJCC stage 1 (+11%, p <.001) and a decrease in stages II-IV. This was reflected in earlier T and N classifications. There was a 14% increase in T1 disease, and corresponding decrease in T2 and T3. Similarly, there was an increase in N0 and N1 (+6% and +5.3%, respectively), with a concomitant decrease in N2 and N3. M classification remained relatively stable over the study period at roughly 2% M1 disease.

We further examined these staging trends for each individual subsite within the oral cavity. We found that oral tongue tumors showed a dramatic trend towards earlier

staging from the first to the final time periods, as the proportion diagnosed at stage 1 disease increased from 27% to 44%, resulting in an absolute increase of 17% and a relative increase of 63%. Other subsites showed mixed results with staging trends when observing absolute change in stage 1 disease (lip: -4%, gum: +10%, floor of mouth: +7%, hard palate: -9%, buccal: +6%, retromolar: +9%). Overall, 62.9% of OT cancers were stage 1 or 2, compared to just 44.5% of other oral cavity cancers. Other oral cavity cancers were also twice as likely as OT cancers to be stage 4 (40.4% vs. 19.7%).

When examining T classification across individual subsites, there was a near unanimous increase in earlier staged disease, indicating that tumors throughout the oral cavity were routinely being detected at smaller sizes. In the first time period, 39% of oral tongue tumors were 'T1' disease (0-2 cm), compared with 55% in the final time period. Other subsites showed a similar increase in T1 disease from the first to the last periods (34% to 46%). N classifications followed a similar trend, with nearly all oral cavity subsites showing an increase in stage N0 and N1 disease, and a concomitant decrease in N2 and N3 disease. M staging remained relatively stable for each subsite, changing less than 1% over time.

Tumor grade remained fairly stable over time (table 1), though there was a 7% increase in Grade II (moderate differentiation), with a corresponding decrease in Grade I (well differentiated). Notably, this increase in the 'moderately differentiated' group over time was consistently found in each subsite throughout the oral cavity.

Treatment Trends in OCSCC

In recent years, more patients with OCSCC received surgery, while fewer received only radiation (Table 1). Since disease stage is one of the predominant determinants of treatment, we analyzed treatment trends within each stage (Table 3). For stage 1 disease, there was an increase in 'surgery only' (79.5% to 85.1%, $p<.01$) and a decrease in the use of radiation in treatment (20.5% to 14.9%, $p<.01$). This might represent an awareness of the morbidity associated with radiation therapy and a desire to spare the patient when a cure is achievable through surgical excision alone. There was relatively little change in the management of stages 2 & 3, where both surgery and radiation are frequently used as standard of care. Meanwhile, for stage 4 disease, there was an increase in radiation as the sole therapy (12.1% to 24.3%), with a concomitant decrease in surgical therapy (87.9% to 75.7%). This could represent the decision by care providers not to perform surgery in a select group of late-disease stage patients in whom a cure is highly unlikely, who may not tolerate the procedure, and in whom quality of life is a priority.

Over the past 20 years, there was also a gradual increase in the proportion of patients receiving neck dissections (+7.5%, $p < .001$) (Table 4). This was especially pronounced for patients diagnosed with early staged disease (stages 1&2); this cohort was treated with a neck dissection only 19% of the time in 1988-92, and 26% in 2003-07, a relative increase of 36%. There was no significant change in the percentage of late stage (stage 3&4) cases being treated with neck dissection, with roughly 50% of this cohort receiving neck dissections throughout the study periods.

Distinct Demographic Affected by Oral Tongue (OT) Cancers

Based on the paradoxical increase in OTSCC incidence, as well as previous reports of a shift in the epidemiology, we chose to further examine the characteristics of OTSCC. Accordingly, oral tongue cancer was found to be quite different from other oral cavity subsites in its demographic characteristics. First, it was seen to affect a younger population than other oral subsite cancers. The median age of OTSCC patients at diagnosis was 60 years, vs. 65 years for other oral subsites combined. There was also a trend toward a younger population over time for OTSCC, as the median patient age decreased from 63 years of age from 1988-92 to 59 years during the 2003-07 time period. A larger proportion of OTSCC affected females (43% of OT, 37.5% of other oral cancers), with a trend toward more females over the duration of the study period (+5%). With regard to race, the various subsites showed a similar breakdown of 'White', 'Black, and 'Other' patients affected. The only salient difference between OT cancers and other subsites, was a greater proportion of 'Other' patients affected (in the final time period: 10% for OT, 5.9% for all other subsites combined). Additionally, there was an increase in the number of 'Other' patients affected by OT cancers over time (7.7% to 10% from the first to the last time period). As opposed to prior studies, we did not find an increase in the proportion of White individuals affected.

Trends in Oral Cavity Cancer Survival

Overall 5-year survival in OCSCC showed an absolute increase of 10.5% between 1988-92 and 2003-07, from 39.9% to 50.4% (table 2) and cancer-specific 5-year survival increased 7.5% over the same period. Over the course of the four consecutive time periods examined, there is an incremental increase in overall survival (figure 2). Though

not included here, trends in relative survival calculated using the Ederer II method were very similar, showing a 10.6% increase in 5-year survival over the time periods.

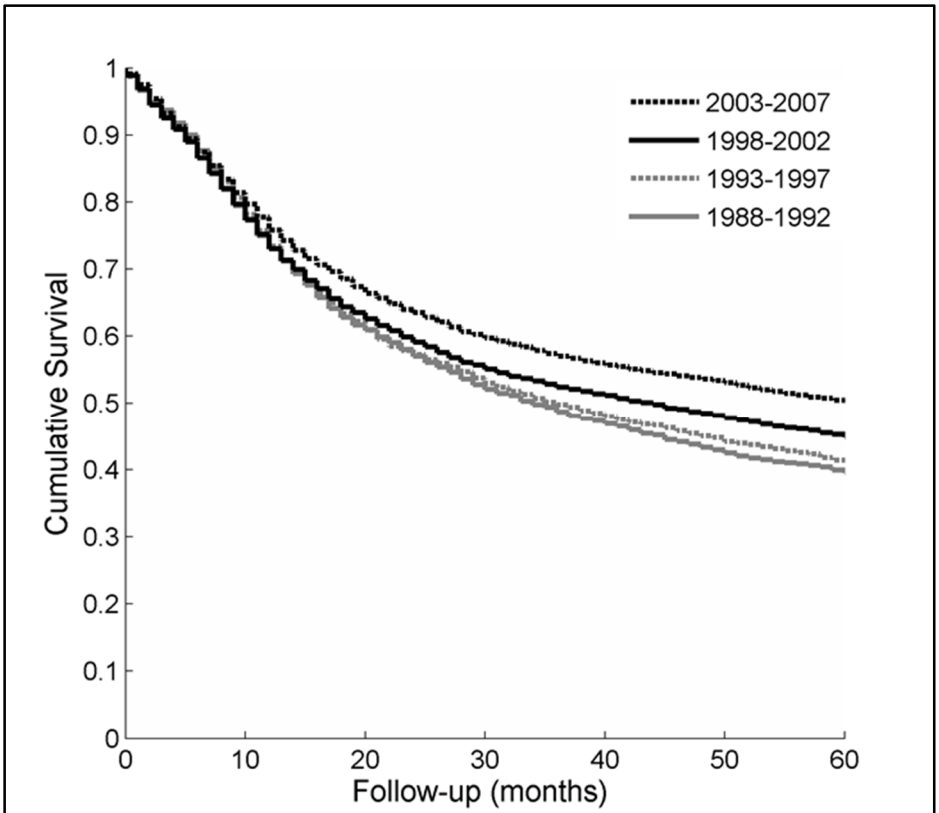


Figure 2 - OCC Survival By Period. Kaplan Meier curves showing overall survival for OCSCC for each study time period. Log rank comparing the first and last time periods was performed (p<.001).

We next examined survival trends within subgroups of OCSCC. The 5-year overall survival trend was examined across each clinical variable included in the study (table 2). The youngest age groups (18-40, 41-55) experienced the greatest improvements in survival. There was minimal change in the survival of the most advanced age group (85+). Males appreciated a 13% increase in 5-year survival, compared to just 6% for women – leading to approximately equivalent survival rates in

the most recent time period. Whites had a higher 5-year survival than blacks in every time period. However, black individuals experienced the highest increase in survival, with a near doubling of five year survival from the first to the last time periods in the study (from 19.4% to 35.3%). Though the increase in survival of black individuals closed the historically large gap in survival of OCSCC between races, blacks still had the lowest five year survival of any racial group during the most recent time period (White: 51.1%, Other: 55.2%, Black: 35.3%). In order to further explore the historically lower survival rates in black individuals, we examined the clinical characteristics of this subgroup. We found that, as compared to other racial groups, they were being diagnosed at later stages overall (>50% stage 4), and with more advanced T, N and M classifications on average.

Survival Trends for Oral Tongue and other subsites

From 1988 to 2007, survival improved for almost every oral subsite. However, oral tongue demonstrated a particularly large increase in 5-year survival (+14.3% vs. +4% for all other subsites). Our findings complicate the question of whether OT has a different prognosis from other oral subsites, as it appears that the answer has changed over time (figure 3). In the first time period, OT cancers had a similar overall survival to other subsites, but by the final time period, it has a significantly higher five year overall survival.

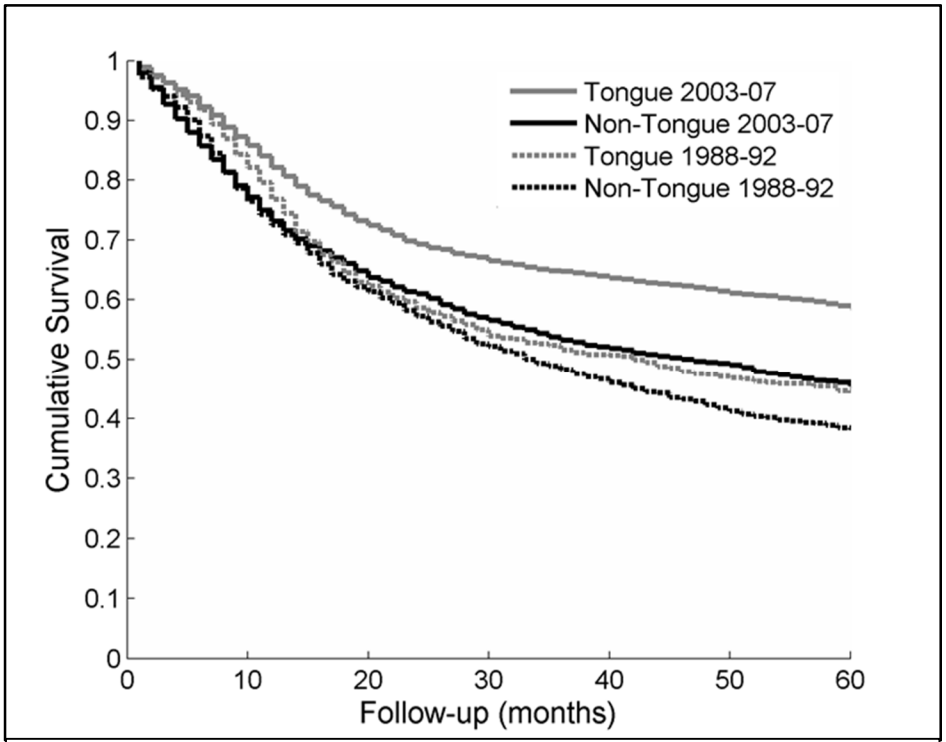


Figure 3 - Tongue vs. Non-Tongue Survival. Kaplan Meier curves comparing overall survival for Tongue and Non-Tongue subsites. 'Non-Tongue' represents all other subsites in the oral cavity.

Treatment and Survival Trends

While improved survival was found across all treatment categories, the increases were not uniform. Patients who received surgery showed a greater improvement in 5-year survival than those who received radiation only (+11.4% vs. +7.1%). Additionally, patients who were treated with neck dissections experienced a larger increase in survival than those who did not (+15.9% vs. +8.1%). It is not possible using the SEER database to determine whether this is due to improvement in therapy or simply changes in patient selection. Although a randomized control trial would be a more appropriate way to answer such a question, it is unlikely that such a study could be performed.

Multivariate Analysis

A univariate analysis using Cox proportional hazards regression was first used to determine which clinical variables correlated with survival. Multivariate analysis was performed on the clinical variables which reached significance on univariate analysis (figure 4). Even with all variables taken into account, there was a significant trend toward improved survival with subsequent time periods, as represented by decreasing Hazard Ratios (HR). Meanwhile, age and stage classifications showed the largest, incremental effects on hazard ratios (HR). Sex, race, and grade were also significant covariates. Males had a higher HR than females, and blacks had the highest HR among races.

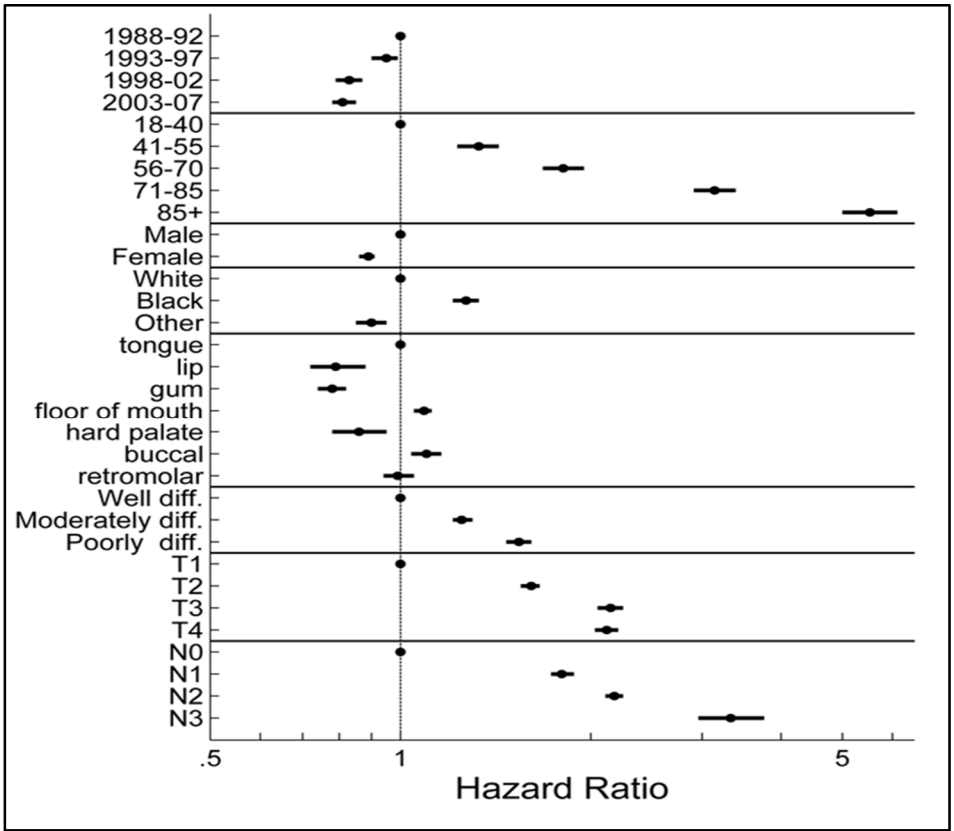


Figure 4 - Multivariate Analysis. Forest Plot demonstrating hazard ratios derived from multivariate Cox Proportional Hazards regression model. All variables shown were included in the model. M stage was not included in the model, as only M0 patients

The location of the primary within the oral cavity had very little effect on survival – visualized by the tight clustering of HR's between different subsites (figure 4). While OTSCC had a better survival than other subsites in the univariate analysis, this did not hold true in the multivariate model. Interestingly, the OTSCC had a slightly higher HR than other oral cancers combined (figure 4). Since we have already shown age and stage to exist in more favorable proportions in tongue cancers, we hypothesized that these variables were confounders for the subsite analysis. To test this theory, when age and stage were subtracted from the model one at a time, the HR for OTSCC indeed fell below other that of sites.

Genomic Analysis of Oral Tongue tumors

Clinical Attributes of TCGA cohort

We identified 511 HNSCCs available for analysis from the TCGA cohort. This was far more than the cohort of 279 tumors which was used in the recent HNSCC molecular characterization paper published by the TCGA [9]. There were 129 oral tongue tumors, and 178 other oral cavity tumors ('oral cavity' tumors herein include only oral subsites which are not the oral tongue). The YNT (young, non-smoking, oral tongue) group accounted for 21 tumors, while the OST (old, smoking, oral tongue) group consisted of 47 tumors. Other subsite totals and characteristics can be found in Table 5.

Similar to the SEER cohort, the TCGA OT cancers affected a younger population than other oral cavity subsites (mean age 58 vs 64). Also corroborating SEER data, when compared with other anatomical sites, a larger proportion of the OT cancers affected

females and non-smokers. Notably, HPV testing was performed on 27 oral tongue, 28 oral cavity, and 23 larynx/hypopharynx samples, and was reported positive by either p16 or ISH testing in only 11%, 7%, and 17% of samples, respectively. Conversely, oropharyngeal cancers were positive for HPV in 81% of samples tested. This reaffirms the anatomical segregation of HPV involvement. Since many other oropharyngeal tumors had not yet been tested at the time of this study, we group all OP tumors together, rather than only HPV+ tumors, assuming a high rate of HPV presence for the entire group.

Somatic Mutations

Of the 511 tumor cohort, 493 contained whole exome sequencing data. Previous TCGA HNSCC studies used only 279 samples, so this represents the largest sample size to date for mutational analysis. Because of the small sample size in the YNT cohort, we limited our initial search to genes reported by the TCGA as being most commonly mutated in HNSCC. Table 6 shows the frequency of non-silent mutations in these genes. The average mutation rate in the YNT cohort was 65, compared to 120 mutations for the OST group. This is not necessarily surprising given that mutation rates have been shown to increase with age and tobacco exposure. Meanwhile, the overall oral tongue group (OT) had far fewer mutations than the rest of the oral cavity (94 vs. 175, $p < .01$), even when accounting for smoking status/age differences. This finding has not been previously established.

The YNT group actually displayed a significant increase in the rate of mutations to cell cycle regulators p53 and CDKN2A (91% and 43%, respectively; $p < .05$ for each), when

compared to all other tumors. This is in stark contrast to HPV-related oropharyngeal (OP) cancers, which showed dramatically decreased rates of 29% and 7% for p53 and CDKN2A, respectively. The YNT group thus lacks important, classic signatures of HPV+ tumors (absence of CDKN2A and p53 mutations), which results from expression of the HPV oncoproteins E6 and E7. With the exception of p53 and CDKN2A, almost every other gene examined here showed a decreased mutation rate in the YNT group, although none reached significance, possibly due to the small cohort size of YNT. Several genes showed a trend toward decreased mutation rate (FAT1, CASP8, PIK3CA, NSD1) which could be better evaluated with a larger sample size.

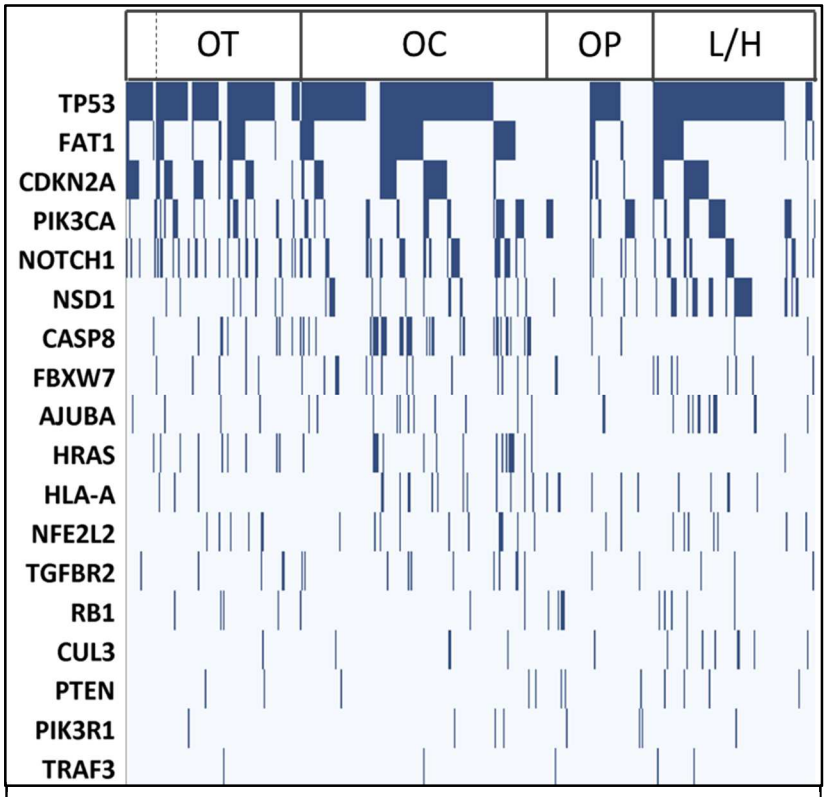


Figure 5 - Mutation heat map illustrating non-silent mutations (Blue stripes) in the most commonly mutated genes of the TCGA published cohort. Samples are aligned horizontally. OT = Oral Tongue, OC = Oral Cavity (not including oral tongue), OP = Oropharynx, L/H = Larynx/Hypopharynx. The dotted line at the top left borders the left-most, YNT group.

When comparing OT to other OC subsites, there was a significantly decreased mutation rate in FAT1 and CASP8. This is an interesting finding in the context of the recent TCGA publication which showed a unique sub-population of tumors with mutated CASP8 and HRAS mutations and possessing a superior

prognosis [9]. Based on our results, this subtype may be segregated to another oral cavity subsite, possibly the buccal mucosa (8/15 with CASP8 mutation) and hard palate (2/4 with CASP8 mutation). These rates are extremely high, and reach significance when compared to the rest of the head and neck (14/318 = 4%). Figure 5 illustrates the commonly mutated genes across the samples included in each cohort.

To conclude our mutational analysis, we searched for mutations enriched in our cohorts which have not been previously well described. To do so, we examined the whole exome sequences of each cohort to find the most highly mutated genes in each anatomical subgroup. The results can be found in table 7. Briefly, this simply shows that very few genes are recurrently mutated in the YNT group. Only a few genes were mutated in >20% of cases, and none of these were unique to the YNT cohort and/or more commonly mutated than in other cohorts.

Copy Number Analysis

Copy number data were available in all 511 tumors in our cohort. Previous analysis by the TCGA group and others had shown recurring copy number alterations (CNAs), consisting of losses of 3p and 8p, and gains of 3q, 5p and 8q chromosomal regions. These CNAs are driven by loss of tumor suppressors or gains in oncogenes, resulting in a competitive advantage to cells containing these defects. With a mean of 141 CNAs per tumor, HNSCC shows a high degree of instability relative to most types of cancers. Figure 6 displays the copy number profile of the YNT and OST cohorts, as well as the parent OT cohort and OC cohort. Across the head and neck subsites, the copy number

profiles (proportion of each locus which is amplified, deleted, or normal) are very similar to one another. The profile of the oropharyngeal subset is noticeably different – likely due to its unique pathogenesis and selective pressures. The copy number profile of the YNT cohort resembles that of the traditional head and neck cohorts more closely than it resembles oropharyngeal cancers, indicating that its causative mechanism may be dissimilar from HPV-related tumors. There was an increased incidence of 11p13 amplification in YNT compared to all other cohorts. Meanwhile, OT tumors overall showed an increased loss of 4q and 5q compared to other oral cavity tumors ($p < .05$, using fisher's exact test at each chromosomal locus).

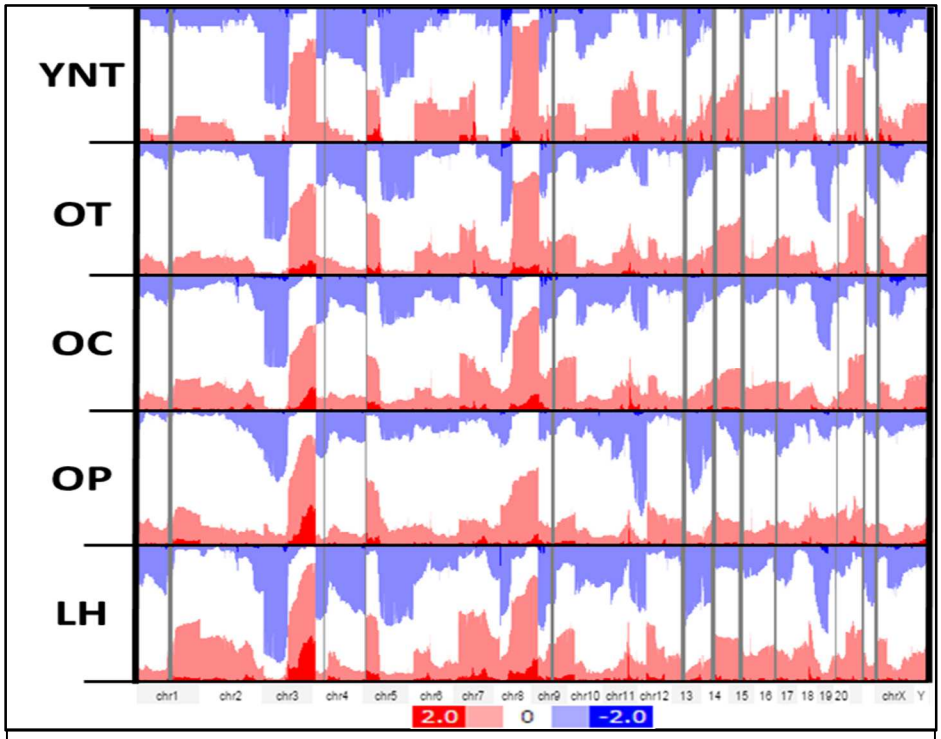


Figure 6 – Copy number changes (Gistic 2.0, thresholded), viewed grossly across all chromosomes to highlight commonly recurring amplification (red) and deletions (blue). Colors illustrate the proportion of samples which are amplified (red), deleted (blue), or normal (white) at each locus. Darker blue indicates homozygous deletion, lighter blue heterozygous. Darker red indicates exceedingly high copy number. CNAs are shown for each anatomical group

Mutation+CNA Analysis

Several different types of genetic alterations can lead to a similar outcome (loss of a tumor suppressor or activation of an oncogene). We integrated mutational and CNA data in order to identify genes most commonly affected in HNSCC irrespective of the type of defect. We segregated samples as in the previous analyses, and focused on genes previously implicated in HNSCC[9]. Figure 7 shows the frequency and type of alteration in each gene/pathway. The YNT cohort can be found at the very left of the figure, and shows an overall decreased frequency of alterations. Specifically, the PIK3CA pathway, cell death pathway, and oxidative stress pathway show almost no defects for this cohort.

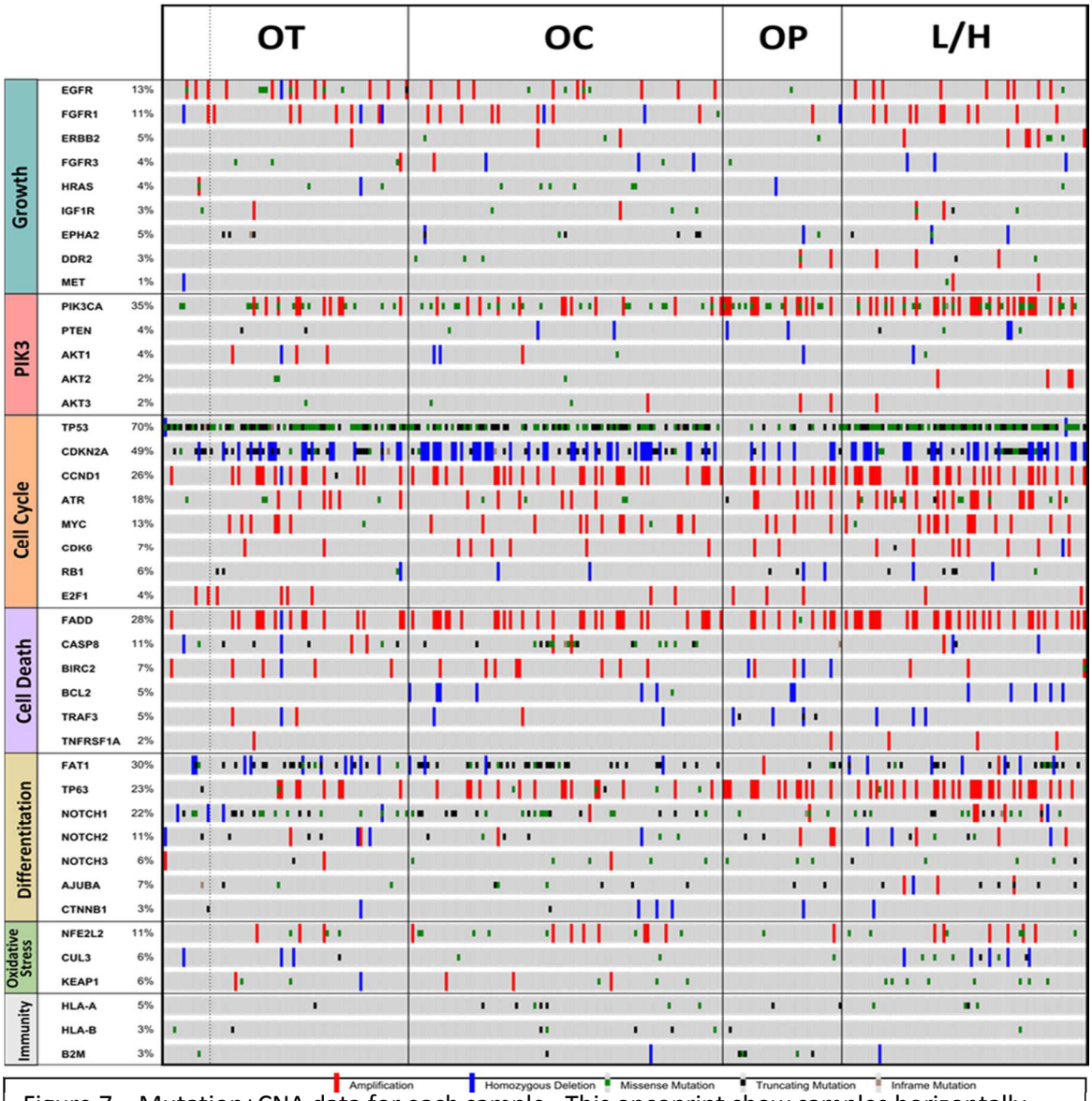


Figure 7 – Mutation+CNA data for each sample. This oncoprint show samples horizontally, divided by anatomic origin (same as figures 5/6). The gene listing is vertically divided according to common biological pathways involved in cancer.

Expression Analysis

We sought to explore differences in gene expression among our clinical cohorts, focusing on the differences between oral tongue and other oral cavity subsites (OT v OC) as well as distinct expression signatures within the oral tongue subgroups (YNT v OST). With expression data available for 20530 genes across 511 samples, we first found the most

differentially expressed genes (DEGs) between YNT and OST. 165 genes were significantly more highly expressed in YNT and 116 genes in the OST group (FDR < .05, p < .001). We examined these genes for over-representation of biological pathways using the Gene Ontology (GO) enrichment tool. Ion binding was the only enriched function in YNT tumors, and no functions or pathway were enriched in OST.

Since there was the gene expression among the oral tongue subgroups was fairly homogenous based on pathway analysis, we then found the DEGs between OT and OC. 1500 genes were significantly more highly expressed in OT and 673 genes in the OC group (FDR < .05, p < .001). We again examined these genes for over-representation of biological pathways using the Gene Ontology (GO) enrichment tool. The most upregulated genes in the OT group were almost exclusively muscle-related genes (p < .00001, GO Enrichment), likely due to the presence of muscle cells in the oral tongue cancer specimen. To screen for relevant genes, we performed the same analysis without muscle-related genes – and found that cell differentiation, nervous system development, cellular response to DNA damage, DNA repair, DNA metabolic process, and mitotic functions were each enriched. Those genes which were upregulated in OC cancers compared to OT cancers were enriched for: cell cycle processes, G2/M phase transition, mitosis, DNA replication, DNA repair, assembly of cellular components. Olfactory response and response to bacterium were also upregulated, likely as a by-product of the normal functions of oral cavity tissue (i.e. contamination).

We next performed an unsupervised hierarchical clustering of all 511 samples using the 3600 genes which were most variably expressed in the samples, as described previously [47, 48]. Figure 8 shows the resulting dendrogram. Roughly four distinct expression

subtypes were found, and anatomical sites segregate highly into each of these. Oropharynx, Larynx, Oral Tongue and Oral Cavity tumors (from left to right) comprised the bulk of each subtype. The fact that OT tumors clustered separately from most other oral cavity tumors suggest they possess a distinct expression profile. The YNT did not cluster tightly within the OT tumors, indicating that they do not have a unique or distinct profile from other OT tumors.

To examine whether OT was comprised of multiple sub-types, and determine whether the YNT group possessed a unique expression signature, we performed an unsupervised clustering of only OT tumors. Figure 9 shows that while multiple subgroups were found, they were much more closely related than the subgroups of figure 8. YNT did not cluster with one another, and in fact no clinical markers examined (tobacco use, gender, race, age) were significantly associated with a single expression profile. OT cancers appear to consist of a fairly homogenous expression profile.

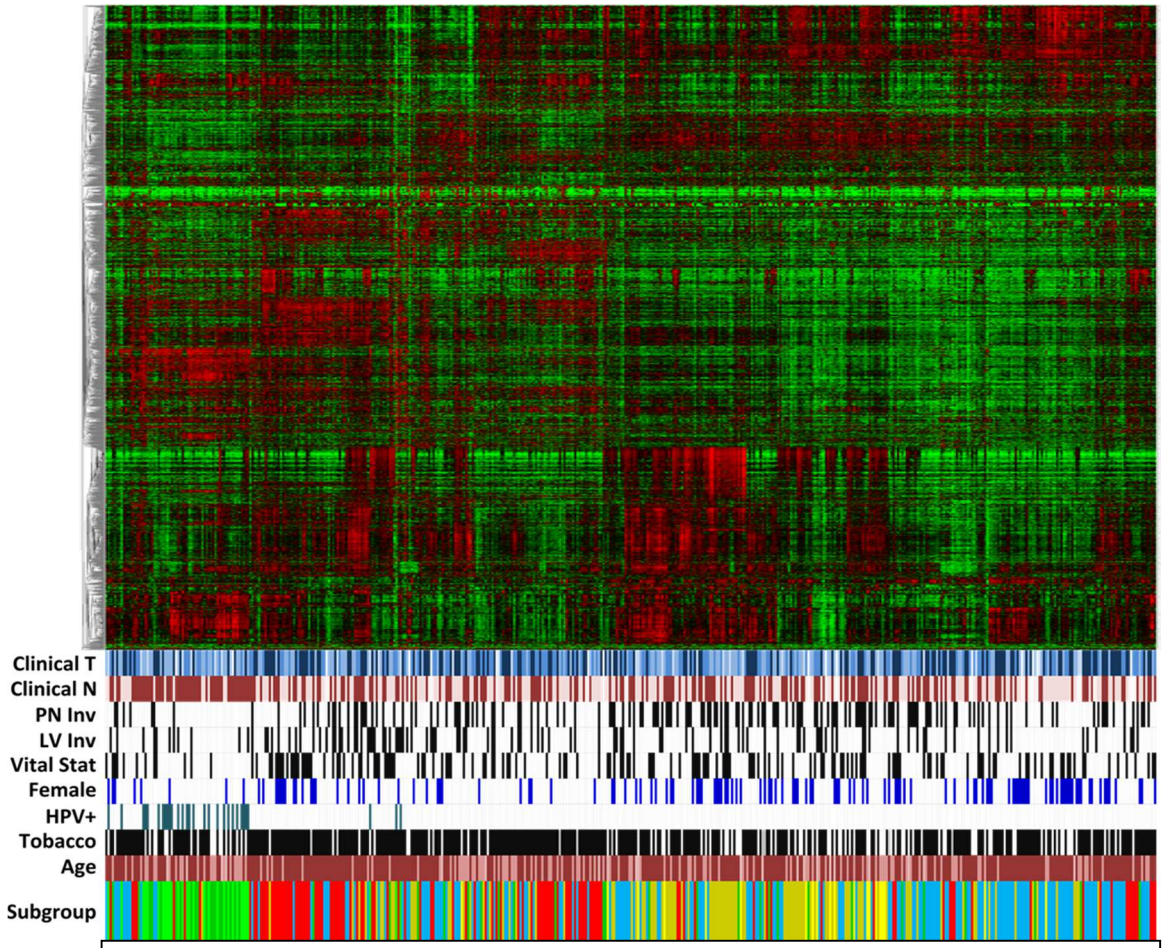


Figure 8 – Unsupervised mRNA expression analysis of HNSCC

1. Clinical T: Darker blue = higher stage (T1-T4)
2. Clinical N: Darker Red = positive nodes
3. PN Inv = Perineural Invasion: Black = Yes
4. LV Inv = Lympho-Vascular Invasion: Black = Yes
5. Vital Status: Black = Dead
6. Female: Blue = yes
7. HPV+: Blue = positive p16 or ISH
8. Tobacco Use: Black = smoker, gray = unknown
9. Age: Light Red = < 55, Dark red = 55+
10. Subgroup Category: Light Yellow = YNT, Dark Yellow = OT, Light blue = OC, Light Green = HPV+, Dark Green = OP, Red = LH

Following unsupervised clustering using 3600 genes, it can be observed that anatomical sited, cluster quite closely to one another, including OT vs OC.

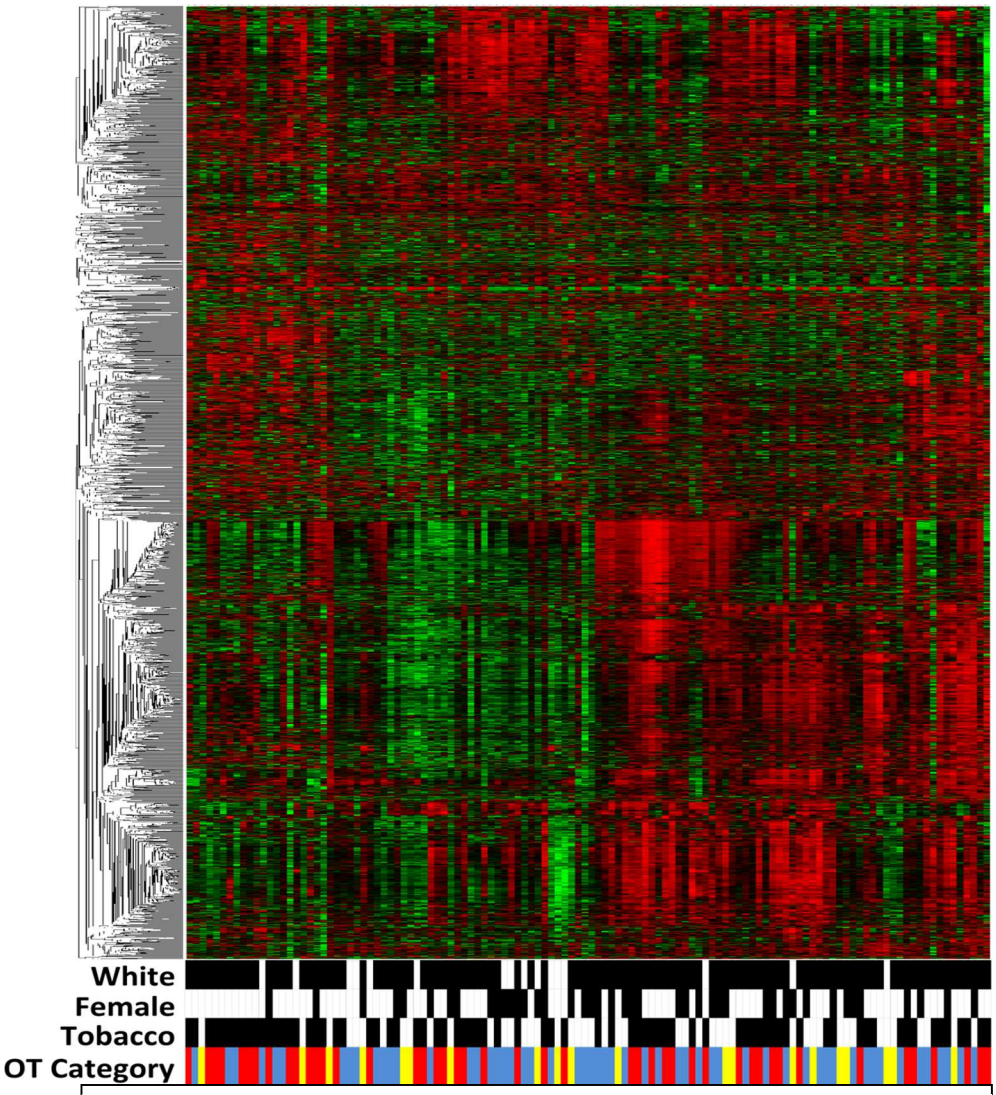


Figure 9 – Unsupervised mRNA expression analysis of OTSCC

OT Category: Yellow – YNT, Red = OST, Blue = Remaining OT

Following unsupervised clustering using 3600 genes, it can be observed that OTSCC does not cluster into well defined clinical groups, as visualized by the scatter of blue, red and yellow at the bottom of the figure. Overall, OTSCC appears to represent a fairly homogenous expression group.

Discussion

Changing Characteristics of OCSCC

Using the SEER database, we found that the OCSCC characteristics and survival changed over the past 20 years. The affected patient population became younger and

more female, while tumors were diagnosed at earlier stages. Specifically, oral tongue cancers were largely responsible for the observed changes. Additionally, while the incidence of SCC within most oral subsites declined, OTSCC incidence increased over the study period and this finding is supported by previous reports of a change in the epidemiology of OTSCC [22]. However, despite implications that this epidemiological shift may entail a new causative mechanism, tumors from the oral tongue subsite could not be distinguished from other tumors of the oral cavity based on current molecular data. The decrease in incidence of floor of mouth cancers has not been previously described, but interestingly the decrease in floor of mouth cancer closely mirrors the rise in oral tongue cancers. It is possible that a change in environmental risk factors, such as a decline in the usage of chewing tobacco, has led to this decline in incidence of floor of mouth cancers. Alternatively, it is possible that changes in coding patterns of the primary tumor location has led to fewer cancers being labeled as 'floor of mouth' and instead being attributed to adjacent subsites. In addition, with OTSCC being diagnosed at smaller sizes, it is possible that ventral tongue cancers are less frequently being labeled as floor of mouth cancers. More study is required to further detail this trend and explore these possibilities.

Improving OCSCC Survival

While traditional thought has been that survival in oral cavity cancers has not improved during the last several decades, data reported here show that survival for OCSCC has gradually improved over the past 20 years (HR = 0.81). In opposition, Carvahlo et al used the SEER database to show that while many types of head and neck cancers showed improved prognosis from 1973-1999, oral cavity cancer survival did not

[28]. However, several more recent studies support our analyses [23, 26, 27]. Pulte et al performed a period analysis of the SEER database (1982-2006) revealing a significant increase in survival of oral cavity and oral tongue cancers[26]. Mehta et al used the SEER database (1975-2006) to stratify tumors by histologic grade and came to similar findings[27]. Importantly, both of these studies used the 'Tongue' category site code, which includes tumors of the base of tongue and other lesions as possible confounders. To understand the significance of this inclusion, we analyzed the ICD codes of cases within the 'Tongue' category. We found that over half (66%) of these cancers consisted of likely oropharyngeal subsites including: 'base of tongue' (47%), 'lingual tonsil' (1%), 'overlapping lesions' (3%), or 'tongue, NOS' (15%). Exclusion of non-oral cavity carries particular significance considering the distinct epidemiology, molecular characteristics, clinical characteristics, and prognosis of oropharyngeal cancers, which are mostly HPV-related. More recently, Saba et al performed an analysis of survival using ICD codes more precisely distinguishing between oral cavity and oropharynx [23], and found improved survival for both oral tongue and base of tongue lesions. In a retrospective, international study of 2,738 patients who underwent resection for OCSCC, Amit et al found a significant improvement in survival from the 1990's – 2000's[30]. Our findings build on these recent reports using a larger and more recent cohort from SEER data, and with a detailed trend analysis of patient, clinical, and treatment characteristics.

Prognosis of Oral Tongue vs. Other Subsites

Though nearly all subsites experienced an improvement in five year survival, oral tongue cancers experienced the greatest jump. In the final time period, oral tongue

cancers had a five year survival of 58.6%, versus just 42.8% for all other oral cavity subsites. However, the apparent superior prognosis of OTSCC was found to be attributable to a favorable stage and age profile, but not specifically to the oral tongue subsite based on multivariate analysis. The most important covariates correlating with improved survival of OTSCC were stage, age, and grade (figure 4). Also using multivariate analysis, we show that the prognosis of oral tongue cancers compared to other oral subsites has not significantly changed over time. We are the first to examine subsite survival trends to show this, using rigorous criteria which exclude oropharyngeal subsites in the SEER database. The analyses here help clarify a widely debated topic in the literature, as to whether prognosis for oral cavity cancers vary based on subsite. Single institutional studies relying on smaller cohorts have reported mixed results with regard to differences in prognosis based on subsite within the oral cavity [13, 16-18]. Using the SEER database, Rusthoven et al [15] concluded that earlier stage tongue cancers had a poorer 5-year overall and disease-specific survival compared to other early stage oral cavity cancers. Our univariate analysis, by examining this relationship over time, found that survival patterns have been changing over the past twenty years. Twenty years ago, OT had an overall worse prognosis than many other oral cavity subsites, and this may be contributing to the conflicting reports. This trend also coincides in time with rising incidence of OTSCC and its association with a younger cohort being diagnosed at earlier stages. While our analyses and results confirm previous studies identifying a shift in the epidemiology of oral tongue cancers [1, 22, 49, 50], the correlations identified are unable to explain why oral tongue cancers are being diagnosed at earlier stages and in a younger population.

Clinical Relevance of the Epidemiological Shift

As the characteristics and patient population of OCSCC evolves over time, it is important for care providers to stay up to date regarding these changes. For example, OCSCC should no longer be considered a disease restricted to an older population or males. When a suspicion of malignancy arises, the same care and precaution should be taken regardless of age or gender. Care providers should also understand the prognosis and natural course of the disease, which includes understanding the clinical factors which affect survival. This would allow them to adjust their treatment strategies, and keep their patients up to date and informed. Additionally, with the various changes to treatment protocols over the past twenty years, it remains important to determine the effect of these changes on the survival of patients, and measure our progress in combatting OCSCC. In the same vein, by capturing incidence and staging trends over time, we can also measure the effectiveness of primary cancer prevention and early detection methods.

If the epidemiological shift of OCSCC or OTSCC is due to a new causative agent, identification of this agent through follow-up study may allow the development of novel strategies for primary prevention, as well as diagnostic or treatment methods specific to the etiology. The identification of HPV as a causative agent of oropharyngeal cancer spurred the development of diagnostic tests such as p16 expression assays and PCR based viral detection methods. It also has driven intense research into novel targeted drugs and alternative/de-escalated treatment protocols. Studying the genetic and molecular characteristics can help identify new attributes in this subset of tumors, but further study of these individual tumors, the oral microbiome, and other potential exogenous carcinogens would be necessary to identify any new causative agents. An appropriate

first step might be testing these samples for the presence of known oncogenic viruses or other strains of HPV.

Genomic Analysis of OTSCC

Previous genomic analyses of HNSCC have examined characteristics of the oral cavity as a whole, but very few have focused efforts on the oral tongue or individual oral subsites. The recently published TCGA paper [9] performs similar analyses for HNSCC as a whole, but without focusing on OT or YNT cohorts, as we have. One analysis, published very recently, examines a similar OT and YNT cohort, but focuses more on the types of mutations (e.g. C->A vs C->T), concluding that the OT group mutation profile resembles a 'smoking' signature, despite being mostly non-smokers[34]. They also utilize a smaller sample size (n = 323), which did not allow statistical power to find uniquely affected genes. Having access to the largest cohort of samples yet reported (n = 511), here we show that oral tongue tumors in the TCGA cohort carried similar clinical characteristics to OT patients in the SEER population (e.g. younger, more female, more non-smokers). We demonstrate that the overall mutation rate is significantly lower in OT cancers compared to other cancers of the OC and head and neck. The YNT cohort has the lowest mutation rate of all, and carries fewer mutations in most cancer-related genes – with the fascinating exception of increased mutation rates in p53 and CDKN2A, both of which play critical roles in cell cycle regulation.

There was a similar copy number profile amongst OT tumors regardless of age and smoking status, and this profile resembled that of other 'traditional' head and neck cancers at almost every locus, with an additional amplification at 11p. This profile is

very different from HPV-related tumors of the oropharynx, and could indicate that the selection pressures for CNAs in OT cancers is similar to that of traditional HNSCC related to smoking and environmental exposures. Lastly, the expression profile of OT tumors was distinct from that of other sites in the oral cavity. This is a previously unrecognized phenomenon, though the significance of this finding is unclear, and it is difficult to determine how much of this effect is due to contamination by native tissue (e.g.. muscle in the tongue). A search for a unique molecular trademarks (e.g. enriched mutations or CNAs) in this expression subgroup, as well as characterization of the gene ontology of the DEGs would be prudent.

While the genomic analyses do not have the ability to directly identify an etiological agent, they do show that OT cancers are a fairly homogenous group (even YNT closely resembled OST), and do resemble traditional, smoking-related oral cavity cancers. Given the molecular profile similarities amongst OT cancers and between older smokers and younger non-smokers, and their difference from the molecular profile of HPV-associated HNSCC, it is unlikely that HPV or a similar-acting oncogenic virus is involved in the etiology. It is more likely that secondhand smoke or another environmental exposure is involved in the etiology of OT cancers in non-smokers.

Changes in Five Year Survival

There are many possible causes of a change in 5-year survival, though many are beyond the scope of analysis using the SEER database. Conceptually, an improvement in 5-year disease survival can be due to changes in 1) the underlying disease (etiology), 2) changes in diagnosis/early detection (lead time), 3) more effective treatment of the

disease (management), 4) or changes in the affected population (age or comorbidities). Though the majority of these possibilities cannot be fully explored through the SEER database, we discuss them here for the possibility of future investigation into these trends

Regarding etiology and the affected population, the primary causes of head and neck SCC have long been considered environmental carcinogens such as tobacco and alcohol. However, an increasing number of non-smoking, non-drinking patients are being affected. The finding that OTSCC affects a different demographic [50] could be an indication that the etiologic landscape of OCSCC (especially oral tongue cancer) is changing. Rates of HPV related cancers are clearly on the rise [19, 51], and have an improved prognosis compared to HPV-negative HNSCC [52, 53], but HPV is rarely found in oral cavity cancers. As speculated above, other HPV strains, tumorigenic viruses, or even other environmental causes could all be possible drivers of this change. Myers et al ([34]) recently analyzed the molecular characteristics of OTSCC in young, non-smoking patients and found that cancers from this younger, non-smoking cohort were very similar to OTSCC in older, smoking patients, and very different from HPV-related cancers of the head and neck. This makes it unlikely that HPV is causing these tumors. Regardless, it may be that diagnosis of OCSCC in a younger non-smoking cohort will improve overall survival based on natural lifespan and fewer recognized or unrecognized comorbid conditions in the non-smokers.

Regarding early detection, we observed improvements in survival for every stage, so it is unlikely that earlier detection or stage migration alone accounts for all of the effect seen. However, we did find a trend towards lower stages at diagnosis over time, driven by lower T and N classification upon diagnosis. Diagnosis at lower stage was

found despite rigorous guidelines and newer technologies that would tend to lead to higher staging of patients. No population-based screening programs have been implemented in the US, and existing smaller screening programs (such as OHANCAW [54]) have not been studied to determine their effectiveness on a population scale.

With respect to management, the treatment of OCSCC has evolved as well. We found that more surgery and neck dissections were being performed, especially for early stage disease. Later stages and older age groups were being treated with less surgery and more radiation, possibly indicating recognition of when palliation is the best approach. Treatment has also evolved in many ways which we could not evaluate. Some of these include: emphasis on negative margin surgical resection, improvements in reconstructive surgical techniques, better chemo-radiotherapy regimens, a defined role for selective and elective neck dissections [55], new targeted drugs [56, 57], and advances in radiotherapy such as intensity modulated radiotherapy and hyperfractionation [58, 59]. Though speculative, it also is possible that greater access to these standards, or a transition to care at higher-volume academic centers has resulted in better outcomes overall.

Limitations

While SEER provides a large sample size from which to calculate incidence and survival, it has limitations as well [60]. Though SEER records data from roughly 26% of the US population, it does not necessarily represent the entire United States. Use of a national database is also limited by variations in data reporting among sites, incomplete data entry, and migration of patient populations, though SEER is audited to limit such errors [60]. For our purposes, there was no information on tobacco/alcohol use,

chemotherapy, HPV status, and many other clinical/treatment variables which might influence incidence and survival. Single or multi-institutional studies would be better suited for examining trends in these variables with respect to survival. It should also be noted that trends in 5-year survival do not directly equate to trends mortality or prognosis, due to confounding factors such as changing diagnostic habits [61] and changes in incidence. Lastly, because of a shorter follow up time, the number of censored patients increases with more recent time periods, which may increase the likelihood of bias.

Conclusion

In conclusion, survival has improved in OCSCC. Our analysis suggests that this may be partially related to detection at earlier stages and a younger/healthier population affected. It is possible that improvements in diagnostics and treatment, or even a change in the disease etiology are also playing a role. Further study of each of these individual elements, such as a retrospective analysis of the treatment modalities used at a single institution over the years, may help delineate the specific causes for the increase in survival.

Oral tongue SCC survival has improved even more dramatically than OCSCC over the past 20 years, and now carries a significantly better prognosis compared to other subsites. This improvement correlated to diagnosis at earlier stages and in a younger population. Analysis of genomic data revealed that OT cancers form a distinct molecular subtype. Further study of this subset of tumors, the oral microbiome, and other potential exogenous carcinogens would be necessary to identify any new causative agents.

References

1. Chaturvedi, A.K., et al., *Incidence trends for human papillomavirus-related and -unrelated oral squamous cell carcinomas in the United States.* J Clin Oncol, 2008. **26**(4): p. 612-9.
2. Siegel, R., D. Naishadham, and A. Jemal, *Cancer statistics, 2012.* CA Cancer J Clin, 2012. **62**(1): p. 10-29.
3. Shiboski, C.H., S.C. Shiboski, and S. Silverman, Jr., *Trends in oral cancer rates in the United States, 1973-1996.* Community Dent Oral Epidemiol, 2000. **28**(4): p. 249-56.
4. Silverman, S., Jr., *Demographics and occurrence of oral and pharyngeal cancers. The outcomes, the trends, the challenge.* J Am Dent Assoc, 2001. **132 Suppl**: p. 7S-11S.
5. Hashibe, M., et al., *Socioeconomic status, lifestyle factors and oral premalignant lesions.* Oral Oncol, 2003. **39**(7): p. 664-71.
6. Decker, J. and J.C. Goldstein, *Risk factors in head and neck cancer.* N Engl J Med, 1982. **306**(19): p. 1151-5.
7. Vokes, E.E., et al., *Head and neck cancer.* N Engl J Med, 1993. **328**(3): p. 184-94.
8. Chaturvedi, A.K., et al., *Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers.* J Clin Oncol, 2013. **31**(36): p. 4550-9.
9. The Cancer Genome Atlas, N., *Comprehensive genomic characterization of head and neck squamous cell carcinomas.* Nature, 2015. **517**(7536): p. 576-582.
10. Dahlstrom, K.R., et al., *Differences in history of sexual behavior between patients with oropharyngeal squamous cell carcinoma and patients with squamous cell carcinoma at other head and neck sites.* Head Neck, 2011. **33**(6): p. 847-55.
11. Rettig, E., A.P. Kiess, and C. Fakhry, *The role of sexual behavior in head and neck cancer: implications for prevention and therapy.* Expert Rev Anticancer Ther, 2015. **15**(1): p. 35-49.
12. Friedman, J.M., M.J. Stavas, and A.J. Cmelak, *Clinical and scientific impact of human papillomavirus on head and neck cancer.* World J Clin Oncol, 2014. **5**(4): p. 781-91.
13. Rusthoven, K.E., et al., *Survival and patterns of relapse in patients with oral tongue cancer.* J Oral Maxillofac Surg, 2010. **68**(3): p. 584-9.
14. Bello, I.O., Y. Soini, and T. Salo, *Prognostic evaluation of oral tongue cancer: means, markers and perspectives (I).* Oral Oncol, 2010. **46**(9): p. 630-5.
15. Rusthoven, K., et al., *Poor prognosis in patients with stage I and II oral tongue squamous cell carcinoma.* Cancer, 2008. **112**(2): p. 345-51.
16. Bell, R.B., et al., *Tongue cancer: Is there a difference in survival compared with other subsites in the oral cavity?* J Oral Maxillofac Surg, 2007. **65**(2): p. 229-36.
17. Yao, M., et al., *The failure patterns of oral cavity squamous cell carcinoma after intensity-modulated radiotherapy-the university of iowa experience.* Int J Radiat Oncol Biol Phys, 2007. **67**(5): p. 1332-41.
18. Listl, S., et al., *Survival of patients with oral cavity cancer in Germany.* PLoS One, 2013. **8**(1): p. e53415.
19. Society, A.C., *Cancer Facts & Figures 2012.* Atlanta: American Cancer Society, 2012.
20. (CDC), C.f.D.C.a.P., *Alcohol-Related Disease Impact (ARDI).* Centers for Disease Control and Prevention (CDC), Atlanta, 2012.
21. Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014, U.S.D.o.H.a.H.S.T.H.C.o.S.Y.o.P.A.R.o.t.S.G.A.U.S.D.o.

22. Patel, S.C., et al., *Increasing incidence of oral tongue squamous cell carcinoma in young white women, age 18 to 44 years.* J Clin Oncol, 2011. **29**(11): p. 1488-94.
23. Saba, N.F., et al., *Gender and ethnic disparities in incidence and survival of squamous cell carcinoma of the oral tongue, base of tongue, and tonsils: a surveillance, epidemiology and end results program-based analysis.* Oncology, 2011. **81**(1): p. 12-20.
24. Myers, J.N., et al., *Squamous cell carcinoma of the tongue in young adults: increasing incidence and factors that predict treatment outcomes.* Otolaryngol Head Neck Surg, 2000. **122**(1): p. 44-51.
25. Brown, L.M., D.P. Check, and S.S. Devesa, *Oral cavity and pharynx cancer incidence trends by subsite in the United States: changing gender patterns.* J Oncol, 2012. **2012**: p. 649498.
26. Pulte, D. and H. Brenner, *Changes in survival in head and neck cancers in the late 20th and early 21st century: a period analysis.* Oncologist, 2010. **15**(9): p. 994-1001.
27. Mehta, V., G.P. Yu, and S.P. Schantz, *Population-based analysis of oral and oropharyngeal carcinoma: changing trends of histopathologic differentiation, survival and patient demographics.* Laryngoscope, 2010. **120**(11): p. 2203-12.
28. Carvalho, A.L., et al., *Trends in incidence and prognosis for head and neck cancer in the United States: a site-specific analysis of the SEER database.* Int J Cancer, 2005. **114**(5): p. 806-16.
29. Hoffman, H.T., et al., *Laryngeal cancer in the United States: changes in demographics, patterns of care, and survival.* Laryngoscope, 2006. **116**(9 Pt 2 Suppl 111): p. 1-13.
30. Amit, M., et al., *Improvement in survival of patients with oral cavity squamous cell carcinoma: An international collaborative study.* Cancer, 2013. **119**(24): p. 4242-8.
31. Salem, A., *Dismissing links between HPV and aggressive tongue cancer in young patients.* Ann Oncol, 2010. **21**(1): p. 13-7.
32. Combes, J.D. and S. Franceschi, *Role of human papillomavirus in non-oropharyngeal head and neck cancers.* Oral Oncol, 2013.
33. Mirghani, H., et al., *Do high-risk human papillomaviruses cause oral cavity squamous cell carcinoma?* Oral Oncol, 2015. **51**(3): p. 229-236.
34. Pickering, C.R., et al., *Squamous cell carcinoma of the oral tongue in young non-smokers is genomically similar to tumors in older smokers.* Clin Cancer Res, 2014. **20**(14): p. 3842-8.
35. Agrama, M.T., et al., *Node counts in neck dissection: are they useful in outcomes research?* Otolaryngol Head Neck Surg, 2001. **124**(4): p. 433-5.
36. Ederer, F., L.M. Axtell, and S.J. Cutler, *The relative survival rate: a statistical methodology.* Natl Cancer Inst Monogr, 1961. **6**: p. 101-21.
37. Goldman, M., et al., *The UCSC Cancer Genomics Browser: update 2013.* Nucleic Acids Res, 2013. **41**(Database issue): p. D949-54.
38. Cline, M.S., et al., *Exploring TCGA Pan-Cancer data at the UCSC Cancer Genomics Browser.* Sci Rep, 2013. **3**: p. 2652.
39. Reich, M., et al., *GenePattern 2.0.* Nat Genet, 2006. **38**(5): p. 500-1.
40. Lawrence, M.S., et al., *Mutational heterogeneity in cancer and the search for new cancer-associated genes.* Nature, 2013. **499**(7457): p. 214-8.
41. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. **6**(269): p. pl1.
42. Gould, J., et al., *Comparative gene marker selection suite.* Bioinformatics, 2006. **22**(15): p. 1924-5.

43. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
44. Eden, E., et al., *GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists.* BMC Bioinformatics, 2009. **10**: p. 48.
45. Forbes, S.A., et al., *The Catalogue of Somatic Mutations in Cancer (COSMIC).* Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.], 2008. **CHAPTER**: p. Unit-10.11.
46. Eisen, M.B., et al., *Cluster analysis and display of genome-wide expression patterns.* Proc Natl Acad Sci U S A, 1998. **95**(25): p. 14863-8.
47. Chung, C.H., et al., *Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression.* Cancer Cell, 2004. **5**(5): p. 489-500.
48. Walter, V., et al., *Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes.* PLoS One, 2013. **8**(2): p. e56823.
49. Shiboski, C.H., B.L. Schmidt, and R.C. Jordan, *Tongue and tonsil carcinoma: increasing trends in the U.S. population ages 20-44 years.* Cancer, 2005. **103**(9): p. 1843-9.
50. Li, R., et al., *Distinct epidemiologic characteristics of oral tongue cancer patients.* Otolaryngol Head Neck Surg, 2013. **148**(5): p. 792-6.
51. Hammarstedt, L., et al., *Human papillomavirus as a risk factor for the increase in incidence of tonsillar cancer.* Int J Cancer, 2006. **119**(11): p. 2620-3.
52. Dahlgren, L., et al., *Human papillomavirus is more common in base of tongue than in mobile tongue cancer and is a favorable prognostic factor in base of tongue cancer patients.* Int J Cancer, 2004. **112**(6): p. 1015-9.
53. Ang, K.K., et al., *Human papillomavirus and survival of patients with oropharyngeal cancer.* N Engl J Med, 2010. **363**(1): p. 24-35.
54. http://www.headandneck.org/.
55. Capote, A., et al., *Elective neck dissection in early-stage oral squamous cell carcinoma--does it influence recurrence and survival?* Head Neck, 2007. **29**(1): p. 3-11.
56. Rampino, M., et al., *Efficacy and feasibility of induction chemotherapy and radiotherapy plus cetuximab in head and neck cancer.* Anticancer Res, 2012. **32**(1): p. 195-9.
57. Acevedo-Henao, C.M., et al., *Radiotherapy combined with cetuximab for locally advanced head and neck cancer: results and toxicity.* Cancer Radiother, 2012. **16**(7): p. 601-3.
58. Budach, W., et al., *A meta-analysis of hyperfractionated and accelerated radiotherapy and combined chemotherapy and radiotherapy regimens in unresected locally advanced squamous cell carcinoma of the head and neck.* BMC Cancer, 2006. **6**: p. 28.
59. Chen, P.Y., et al., *Intensity-modulated radiotherapy improves outcomes in postoperative patients with squamous cell carcinoma of the oral cavity.* Oral Oncol, 2012. **48**(8): p. 747-52.
60. Harlan, L.C. and B.F. Hankey, *The surveillance, epidemiology, and end-results program database as a resource for conducting descriptive epidemiologic and clinical studies.* J Clin Oncol, 2003. **21**(12): p. 2232-3.
61. Welch, H.G., L.M. Schwartz, and S. Woloshin, *Are increasing 5-year survival rates evidence of success against cancer?* JAMA, 2000. **283**(22): p. 2975-8.

Tables

Table 1 – Patient, Tumor, and Treatment Characteristics By Period

	1988-92		1993-97		1998-02		2003-07		change	p value
	%	n	%	n	%	n	%	n		
Age										**p < 0.001**
18-40	5.2%	(114)	5.9%	(169)	6.3%	(306)	4.9%	(313)	-0.3%	
41-55	20.7%	(455)	23.1%	(661)	26.6%	(1,299)	26.5%	(1,688)	+5.8%	
56-70	39.2%	(861)	34.4%	(983)	32.1%	(1,570)	35.4%	(2,254)	-3.8%	
71-85	29.1%	(639)	30.9%	(885)	28.6%	(1,397)	26.5%	(1,683)	-2.7%	
85+	5.7%	(125)	5.7%	(162)	6.4%	(312)	6.6%	(422)	+0.9%	
Total		(2,194)		(2,860)		(4,884)		(6360)		
Sex										**p < 0.001**
Female	35.6%	(782)	41.6%	(1,189)	40.3%	(1,968)	39.3%	(2,499)	+3.6%	
Male	64.4%	(1,412)	58.4%	(1,671)	59.7%	(2,916)	60.7%	(3,861)	-3.6%	
Total		(2,194)		(2,860)		(4,884)		(6,360)		
Race										**p = 0.001**
Black	9.5%	(207)	9.6%	(273)	8.9%	(432)	8.1%	(508)	-1.4%	
Other	5.0%	(110)	7.1%	(203)	6.4%	(307)	7.5%	(469)	+2.5%	
White	85.5%	(1,872)	83.3%	(2,366)	84.7%	(4,094)	84.4%	(5,297)	-1.1%	
Total		(2,189)		(2,842)		(4,833)		(6,274)		
Primary site										**p < 0.001**
lip	4.2%	(93)	4.0%	(113)	5.4%	(266)	4.8%	(306)	+0.6%	
tongue	30.1%	(660)	32.5%	(929)	35.7%	(1,744)	37.8%	(2,405)	+7.7%	
gum	12.3%	(269)	13.3%	(381)	12.5%	(611)	13.1%	(830)	+0.8%	
floor of mouth	34.2%	(751)	30.3%	(867)	26.5%	(1,296)	23.7%	(1,508)	-10.5%	
hard palate	2.1%	(46)	3.3%	(94)	3.3%	(160)	3.2%	(202)	+1.1%	
buccal	7.6%	(167)	7.6%	(218)	8.8%	(428)	9.5%	(605)	+1.9%	
retromolar	9.5%	(208)	9.0%	(258)	7.8%	(379)	7.9%	(504)	-1.6%	
Total		(2,194)		(2,860)		(4,884)		(6,360)		
Grade										**p < 0.001**
Well diff.	32.6%	(588)	26.7%	(659)	25.7%	(1,077)	26.0%	(1,416)	-6.6%	
Mod. diff.	49.2%	(887)	52.9%	(1,306)	56.1%	(2,353)	56.2%	(3,062)	+7.0%	
Poorly diff.	17.7%	(328)	19.9%	(502)	17.7%	(768)	17.5%	(971)	-0.3%	
Total		(1,803)		(2,467)		(4,198)		(5,449)		
Stage										**p < 0.001**
1	22.5%	(269)	24.7%	(419)	28.5%	(906)	34.0%	(1,685)	+11.4%	
2	24.4%	(291)	24.2%	(409)	21.1%	(671)	19.3%	(959)	-5.0%	
3	19.5%	(233)	14.8%	(251)	14.9%	(474)	15.6%	(776)	-3.9%	
4	33.6%	(401)	36.3%	(614)	35.4%	(1,126)	31.0%	(1,539)	-2.6%	
Total		(1,194)		(1,693)		(3,177)		(4,959)		
T stage										**p < 0.001**
T1	27.3%	(398)	30.7%	(613)	34.2%	(1,198)	40.7%	(2,045)	+13.4%	
T2	35.4%	(516)	33.5%	(669)	30.4%	(1,065)	30.2%	(1,516)	-5.2%	
T3	14.2%	(207)	12.1%	(242)	11.7%	(408)	11.6%	(581)	-2.6%	
T4	23.1%	(337)	23.8%	(476)	23.7%	(831)	17.6%	(882)	-5.6%	
Total		(1,458)		(2,000)		(3,502)		(5,024)		
N stage										**p < 0.001**
N0	62.6%	(1,015)	66.4%	(1,477)	67.7%	(2,833)	68.4%	(3,989)	+5.8%	
N1	7.5%	(122)	7.4%	(164)	7.7%	(324)	12.9%	(751)	+5.4%	
N2	28.6%	(464)	25.1%	(559)	23.7%	(990)	17.6%	(1,027)	-11.0%	
N3	1.3%	(21)	1.2%	(26)	0.9%	(38)	1.1%	(67)		
Total		(1,622)		(2,226)		(4,185)		(5,834)		
M stage										**p = 0.739**
M0	98.1%	(1,999)	97.8%	(2,618)	97.7%	(4,542)	97.9%	(5,794)	-0.1%	
M1	1.9%	(39)	2.2%	(58)	2.3%	(107)	2.1%	(122)	+0.1%	
Total		(2,038)		(2,676)		(4,649)		(5,916)		
Treatment										**p < 0.001**
surgery	49.1%	(981)	47.2%	(1,219)	49.9%	(2,160)	52.7%	(2,972)	+3.5%	
surgery+rad	34.0%	(678)	36.5%	(944)	34.5%	(1,493)	32.5%	(1,837)	-1.4%	
rad only	16.9%	(338)	16.3%	(420)	15.6%	(676)	14.8%	(835)	-2.1%	
Total		(1,997)		(2,583)		(4,329)		(5,644)		
Neck dissection										**p < 0.001**
no	75.4%	(1,624)	73.9%	(2,095)	69.2%	(3,330)	67.9%	(4,187)	-7.5%	
yes	24.6%	(530)	26.1%	(740)	30.8%	(1,480)	32.1%	(1,981)	+7.5%	
Total		(2,154)		(2,835)		(4,810)		(6,168)		

Table 1. Temporal trends of patient, tumor, and treatment characteristics, shown as proportions of the total case-load. The number of cases (n) are shown in parenthesis, and the absolute change in proportion from the first to last time period is found in the right-most column. Gender, Other = American Indian/AK Native, Asian/Pacific Islander.

Table 2 – Five Year Overall Survival By Period

	1988-92	1993-97	1998-02	2003-05	change	p value
Overall	39.9%	41.3%	45.2%	50.4%	+10.6%	<0.001
Age						
18-40	64.0%	63.3%	64.9%	74.4%	+10.4%	0.030
41-55	46.1%	52.3%	55.5%	58.6%	+12.5%	<0.001
56-70	42.6%	43.2%	49.0%	50.4%	+7.7%	<0.001
71-85	31.8%	31.0%	33.0%	38.0%	+6.2%	0.005
85+	16.0%	17.3%	14.5%	16.5%	+0.5%	0.448
Sex						
Female	43.9%	44.1%	44.4%	50.1%	+6.3%	0.002
Male	37.8%	40.1%	46.0%	50.6%	+12.7%	<0.001
Race						
Black	19.4%	25.4%	30.9%	35.3%	+15.9%	<0.001
Other	42.1%	50.3%	41.7%	55.2%	+13.1%	0.030
White	41.9%	42.6%	46.9%	51.1%	+9.1%	<0.001
Primary site						
lip	65.4%	74.9%	74.5%	70.1%	+4.7%	0.222
tongue	44.3%	48.5%	52.9%	58.6%	+14.3%	<0.001
gum	40.6%	40.7%	44.2%	47.7%	+7.1%	0.031
floor of mouth	36.2%	36.4%	39.2%	41.6%	+5.5%	0.011
hard palate	45.5%	28.9%	30.4%	43.0%	-2.4%	0.610
buccal	36.8%	34.6%	32.0%	43.5%	+6.7%	0.075
retromolar	28.2%	30.5%	34.3%	39.6%	+11.4%	0.004
Grade						
Well diff.	50.0%	53.2%	57.0%	59.4%	+9.4%	<0.001
Moderately diff.	35.9%	39.2%	43.4%	49.3%	+13.4%	<0.001
Poorly diff.	26.9%	27.9%	29.2%	37.3%	+10.4%	0.005
Stage						
1	66.8%	68.5%	72.7%	73.4%	+6.6%	0.045
2	41.4%	47.9%	54.5%	52.8%	+11.4%	0.009
3	28.8%	27.3%	33.2%	41.0%	+12.2%	0.005
4	20.3%	22.9%	27.0%	24.9%	+4.6%	0.256
T stage						
T1	41.9%	42.6%	46.9%	51.1%	+9.1%	0.042
T2	38.3%	38.5%	45.2%	46.7%	+8.3%	0.020
T3	24.5%	29.0%	29.2%	29.7%	+5.2%	0.466
T4	24.2%	26.1%	27.6%	25.4%	+1.2%	0.709
N stage						
N0	50.8%	53.8%	57.2%	61.1%	+10.3%	<0.001
N1	20.7%	20.5%	27.7%	36.5%	+15.8%	<0.001
N2	17.5%	16.6%	20.3%	22.9%	+5.4%	0.057
N3	0.0%	4.0%	13.6%	12.6%	+12.6%	0.648
M stage						
M0	41.2%	43.0%	46.4%	51.5%	+10.2%	<0.001
M1	5.4%	10.3%	11.0%	3.8%	-1.6%	0.699
Treatment						
surgery	57.9%	60.6%	64.9%	69.3%	+11.4%	<0.001
surgery+rad	32.8%	32.1%	36.9%	44.2%	+11.4%	<0.001
rad only	11.3%	15.8%	15.6%	18.4%	+7.1%	0.048
Neck dissection						
no	41.8%	43.5%	46.1%	49.9%	+8.1%	<0.001
yes	34.8%	37.3%	43.7%	50.7%	+15.9%	<0.001

Table 2. Overall 5-year survival stratified by all patient, tumor, and treatment characteristics. In the right-most column is the change in survival rate from the first to last time period. Gender: Other = American Indian/AK Native, Asian/Pacific Islander. Note that survival only includes the years 2003-05 for the final time period.

Table 3 - Treatment by stage over time

Stage 1		**1988-92**	**1993-97**	**1998-02**	**2003-07**	**Change**
	Surgery	79.5%	80.2%	82.7%	85.1%	**5.7%**
	Surgery+Rad	16.8%	17.3%	14.7%	11.9%	**-4.9%**
	Radiation Only	3.7%	2.4%	2.6%	2.9%	**-0.8%**
Stage 2		**1988-92**	**1993-97**	**1998-02**	**2003-07**	
	Surgery	55.2%	59.3%	57.5%	51.3%	**-3.9%**
	Surgery+Rad	31.3%	27.5%	31.5%	34.7%	**3.5%**
	Radiation Only	13.5%	13.1%	11.0%	14.0%	**0.4%**
Stage 3		**1988-92**	**1993-97**	**1998-02**	**2003-07**	
	Surgery	27.2%	18.9%	27.0%	28.3%	**1.1%**
	Surgery+Rad	54.8%	60.1%	55.0%	55.1%	**0.2%**
	Radiation Only	18.0%	21.0%	18.0%	16.6%	**-1.4%**
Stage 4		**1988-92**	**1993-97**	**1998-02**	**2003-07**	
	Surgery	27.5%	20.4%	20.9%	18.1%	**-9.3%**
	Surgery+Rad	60.4%	63.4%	63.8%	57.6%	**-2.8%**
	Radiation Only	12.1%	16.2%	15.3%	24.3%	**12.1%**

Table 4 - Neck Dissection by stage over time

Stage 1		**1988-92**	**1993-97**	**1998-02**	**2003-07**	**Change**
	no	89.0%	84.0%	81.1%	78.5%	-10.4%
	yes	11.0%	16.0%	18.9%	21.3%	10.3%
Stage 2		**1988-92**	**1993-97**	**1998-02**	**2003-07**	
	no	73.6%	72.6%	70.5%	66.5%	-7.1%
	yes	26.1%	27.4%	29.5%	33.5%	7.4%
Stage 3		**1988-92**	**1993-97**	**1998-02**	**2003-07**	
	no	48.4%	51.9%	46.8%	47.0%	-1.3%
	yes	51.6%	48.1%	53.2%	52.6%	1.0%
Stage 4		**1988-92**	**1993-97**	**1998-02**	**2003-07**	
	no	47.3%	42.7%	41.9%	49.5%	2.3%
	yes	52.2%	57.1%	58.0%	50.1%	-2.2%

Table 5 – Clinical Attributes of TCGA Cohort

	Oral Tongue	Other Oral Cavity	Oropharynx	Larynx/Hypopharynx	Total
N	129	178	81	123	511
Age < 55	49	47	38	28	162
Age > 55	78	133	43	95	349
Mean Age	58.0	64.2	55.9	62.1	60.8
Female %	36.4%	31.5%	14.8%	17.9%	26.8%
White %	85.3%	85.4%	92.6%	80.5%	85.3%
Smoker	74	130	51	111	366
Non-Smoker	52	41	29	9	131
Unknown	3	7	1	3	14
% Smokers	58.7%	76.0%	63.8%	92.5%	73.6%
HPV+ (p16 or ISH)	3	2	34	4	43
HPV-	24	26	8	19	77
Not Reported	102	150	39	100	391
% Tested Positive	11.1%	7.1%	81.0%	17.4%	35.8%

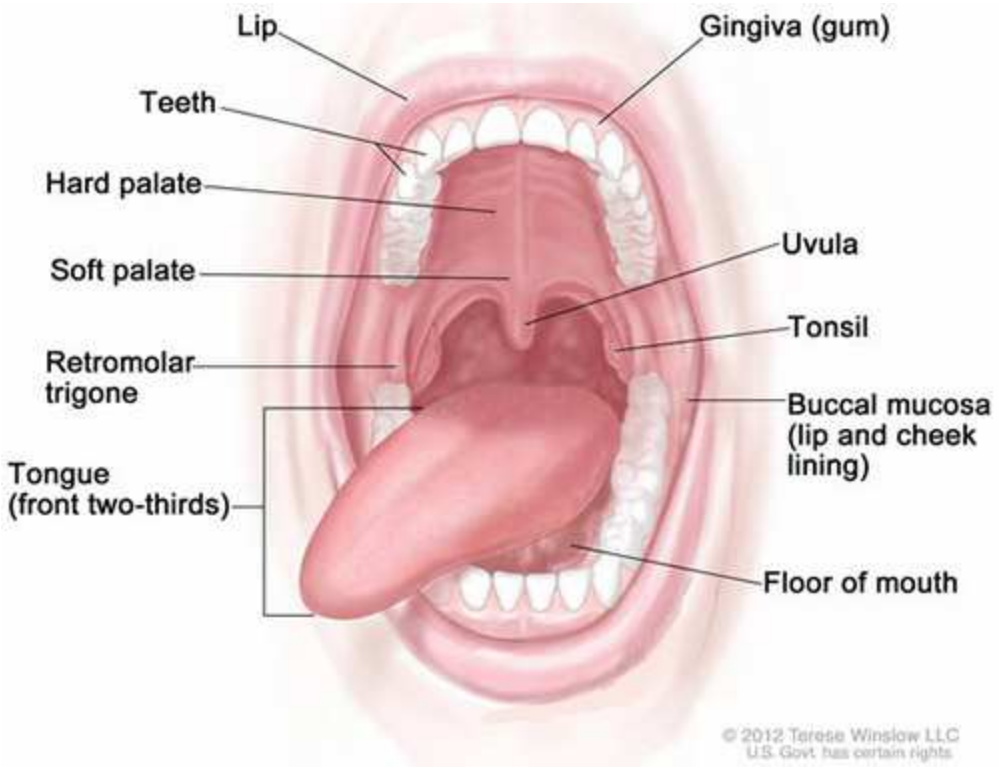
Table 6 – Commonly Mutated Genes (TCGA HNSCC)

Cohort (N)	YNT (21)	OST (46)	OT (125)	OC (175)	OP (76)	LH (117)	Total HN (493)
Mean # Mutations	**64.6**	**120.2**	**93.8**	**175.4**	**113.1**	**224.2**	**156.7**
TP53	90.5%	73.9%	80.8%	72.6%	28.9%	84.6%	70.8%
FAT1	14.3%	30.4%	21.6%	32.0%	7.9%	21.4%	23.1%
CDKN2A	42.9%	21.7%	29.6%	22.9%	6.6%	23.1%	22.1%
PIK3CA	9.5%	21.7%	17.6%	17.1%	19.7%	22.2%	18.9%
NOTCH1	19.0%	17.4%	20.0%	20.6%	6.6%	17.1%	17.4%
NSD1	0.0%	10.9%	5.6%	9.7%	6.6%	27.4%	12.4%
CASP8	4.8%	8.7%	8.0%	21.1%	2.6%	1.7%	10.3%
FBXW7	0.0%	4.3%	4.0%	8.6%	3.9%	6.8%	6.3%
AJUBA	4.8%	2.2%	3.2%	6.3%	2.6%	10.3%	5.9%
HRAS	4.8%	8.7%	7.2%	10.3%	0.0%	0.9%	5.7%
HLA-A	0.0%	0.0%	2.4%	7.4%	7.9%	4.3%	5.5%
NFE2L2	0.0%	8.7%	4.8%	6.3%	2.6%	6.0%	5.3%
TGFBR2	4.8%	6.5%	4.0%	6.3%	2.6%	2.6%	4.3%
RB1	0.0%	2.2%	4.0%	1.1%	6.6%	4.3%	3.4%
CUL3	0.0%	2.2%	0.8%	2.3%	1.3%	6.8%	2.8%
PTEN	0.0%	2.2%	1.6%	1.7%	3.9%	3.4%	2.4%
PIK3R1	0.0%	0.0%	0.8%	1.7%	3.9%	0.9%	1.6%
TRAF3	0.0%	0.0%	0.8%	0.6%	1.3%	1.7%	1.0%

Table 7 – Commonly Mutated Genes (Whole Exome)

YNT N = 21		OST N = 46		OT N = 125		OC N = 175		OP N = 76		L/H N = 117	
TP53	90%	TP53	74%	TP53	81%	TP53	73%	TTN	36%	TP53	85%
CDKN2A	43%	TTN	35%	TTN	30%	TTN	44%	TP53	29%	TTN	62%
MUC16	24%	FAT1	30%	CDKN2A	30%	FAT1	32%	PIK3CA	20%	CSMD3	41%
TTN	24%	CDKN2A	22%	FRG1B	23%	CDKN2A	23%	MUC16	18%	SNHG14	38%
NOTCH1	19%	PIK3CA	22%	FAT1	22%	CASP8	21%	SYNE1	16%	SYNE1	30%
RRN3P2	19%	FRG1B	22%	NOTCH1	20%	NOTCH1	21%	HSD17B7P2	16%	MUC16	29%
DYNC1H1	19%	CSMD3	20%	PIK3CA	18%	MUC16	21%	LRP1B	16%	LRP1B	29%
PCSK5	14%	BAGE2	20%	PCLO	17%	FRG1B	19%	FRG1B	14%	NSD1	27%
FAT1	14%	PCLO	20%	CSMD3	15%	SYNE1	18%	RP11-798G7.5	14%	KMT2D	26%
COL1A2	14%	AC008103.5	20%	BAGE2	15%	SNHG14	18%	KMT2D	13%	USH2A	26%
RASA1	14%	NOTCH1	17%	AC008103.5	14%	PIK3CA	17%	LINC00969	13%	FRG1B	26%
NBPF1	14%	USH2A	17%	MUC16	14%	BAGE2	17%	SNHG14	13%	BAGE2	26%
SNHG14	14%	DNAH5	15%	DNAH5	13%	CSMD3	16%	TSSC2	13%	PCLO	24%
SZT2	14%	CROCCP2	15%	SYNE1	13%	KMT2D	16%	DST	12%	CDKN2A	23%
HUWE1	14%	ZFHX4	15%	HSD17B7P2	11%	FLG	16%	CSMD3	12%	RYR2	23%
PMS2CL	14%	AHNAK	15%	USH2A	10%	LRP1B	15%	MUC4	12%	PKHD1L1	23%
IKBKB	14%	MUC16	13%	TSSC2	10%	PCLO	14%	NIPBL	11%	PIK3CA	22%
PCLO	14%	SYNE1	13%	LINC00969	10%	LL22NC03-80A10.6	14%	MACF1	11%	LAMA2	22%
ZMYND11	14%	HSD17B7P2	13%	LRP1B	10%	LINC00969	13%	FLG	11%	FAT1	21%
LL22NC03-80A10.6	14%	TSSC2	13%	TUBB8P7	10%	DNAH5	13%	BAGE2	11%	FLG	21%

Appendix A



http://seer.cancer.gov/statfacts/html/oralcav.html