Title: Identification and characterization of novel human papillomaviruses in oral rinse samples from oral cavity and oropharyngeal cancer patients.

Running title: Novel HPVs in oral cancer patients

Individual entry for Research Category

Juliet Dang, PhD, MS, RDH
jhtdang@uw.edu
206.359.0134
University of Washington, Seattle, WA, USA

Contributing authors:
Qing Zhang, PhD
Fred Hutch, Seattle, WA, USA

Gregory A. Bruce, PhD
Seattle Children's Research Institute, Seattle, WA

This manuscript has not been previously published.

Key words: HPV, oral cancer, oropharyngeal cancer, NGS, oral rinse samples

Abstract

Objectives
The objectives of this study were to: I) Discover novel HPVs using Next Generation Sequencing (NGS) technology in oral rinse samples collected from OCC and OPC patients; II) Determine prevalence of novel HPVs in archived OCC and OPC tissue samples; and III) Examine frequency of novel oncogenic HPVs in cancer and non-cancer oral rinse samples using real-time PCR.

Statement of Purpose
To identify and characterize novel HPVs, in oral rinse samples from OCC and OPC patients, in order to gain a better understanding of HPV-associated cancers.

Methods
Oral rinse samples were collected from 100 head and neck cancer patients, and 110 healthy patients. NGS techniques were used to detect for novel HPVs.

Results & Conclusions
3 new potential types of HPV were discovered. Novel virus (NV) 14.4 was closely related to HPV76 with an 89% homology, and is part of the betapapillomavirus (β-PV) genus; NV69.1 was distantly related to the alphapapillomavirus (α-PV) genus; and NV95 was closely related to HPV147 with a 65-77% homology, and is part of the gammapapillomavirus (γ-PV) genus. In archived oral tissue samples, NV14.4 was detected from a single patient with OCC. Of the oral rinse samples, NV69.1 was more prevalent than the other two novel viruses. Our results demonstrate there are novel HPVs present in oral rinse samples that may be associated with OCC and OPC.
In conclusion, novel HPVs can be identified from oral rinse samples of cancer patients using NGS.

Clinical Relevance

Scientific rationale for study
Due to the rise of HPV-associated oral and oropharyngeal cancers, which is following the same paradigm as cervical cancers, up-to-date research is imperative.

Principal Findings
We identified 3 novel HPVs in the oral rinse samples of cancer patients.

Practical implications
We currently know that HPV types 16 and 18 are associated with oral and oropharyngeal cancers. However, with the rise of HPV infection new types should be identified to gain a better understanding of the natural history of HPV. Such knowledge will lead to better preventative and diagnostic measures, as well as treatments.

Introduction

The role of the dental hygienist is crucial in overall oral health, oral disease prevention, and education for patients. The importance of cancer screenings is ever increasing, as well as the need for dental hygienists to be up-to-date with such topics, and to be involved in the research as well. This manuscript provides details of our original and innovative clinical/translational study.

Background

Human papillomavirus (HPV) is a circular, double-stranded DNA molecule (1). 79 million Americans are currently infected with HPV, and every year 14 million new infections will occur (2). HPVs that affect mucosal regions are classified as low-risk, which usually produce warts, or high-risk, which are associated with cervical cancer (3, 4). High-risk HPVs contain the oncoproteins E6 and E7, which are responsible for inhibition of apoptosis, deactivation of tumor suppressor proteins, and creating an environment for genome instability, thus increasing risk for malignancy (4).

Annually, it is estimated that 263,000 oral and 135,000 pharyngeal cancers occur globally (5). Grouping the two cancers together, the result is the sixth most common cancer in the world (6). Oral HPV infection is increasing at considerable rates and the projected number of HPV-positive oropharyngeal cancer cases (OPC) is expected to surpass the annual number of cervical cancer cases by 2020 (7). The most prevalent type of HPV associated with oral infection is type 16 (8, 9). Both type 16 and 18 have been demonstrated to be oncogenic in OPCs (10-12).

According to the PaVE search database there are almost 200 different HPV types (http://pave.niaid.nih.gov/#search/search_database). With the modern

techniques and methods available, it should not be impossible to identify novel HPVs within the head and neck. The guidelines for determining whether an HPV is novel can be found on the PaVE submission process page (http://pave.niaid.nih.gov/#explore/taxonomy/submission_process). When studying an unknown HPV, if there is a greater than 10% difference in homology of the L1 region, when compared to all existing HPVs, the unknown HPV is potentially new.

From our previous study, we identified and characterized three novel HPVs from the oral rinse samples of healthy individuals (13). HPV 171, 172, and 173 were all of the γ-HPV type.

We hypothesize there are unidentified novel HPVs in the oral rinse samples of oral and oropharyngeal cancer patients.

Study Population

Between 2011 and 2013, we recruited 100 cancer patients from the Seattle Cancer Care Alliance (Seattle, WA), and 110 healthy subjects from University of Washington Dental Clinic (Seattle, WA). The cancer cases included patients with oropharyngeal cancer (OPC), oral cavity cancer (OCC), laryngeal cancer, sinus cancer, and supraglottis cancer. We screened the schedules of five oncologists in order to identify eligible cancer patients, and discussed our study at their appointment. One patient declined due to mouth sores and sensitivity. 21/100 cancer patients had already begun treatment: 18 patients had treatment less than 21 days before sampling, 2 had over 30 days of treatment, and one patient had treatment for 7 months. Healthy subjects were randomly selected within the student dental clinic, one patient declined to take part in the study. Inclusion requirements for the healthy population included being: cancer-free, not pregnant, HIV-free, and over the age of 16. Each patient signed consent forms and answered a simple health questionnaire. Gender, age, race, smoking and alcohol history were recorded for all subjects. All human subjects IRB protocols and regulations were followed under the Fred Hutchinson Cancer Research Center guidelines (IRB #7490 approved April 9, 2014)

Collection & DNA Purification Methods

For sample collection, all patients rinsed and gargled for 30 seconds with Original Mint Scope® mouthwash. Four normal, healthy individuals requested to use Crest® Alcohol-free mouthwash due to a history of alcoholism. Oral rinse samples were centrifuged for 15 min. at 4°C to form a pellet, the supernatant was discarded, and the pellet was placed in -80°C until further processing. The Puregene® DNA Purification Kit was used to isolate genomic DNA from the buccal cell pellet within the mouthwash samples (Qiagen item #158467, manufacturer's protocol was followed).

HPV & Analytic Methods

Identification of novel HPVs

We first determined whether our samples were infected with HPV16 and 18. Taqman real-time PCR (RT-PCR) assays were used for detection on the ABI Prism 7900 Sequence Detection System with 40 cycles in a reaction (denaturation at 95°C, annealing and extension at 60°C). Absolute quantification was used to determine HPV16 and 18 viral load, and total human genomic DNA in the sample was determined on Alu sequences. Serial dilutions of human genomic DNA and full length HPV16 and 18 plasmids, of known concentrations, were used as standard curves.

Table 1. Primers and probes used for RT-PCR.

	HPV16 E7	HPV18 E7	Alu
Forward primer	CGGACAGAGCCCATTACAATATT	CCGACGAGCCGAACCA	GGCCAACACGGTGAAACC
Reverse primer	CGCACAACCGAAGCGTAGA	TGGCTTCACACTTACAA CACATACA	CCACGCCCGGCTAATTTT
Probe	TAACCTT(T/C)TGTTGCAAGTGT	AACGTCACACAATGTT	CGTCTCTACTAAAAATAC

In order to increase the efficiency of HPV detection, we used the multiply-primed rolling-circular amplification technique (MP-RCA) to preferentially amplify unknown, circular HPV DNA. MP-RCA has been demonstrated to amplify circular DNA templates up to 10^7-fold (14). The TempliPhi 100 Amplification Kit (Amersham Biosciences) protocol was followed.

The fluorescent arbitrarily primed (FAP) PCR published protocol was followed to detect for a broad range of HPV types where primers were developed from conserved L1 regions (15). We only performed this technique on the cancer case samples.

NGS was performed on 7-pooled samples using the Illumina Hi-Seq 2500 platform (Figure 1). As for supervised assembly, a total of 189 HPV L1 gene sequences and whole genome sequences were downloaded from the papillomavirus knowledge source at http://pave.niaid.nih.gov/. This data set includes a number of non-reference genomes. Shot reads were aligned to the L1 gene region using BWA (http://bio-bwa.sourceforge.net/, version 0.7.12) with default settings. Subsequently, the whole genome sequences were used as the reference database if majority of the short reads did not map to the specific L1 gene region.

De novo assembly strategy was applied for cases where the majority of the short reads did not align to HPV genome sequences at all. The short reads were first aligned to the human genome UCSC hg19 (http://support.illumina.com/sequencing/sequencing_software/igenome.html), and then unmapped reads were selected to run Velvet (https://www.ebi.ac.uk/~zerbino/velvet/, version 1.2.10) to construct high quality unique contigs (short DNA fragments). Contigs with 200 bp or larger in length and with a minimum 100 coverage were aligned (BLASTn) to both HPV Specific database at http://pave.niaid.nih.gov/#search/pv_specific_blast, and to NCBI nucleotide collection (nt) database at http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome.

BLASTn and PaVE databases were used to determine homology of the sequences to HPV. A homology of <90% of the L1 region indicates a novel HPV (3, 16).

Ten samples that were FAP PCR positive but not HPV16 positive, underwent Sanger sequencing to determine if novel HPVs were present. Five samples had ambiguous and poor quality results, thus cloning was performed using CloneJet PCR Cloning Kit by Thermo Scientific (manufacturer's protocol followed). Sanger sequencing by Genewiz was executed after cloning followed by BLASTn (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins/) and PaVE (http://pave.niaid.nih.gov/) database search for similar HPVs. Chen et al. recommend using both databases in order to determine homology (17).

Clustal Omega was used to align multiple sequences to produce a phylogenetic tree for observation of evolutionary relationships (http://www.ebi.ac.uk/Tools/msa/clustalo/). Novel viruses were compared to all HPVs in the PaVE database and sequences from the L1 region were obtained.

Determining prevalence of novel HPVs in oral rinse samples

OligoArchitect by Sigma (http://www.sigmaaldrich.com/technical-documents/articles/biology/oligoarchitect-online.html) was used to create custom primers and probes for the three potential novel HPVs. Using RT-PCR Taqman assays we determined the frequency of the novel oncogenic HPVs in cancer and non-cancer oral rinse samples.

We categorized smoking history as follows: non-smoker (0 packs); light smoker (<1pack/week); moderate smoker (≥1pack/week ≤1 pack/day); heavy smoker (≥1pack/day). For those who smoked cigars or chewed tobacco we calculated to the equivalent of packages of cigarettes smoked. Alcohol history was

categorized as follows: none (never drinks); rarely/occasionally (1 drink every 1-2 months); light (1-6 drinks/week for females, 1-13 drinks/week for males); moderate (7 drinks/week for females, 14 drinks/week for males); heavy (>7drinks/week for females, >14 drinks/week for males).

Determining prevalence of novel HPVs in archived tissue blocks

Using the same newly created primers and probes, the presence of the new viral sequences in 106 normal and 115 malignant (OCC/OPC) oral tissues was determined by HPV-type specific quantitative RT-PCR Taqman assays. Archived oral tissue blocks were accessed through the Department of Pathology's repository (University of Washington). A total of 80μm of tissue was cut with a microtome, with a new blade for each tissue block to eliminate contamination. Genomic DNA was isolated using RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE Tissues (Applied Biosystems) according to manufacturer's protocol.

All data analysis was completed using Stata MP 13.1 (StataCorp LP, Texas, USA).

Results

We had substantial amplification present after MP-RCA as our HPV16 positive control had enrichment of almost 800,000-fold compared to only a 0.5-fold amplification with Siha (cervical cancer cell line with integrated HPV16), which was our negative control (Table 2). Thus, circular DNA should have amplified substantially whereas linear DNA should have had minimal amplification.

From NGS, we only had reads from sample 1 where 82% of the reads mapped onto the L1 gene of all 189 HPV types in the PaVE database (Table 3). For sample 2, we should have had reads mapped onto the HPV L1 gene as this was our positive control, but no results were seen. We did not see any results in sample 3 and 4. Sample 5 had 31% of reads mapped onto the L1 gene; sample 6 had 55%; and sample 7 had 51% of reads mapped onto the L1 gene (Table 3).

With the results from Sanger sequencing of the ten pre-selected samples, we matched the HPV type results from the BLASTn and PaVE search with the NGS mapped reads (Table 4). OCL3, 40, and 42 were positive for HPV172, which was one of the novel HPVs that our lab discovered. OCL14 and 93 presented with multiple HPV infection. Our HPV16 control samples demonstrated multiple HPV infection as well (Table 5).

Five samples that did not have clear Sanger sequencing results were cloned (Table 6). For each sample we chose five different clones to be sequenced. The

samples that we chose to investigate further were OCL14.4, 69.1, and 95 because their % homology difference was greater than 10%.

Each potential novel HPV was isolated from all males; type of cancer was OPC, OCC, or laryngeal cancer; patients were 60 or over; two patients had a heavy smoking history while the other had a light history; alcohol history ranged from rarely to heavy (Table 7).

Phylogenetic tree analysis was completed for each of the novel HPVs (NV) (Figures 2-4). NV14.4 was closely related to HPV76 with an 89% homology, and is part of the betapapillomavirus (β-PV) genus. NV 69.1 was distantly related to the alphapapillomavirus (α-PV) genus. NV95 was closely related to HPV147 with a 65-77% homology, and is part of the gammapapillomavirus (γ-PV) genus.

Newly created primers and probes were created for the three NVs (Table 8). Out of the three NVs, only NV14.4 was found to be positive in an OCC archived tissue biopsy (Table 9). Within the oral rinse samples NV14.4 was detected in 1 OPC and 1 OCC sample; NV69.1 was detected in 10 OPC, 2 OCC, and 1 head and neck cancer (HNC) sample; NV95 was detected in 1 OPC and 1 OCC sample (Table 10).

Discussion

Through NGS we were able to identify HPVs in oral rinse samples. However, in order to obtain an accurate sequence, Sanger sequencing was used. Regarding amplification with MP-RCA, our control seemed to have significant enrichment, however it is possible that other samples with HPV may not have amplified as well. MP-RCA has been demonstrated to amplify circular DNA templates up to 10^7-fold (14). Thus, if there was sufficient enrichment, NGS should have allowed us to discover the novel HPVs without resorting to Sanger sequencing. Pooling of samples may also have caused more confusion when analyzing, as the amount of data obtained is extremely large.

From the NGS results, OCL3, 40, and 42 were positive for HPV172, which was one of the HPVs our lab identified. What is interesting is that HPV172 was predominately seen in non-cancer tissue biopsies (13). It could be possible that HPV172 is becoming more prominent in oral HPV infection in general.

The nucleotide searches accomplished using BLASTn and PaVE coincided with each other for the majority of the time. For OCL95 there was a notable difference in results where BLASTn demonstrated a 77% homology compared to PaVE, which showed a 65% homology to HPV147. We see an even greater difference between databases for OCL14, 80% vs. 49%. Each database may contain different variants of HPVs, which would result in inconsistencies between the two.

Each of the novel HPVs belonged to separate genera although OCL69.1 was possibly distantly related to α-PVs. Obtaining the entire genome would give us a clearer picture of its HPV homology.

In the archived oral tissue biopsies we only detected OCL14.4 in one of the OCC samples. DNA degradation within the tissue blocks is possible for older samples, which limits our results. The oral rinse samples, on the other hand, demonstrated more frequent infection of the NVs compared to the tissue blocks. All tissue and oral rinse samples were quantified for human genomic DNA, and all samples had adequate amounts for detection (Figures 5 and 6).

One major limitation to our study was budget. NGS is expensive, thus we opted to make the most use of our runs by pooling our samples. However, if budget was not an issue, less pooling would be preferred.

Through NGS we were able to identify HPVs in oral rinse samples. With Sanger sequencing we were able to identify and characterize 3 different novel HPVs in oral rinse samples. We obtained similar results with both sequencing platforms when detecting for HPVs, and propose that NGS is a more modern technique, which should be investigated further.

Conflict of interests

None

Sources for funding

Warren G. Magnuson Scholarship
University of Washington Royalty Research Fund
ITHS TL1 Training Program

References

1. Dell G, Gaston K. Human papillomaviruses and their role in cervical cancer. Cellular and molecular life sciences : CMLS. 2001;58(12-13):1923-42.
2. CDC. Genital HPV Infection - Fact Sheet 2015 [updated February 23, 2015; cited 2015 Nov 25th]. Available from: http://www.cdc.gov/std/HPV/STDFact-HPV.htm - a7.
3. Bernard HU, Burk RD, Chen Z, van Doorslaer K, zur Hausen H, de Villiers EM. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology. 2010;401(1):70-9.
4. Doorbar J, Quint W, Banks L, Bravo IG, Stoler M, Broker TR, et al. The biology and life-cycle of human papillomaviruses. Vaccine. 2012;30 Suppl 5:F55-70.

5. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. International journal of cancer Journal international du cancer. 2010;127(12):2893-917.
6. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. Oral oncology. 2009;45(4-5):309-16.
7. Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, Kim E, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2011;29(32):4294-301.
8. Gillison ML, Broutian T, Pickard RK, Tong ZY, Xiao W, Kahle L, et al. Prevalence of oral HPV infection in the United States, 2009-2010. JAMA : the journal of the American Medical Association. 2012;307(7):693-703.
9. Hariri S, Unger ER, Sternberg M, Dunne EF, Swan D, Patel S, et al. Prevalence of genital human papillomavirus among females in the United States, the National Health And Nutrition Examination Survey, 2003-2006. The Journal of infectious diseases. 2011;204(4):566-73.
10. Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. Journal of the National Cancer Institute. 2000;92(9):709-20.
11. Mineta H, Ogino T, Amano HM, Ohkawa Y, Araki K, Takebayashi S, et al. Human papilloma virus (HPV) type 16 and 18 detected in head and neck squamous cell carcinoma. Anticancer research. 1998;18(6B):4765-8.
12. Michaud DS, Langevin SM, Eliot M, Nelson HH, Pawlita M, McClean MD, et al. High-risk HPV types and head and neck cancer. International journal of cancer Journal international du cancer. 2014;135(7):1653-61.
13. Martin E, Dang J, Bzhalava D, Stern J, Edelstein ZR, Koutsky LA, et al. Characterization of three novel human papillomavirus types isolated from oral rinse samples of healthy individuals. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology. 2014;59(1):30-7.
14. Nelson JR, Cai YC, Giesler TL, Farchaus JW, Sundaram ST, Ortiz-Rivera M, et al. TempliPhi, phi29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. BioTechniques. 2002;Suppl:44-7.
15. Forslund O, Antonsson A, Nordin P, Stenquist B, Hansson BG. A broad range of human papillomavirus types detected with a general PCR method suitable for analysis of cutaneous tumours and normal skin. The Journal of general virology. 1999;80 (Pt 9):2437-43.
16. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. Virology. 2004;324(1):17-27.
17. Chen Z, Freitas LB, Burk RD. Evolution and classification of oncogenic human papillomavirus types and variants associated with cervical cancer. Methods in molecular biology. 2015;1249:3-26.

Table 2. MP-RCA preferential amplification. HPV16 acted as a positive control, and Siha a negative control.

	HPV16 gene		Alu gene		
Sample	Before MP-RCA (Ct)	After MP-RCA (Ct)	Before MP-RCA (Ct)	After MP-RCA (Ct)	Enrichment*
HPV16	32	16	34	35	786,431
Siha	29	30	26	24	0.5

*Enrichment was calculated using the $2^{-\Delta\Delta Ct}$ method, where $\Delta\Delta Ct = [Ct_{HPV}(\text{after RCA}) - Ct_{HPV}(\text{before RCA})] - [Ct_{ALU}(\text{after RCA}) - Ct_{ALU}(\text{before RCA})]$

Table 3. NGS data analysis. Mapped reads indicate NGS reads, which were mapped onto the extracted L1 region of all HPV types.

Pooled sample	Total Reads	Mapped Reads	% Mapped Reads to L1 gene
1. FAP PCR	338374994	277598421	82.04
2. HPV16 RCA	394607770	9	0
3. FAP PCR RCA	380204724	67	0
4. RCA	382376134	87	0
5. FAP PCR RCA	196273326	61518757	31.34
6. FAP PCR RCA2	229994340	126833119	55.15
7. FAP PCR RCA16	199750516	101856187	50.99

Table 4. L1 sequence analysis. Matched OCL sample indicates patient sample where Sanger sequencing results coincided with results from NGS.

Pooled sample	HPV type	NGS total reads	Matched OCL sample
1. FAP PCR	172	109936978	3, 40, 42
	8	39473771	14
	23	37762624	14
	76	4134700	14
5. FAP PCR RCA	62	26223379	93
	152	15993736	69
	32	6799264	59
	122	461	93
6. FAP PCR RCA2	90	35679275	52
	23	26865855	14
	11	26393016	96
	8	23967471	14
	33	1390467	72
	105	1112699	63
	76	490440	14
	147	15910	95

Table 5. L1 sequence analysis from NGS in pooled sample 7, which acted as positive control.

HPV type	Total reads
32	25896574
24	23472631
80	16758587
111	14702114
143	8484485
16	4781520
92	4685275
138	660990
135	582723
134	552801
76	137580
12	39851

172	39768
121	6491

Table 6. Sanger sequencing results. BLASTn and PaVE database search results along with HPV types associated with the samples. % homology was the product of the query identity and the max identification multiplied by 100.

Sample	Clone	HPV type	% Homology BLASTn	% Homology PaVE
OCL 3.1	Yes	172	97.02	96.51
OCL 3.2	Yes	172	95.04	96.51
OCL 3.6	Yes	172	96.00	96.70
OCL 3.8	Yes	172	92.07	89.20
OCL 3.10	Yes	172	97.02	96.51
OCL 14	No	23	80.00	49.00
OCL 14.1	Yes	8	95.04	95.63
OCL 14.2	Yes	8	95.04	93.74
OCL 14.3	Yes	76	91.08	86.67
OCL 14.4	Yes	76	89.24	88.51
OCL 14.6	Yes	76	87.40	86.70
OCL 40	No	172	98.01	98.00
OCL 42	No	172	98.01	95.60
OCL 52	No	90	98.01	99.60
OCL 63.2	Yes	105	96.03	96.42
OCL 63.3	Yes	105	98.01	97.91
OCL 63.4	Yes	105	96.03	94.01
OCL 63.5	Yes	105	96.03	96.42
OCL 63.6	Yes	105	96.03	96.43
OCL 69.1	Yes	152	84.63	85.37
OCL 69.2	Yes	152	89.18	89.18
OCL 69.3	Yes	152	83.72	83.27
OCL 69.4	Yes	152	89.18	89.55
OCL 69.5	Yes	152	88.27	89.19
OCL 93.1	Yes	62	95.04	94.94
OCL 93.2	Yes	62	96.03	96.43
OCL 93.3	Yes	122	99.00	99.30
OCL 93.4	Yes	122	94.20	94.20
OCL 93.5	Yes	122	96.03	96.72
OCL 95	No	147	76.80	64.55
OCL 96	No	11	98.01	99.60

Table 7. Medical history for patients where potential novel HPVs were isolated.

Sample	Cancer location	Gender	Age	Smoking history	Alcohol history
OCL 14	Larynx	Male	60	Heavy	Rarely
OCL 69	Base of tongue	Male	70	Light	Light
OCL 95	Lateral tongue	Male	63	Heavy	Heavy

Table 8. Primers and probe sequences for novel HPVs.

NV 14.4 forward primer	GACACTGTGGATCAAACA
NV 14.4 reverse primer	GCCTTTCCTTTTCAGGATTA
NV 14.4 probe (6FAM, BHQ-1)	AGACATTCATATCTACCAAGGCGAA
NV 69.1 forward primer	GGAGTTTACTAGTTCTATTGG
NV 69.1 reverse primer	ACTGCAAATGTTTATTATTGG
NV 69.1 probe (6FAM, BHQ-1)	ATCATCAGCAACGCAGGCAG
NV 95 forward primer	CAGACAGTGAGCGGTTAGTGTG
NV 95 reverse primer	CAATGTTCACCGATAGCAGGTT
NV 95 probe (6FAM, BHQ-1)	ATTTGAACCTGTTTAGGATCCATGCTA

NV - novel virus

Table 9. Frequency of HPV infection in archived oral cavity and oropharyngeal tissue biopsies.

HPV type	Oral cavity		Oropharynx	
	Cancer- (n=56)	Cancer+ (n=65)	Cancer- (n=50)	Cancer+ (n=50)
HPV16	9, 16%	6, 9%	1, 2%	39, 79%
HPV18	-	-	-	-
NV14.4	-	1, 1.5%	-	-
NV69.1	-	-	-	-
NV95	-	-	-	-

Table 10. Frequency of HPV infection in oral rinse samples.

HPV type	No cancer (n=110)	OPC (n=76)	OCC (n=16)	Other HNC (n=8)
HPV16	-	19, 25%	2, 12.5%	2, 25%
HPV18	1, 0.9%	-	-	-
NV14.4	-	1, 1%	1, 6%	-
NV69.1	-	10, 13%	2, 12.5%	1, 12.5%
NV95	-	1, 1%	1, 6%	-

OPC – oropharyngeal cancer; OCC – oral cavity cancer, HNC – head and neck cancer

Figure 1. Details for process of 7-pooled samples during NGS OCL represents patient sample numbers.

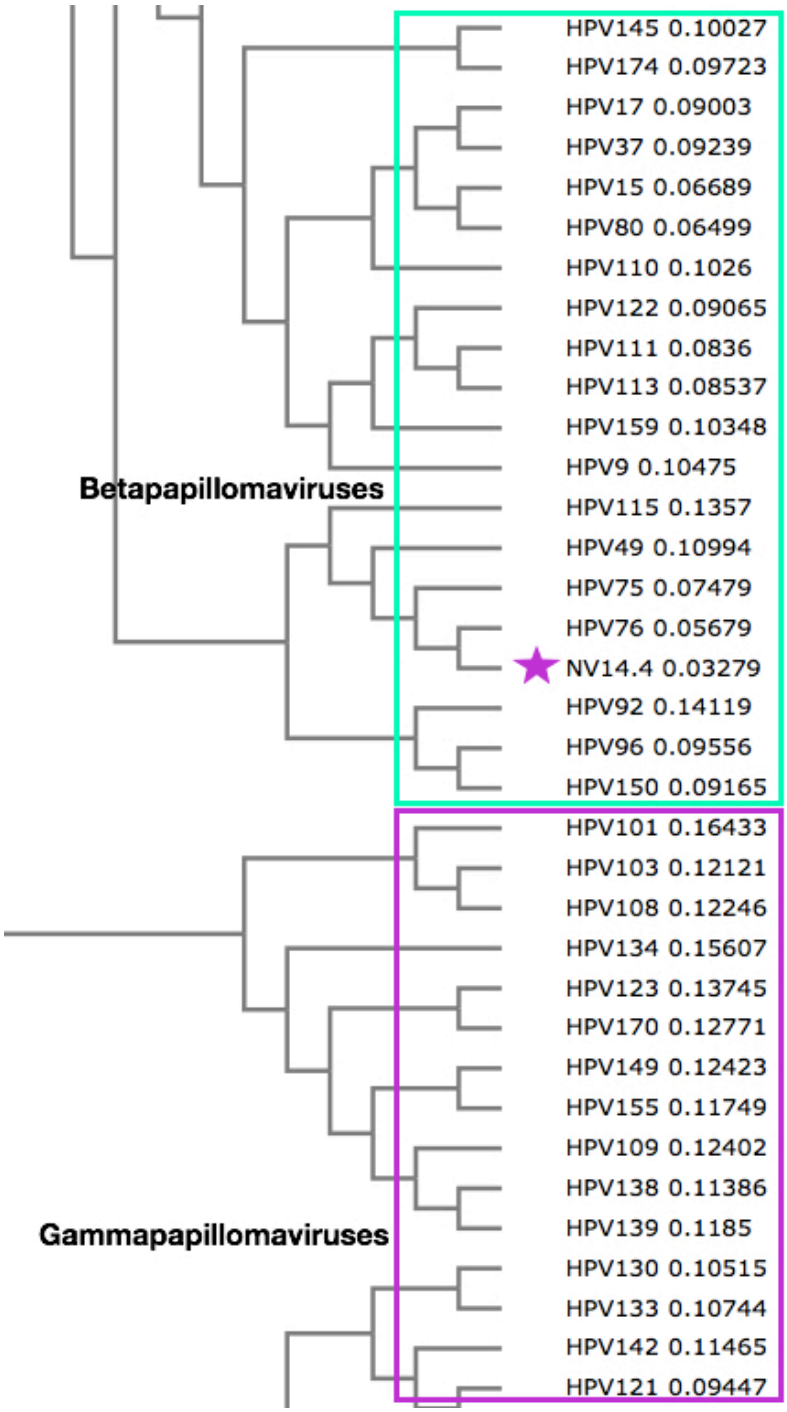

Figure 2. NV14.4 (purple star) is related to HPV76, which belongs to the betapapillomavirus genus. The % homology between the L1 region of NV14.4 and HPV76 is 89%.

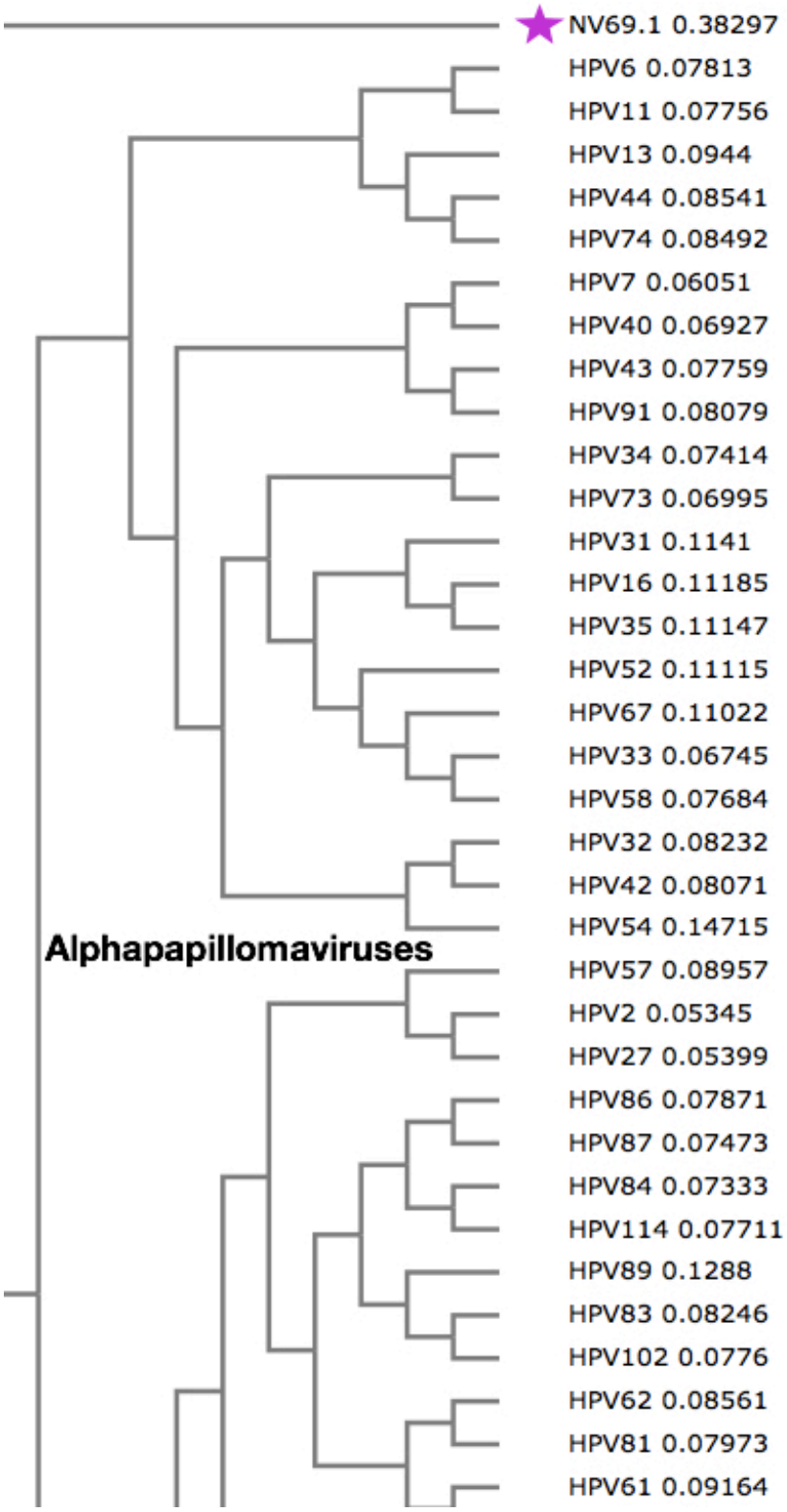

Figure 3. NV69.1 (purple star) is distantly related to the alphapapillomavirus genus.

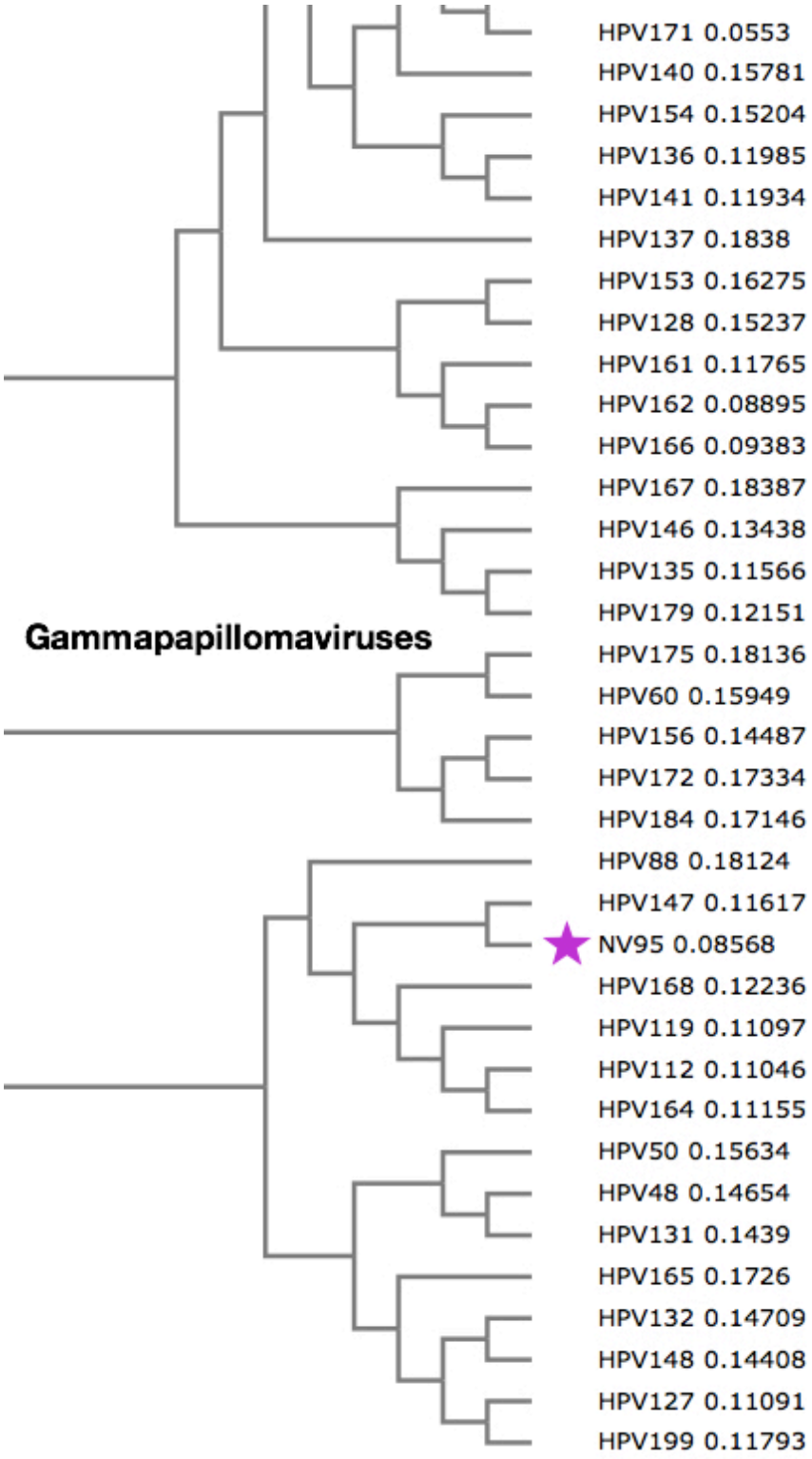

Figure 4. NV95 (purple star) is related to HPV147, which is part of the gammapapillomavirus genus. The % homology between NV95 and HPV147 ranged between 65-77%.

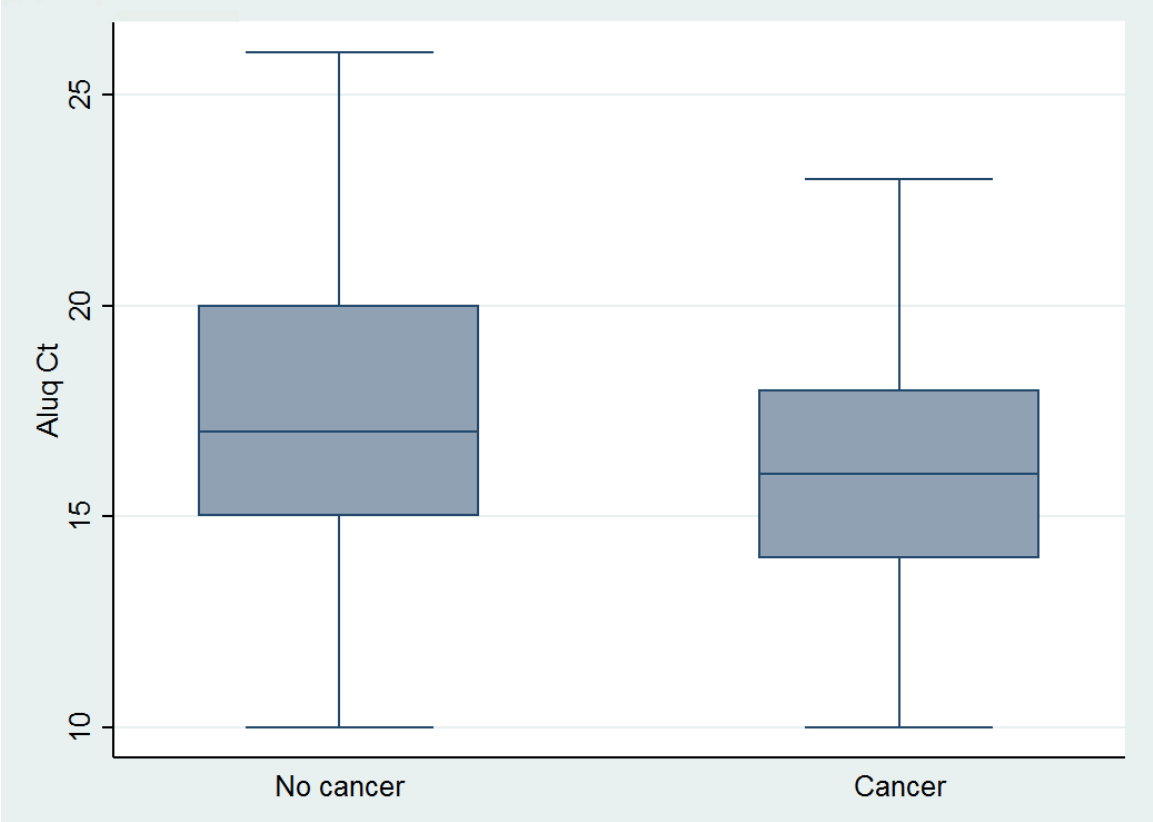

Figure 5. RT-PCR results for Alu sequence demonstrating human genomic DNA content. Ct values of less than 30 indicates detectable DNA at sufficient levels.

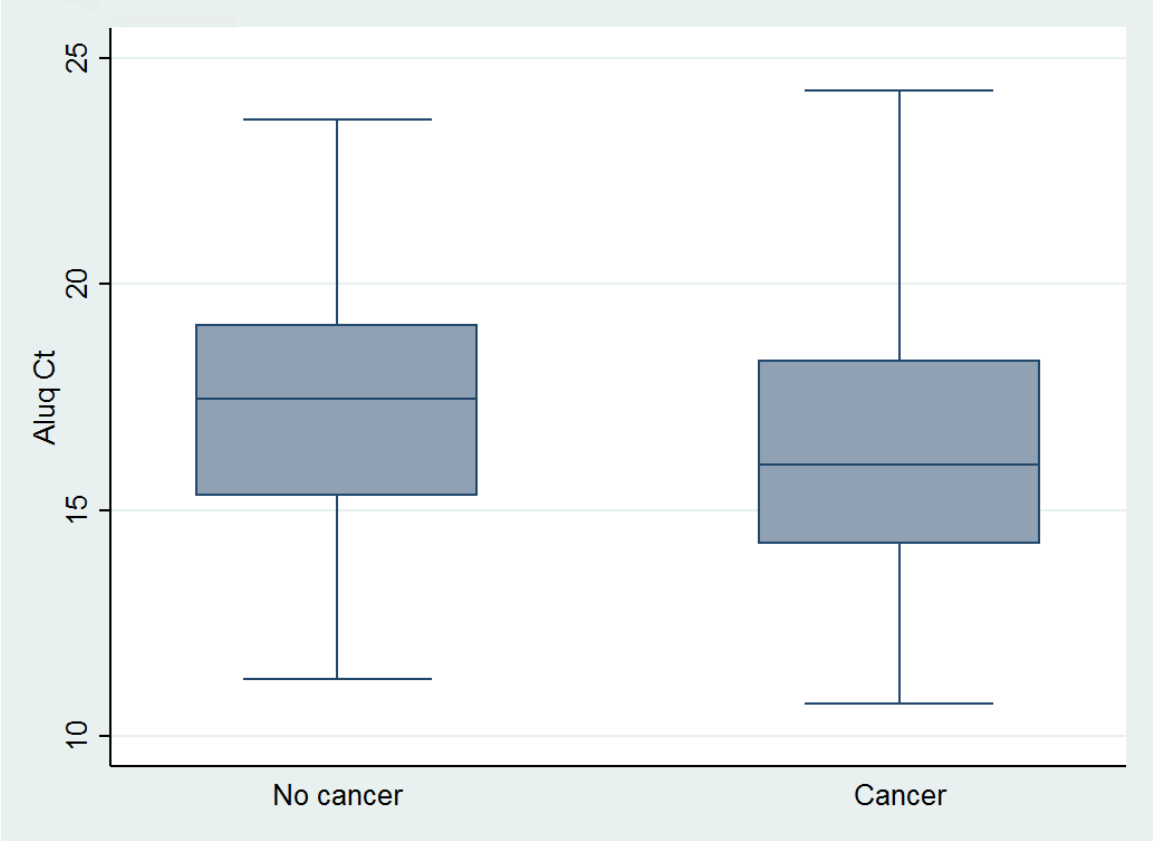

Figure 6. RT-PCR results for Alu sequence demonstrating human genomic DNA content. Ct values of less than 30 indicates detectable DNA at sufficient levels.