NIH Public Access
Author Manuscript
Head Neck. Author manuscript; available in PMC 2015 December 21.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Clinical, Genomic and Metagenomic Characterization of Oral Tongue Squamous Cell Carcinoma in Patients Who Do Not Smoke

Ryan Li, MD[1,*], **Daniel L. Faden, MD**[3,*], **Carole Fakhry, MD, MPH**[1,2,*], **Chaz Langelier, MD**[4], **Yuchen Jiao**[5], **Yuxuan Wang, PhD**[5], **Matthew D. Wilkerson, PhD**[6,7], **Chandra Sekhar Pedamallu, PhD**[8,9], **Matthew Old, MD**[10], **James Lang, PhD**[10], **Myriam Loyo, MD**[1], **Sun Mi Ahn, MD**[1], **Marietta Tan, MD**[1], **Zhen Gooi, MD**[1], **Jason Chan, MD**[1], **Jeremy Richmon, MD**[1], **Laura D. Wood, MD, PhD**[11], **Ralph H. Hruban, MD**[11], **Justin Bishop, MD**[11], **William H. Westra, MD**[11], **Christine H. Chung, MD**[12], **Joseph Califano, MD**[2], **Christine G. Gourin, MD, MPH**[1], **Chetan Bettegowda, MD, PhD**[13], **Matthew Meyerson, MD, PhD**[8,9,14], **Nickolas Papadopoulos, PhD**[5], **Kenneth W. Kinzler, PhD**[5], **Bert Vogelstein, MD**[5], **Joseph L. DeRisi, PhD**[15,16], **Wayne M. Koch, MD**, and **Nishant Agrawal, MD**[1,5]

[1]Department of Otolaryngology- Head and Neck Surgery, Johns Hopkins University School of Medicine, Baltimore, MD 21231, USA

[2]Milton J. Dance Head and Neck Center, Greater Baltimore Medical Center, Baltimore, Maryland, 21204

[3]Department of Otolaryngology- Head and Neck Surgery, University of California, San Francisco, San Francisco CA 94131, USA

[4]Department of Medicine, University of California San Francisco, San Francisco, CA 94131, USA

[5]The Ludwig Center and the Howard Hughes Medical Institute at Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21231, USA

[6]Department of Genetics, University of North Carolina, Chapel Hill, NC

[7]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA

[8]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA

[9]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

[10]Department of Otolaryngology-Head and Neck Surgery, Ohio State University, Columbus, OH 42312, USA

Correspondence to: Joseph L. DeRisi; Wayne M. Koch; Nishant Agrawal.

[*]Designates co-authorship

Disclosure of Potential Conflicts of Interest

Under agreements between the Johns Hopkins University, Genzyme, Exact Sciences, Inostics, Qiagen, Invitrogen and Personal Genome Diagnostics, NP, BV, and KWK are entitled to a share of the royalties received by the University on sales of products related to genes and technologies described in this manuscript. NP, BV, and KWK are co-founders of Inostics and Personal Genome Diagnostics, are members of their Scientific Advisory Boards, and own Inostics and Personal Genome Diagnostics stock, which is subject to certain restrictions under Johns Hopkins University policy. The terms of these arrangements are managed by the Johns Hopkins University in accordance with its conflict-of-interest policies.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

[11]Department of Pathology, Johns Hopkins University, Baltimore, MD 21231, USA

[12]Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD 21231, USA

[13]Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, MD 21231, USA

[14]Department of Pathology, Harvard Medical School, Boston, MA, 02115, USA

[15]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

[16]Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA 94131, USA

Abstract

Background—Evidence suggests the incidence of oral tongue squamous cell carcinoma is increasing in young patients, many who have no history of tobacco use.

Methods—We clinically reviewed 89 oral tongue cancer patients. Exomic sequencing of tumor DNA from 6 non-smokers was performed and compared to previously sequenced cases. RNA from 20 tumors was evaluated by massively parallel sequencing to search for potentially oncogenic viruses.

Results—Non-smokers (53 of 89) were younger than smokers (36 of 89) mean 50.4 vs. 61.9 years, P<0.001), and appeared more likely to be female, (58.5% vs. 38.9%, P=0.069). Non-smokers had fewer *TP53* mutations (P=0.02) than smokers. No tumor-associated viruses were detected.

Conclusions—The young age of non-smoker oral tongue cancer patients, and fewer *TP53* mutations suggest a viral role in this disease. Our efforts to identify such a virus were unsuccessful. Further studies are warranted to elucidate the drivers of carcinogenesis in these patients.

Keywords

Head and neck cancer; oral tongue; squamous cell carcinoma; next-generation sequencing; nonsmokers

Introduction

Historically, tobacco consumption has been the most significant risk factor for developing head and neck squamous cell carcinoma (HNSCC), while a multiplicative increase in cancer risk has been shown with concomitant alcohol abuse [1–3]. An overall decline in tobacco consumption has led to a decreasing incidence of head and neck squamous cell carcinoma. The most notable exception is the increasing incidence of oropharyngeal SCC, attributable to infection with Human Papilloma Virus (HPV). More recently, an increasing incidence of oral tongue cancers in young patients has also been described [4]. This subsite is distinct from the oropharynx, and a robust, multi-institutional study confirmed that HPV does not play a significant role in the etiology of oral tongue SCCs [5]. Epidemiologically it appears that oral

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

tongue SCC is most commonly arising in young, white, female patients [4]. In fact, in young HNSCC patients without a history of tobacco use, or association with HPV, the oral tongue appears to be the most common subsite of origin [6].

This important clinical entity has ignited interest in identifying undiscovered environmental and biological risk factors beyond tobacco use. We, therefore, were interested in comparing clinico-demographic differences between non-smokers and smokers with oral tongue cancer and sought to identify differences in the genomic and metagenomic profiles between these two subgroups of oral tongue cancer patients.

Materials and Methods

Patients

After gaining approval of the Institutional Review Board (IRB) at the Johns Hopkins Medical Institutions, a retrospective review of patients treated for oral tongue SCC from 1984 to 2011, at Johns Hopkins Hospital (JHH) was performed. All patients were surgically treated and had no prior history of radiation or chemotherapy. All patients provided written informed consent for obtaining medical information and tissue for research purposes.

Clinical Data

Clinical and demographic data including age, gender, race, tobacco and alcohol use were abstracted from electronic medical records. Patients with no tobacco smoking or chewing history were classified as never-smokers. Subjects with any history of tobacco consumption were categorized as smokers. Clinicopathologic variables of interest included tumor histologic grade[7], TNM classification[8], treatment regimen, and disease status (no recurrence, local, regional, or distant recurrence).

Preparation of clinical samples

Fresh-frozen surgically resected carcinoma and matched blood were obtained from patients under IRB-approved protocols from Johns Hopkins University (JHH), University of California, San Francisco (UCSF) and Ohio State University (OSU). All samples were snap frozen at the time of surgery and stored at −80 until the time of processing. Six tumors were selected from JHU for exome sequencing based on lack of tobacco use and stringent quality assessment of normal and tumor DNA. In our tumor bank, the availability of oral tongue tumors from non-smokers with high tumor cellularity and sufficient high quality tumor and matched normal DNA is limited precluding a larger cohort. Ten samples from UCSF (8 tumors, 1 healthy control and 1 HPV positive oropharyngeal tumor) and seven tumor samples from OSU, in addition to five of the six samples from JHU, were selected for transcriptomic analysis. Prior to analysis, the diagnosis of each specimen underwent central pathological review. Tumor tissue was analyzed by frozen section to assess neoplastic cellularity. Tumors were macrodissected (JHU and OSU) to remove residual normal tissue and enhance neoplastic cellularity, as confirmed by serial frozen sections. An estimated average of 60% neoplastic cellularity was obtained.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Preparation of genomic DNA and cDNA libraries

Genomic DNA and cDNA libraries were prepared following Illumina's (Illumina, San Diego, CA) suggested protocol with the following modifications. **(1)** gDNA from tumor, gDNA from normal cells, or cDNA in 100 microliters (µl) of buffer was fragmented in a Covaris sonicator (Covaris, Woburn, MA) to a size of 100–500 bp. DNA was purified with a PCR purification kit (Cat # 28104, Qiagen, Valencia, CA) and eluted in elution buffer included in the kit. **(2)** Purified, fragmented DNA was mixed with 40 µl of H_2O, 10 µl of 10 × T4 ligase buffer with 10 mM ATP, 4 µl of 10 mM dNTP, 5 µl of T4 DNA polymerase, 1 µl of Klenow Polymerase, and 5 µl of T4 polynucleotide kinase. All reagents used for this step and those described below were from New England Biolabs (NEB, Ipswich, MA) unless otherwise specified. The 100 µl end-repair mixture was incubated at 20°C for 30 min, purified by a PCR purification kit (Cat # 28104, Qiagen) and eluted with 32 µl of elution buffer (EB). **(3)** To A-tail, all 32 µl of end-repaired DNA was mixed with 5 µl of 10 × Buffer (NEB buffer 2), 10 µl of 1 mM dATP and 3 µl of Klenow (exo-). The 50 µl mixture was incubated at 37°C for 30 min before DNA was purified with a MinElute PCR purification kit (Cat # 28004, Qiagen). Purified DNA was eluted with 12.5 µl of 70°C EB and obtained with 10 µl of EB. **(4)** For adaptor ligation, 10 µl of A-tailed DNA was mixed with 10 µl of PE-adaptor (Illumina), 25 µl of 2x Rapid ligase buffer and 5 µl of Rapid Ligase. The ligation mixture was incubated at room temperature (RT) or 20°C for 15 min. **(5)** To purify adaptor-ligated DNA, 50 µl of ligation mixture from step (4) was mixed with 200 µl of NT buffer from NucleoSpin Extract II kit (cat# 636972, Clontech, Mountain View, CA) and loaded into a NucleoSpin column. The column was centrifuged at 14,000 g in a desktop centrifuge for 1 min, washed once with 600 µl of wash buffer (NT3 from Clontech), and centrifuged again for 2 min to dry completely. DNA was eluted in 50 µl elution buffer included in the kit. **(6)** To obtain an amplified library, ten PCRs of 25 µl each were set up, each including 13.25 µl of H_2O, 5 µl of 5 × Phusion HF buffer, 0.5 µl of a dNTP mix containing 10 mM of each dNTP, 0.5 µl of Illumina PE primer #1, 0.5 µl of Illumina PE primer #2, 0.25 µl of Hotstart Phusion polymerase, and 5 µl of the DNA from step (5). The PCR program used was: 98°C 1 minute; 6 cycles of 98°C for 20 seconds, 65°C for 30 seconds, 72 °C for 30 seconds; and 72 °C for 5 min. To purify the PCR product, 250 µl PCR mixture (from the ten PCR reactions) was mixed with 500 µl NT buffer from a NucleoSpin Extract II kit and purified as described in step (5). Library DNA was eluted with 70°C elution buffer and the DNA concentration was estimated by absorption at 260 nm.

Exome DNA Capture

Human exome capture was performed following a protocol from Agilent's SureSelect Paired-End Version 2.0 Human Exome Kit (Agilent, Santa Clara, CA) with the following modifications. **(1)** A hybridization mixture was prepared containing 25 µl of SureSelect Hyb # 1, 1 µl of SureSelect Hyb # 2, 10 µl of SureSelect Hyb # 3, and 13 µl of SureSelect Hyb # 4. **(2)** 3.4 µl (0.5 µg) of the PE-library DNA described above, 2.5 µl of SureSelect Block #1, 2.5 µl of SureSelect Block #2 and 0.6 µl of Block #3; was loaded into one well in a 384-well Diamond PCR plate (cat# AB-1111, Thermo-Scientific, Lafayette, CO), sealed with microAmp clear adhesive film (cat# 4306311; ABI, Carlsbad, CA) and placed in GeneAmp PCR system 9700 thermocycler (Life Sciences Inc., Carlsbad CA) for 5 minutes at 95°C,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

then held at 65°C (with the heated lid on). **(3)** 25–30 μl of hybridization buffer from step (1) was heated for at least 5 minutes at 65°C in another sealed plate with heated lid on. **(4)** 5 μl of SureSelect Oligo Capture Library, 1 μl of nuclease-free water, and 1 μl of diluted RNase Block (prepared by diluting RNase Block 1: 1 with nuclease-free water) were mixed and heated at 65°C for 2 minutes in another sealed 384-well plate. **(5)** While keeping all reactions at 65°C, 13 μl of Hybridization Buffer from Step (3) was added to the 7 μl of the SureSelect Capture Library Mix from Step (4) and then the entire contents (9 μl) of the library from Step (2). The mixture was slowly pipetted up and down 8 to 10 times. **(6)** The 384-well plate was sealed tightly and the hybridization mixture was incubated for 24 hours at 65°C with a heated lid.

After hybridization, five steps were performed to recover and amplify the captured DNA library: **(1)** Magnetic beads for recovering captured DNA: 50 μl of Dynal MyOne Streptavidin C1 magnetic beads (Cat # 650.02, Invitrogen Dynal, AS Oslo, Norway) was placed in a 1.5 ml microfuge tube and vigorously resuspended on a vortex mixer. Beads were washed three times by adding 200 μl of SureSelect Binding buffer, mixing on a vortex for five seconds and then removing the supernatant after placing the tubes in a Dynal magnetic separator. After the third wash, beads were resuspended in 200 μl of SureSelect Binding buffer. **(2)** To bind captured DNA, the entire hybridization mixture described above (29 μl) was transferred directly from the thermocycler to the bead solution and mixed gently; the hybridization mix/bead solution was incubated in an Eppendorf thermomixer at 850 rpm for 30 minutes at room temperature. **(3)** To wash the beads, the supernatant was removed from beads after applying a Dynal magnetic separator and the beads were resuspended in 500 μl SureSelect Wash Buffer #1 by mixing on vortex mixer for 5 seconds, then incubated for 15 minutes at room temperature. Wash Buffer #1 was then removed from beads after magnetic separation. The beads were further washed three times, each with 500 μl pre-warmed SureSelect Wash Buffer #2 after incubation at 65°C for 10 minutes. After the final wash, SureSelect Wash Buffer #2 was completely removed. **(4)** To elute captured DNA, the beads were suspended in 50 μl SureSelect Elution Buffer, vortex-mixed and incubated for 10 minutes at room temperature. The supernatant was removed after magnetic separation, collected in a new 1.5 ml microcentrifuge tube, and mixed with 50 μl of SureSelect Neutralization Buffer. DNA was purified with a Qiagen MinElute column and eluted in 17 μl of 70°C EB to obtain 15 μl of captured DNA library. **(5)** The captured DNA library was amplified in the following way: 15 PCR reactions each containing 9.5 μl of H_2O, 3 μl of 5 × Phusion HF buffer, 0.3 μl of 10 mM dNTP, 0.75 μl of DMSO, 0.15 μl of Illumina PE primer #1, 0.15μl of Illumina PE primer #2, 0.15 μl of Hotstart Phusion polymerase, and 1 μl of captured exome library were set up. The PCR program used was: 98°C for 30 seconds; 14 cycles of 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds; and 72°C for 5 min. To purify PCR products, 225μl PCR mixture (from 15 PCR reactions) was mixed with 450 μl NT buffer from NucleoSpin Extract II kit and purified as described above. The final library DNA was eluted with 30 μl of 70°C elution buffer and DNA concentration was estimated by OD260 measurement.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Somatic Mutation Identification by Massively Parallel Sequencing

Captured DNA libraries were sequenced with the Illumina GAIIx Genome Analyzer, yielding 150 (2 × 75) base pairs from the final library fragments. Sequencing reads were analyzed and aligned to human genome hg18 with the Eland algorithm in CASAVA 1.6 software (Illumina), as has previously been described[9–11]. A mismatched base was identified as a mutation only when (i) it was identified by more than five distinct reads; (ii) the number of distinct reads containing a particular mismatched base was at least 10% of the total distinct reads; and (iii) it was not present in >0.5% of the reads in the matched normal sample. “Distinct reads” were defined as reads that had different sequences at either the 5′ or 3′ end of the sequenced fragment, thereby indicating that they originated from different template molecules. From our previous studies using these quality criteria, more than 90% of mutations are confirmed as true somatic mutations using independent platforms. SNP search databases included the NCBI’s database (http://www.ncbi.nlm.nih.gov/projects/SNP/).

Purification of mRNA and synthesis of cDNA

JHU—From 5 tumors with available RNA, 10ug total RNA were purified with Dynal Oligo(dT) beads twice following suggested protocol. Briefly, 50ul Oligo(dT) beads(cat# 61002, Invitrogen) were washed with 250ul L/B buffer. Total RNA were mixed to 500ul L/B buffer and added to the isolated beads. The mixture was rocked in room temperature for 15 minutes. The beads were isolated and washed with 2X wash buffer A and 1X wash buffer B. Then the beads were resuspended in 25ul DEPC water and incubated in 85 °C for 2 minutes. Place the tube on the magnet for 1 min, then immediately collect the supernatant. The supernatant was purified one more time and collected in 13ul DEPC water. 12ul of Oligo(dT) selected mRNA was mixed with 2ul 50ng/ul (N)6 random primers and used to synthesize double strand cDNA with the SuperScript Double-Stranded cDNA Synthesis Kit(cat#11917-010, Invitrogen). The cDNA reaction mixture was cleaned with Qiagen PCR purification kit and eluted in 100ul Elution buffer.

UCSF and OSU—RNA was extracted from fresh frozen tissue samples by homogenization, guanidine thiocyanate lysis and column purification using the ZR Viral RNA Kit (Zymo Research) in a laminar flow cabinet. RNA was then treated with DNAse-1 (New England Biolabs) before being subject to rRNA depletion with Terminator™ 5′-Phosphate-Dependent Exonuclease (Illumina) according to manufacturer instructions.

Complementary DNA (cDNA) was subsequently prepared using a random-hexamer PCR amplification method. Reverse transcription was carried out by mixing 100ng RNA with 1000pmol of random hexameric primer in a total volume of 6μl, heating to 65°C for 5 min and then cooling to 4°C. One μL of 12.5 mM dNTPs, 2 μL of 5X RT buffer, 0.5 μL 0.1 M DTT and 0.5 μL of Superscript III® reverse transcriptase (Life Technologies) were then added and the reaction was incubated at 42°C for 60 min, then 94°C for 4 min. To this, 7.5 μL ddH$_2$O, 2.25 μL RT buffer (Life Technologies) and 0.25 μL RNAseH/RNAseA were added and the reaction was incubated at 37°C for 15 min, then at 94°C for 2 min. One μL of RT buffer, 0.5 μL Klenow fragment exonuclease (New England Biolabs), 1 μL of 12.5 mM dNTPs and 2.5 μL H$_2$O were then combined and the reaction was incubated at 37°C for 15

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

minutes. DNA was extracted from this reaction by spin-column purification (DNA Clean & Concentrator-5, Zymo Research Inc.) using 75 µL of binding buffer and eluting in 10 µL H_2O. Metagenomic sequencing libraries were constructed using Nextera sample preparation kits with 20ng input DNA according to manufacturer's instructions (Illumina). Individual samples were then pooled in a 1.5 mL microfuge tube containing 10 µl of 0.5 M EDTA. The ZR DNA Clean-Up Kit (Zymo Research) was employed to purify the DNA, which was subsequently eluted in 20 µl ddH_2O. Size selection for fragments 500–600 bp in length was performed on 10 µl of this sample using the LabChip XT DNA system (Perkin Elmer). This was followed by a 20-cycle PCR amplification using the KAPA Real-Time Library Amplification Kit (Kapa Biosystems) according to manufacturer instructions, and followed by a final DNA purification using the ZR DNA Clean-Up Kit (Zymo Research). Fifteen nM of DNA was then utilized for sequencing on an Illumina HiSeq 2500instrument as previously described [12–16].

Transcriptome Analysis for Viral Pathogen Discovery

First, raw Illumina sequences consisting of 75 bp paired-end reads were quality filtered to exclude low-complexity, homopolymeric, low-quality sequences and Illumina adapter sequences as described previously [12–16]. Filtered sequences were then iteratively aligned against the NCBI NT database utilizing Bowtie2, The Basic Local Alignment Search Tool (BLAST) and Blast-Like Alignment Tool (BLAT) algorithms to subtract reads originating from human, bacteriophage and bacterial genomes [12–17]. The remaining sequences were then BLAST aligned against the NCBI viral genome database as described previously [12–17]. De novo assembly of the remaining unaligned reads was attempted using the Trinity algorithm followed by BLASTX alignment of contigs to the NCBI viral genome database [18]. Second, an independent viral detection procedure was also performed, consisting of MapSplice alignment of reads to a database of human genome and virus genome sequence and quantification of viral expression presence [19, 20]. Third, the PathSeq algorithm was used to perform computational subtraction of human reads, followed by alignment of residual reads aligned to microbial reference genomes including bacterial, viral, archaeal, and fungal sequences (downloaded from NCBI in June, 2012). These alignments result in the identification of reads mapping with viral genomes in RNA sequencing of oral tongue squamous cell cancer samples. Human reads were subtracted by first mapping reads to a database of human database sequences using BWA (Release 0.6.1, default settings), Megablast (Release 2.2.25, cut-off E-value 10^{-7}, word size 16) and Blastn (Release 2.2.25, cut-off E-value 10^{-7}, word size 7, nucleotide match reward 1, nucleotide mismatch −3, gap open cost 5, gap extension cost 2) [21–23]. Human sequence databases used for the subtraction were the Ensembl human cDNA database followed by human genome and transcriptome database (downloaded from Ensembl and NCBI in November 2011). Only sequences with perfect or near perfect matches to the human genome were removed in the subtraction process. In addition, low complexity and highly repetitive reads were removed using Repeat Masker (version open-3.3.0, libraries dated 2011-04-19)[24]. To identify microbial reads, the resultant potential non-human reads were aligned with Megablast (Release 2.2.25, cut-off E-value 10^{-8}, word size 16) to a database of microbial and human reference genomes. Then, the unclassified reads were used to identify presence of new or novel viruses through de novo assembly of reads using Trinity (Release 03122011)[18].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Contigs assembled from the Trinity assembler were aligned to human genome plus microbial databases using Blastn (Release 2.2.25, word size 7, nucleotide match reward 1, nucleotide mismatch −3, gap open cost 5, gap extension cost 2) and aligned to NR database (downloaded from NCBI in May 2011) using Blastx (Release 2.2.25, word size 3, nucleotide match reward 1, nucleotide mismatch −3, gap open cost 11, gap extension cost 1) to identify the new or novel viral species in the dataset. Then, the contigs (that are >=500 bases) that showed weak or good homology to viruses were manually reviewed using online Blastx algorithm (with default settings) from NCBI against the NR database available at NCBI homepage. (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Sanger Sequencing of p53 cDNA

All samples that underwent transcriptomic analysis from UCSF and Ohio State were also subjected to Sanger Sequencing of exons 4–10 of p53 cDNA. Exons were first amplified by PCR and then sequenced using primers and run conditions listed in Supplemental Index Section 2. Samples were sequenced on a 3730xl DNA Analyzer (Applied Biosystems) and resulting data was analyzed with Sequencher (GeneCodes) to genotype for variations.

Statistical Analysis

Descriptive statistics were used to summarize the clinicopathologic characteristics of our cohort of oral tongue SCC patients. Comparative statistics were used to describe differences between the never-smokers and smokers groups with respect to clinicopathologic parameters. The Fisher's exact test was used with statistical significance designated by $P < 0.05$. Comparison of never-smokers and smokers with respect to exomic sequencing results was also performed using Fisher's exact test for categorical variables, and Student's t-test for continuous variables, again with statistical significance designated by $P < 0.05$.

Results

Comparison of clinical and demographic characteristics

Clinicopathologic data were analyzed for 94 patients treated for oral tongue SCC. Of the 94 oral tongue cancer patients, 89 had sufficient tobacco usage data for analysis (Table 1). The non-smokers were significantly younger on average (50.4 years of age) when compared to smokers (61.9 years of age, P=0.003). Non-smokers appeared more likely to be female than smokers, although this did not reach statistical significance (58.5% female vs. 38.9%, respectively; P=0.069). The patients within this study were predominantly white (79 of 94). Therefore we could not make meaningful comparisons regarding racial differences. There were no differences in pathological TNM classification or treatment regimen employed. However, the non-smokers appeared to have poorer tumor differentiation on histopathology (86.4% vs. 67.7% moderate to poorly differentiated respectively, P=0.053), and a higher rate of tumor recurrence (43.1% vs. 20.0% respectively, P=0.026), after a mean follow up interval of 52 months (range 1 month to 27 years).

Exomic Sequencing Data

Table 2 summarizes the results of 6 lateral tongue tumors analyzed by next-generation exomic sequencing. Five lateral tongue tumors from smokers were selected as controls from

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

previously sequenced cases [20]. Other than smoking status, there were no other differences in clinical characteristics between non-smokers that underwent exomic sequencing and the historical smoker controls. All of the tumors from non-smokers were tested for the presence of HPV and EBV, and these viruses were not detectable.

In the non-smoker tumors, the average distinct coverage of each base in the targeted region was 116-fold and 92.5% of targeted bases were represented by at least ten reads. In the previously sequenced tumors from smokers, the average distinct coverage of each base in the targeted region was 77-fold and 91.5% of targeted bases were represented by at least 10 reads. After applying the same stringent quality filtering (see Materials and Methods) and visual inspection of both groups of tumors, the sequencing results of the cancers from the 6 non-smokers revealed 510 high confidence somatic mutations in 482 genes (Table S1). Using the same criteria, we identified 342 high confidence somatic mutations in 324 genes in the 5 cancers from smokers.

The tumors from nonsmokers contained an average of 84.8 ± 71.7 high confidence mutations (range 10–193). The tumors from smokers contained an average of 68.4 ± 32.4 high confidence mutations (range 18–113). The mean number of mutations was not statistically different between cancers from smokers and cancers from non-smokers (P=0.65, Student's t-test).

The recurrently mutated genes or genes in a common pathway in our cohort of cancers from non-smokers were *CTNNA3, EIF3A, EP300, FXR1, NEK8, NOTCH1 (NOTCH2* and *NOTCH3* in individual tumors)*, PIK3CA, PKHD1L1, PTCHD2, RALGAPB, SPEN, and UBR4*. In the cancers from smokers, the recurrent mutations were in *CD101, LAMA1, NCAPD3, RIMBP2, SI, SYNE, and TP53*. All 5 tumors from smokers had *TP53* mutations (5 of 5), whereas only one cancer from a non-smoker had a *TP53* mutation (1 of 6, P=0.02, Fisher's exact test).

The difference in frequency of *NOTCH* gene family mutations between non-smokers (3 of 6 tumors) and smokers (zero tumors) was not significant. In a previous study, we also detected telomerase reverse transcriptase (*TERT)* promoter mutations in 5 of 11 (45%) tumors sequenced [25]. There was no difference in frequency of *TERT* mutations between non-smokers (3 of 6 tumors) and smokers (2 of 6).

The mutational spectrum for the two groups was significantly different although the most common substitution in cancers from both non-smokers and smokers was C:G>T:A (Table 3). Compared to cancers from non-smokers, a greater proportion of mutations in cancers from smokers were either A:T>G:C (P<0.0001) or A:T>T:A substitutions (P<0.0001). Conversely—compared to cancers from smokers, a greater proportion of mutations in cancers from non-smokers were C:G>G:C transversions (P<0.0001).

Viral Discovery

Three separate transcriptomic analyses were undertaken to search for the presence of potentially causative viral pathogens. First, RNA from fifteen tumors from non-smoking patients, five tumors from smoking patients, one healthy tongue sample (negative control)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

and one HPV positive oropharyngeal tumor (positive control) from UCSF, OSU and JHU was analyzed through an established iterative approach. The patients for this portion of the study were young, with an average age of 40 years (38 years for smoking patients and 41 for non-smoking patients). 45% were female (40% for smoking patient and 46% for non-smoking patients) and 75% were Caucasian.

Raw Illumina sequences ranged in number from 4 – 63 × 10^6 (average 18 × 10^6) reads per tissue sample. To search for the presence of non-host derived sequences, we removed human derived sequences through iterative alignment using Bowtie2 and BLAT alignments to the human genome (hg19) followed by low complexity filtering. This resulted in an average of 0.16% (range 0.01%–0.44%) of initial reads. Of these, an average of 0.0017% (range 0.0001%–0.0116%) of reads aligned to the viral database. As expected, the HPV 16 positive oropharyngeal sample, used as a positive control, demonstrated the greatest number of aligned viral reads with 0.00025% (n=14) mapping to HPV 16. This compares to 0.018% of reads that aligned to HPV-16 in a recent report that evaluated HPV 16-positive cervical squamous cell carcinoma [26]. None of the sample libraries contained a higher percentage of total viral-matching reads than the HPV-derived read count from the HPV-positive control sample. One of the non-smoking samples contained 0.00016% of reads (n=6) that aligned to HHV-1 (Table S4). All other hits were below meaningful levels. Remaining unaligned reads assembled utilizing the Trinity algorithm did not show any additional alignments to the viral genome database [18].

JHU samples were then analyzed with MapSplice[19], which did not reveal any meaningful viral reads. Lastly, JHU samples were again analyzed using PathSeq and Trinity for denovo assembly from unclassifiable sequences (See Supplementary Index Section 3, Methods and Tables S5 and S6 for detailed analysis). Two samples showed the presence of low levels (12 – 25 reads) of human herpes virus 4 (HHV4) (Table S7). No additional alignments to the viral genome database were found. In summary, three independent search algorithms identified no potentially causative viruses. No HPV was found in any of the tumors other than the HPV positive control.

TP53—All samples that underwent transcriptomic analysis from UCSF and Ohio State underwent Sanger Sequencing of exons 4–10 of *TP53* cDNA. In concordance with the exome sequencing data, 1 of 10 (10%) of the cancers from non-smoking patients had a deleterious *TP53* mutation while 2 out of 5 (40%) of the cancers from patients who smoked had deleterious *TP53* mutations (Table S3).

Discussion

This examination of oral tongue cancer patients stratified by smoking status revealed distinct clinico-demographic and biological differences between these groups. Compared to smokers, the non-smokers with oral tongue SCC were more likely to be less than 50 years of age, and there was a trend towards a higher proportion of non-smokers being female. While our cohort was small, there was also a suggestion that non-smokers with oral tongue SCC may have more poorly differentiated carcinomas, and a higher rate of recurrence after primary treatment. Non-smokers, however, did not exhibit more advanced tumor staging at

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

time of diagnosis. Next-generation exomic sequencing and Sanger Sequencing data suggest that cancers from non-smokers have a lower prevalence of *TP53* mutations. This preliminary evidence suggests a biological difference in the development of oral tongue SCC in these two groups.

In contrast to an overall declining incidence of HNSCC, due in part to reduced tobacco consumption worldwide, several subsites of HNSCC have experienced an increase in incidence [27]. The identification of high-risk HPV-associated oropharyngeal cancers as a distinct subgroup has led to recognition of HNSCC as a biologically heterogeneous disease [5, 28]. The oral tongue subsite likewise appears to have a rising incidence in young patients, most evident in white females, based on the Surveillance, Epidemiology, and End Results (SEER) database [4, 29].

More recently we reported an analysis of the Nationwide Inpatient Sample (NIS) all-payer database, comparing the demographic characteristics of 15,083 oral tongue cancer patients to 29,268 patients with oral cavity cancer from other subsites [30, 31]. In multivariate regression analysis, oral tongue cancer patients were more likely to be younger than 40 years of age, white, and female. Although we did not compare our oral tongue SCC cohort to other oral cancer subsites, we have found that the subgroup of non-smokers is similarly younger and more likely to be female than smokers [31, 32]. Whether the rising incidence of oral tongue SCC reported in SEER analysis is primarily occurring in non-smokers is unclear since the SEER database does not report tobacco consumption. In our study, non-smokers comprised over half of our oral tongue SCC cohort, making this subgroup a significant proportion of patients and provoking scientific investigation into biological differences between these two exposure subgroups. Furthermore, there is a unique clinical pattern to oral tongue cancer in nonsmokers. Virtually all cases arise on the lateral border of the tongue and on one side only. All cases of recurrence/second tongue cancer in our series to date have occurred on the same side as the incident cases, often, although not always preceded by a prolonged period of “smoldering” erythroleukoplakia (showing a fluctuating degree of severity and biopsy evidence of fluctuating degree of epithelial dysplasia). In contrast, oral cavity cancers in smokers occur more commonly on the floor of mouth, and second primary cancers may appear anywhere within the upper aerodigestive tract mucosa.

Given these clinico-demographic differences, we were interested in determining the molecular profiles of oral tongue SCC from non-smokers and smokers. The most striking differences in mutation patterns between cancers from non-smokers and smokers were seen in the prevalence of *TP53* mutations. Both Agrawal et al and Stransky et al have independently reported *TP53* as the most frequently mutated genes on exomic sequencing of all sites of HNSCC. The majority of tumors reported by Agrawal et al and Stransky et al harbored *TP53* mutations or *TP53* inactivation through HPV [33, 34]. In our study only 2 of 16 cancers from non-smokers had a *TP53* mutation, whereas 7 of 10 cancers from smokers had *TP53* mutations. Given that the vast majority of HNSCC have inactivation of *TP53* either through somatic mutation or HPV E6, it is possible that *TP53* inactivation occurs via a mechanism other than somatic mutation in non-smokers. Alternatively, a *TP53* independent pathway may drive oncogenesis in non-smokers that is not currently understood. Larger

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

cohorts of oral tongue SCC tumors need to be studied in patients with accurate tobacco use to validate the results of this genomic sequencing analysis.

The young age of the patients, lack of carcinogen exposure and lack of *TP53* mutations raises the possibility that a known or novel virus could be driving tumor development. Considerable interest has circulated around this topic as a possible explanation for the increasing incidence of oral tongue SCC in non-smokers. We were unable to identify any potentially causative viruses, including HPV, in 19 tongue tumor samples through transcriptomic analysis with three separate approaches. Although our current analysis does not support the role of an infectious agent in these cases, it is important to note that this work does not definitively rule out the involvement of such an agent. The concept of "hit and run" viral transformation could account for a viral role in oncogenesis without persistence of viral genomic DNA within cancer cells [33, 34]. The hit and run hypothesis suggests that oncogenic viruses can either integrate into the host cell genome or remain episomal at a particular point in the progression to cancer. The viral genome may be subsequently lost and become undetectable by the time of clinical diagnosis. Experimental evidence for such a mechanism has been sought in certain Hodgkin's lymphoma subtypes [35, 36, 37] and nasopharyngeal carcinoma [38, 39, 40], while the clinical relevance remains unknown. Future directions may utilize methods to detect epigenetic signatures of previous viral infection at a point when viral DNA has been lost from cancer cells.

Two landmark studies independently demonstrated that non-smokers with HNSCC harbored fewer mutations compared to smokers [33, 3441,42]. We did not observe a statistically significant difference between the number of somatic mutations in cancers from non-smokers and smokers. However, our study may not have had sufficient power to detect differences in overall mutation rates between cancers from smokers and non-smokers. Nevertheless, we did identify other evidence of tumor heterogeneity based on tobacco exposure. Surprisingly, the mutational spectrum in cancers from smokers was not enriched in the C:G>A:T transversions that are typically associated with smoking[3543]. Cancers from smokers demonstrated more A:T>T:A and A:T>G:C transitions, whereas cancers from non-smokers showed higher rates of C:G>G:C transversions. Various oxidative stresses induce C:G>G:C transversions suggesting oxidative DNA damage in non-smokers may be coming from an unique, yet unidentified, source [36, 3744, 45].

This is the first study to examine the genomic and metagenomic profiles of oral tongue SCC in non-smoking patients compared to smoking patients. There are several limitations to this study. We acknowledge the relatively small size of samples undergoing exome sequencing as a limitation. This study did not investigate the role of other genomic alterations including chromosomal rearrangements and copy number variations in oral tongue SCC, nor did we characterize the potential epigenetic changes that may contribute to oncogenesis in this tumor type. At present time no large studies exist that compare long-term prognosis between smokers and non-smokers with oral tongue SCC and our study was not designed for robust survival analysis. While non-smokers had a higher rate of tumor recurrence, we recognize the limitation of small sample size in concluding that non-smokers with oral tongue SCC have more aggressive tumors. Given the significant and novel clinical and genetic differences, we feel that it is important to report this early, but significant data to draw

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

attention to an emerging clinical entity. Further epidemiologic, microbiologic and molecular study of lateral tongue cancers is necessary.

Conclusion

This study provides preliminary evidence that oral tongue SCC patients represent a heterogeneous group with unique biologic and clinical characteristics. Although there are genetic differences in the young non-smokers, given the limited samples size, there may exist other unidentified drivers of carcinogenesis. Acquiring larger cohorts for molecular studies will strengthen the evidence for differing tumor biology, and facilitate the search for unidentified causal factors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank our patients for their courage and generosity. We thank J. Ptak, N. Silliman, L. Dobbyn, J. Schaeffer, C. Blair, and K. Judge for expert technical assistance. This work was supported by the National Institutes of Health grant RC2DE020957 and the Virginia and D.K. Ludwig Fund for Cancer Research, and the Burroughs Wellcoma Award for Medical Scientists, Johns Hopkins Hospital Clinician Scientist Award. This work was also supported by the Johns Hopkins Head and Neck SPORE NIH/NIDCR Grant 5P50DE019032 The manuscript/analysis of this article is based on a web database application provided by Research Information Technology Systems (RITS)–https://www.rits.onc.jhmi.edu/.

References

1. Blot WJ, McLaughlin JK, Winn DM, et al. Smoking and drinking in relation to oral and pharyngeal cancer. Cancer Res. 1988; 48(11):3282–7. [PubMed: 3365707]
2. Mahboubi, E.; Sayed, GM. Oral cavity and pharynx. Philadelphia: W. B. Saunders Co; 1982.
3. Hashibe M, Brennan P, Benhamou S, et al. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. J Natl Cancer Inst. 2007; 99(10):777–89. [PubMed: 17505073]
4. Patel SC, Carpenter WR, Tyree S, 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):1488–94. [PubMed: 21383286]
5. Lingen MW, Xiao W, Schmidt A, et al. Low etiologic fraction for high-risk human papillomavirus in oral cavity squamous cell carcinomas. Oral Oncol. 2013; 49(1):1–8. [PubMed: 22841678]
6. Harris SL, Kimple RJ, Hayes DN, Couch ME, Rosenman JG. Never-smokers, never-drinkers: unique clinical subgroup of young patients with head and neck squamous cell cancers. Head Neck. 2010; 32(4):499–503. [PubMed: 19691028]
7. Shanmugaratnam K, Sobin LH. The World Health Organization histological classification of tumours of the upper respiratory tract and ear. A commentary on the second edition. Cancer. 1993; 71(8):2689–97. [PubMed: 8453591]
8. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol. 2010; 17(6):1471–4. [PubMed: 20180029]
9. Agrawal N, Jiao Y, Sausen M, et al. Exomic sequencing of medullary thyroid cancer reveals dominant and mutually exclusive oncogenic mutations in RET and RAS. J Clin Endocrinol Metab. 2013; 98(2):E364–9. [PubMed: 23264394]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

10. Bettegowda C, Agrawal N, Jiao Y, et al. Mutations in CIC and FUBP1 contribute to human oligodendroglioma. Science. 2011; 333(6048):1453–5. [PubMed: 21817013]
11. Bettegowda C, Agrawal N, Jiao Y, et al. Exomic sequencing of four rare central nervous system tumor types. Oncotarget. 2013; 4(4):572–83. [PubMed: 23592488]
12. Yu G, Greninger AL, Isa P, et al. Discovery of a novel polyomavirus in acute diarrheal samples from children. PLoS One. 2012; 7(11):e49449. [PubMed: 23166671]
13. Greninger AL, Chen EC, Sittler T, et al. A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. PLoS One. 2010; 5(10):e13381. [PubMed: 20976137]
14. Chen EC, Yagi S, Kelly KR, et al. Cross-species transmission of a novel adenovirus associated with a fulminant pneumonia outbreak in a new world monkey colony. PLoS Pathog. 2011; 7(7):e1002155. [PubMed: 21779173]
15. Stenglein MD, Sanders C, Kistler AL, et al. Identification, characterization, and in vitro culture of highly divergent arenaviruses from boa constrictors and annulated tree boas: candidate etiological agents for snake inclusion body disease. MBio. 2012; 3(4):e00180–12. [PubMed: 22893382]
16. Yozwiak NL, Skewes-Cox P, Gordon A, et al. Human enterovirus 109: a novel interspecies recombinant enterovirus isolated from a case of acute pediatric respiratory illness in Nicaragua. J Virol. 2010; 84(18):9047–58. [PubMed: 20592079]
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215(3):403–10. [PubMed: 2231712]
18. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011; 29(7):644–52. [PubMed: 21572440]
19. Wang K, Singh D, Zeng Z, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. 2010; 38(18):e178. [PubMed: 20802226]
20. Agrawal N, Frederick MJ, Pickering CR, et al. Exome Sequencing of Head and Neck Squamous Cell Carcinoma Reveals Inactivating Mutations in NOTCH1. Science. 2011; 333(6046):1154–7. [PubMed: 21798897]
21. Kostic AD, Ojesina AI, Pedamallu CS, et al. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. Nature Biotechnol. 2011; 29(5):393–6. [PubMed: 21552235]
22. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). 2009; 25(14):1754–60.
23. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25(17):3389–402. [PubMed: 9254694]
24. Aea, Smit. RepeatMasker Open-3.0. 1996–2010
25. Killela PJ, Reitman ZJ, Jiao Y, et al. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. Proc Natl Acad Sci U S A. 2013; 110(15):6021–6. [PubMed: 23530248]
26. Arron ST, Ruby JG, Dybbro E, Ganem D, Derisi JL. Transcriptome sequencing demonstrates that human papillomavirus is not active in cutaneous squamous cell carcinoma. J Invest Dermatol. 2011; 131(8):1745–53. [PubMed: 21490616]
27. Chaturvedi AK, Engels EA, Anderson WF, Gillison ML. Incidence trends for human papillomavirus-related and -unrelated oral squamous cell carcinomas in the United States. J Clin Oncol. 2008; 26(4):612–9. [PubMed: 18235120]
28. Gillison ML, Koch WM, Capone RB, et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer Inst. 2000; 92(9):709–20. [PubMed: 10793107]
29. Shiboski CH, Schmidt BL, Jordan RC. Tongue and tonsil carcinoma: increasing trends in the U.S. population ages 20–44 years. Cancer. 2005; 103(9):1843–9. [PubMed: 15772957]
30. HCUP Databases. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; Rockville, MD: Aug. 2013 www.hcup-us.ahrq.gov/nisoverview.jsp
31. Li R, Fakhry C, Koch WM, Gourin CG. The Effect of tumor subsite on short-term outcomes and costs of care after oral cancer surgery. Laryngoscope. 2013; 123(7):1652–9. [PubMed: 23686386]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

32. Li R, Koch WM, Fakhry C, Gourin CG. Distinct epidemiologic characteristics of oral tongue cancer patients. Otolaryngol Head Neck Surg. 2013; 148(5):792–6. [PubMed: 23396594]

33. McDougall JK. “Hit and run” transformation leading to carcinogenesis. Dev Biol. 2001; 106:77–82.

34. Niller HH1, Wolf H, Minarovits J. Viral hit and run-oncogenesis: genetic and epigenetic scenarios. Cancer Lett. 2011; 305(2):200–17. [PubMed: 20813452]

35. Delecluse HJ, Marafiotic T, Hummel M, Dallenbach F, Anagnostopoulos I, Stein H. J Pathol. 1997; 182(4):475–9. [PubMed: 9306970]

36. Brousset P, Schlaifer D, Meggetto F, Bachmann E, Rothenberger S, Pris J, Delsol G, Knecht H. Persistence of the same viral strain in early and late relapses of Epstein-Barr virus-associated Hodgkin’s disease. Blood. 1994; 84(8):2447–51. [PubMed: 7919364]

37. Gan YJ, Razzouk B, Su T, Sixbey JW. A defective, rearranged Epsteain-Barr virus genome in EBER-negative and EBER-positive Hodgkin’s disease. Am J Pathol. 2002; 160(3):781–6. [PubMed: 11891176]

38. Lo KW, To KF, Huang DP. Focus on nasopharyngeal carcinoma. Cancer Cell. 2004; 5(5):423–8. [PubMed: 15144950]

39. Cheung ST, Huang DP, Hui AB, et al. Nasopharyngeal carcinoma cell line (C666-1) consistently harbouring Epstein-Barr virus. Int J Cancer. 1999; 81(1):121–6. [PubMed: 10449618]

40. Dittmer DP, Hilscher CJ, Gulley ML, Yang EV, Chen M, Glaser R. Multiple pathways for Epstein-Barr virus episome loss from nasopharyngeal carcinoma. Int J Cancer. 2008; 123(9):2105–12. [PubMed: 18688856]

41. Agrawal N, Frederick MJ, Pickering CR, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science. 2011; 333(6046):1154–7. [PubMed: 21798897]

42. Stransky N, Egloff AM, Tward AD, et al. The mutational landscape of head and neck squamous cell carcinoma. Science. 2011; 333(6046):1157–60. [PubMed: 21798893]

43. Vahakangas KH, Bennett WP, Castren K, et al. p53 and K-ras mutations in lung cancers from former and never-smoking women. Cancer Res. 2001; 61(11):4350–6. [PubMed: 11389059]

44. Kino K, Sugiyama H. UVR-induced G-C to C-G transversions from oxidative DNA damage. Mutat Res. 2005; 571(1–2):33–42. [PubMed: 15748636]

45. Kino K, Sugiyama H. Possible cause of G-C-->C-G transversion mutation by guanine oxidation product, imidazolone. Chem Biol. 2001; 8(4):369–78. [PubMed: 11325592]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Clinical comparison of nonsmokers and smokers with oral tongue cancer.

	Non-smokers (n=53)	Smokers (n=36)	Significance
Age	**%**	**%**	P=0.003
< 50 Years	50.9	19.4	
> 50 Years	49.1	80.6	
Gender			P=0.069
Male	41.5	61.1	
Female	58.5	38.9	
Tumor Histologic Grade			0.053
Well	13.6	32.3	
Moderate to Poor	86.4	67.7	
Pathologic Stage			P=0.499
Early	65.3	72.2	
Advanced	34.7	27.8	
Treatment			P=0.543
Surgery	65.2	75	
Surgery + RT	26.4	19.4	
Surgery + CRT	9.4	5.6	
Recurrence			P=0.026
Yes	43.1	20	
No	56.9	80	

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Genetic comparison of nonsmokers and smokers (historic data) with oral tongue cancer.

TUMOR SAMPLE	TOBACCO	TNM Classification	GENDER	RACE	HPV	EBV	*TP53* mutation
Tongue 1T	N	T1N0M0	female	Arab	N	N	No
Tongue 2T	N	T1N0M0	male	White	N	N	No
Tongue 3T	N	T3N0M0	female	Latino	N	N	No
Tongue 4T	N	T2N0M0	male	White	N	N	No
Tongue 6T	N	T2N2bM0	male	White	N	N	Yes
Tongue 7T	N	T2N0M0	female	Asian	N	N	No
Historic data	Y	T3N2bM0	male	White	N	N/A	Yes
Historic data	Y	T2N0M0	female	White	N	N/A	Yes
Historic data	Y	T3N2cM0	male	White	N	N/A	Yes
Historic data	Y	T1N1M0	female	White	N	N/A	Yes
Historic data	Y	T2N2bM0	male	Black	N	N/A	Yes

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Spectra of all mutations

	Nonsmokers		Smokers	
Mutations	Number	Percent (%)	Number	Percent (%)
A:T>C:G	19	3.7	15	4.4
A:T>G:C	34	6.7	52	15.2
A:T>T:A	15	2.9	35	10.2
C:G>A:T	47	9.2	44	12.9
C:G>G:C	174	34.1	52	15.2
C:G>T:A	213	41.8	133	38.9
Indel	8	1.6	11	3.2
Total	510	100.0	342	100.0