Fall 12-7-2016

Genetic Differentiation of Oral and Oropharyngeal Carcinoma Based on Human Papillomavirus Status and Race

Aastha Vashist

ABSTRACT

Genetic Differentiation of oral and oropharyngeal carcinoma based on Human Papillomavirus Status and Race

By

Aastha Vashist

December 7th, 2016

INTRODUCTION: Head and neck cancer is one of the most common malignancy in the world. While it has been associated with several factors like alcohol consumption and smoking, there is approximately 25% of head and neck cancer that can be attributed to Human Papillomavirus (HPV) especially HPV 16. HPV associated cancer has been associated with a better prognosis as compared to HPV negative cancers. It has also been shown in previous studies that HPV-negative African Americans have a higher mortality rate as compared to HPV associated cancers in European Americans and HPV-negative European Americans patients. The three states of HPV associated cancers have been compared, which included HPV active, HPV inactive and HPV negative.

AIM: The study aims include: 1) Compare the differences in the gene expression profiles of HPV negative HNSCC in AA from EA patients, and determine the differences in their biological make up. 2) Explore and compare the genetic expression profiles of HPV-active, HPV-inactive and HPV-negative head and neck cancer patients.

METHODS: A secondary data analysis was conducted on 36 oropharyngeal cancer tissues samples with different HPV status (HPV-active, HPV-inactive and HPV- negative). ANOVA was conducted in R to compare all the three groups from each other and identify the genes that were differentially expressed. Bayes Moderated paired t-test was used to compare two groups of HPV-negative European Americans with HPV-negative African Americans.

RESULTS: Our analysis revealed that the genes that were differentially expressed in HPV-active and HPV-negative analysis were different from HPV-active and HPV-inactive analysis. Our analysis also identified genes that were differentially expressed in African Americans as compared to European Americans.

DISCUSSION: This study provides the genetic expression profiles in different groups (European Americans and African Americans) based on different HPV stages. Despite the small sample size of our data, we were able to identify the genes that were differentially expressed amongst different conditions in patients who had oropharyngeal carcinoma. We were also able to identify the genes involved in HPV-negative oral cancer comparing the African Americans to the European Americans.

Genetic Differentiation of oral and oropharyngeal carcinoma based on Human Papillomavirus Status and Race

by

Aastha Vashist

B.D.S, PANJAB UNIVERSITY

(List other degrees awarded in the same format)

A Thesis Submitted to the Graduate Faculty
of Georgia State University in Partial Fulfillment
of the
Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GEORGIA
30303

Genetic Differentiation of oral and oropharyngeal carcinoma based on Human Papillomavirus Status and Race

APPROVAL PAGE

by

Aastha Vashist

Approved:

Dr. Ruiyan Luo _____
Committee Chair

Dr. Ike Okosun _____
Committee Member

Date

Acknowledgments

Many people have helped me through this journey and I'm thankful to all these people for all the support that they offered me. Dr. Ike Okosun, I'm grateful for your valuable feedbacks on my thesis and helping me learn new concepts. Dr. Paul Katz, without your support I would not have been able to pursue my passion for Neurogenomics. I'm really grateful for all the guidance that you provided me throughout my two-years of the program. Andrew Jake Mathews, you have helped me understand the analysis with this dataset and your opinions have been valuable for my professional growth. Rayleen Lewis, for providing me support in my personal life and being my family here. Dr. Ruiyan Luo, thank you for being a wonderful professor and helping me direct my focus towards this thesis topic. Your valuable insights on data analysis helped me understand the concepts better. Lastly, my parents and sister, without whom I would not have been able to see my dream come true. I can't thank you enough for all the financial and emotional support that you provided me. I'm exceptionally happy to have you as my family.

Author's Statement Page

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this thesis may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, School of Public Health. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

Aastha Vashist
Signature of Author

TABLE OF CONTENTS

LIST OF TABLES……7

LIST OF FIGURES…….. 8

INTRODUCTION..... 9

REVIEW OF THE LITERATURE..... 9

METHODS AND
PROCEDURES……..... 12
Tissue Samples and Extraction of Nucleic acids……. 12
Human Papillomavirus Typing and Status……. 13
mRNA labelling and Microarray analysis……..... 13
Statistical Analysis..... 14
Reverse transcriptase-quantitative polymerase chain reaction……... 14

RESULTS..... 15
Gene expression profiles of human papillomavirus…… 15
Figure 1: Volcano plot comparing different HPV conditions…….....16
Table 1: Comparison of all three groups-ANOVA……..20
Figure 2: Hierarchical clustering of the three HPV conditions……21

DISCUSSION.....22
Discussion.....22
Limitations.....23
Future research.....24

REFERENCES.....24

List of Tables

Table 1: Comparison of all three groups-ANOVA…….. 20

List of Figures

Figure 1: Volcano plot comparing different HPV conditions……..... 18
Figure 2: Hierarchical clustering showing different groups……. 21

Introduction

Head and neck cancer (HNC) is the fifth most common malignancy worldwide with an annual mortality rate of 200,000. Approximately 90% of HNC can be classified as head and neck squamous cell carcinomas (HNSCC), of which approximately 75% are attributed to alcohol and tobacco consumption and 25% are associated with human papillomavirus (HPV), predominantly HPV16 (Tomar et al., 2016). It has been shown in previous studies that HPV associated with Oro Pharyngeal Carcinoma have a better prognosis and respond better to therapy as compared to HPV- negative tumors (Weinberger et al., 2006). It can be seen from the differences in risk factors, age of presentation, clinical behavior and gene expression profiles that HPV-positive and HPV-negative tumors develop via different molecular mechanisms and are biologically distinct (Deng et al., 2013). African Americans (AA) males have shown to have higher incidence of HNC than any other racial/gender group, and have a three-times mortality rate than that observed in European Americans (EA) males (Cole, Polfus, & Peters, 2012a). Previous studies have indicated that AA tend to present with more HPV-negative OPC and have worse prognosis as compared to both HPV-positive and HPV-negative HNSCC in EA patients (Zandberg et al., 2016). This study aims to compare the differences in the gene expression profiles of HPV negative HNSCC in AA from EA patients, and determine the differences in their biological make up. We also aim at exploring and comparing the genetic expression profiles of HPV-active, HPV-inactive and HPV-negative head and neck cancer patients

Literature Review:

Head and neck cancer (HNC), which is defined as the cancer of the oral cavity, pharynx, larynx, paranasal sinuses and nasal cavity and the salivary glands, is the ninth most common cancer in the USA and 14th most common cause of death (Daraei & Moore, 2015). Even though

the incidence of head and neck cancer has plateaued now, its mortality and morbidity continue to remain high (Curado & Boyle, 2013). The incidence, mortality, and the relative survival rates of HNC vary with respect to sex and race. The incidence and mortality rates are more than twice as high in men as in women, and are greatest in men who are older than 50 years of age (Cole, Polfus, & Peters, 2012b). African American men have a higher incidence and mortality rates for HNC as compared to white men (Weinberger et al., 2010). Incidence rates for African American and white women are almost similar, though African American women have a slightly higher mortality rate compared to white women (Gourin & Podolsky, 2006). Both the incidence and mortality rates depend upon the anatomic location of the tumor and vary considerably. It has been shown that tongue and oropharynx are the anatomic sites that have the highest incidence and mortality rates. In addition to the anatomic sites, the stage of cancer also strongly affects the survival rate. The five-year relative survival rate for localized tumors is approximately 83%, in comparison to regional tumors that demonstrate 54% of five-year relative survival rate and distant tumors have a 32% five-year relative survival rates (Cole et al., 2012b).

The HPV-related oropharyngeal cancer results from the ability of the virus to infect the discontinuous reticulated epithelium of tonsillar crypts (Andersen, Koldjær Sølling, Ovesen, & Rusan, 2014). HPV results in anogenital and oropharyngeal cancer by transforming primary human keratinocytes from genital or oral epithelia and also by disrupting cell-cycle regulatory pathways (Yim & Park, 2005). The precise mechanisms due to which HPV results in malignant transformation of the keratinocytes in the upper digestive tract epithelia are unknown. It has been hypothesized that HPV E7 causes the overexpression of p16INK4A (Li, Nichols, Shay, & Xiong, 1994), which is considered a surrogate marker for HPV positivity/activity (Klaes et al., 2001). However, this biomarker alone is not sufficient as a predictor of HPV positivity in different

mucosal subsites of head and neck cancer (Wilson DD, Rahimi AS, Saylor DK, & et al, 2012). Hence, it is important to assess viral load and viral oncogenic expression, leading in further classification of HPV-positive oropharyngeal cancers as HPV-active and HPV-inactive (Deng et al., 2013).

Other risk factors associated with HNC include the intensity and duration of alcohol and tobacco consumption (Dal Maso et al., 2016). Approximately 75% of HNC cases are attributable to alcohol and tobacco exposure, whereas 15-20% of cases consist of non-smokers and non-drinkers (Cole et al., 2012b). This shows that the etiology of HNC has other risk factors besides alcohol and smoking.

The role of HPV, specifically HPV-16, in oropharyngeal cancer is well established(Brawley, 2009). Recent epidemiological studies of oral HPV infection in the US population have observed that the distribution of high-risk HPV infection is similar to the risk for the OPC in the United States (Gillison et al., 2008). The prevalence of oral high-risk and HPV 16 infections in the United States was almost three to five fold higher among men than in women, which is consistent with the three to five-fold relative risk for OPC among compared to women(Gillison et al., 2008). The high prevalence of HPV-related oropharyngeal carcinoma is determined by the ability of the virus to infect the reticular epithelium of tonsillar crypts(Tomar et al., 2016). The tumor HPV status is considered as an important prognostic factor for HNC based on the survival rates associated with HPV-positive versus HPV-negative HNC (Firmino et al., 2016). Patients with HPV-positive oropharyngeal tumors have a better prognosis compared with HPV-negative oropharyngeal patients (Firmino et al., 2016). Previous studies have identified E6/E7-dependent inactivation of p53 and Rb tumor suppressor proteins to be the underlying cause for the oncogenic proteins (Vlashi et al., 2016). However, studies replicating

the clinical radiosensitivity of HPV-positive HNSCC experimentally are sparse and conflicting. Recent reports demonstrated that HPV-positive HNSCC cell lines are more radiosensitive than HPV-negative cell lines (Yim & Park, 2005), but a clear mechanistic link between overexpression of HPV genes and increased radiosensitivity is still missing, thus indicating a complex role of HPV in HNSCC RT. There have been different hypotheses about the differences in the gene expression profiles in HPV-positive and HPV-negative HNC, which needs to be investigated in further details.

HPV has been characterized as a risk factor for oropharyngeal cancer based on race, lifestyle, survival outcomes in both African American and European American patients. Some studies have reported that tumors associated with HPV are much lower in African Americans as compared to the European Americans (Zandberg et al., 2016) whereas African Americans have a higher incidence of head and neck cancer and a mortality rate almost 3-times than that observed in European American men (Murdock & Gluckman, 2001). Our analysis was based of Tomar et al. 2016 dataset and study.

Patient demographics, tumor characteristics and activity in oropharyngeal cancer from Tomar et al. 2016

The sample was collected from the MUSC tissue bank and the inclusion criteria were that the patients had given consent to the use of the samples for the research. The study conducted analysis and found that there was no significant difference between African Americans and European Americans based on clinical characteristics like alcohol consumption and smoking. The oropharyngeal cancer samples were classified into 3 categories: whether the HPV DNA was present and transcribed (HPV-active), HPV DNA was present and not transcribed (HPV-inactive) and HPV DNA was not detected (HPV-negative).

The data presented in the paper (Tomar et al., 2016) suggested that on average, the age of presentation for HPV active oral and oropharyngeal carcinoma was less than the age of presentation for HPV-negative carcinoma especially in men (mean ages in HPV active European American men= 52 years; HPV negative = 71 years; p=0.0003). The paper also reported that the HPV-negative cancers in African American patients presented at a younger age as compared to the European Americans (mean ages of presentation for HPV-negative cancers: African American men=53 years; European American men= 71 years*, p- value= 0.0023).* The authors also suggested that based on the stage of cancer at presentation, there was no difference between African American and European American patients.

The paper also depicted that despite the small sample size, the observed differences were significant between African Americans being more likely to present with HPV-negative tumors as compared to the European American patients. The odds of having HPV-positive tumor in European American patients was 3.4 times in comparison to the African American patients (95% confidence interval [CI] = 1.08–10.7; *p-value*= .035). The odds of presenting with an HPV-active tumor among European American patients were 5.7 times in comparison with African American patients (95% CI = 1.15–28.6; *p-value*= .033).

The purpose of this study was to explore the possible genetic differences in HPV active, inactive and negative cases and to examine the gene expression profiles in the HPV-negative African Americans and European Americans.

Materials and Methods

Tissue Samples and Extraction of Nucleic acids

The data consisted of 36 oropharyngeal cancer tissue samples, which consisted of samples with oral cancer, hypopharyngeal cancer and maxillary carcinoma. The sample collected

comprised of African Americans and European American patients at MUSC (Medical University of South Carolina), Charleston, South Carolina. The dataset was available on GEO Datasets on NCBI. The tissue samples were total RNA and genomic DNA samples that were selected for microarray analysis based upon the quality of frozen specimen available and the resulting RNA. The samples were Microarray experiments and were performed using the Agilent technologies platform. The data obtained were extracted from images with Feature extractor software version 10.7.3.1 (Agilent) where background correction was performed (Tomar et al., 2016).

Human Papillomavirus Typing and Status

Genomic DNA from oropharyngeal tissue samples was analyzed (Tomar et al., 2016) for the presence of specific HPV by the reverse hybridization of L1 region with specific oligonucleotides (INNO-LIPA) HPV Genotyping extra assay, a multiplex polymerase chain reaction (PCR)- based assay based on the reverse line blot hybridization principle. The INNO-LIPA assay targets a 65-bp fragment of the L1 open reading frame to detect and identify the 28 different types of HPV types, which include the 18 high risk type of HPV strains (16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 73, and 82); 6 low-risk HPVs (6, 11, 40, 43, 44, 54, and 70) and 3 other non-classified HPVs (69, 71, and 74) (van Hamont, van Ham, Bakkers, Massuger, & Melchers, 2006). Reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR) assessed the type specific HPV E7 oncogene expression in order to assess the active or inactive status of the virus in the samples (Wang et al., 2015). The primers used in order to detect E7 were specifically designed from each individual HPV type analyzed (Tomar et al. 2015). Approximately 51,000 genes per sample were analyzed for this study.

mRNA labelling and Microarray analysis

The Microarray experiments were performed with the help of the Agilent technologies platform. The total RNA samples were amplified and labelled using Agilent's Low Input Quick Amp Labeling Kit. The mRNA contained in the total(200ng) of RNA was converted into cDNA with the help of a poly-dT primer that also contained the T7 RNA polymerase promoter sequence. Subsequently, T7 RNA polymerase was added to cDNA samples to amplify the original mRNA molecules and to simultaneously incorporate cyanine-3 labeled cricothyroidopexy into the amplification product (cRNA). In the next step, labeled cRNA molecules were purified using Qiagen's RNeasy Mini Kit (Valencia, CA). After spectrophotometric assessment of dye incorporation and cRNA yield, samples were stored at −80°C until hybridization. Labeled cRNA samples (600 ng) were hybridized to SurePrint G3 Human Gene Expression 8 × 60 K v2 microarrays at 65°C for 17 hours using Agilent's Gene Expression Hybridization Kit, in accord with the manufacturer's recommendations. After washes, arrays were scanned using a High Resolution Agilent DNA Microarray Scanner and images saved in TIFF format (Tomar et al., 2016).

Data Analysis

The data were log2 transformed, quantile normalized, and base line transformed using the median of all the samples in R statistical package version 3.3.1 (Sifakis et al. 2012). Differentially expressed genes were determined by the analysis of the data using Bayes moderated paired t test and ANOVA in R. Benjamini & Hochberg's False discovery rate approach was used to calculate the adjusted p-value (Pike, 2011). Cutoff values of 0.05 and 2 were used for p value and fold change respectively. A p value of 0.05 or less was considered significant. We conducted hierarchical cluster analysis using GeneSpring software version 14.5.

Reverse transcriptase-quantitative polymerase chain reaction

The iScript cDNA synthesis kit (BioRad, Hercules, CA) was used to synthesize complementary DNA (cDNA) from the total RNA. Real-Time PCR sciTool (Integrated DNA Technology, Coralville, IA) and Primer-Blast software (National Center for Biotechnology Information) were used to design all other primers except TP53 and TGFB2. iCycler IQ detection system (BioRad) with iScriptTM Sybr Green Supermix Kit (BioRad) were used to perform RT –qPCR.

Results

Gene expression profiles of human papillomavirus-active, human papillomavirus-inactive, and human papillomavirus-negative oropharyngeal cancer samples

The gene expression between different groups was explored. In the first step, we first compared HPV- active samples (11 European Americans, 1 African American, all males) to HPV- negative samples (8 European Americans, 2 females and 6 males), we discovered 400 differentially expressed genes with cut off value of 2 for log fold change (logFC). We just included European Americans for this analysis. We compared the samples using Bayes moderated t-test statistics in R, with the help of Benjamini & Hochberg False discovery rate approach for calculating the adjust P-value. We found the genes with most differential expression by using the logFC cut off value of greater than or equal to 5 and -log10(Adj. P. Value) of greater than 2 (Figure 1A). We found 7 such highly expressed genes. We then compared HPV-active (12 samples, all males, 11 European Americans and 1 African American) with HPV-inactive (8 sample, 6 females and 2 males, 4 African Americans, 4 European Americans). Our cut-off for the log fold change was 2 and we found 440 differentially expressed genes. We found 25 genes that were highly expressed according to our criteria of both logFC cut off of more than 5 and -log10 (Adj. P. Value) of greater than 2 (Figure 1B). The comparison of

16 samples of HPV-negative (8 European Americans and 8 African Americans; 5 females and 11 males) with that of 8 samples of HPV- inactive depicted 48 genes with log fold change (LogFC) cut off 2. None of the genes were highly expressed according to our criteria (Figure 1C). On comparing the 8 samples of HPV- negative European Americans with the 8 samples of African Americans, we found 152 genes that were differentially expressed. The number of genes differentially expressed were a lot but none of the gene showed the high differential expression values (Figure 1D), that is none of the genes gave values greater than 5 for logFC and -log 10(Adj. P.Value) greater than 2.

We also compared all three groups together using moderated F statistics (Table 3). We found the genes that showed the largest F value, also showed the smallest logFC values for the same genes in HPV active-HPV inactive comparison and HPV active-HPV negative comparison. This showed that most of the highly differentially expressed genes were same in the two comparison, which were not as differentially expressed in HPV inactive-HPV negative comparison.

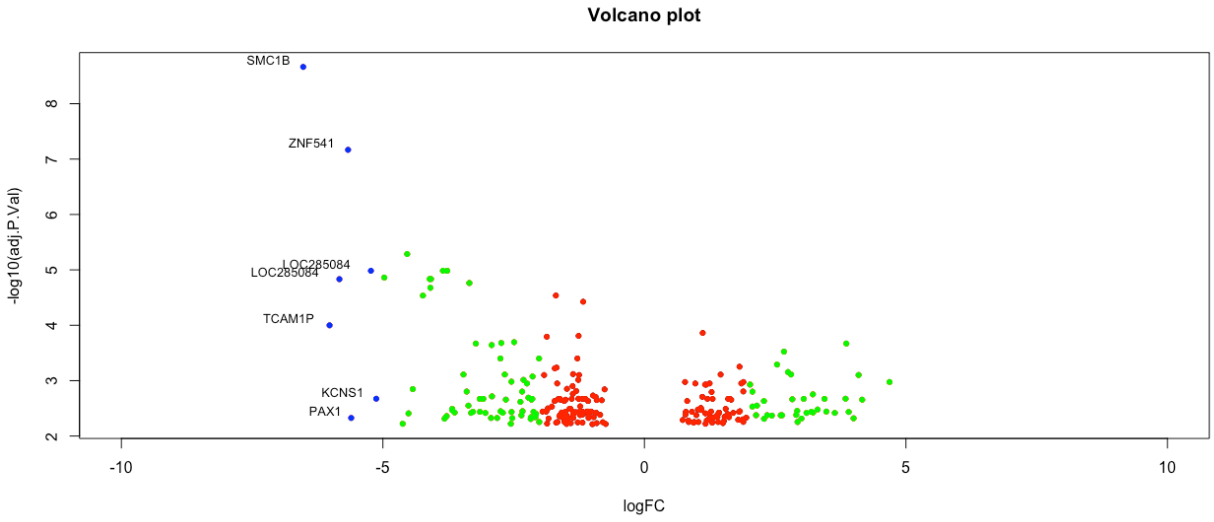
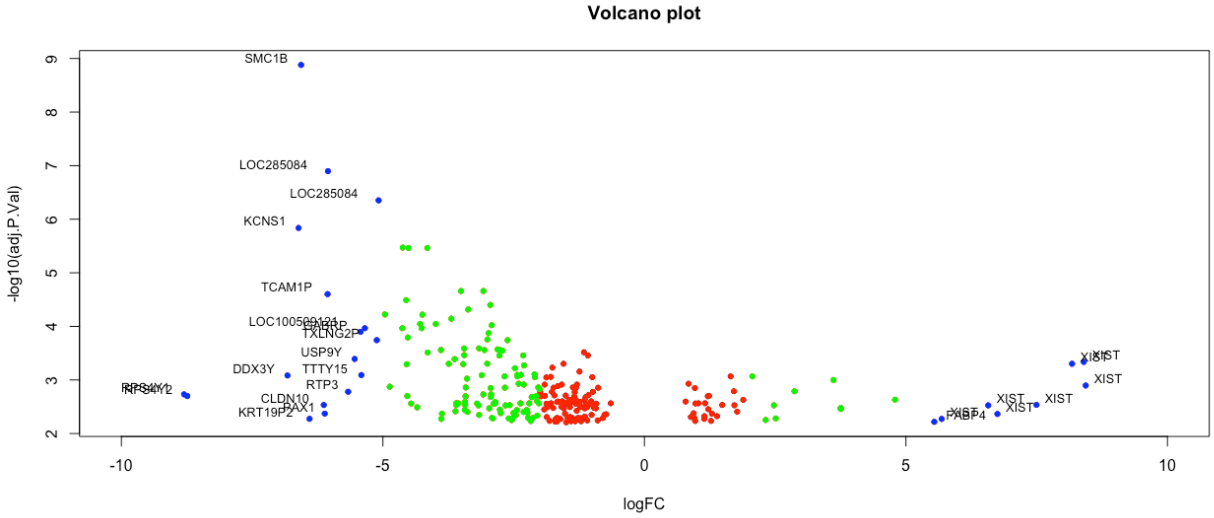


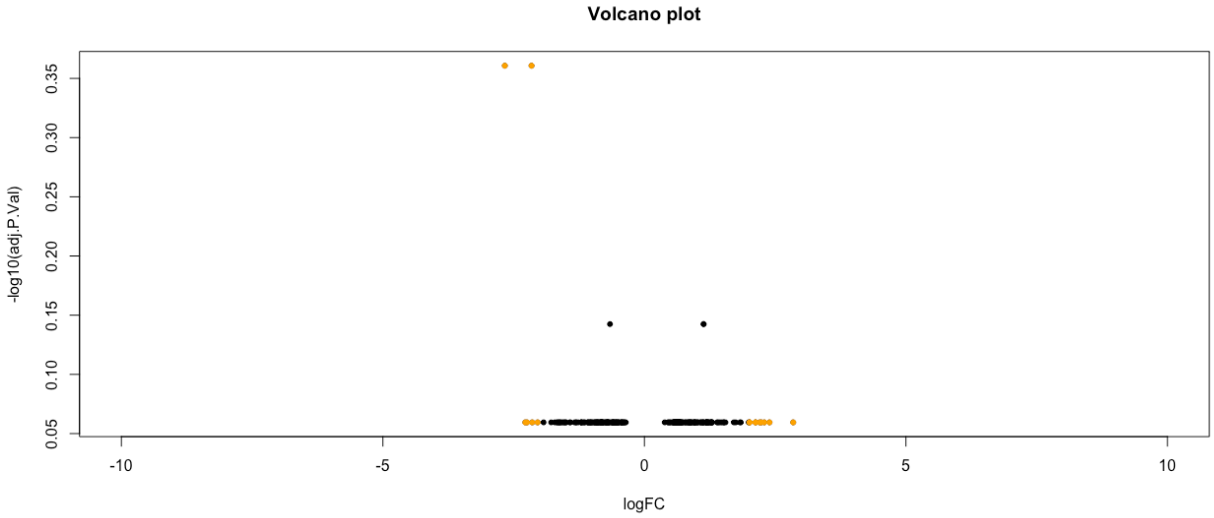
Figure 1

A. Represents a volcano plot where the x- axis is the log fold change of gene expression and the y-axis is the -log 10 base of adjusted p-value between HPV active and HPV negative samples. The red dots in the figure represent the genes that have absolute logFC less than 2 and the green dots represent the genes that have -log 10(Adj.P.Value) of greater than 2 and absolute value of logFC value greater than 2 but less than 5. The Blue dots represent the genes that have -log10(Adj.P.Val) greater than 2 as well as logFC greater than or equal to 5. The genes in blue dots showed the highest expression values. Gene names include: *SMC1B: Structural maintenance of chromosome 1B; ZNF541: Zinc Finger protein 541; LOC285084: uncharacterized gene; LOC254559: uncharacterized gene; TCAM1P: testicular cell adhesion molecule 1; KCNS1: Potassium voltage gated channel; PAX1: Paired box1*

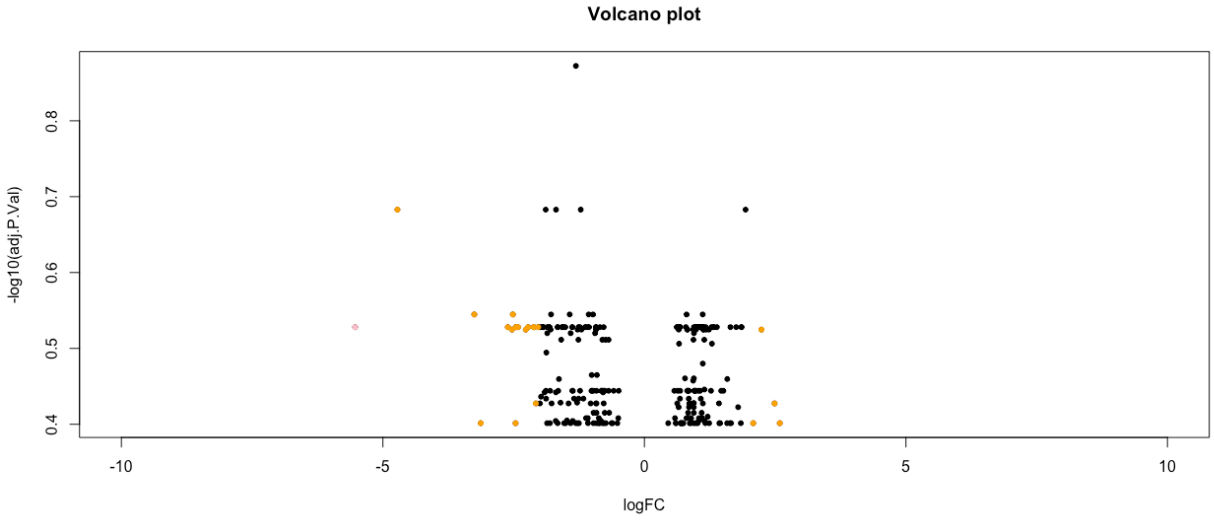


B. Represents a volcano plot where the x- axis is the log fold change and the y- axis is the -log 10 base of adjusted p-value between HPV active and HPV Inactive samples. The red dots in the figure represent the genes that have -log10 (Adj.p.Value) of greater than 2 only and the green dots represent the genes that have -log 10(Adj.P.Value) of greater than 2 and logFC value greater than 2 but less than 5. The Blue dots represent the genes that have both -log 10(Adj. P.Value) greater than 2 and logFC greater than 5. These are the genes that are the most differentially expressed genes. Gene names include: *SMC1B: Structural maintenance of chromosome 1B; ZNF541: Zinc Finger protein 541; LOC285084: uncharacterized gene; LOC254559: uncharacterized gene; TCAM1P: testicular cell adhesion molecule 1; KCNS1: Potassium voltage gated channel; LOC100509121: Uncharacterized protein; TXLNG2P: taxilin gamma 2, pseudogene; USP9Y: ubiquitin specific peptidase 9, Y-linked; TTTY15: testis-specific transcript, Y-linked 15 (non-protein coding); RTP3: receptor (chemosensory) transporter protein 3; DDX3Y: DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked; CLDN10: claudin 10; KRT19P2: keratin 19 pseudogene 2; RPS4Y1: Ribosomal protein S4, Y-linked 1; RPS4Y2:*

Ribosomal protein S4, Y-linked 2; XIST: X (inactive)-specific transcript (non-protein coding); FABP4: fatty acid binding protein 4, adipocyte; PAX1: Paired box1



C. Represents a volcano plot where the x- axis is the log fold change and the y- axis is the -log 10 base of adjusted p-value between HPV negative and HPV Inactive samples. The black dots in the figure represent the genes that are differentially expressed and the orange dots represent the genes that have logFC value greater than 2 but less than 5. None of the genes had -log 10(Adj. P. Value) of greater than 2.



D. Represents a volcano plot where the x- axis is the log fold change and the y- axis is the -log 10 base of adjusted p-value between HPV negative European Americans and African American samples. The black dots in the figure represent the genes that are differentially expressed and the orange dot represents the gene that have logFC greater than 2 but less than 5. The only pink dot represents the gene that has logFC value greater than 5. No gene showed exceptionally high expression values.

The results depicted subtle differences in the gene expression profiles of HPV-inactive and HPV-negative samples and HPV-negative European Americans and African Americas whereas there was more difference to be seen between HPV-active and HPV-inactive, and HPV-active and HPV-negative samples.

GENE_NAME	adj.P.Val	P.Value	F	G2...G0	G2...G1	G1...G0
structural maintenance of chromosomes 1B	4.36E-17	8.6E-22	257.8638	-6.39571	0.16804	-6.563751738
zinc finger protein 541	1.6E-14	6.31E-19	171.9409	-5.75633	-1.13616	-4.620167822
testicular cell adhesion molecule 1 homolog	2.24E-14	1.32E-18	164.1369	-4.14024	0.010051	-4.150289469
uncharacterized LOC285084	4.07E-13	3.35E-17	133.7273	-5.59906	-0.51674	-5.082321134
uncharacterized LOC285084	4.07E-13	4.02E-17	132.1891	-6.38685	-0.33626	-6.050593379
glutamate receptor, ionotropic, N-methyl D-aspartate 2C	1.25E-11	1.48E-15	104.4926	-3.73685	-0.75841	-2.978444336
t-complex 11 homolog (mouse)	1.85E-11	2.56E-15	100.7713	-4.84941	-0.56216	-4.287255909
	1.67E-10	2.64E-14	86.10507	-3.0664	0.011365	-3.077762693
potassium voltage-gated channel, delayed-rectifier	1.39E-09	2.46E-13	73.77498	-5.97744	0.635414	-6.612856547
testicular cell adhesion molecule 1 homolog	3.4E-09	6.7E-13	68.73365	-6.82141	-0.76392	-6.057485568
uncharacterized LOC254559	3.75E-09	8.12E-13	67.79776	-3.55101	-0.54629	-3.004726462
uncharacterized LOC254559	7.24E-09	1.71E-12	64.26441	-3.08785	-0.65628	-2.431573103
kelch-like 35 (Drosophila)	1.91E-08	4.9E-12	59.52089	-3.25501	0.191949	-3.446957787
uncharacterized LOC375196	2.36E-08	6.5E-12	58.29904	-4.24349	-0.10112	-4.142373627
synaptonemal complex protein 2	3.25E-08	9.62E-12	56.63026	-4.91809	-0.36338	-4.554709253
chromosome 19 open reading frame 57	4.01E-08	1.27E-11	55.48514	-2.94324	0.05838	-3.001624707
stromal antigen 3	1.98E-07	6.94E-11	48.76223	-4.00291	-0.57265	-3.430259893
cyclin-dependent kinase inhibitor 2C	1.98E-07	7.02E-11	48.72083	-2.35272	-0.58419	-1.768522982
synaptogyrin 3	3.64E-07	1.36E-10	46.27779	-2.15333	-0.05299	-2.100349285
stromal antigen 3	4.06E-07	1.63E-10	45.6227	-5.22545	-0.69253	-4.53291913
myosin IIIA	4.06E-07	1.68E-10	45.52002	-3.07256	-0.22268	-2.849881918
	4.33E-07	1.88E-10	45.12138	-2.78342	-0.54824	-2.235180382
ring finger protein 165	6E-07	2.72E-10	43.81954	-2.33542	0.610408	-2.94583211
replication protein A2, 32kDa	6.59E-07	3.11E-10	43.34901	-1.23198	0.00905	-1.241027308
uncharacterized LOC100131654	1.18E-06	5.92E-10	41.16942	-2.45023	-0.08147	-2.36875392
synaptonemal complex central element protein 2	1.18E-06	6.03E-10	41.11072	-3.26091	0.246128	-3.507033733
RIB43A domain with coiled-coils 2	1.79E-06	9.53E-10	39.60619	-3.31601	0.94273	-4.25873942
RAN binding protein 17	2.15E-06	1.18E-09	38.90526	-3.2204	0.244448	-3.464844969
CUGBP, Elav-like family member 4	2.68E-06	1.53E-09	38.08054	-3.50987	-1.30723	-2.202636346
Y box binding protein 2	3.12E-06	1.84E-09	37.50608	-3.02746	0.341119	-3.36858182
replication protein A2, 32kDa	7.57E-06	4.63E-09	34.7026	-1.20269	0.108671	-1.311363418
potassium voltage-gated channel	8.45E-06	5.33E-09	34.28561	-3.47844	-0.04446	-3.433973768
	1.02E-05	6.61E-09	33.65712	-1.87172	0.605196	-2.476917865
	1.6E-05	1.1E-08	32.20318	-2.25677	0.662748	-2.919515565
uncharacterized LOC645566	1.6E-05	1.1E-08	32.1916	-2.31606	0.492562	-2.808625944
uncharacterized LOC645566	1.61E-05	1.14E-08	32.08511	-2.57432	0.841031	-3.415347212
uncharacterized LOC100506994	1.87E-05	1.36E-08	31.59803	-2.30499	0.856045	-3.161036803
WAS/WASL interacting protein family, member 3	1.87E-05	1.45E-08	31.41554	-2.34656	-0.47941	-1.86715255
gamma-aminobutyric acid (GABA) A receptor, pi	1.87E-05	1.48E-08	31.36455	-4.50982	0.917503	-5.42732139
Mdm2 p53 binding protein homolog (mouse)	1.87E-05	1.48E-08	31.35567	-1.42658	-0.14907	-1.277513379
	1.87E-05	1.52E-08	31.29815	-2.12247	0.639337	-2.761801803
	1.87E-05	1.55E-08	31.2388	-3.07702	-0.01958	-3.057437836
potassium voltage-gated channel,	1.95E-05	1.68E-08	31.00853	-3.70887	-0.3778	-3.331064956
cyclin-dependent kinase inhibitor 2A	1.95E-05	1.69E-08	30.99482	-3.71174	-0.01939	-3.692344891
ureidopropionase, beta	2.66E-05	2.36E-08	30.09132	-2.48094	0.256689	-2.737624521
chromosome 6 open reading frame 124	2.67E-05	2.42E-08	30.02113	-1.29	0.259187	-1.549187261
LIM domain only 4	2.73E-05	2.53E-08	29.9029	-1.79716	0.520057	-2.31721832
choline dehydrogenase	2.89E-05	2.73E-08	29.6944	-1.53568	-0.15469	-1.380982815
interleukin 17 receptor B	3.14E-05	3.04E-08	29.4104	-3.12023	0.050526	-3.170754283

Table 1:

The table represents the comparison of the three groups and the first 50 differentially expressed genes. ANOVA was performed to compare all the three groups (HPV-active, HPV-inactive and HPV-negative). G2…G0 represents the logFC values for HPV active (G0) with HPV-negative (G2). G2…G1 represents the logFC values for HPV-negative with HPV-inactive (G1). G1…G0 represents the logFC values for HPV-inactive with the HPV active samples.

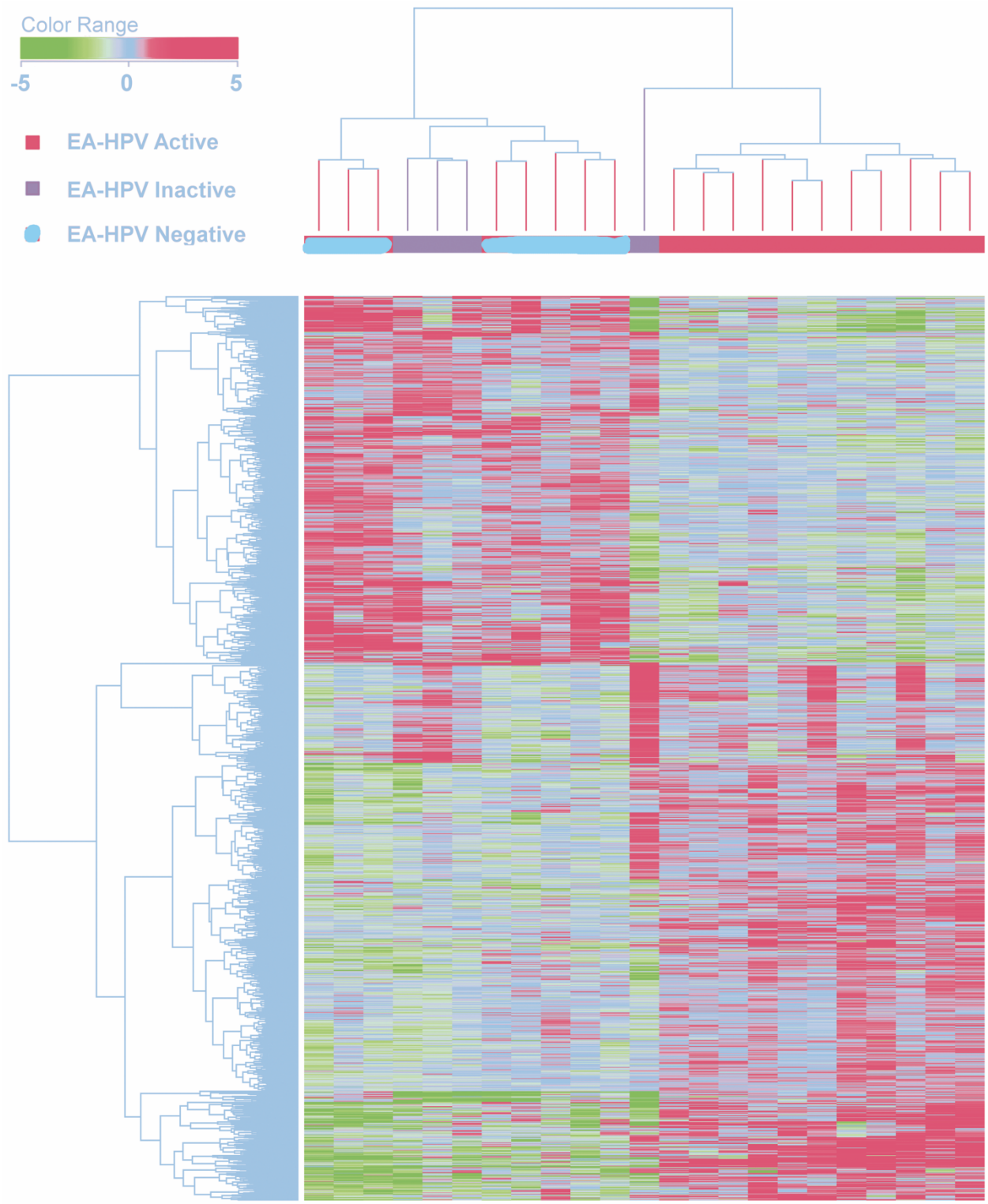


Figure 2

Unsupervised hierarchical clustering for oropharyngeal cancers comparing HPV-active, HPV-inactive and HPV-negative samples.

We also performed a hierarchical cluster analysis on our three groups: HPV-active, HPV-inactive and HPV- negative. We found that our tumor samples, irrespective of the race they belonged to cluster as we saw in our above group comparisons (Figure 2). Our HPV-inactive tumor samples clustered with HPV-negative tumor samples and both the groups segregated from the HPV-active group.

Discussion

The study conducted by Tomar et al. 2015, classified HPV status based on HPV DNA detection and the expression of E7 through RT-qPCR. Hence, the tumors that tested positive for both INNO-LiPA and E7 RT-qPCR were classified as HPV-positive, the ones that tested positive only for INNO-LiPA were classified as HPV-inactive, whereas the ones that tested negative for both were classified as HPV-negative tumors. Hence, this study suggests that HPV DNA is not the sole reason for the HPV status of head and neck cancer.

Looking at the gene expression profile, we found that there were certain genes that were differentially expressed in African Americans as compared to European Americans in HPV negative patients. Some of these included ATP9A (which has been found to be associated with lymphocytic leukemia). Other genes that showed differential expression included ADAM12, MMP7, IL22RA1. MMP7 and ADAM12 have been found to be associated with protein degradation in oral cancers (Kamatani et al., 2013), which explains the gene expression profile difference in cancer progression between HPV-negative cancers in European Americans and African Americans. In our moderated F statistics comparison for all the three groups we found almost the same genes being highly differentially in HPV active-HPV inactive and HPV-active and HPV-negative comparisons whereas the genes that were highly differentially expressed in the two groups showed small absolute logFC value for the group HPV-inactive and HPV-

negative (Table 3). This (along with Figure 1C that shows very few genes that are differentially expressed), indicates towards the possibility that HPV inactive may have originated as HPV active and then lost its dependence on HPV as the disease progressed. This explains the subtle difference between HPV-inactive and HPV- negative cancers. This hypothesis if explored might help in looking at the potential of HPV vaccines in the prevention of certain cancers.

Our analysis showed that most of the genes that were differentially expressed in HPV-active and HPV-inactive sample analysis were similar from the genes that were differentially expressed in HPV-active and HPV- negative sample analysis (Figure 1A and Figure 1B). Some of these gene were not similarly expressed in the two group comparisons. HPV active-HPV-inactive comparison showed more differentially expressed genes (Figure 1B). This showed that E7 expression through RT-qPCR resulted in HPV inactive group that lead to a different molecular process in the causation of cancer. This depicted that HPV-inactive (with E7 expression through RT-qPCR) is a different group and the presence of E7 expression results in a different molecular and biological process in the causation of cancer. We can also see the difference in HPV-inactive from HPV-active by our cluster analysis, which showed that HPV-inactive clustered with HPV-negative (Figure 2).

Limitations

There were certain limitations of the present study that should be noted. First, the sample size of our study was really small. The limited power of the analysis prevented us from comparing certain groups like African Americans HPV-active samples to HPV-negative. Another limitation of this study was it prevented us from getting a differential expression profile for every gene. Since, this was a secondary data, we were limited by the information provided to us and so we couldn't conduct RT-qPCR analysis. Hence, getting the exact value of gene

expression in each group could not be done. Another limitation of this study was the lack of any demographic information in the dataset. Since the data just provided the background intensity values, it was not possible for us to see if there were any differences due to demographic characteristics.

Future Research

The difference seen in African Americans HPV-negative samples and European American HPV negative samples requires further investigation with larger sample size. Our analysis showed that HPV-inactive affects different molecular and biological process need to be further investigated by getting RT-qPCR examination of all the differentially expressed genes and getting the exact value of expression of all the genes instead of comparison of gene values only. It would also be useful to compare each of the HPV associated Oral squamous cell tissue groups (European Americans and African Americans) with their own normal tissue samples. Comparing it with normal tissue on a larger dataset can be helpful in understanding how the process of oral cancer is progressing based on the HPV status.

References

Andersen, A. S., Koldjær Sølling, A. S., Ovesen, T., & Rusan, M. (2014). The interplay between HPV and host immunity in head and neck squamous cell carcinoma. *International Journal of Cancer*, *134*(12), 2755–2763. https://doi.org/10.1002/ijc.28411

Brawley, O. W. (2009). Oropharyngeal Cancer, Race, and the Human Papillomavirus. *Cancer Prevention Research*, *2*(9), 769–772. https://doi.org/10.1158/1940-6207.CAPR-09-0150

Cole, L., Polfus, L., & Peters, E. S. (2012a). Examining the Incidence of Human Papillomavirus-Associated Head and Neck Cancers by Race and Ethnicity in the U.S., 1995–2005. *PLoS ONE*, *7*(3), e32657. https://doi.org/10.1371/journal.pone.0032657

Cole, L., Polfus, L., & Peters, E. S. (2012b). Examining the Incidence of Human Papillomavirus-Associated Head and Neck Cancers by Race and Ethnicity in the U.S., 1995–2005. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0032657

Curado, M. P., & Boyle, P. (2013). Epidemiology of head and neck squamous cell carcinoma not related to tobacco or alcohol. *Current Opinion in Oncology*, *25*(3). Retrieved from http://journals.lww.com/co-oncology/Fulltext/2013/05000/Epidemiology_of_head_and_neck_squamous_cell.5.aspx

Dal Maso, L., Torelli, N., Biancotto, E., Di Maso, M., Gini, A., Franchin, G., … Polesel, J. (2016). Combined effect of tobacco smoking and alcohol drinking in the risk of head and neck cancers: a re-analysis of case–control studies using bi-dimensional spline models. *European Journal of Epidemiology*, *31*(4), 385–393. https://doi.org/10.1007/s10654-015-0028-3

Daraei, P., & Moore, C. E. (2015). Racial Disparity Among the Head and Neck Cancer Population. *Journal of Cancer Education*, *30*(3), 546–551. https://doi.org/10.1007/s13187-014-0753-4

Deng, Z., Hasegawa, M., Kiyuna, A., Matayoshi, S., Uehara, T., Agena, S., … Suzuki, M. (2013). Viral load, physical status, and E6/E7 mRNA expression of human papillomavirus in head and neck squamous cell carcinoma. *Head & Neck*, *35*(6), 800–808. https://doi.org/10.1002/hed.23034

Firmino, N., Martinez, V. D., Rowbotham, D. A., Enfield, K. S. S., Bennewith, K. L., & Lam, W. L. (2016). HPV status is associated with altered PIWI-interacting RNA expression pattern in

head and neck cancer. *Oral Oncology*, *55*, 43–48. https://doi.org/10.1016/j.oraloncology.2016.01.012

Gillison, M. L., D'Souza, G., Westra, W., Sugar, E., Xiao, W., Begum, S., & Viscidi, R. (2008). Distinct Risk Factor Profiles for Human Papillomavirus Type 16–Positive and Human Papillomavirus Type 16–Negative Head and Neck Cancers. *Journal of the National Cancer Institute*, *100*(6), 407–420. https://doi.org/10.1093/jnci/djn025

Gourin, C. G., & Podolsky, R. H. (2006). Racial Disparities in Patients With Head and Neck Squamous Cell Carcinoma. *The Laryngoscope*, *116*(7), 1093–1106. https://doi.org/10.1097/01.mlg.0000224939.61503.83

Kamatani, T., Shiogama, S., Yoshihama, Y., Kondo, S., Shirota, T., & Shintani, S. (2013). Interleukin-1 beta in unstimulated whole saliva is a potential biomarker for oral squamous cell carcinoma. *Cytokine*, *64*(2), 497–502. https://doi.org/10.1016/j.cyto.2013.08.011

Klaes, R., Friedrich, T., Spitkovsky, D., Ridder, R., Rudy, W., Petry, U., … von Knebel Doeberitz, M. (2001). Overexpression of p16INK4A as a specific marker for dysplastic and neoplastic epithelial cells of the cervix uteri. *International Journal of Cancer*, *92*(2), 276–284. https://doi.org/10.1002/ijc.1174

Li, Y., Nichols, M. A., Shay, J. W., & Xiong, Y. (1994). Transcriptional Repression of the D-Type Cyclin-dependent Kinase Inhibitor p16 by the Retinoblastoma Susceptibility Gene Product pRb. *Cancer Research*, *54*(23), 6078.

Pike, N. (2011). Using false discovery rates for multiple comparisons in ecology and evolution. *Methods in Ecology and Evolution*, *2*(3), 278–282. https://doi.org/10.1111/j.2041-210X.2010.00061.x

Tomar, S., Graves, C. A., Altomare, D., Kowli, S., Kassler, S., Sutkowski, N., … Pirisi, L. (2016). Human papillomavirus status and gene expression profiles of oropharyngeal and oral cancers from European American and African American patients. *Head & Neck*, *38*(S1), E694–E704. https://doi.org/10.1002/hed.24072

van Hamont, D., van Ham, M. A. P. C., Bakkers, J. M. J. E., Massuger, L. F. A. G., & Melchers, W. J. G. (2006). Evaluation of the SPF(10)-INNO LiPA Human Papillomavirus (HPV) Genotyping Test and the Roche Linear Array HPV Genotyping Test. *Journal of Clinical Microbiology*, *44*(9), 3122–3129. https://doi.org/10.1128/JCM.00517-06

Vlashi, E., Chen, A. M., Boyrie, S., Yu, G., Nguyen, A., Brower, P. A., … Pajonk, F. (2016). Radiation-Induced Dedifferentiation of Head and Neck Cancer Cells Into Cancer Stem Cells Depends on Human Papillomavirus Status. *International Journal of Radiation Oncology*Biology*Physics*, *94*(5), 1198–1206. https://doi.org/10.1016/j.ijrobp.2016.01.005

Wang, H., Park, S., Lee, D., Kim, S., Kim, G., Park, K. H., & Lee, H. (2015). Prevalence of type-specific oncogenic human papillomavirus infection assessed by HPV E6/E7 mRNA among women with high-grade cervical lesions. *International Journal of Infectious Diseases*, *37*, 135–142. https://doi.org/10.1016/j.ijid.2015.06.018

Weinberger, P. M., Merkley, M. A., Khichi, S. S., Lee, J. L., Psyrri, A., Jackson, L. L., & Dynan, W. S. (2010). HPV-Active Head and Neck Cancer and Ethnic Health Disparities. *The Laryngoscope*, *120*(8), 1531–1537. https://doi.org/10.1002/lary.20984

Weinberger, P. M., Yu, Z., Haffty, B. G., Kowalski, D., Harigopal, M., Brandsma, J., … Psyrri, A. (2006). Molecular Classification Identifies a Subset of Human Papillomavirus–Associated Oropharyngeal Cancers With Favorable Prognosis. *Journal of Clinical Oncology*, *24*(5), 736–747. https://doi.org/10.1200/JCO.2004.00.3335

Wilson DD, Rahimi AS, Saylor DK, & et al. (2012). P16 not a prognostic marker for hypopharyngeal squamous cell carcinoma. *Archives of Otolaryngology–Head & Neck Surgery*, *138*(6), 556–561. https://doi.org/10.1001/archoto.2012.950

Yim, E.-K., & Park, J.-S. (2005). The Role of HPV E6 and E7 Oncoproteins in HPV-associated Cervical Carcinogenesis. *Cancer Research and Treatment : Official Journal of Korean Cancer Association*, *37*(6), 319–324. https://doi.org/10.4143/crt.2005.37.6.319

Zandberg, D. P., Liu, S., Goloubeva, O., Ord, R., Strome, S. E., Suntharalingam, M., … Cullen, K. J. (2016). Oropharyngeal Cancer (OPC) Drives Racial Outcome Disparities in Squamous Cell Carcinoma of the Head and Neck (HNSCC): Ten Year Experience at the University of Maryland Greenebaum Cancer Center (UMGCC). *Head & Neck*, *38*(4), 564–572. https://doi.org/10.1002/hed.23933