# High-risk human papillomavirus in oral cavity squamous cell carcinoma

HPV and oral cavity tumors

Vinayak Palve[1], Jamir Bagwan[1], Neeraja M Krishnan[1], Manisha Pareek[1], Udita Chandola[1], Amritha Suresh[2], Gangotri Siddappa[2], Bonney L James[2], Vikram Kekatpure[3], Moni Abraham Kuriakose[2,3] and Binay Panda[1,4*]

[1]Ganit Labs, Bio-IT Centre, Institute of Bioinformatics and Applied Biotechnology, Bangalore, India

[2]Mazumdar Shaw Centre for Translational Cancer Research, Bangalore, India

[3]Mazumdar Shaw Medical Centre, Bangalore, India

[4]Strand Life Sciences, Bangalore, India

*corresponding author: binay@ganitlabs.in

## Conflict of interests

None of the authors declare any conflicts of interest.

**Abstract**

Human papillomavirus (HPV) is a known risk factor for certain head and neck cancers. Tumors of head and neck region are heterogeneous in nature with different incidences, mortalities and prognosis for different subsites. Unlike oropharynx, where data favors inclusion of HPV status in disease management, role of HPV in oral cavity squamous cell carcinoma (OSCC) is not well understood. The prevalence of HPV in OSCC, although considered lower than oropharynx, vary greatly based on the choice of HPV assay and patient geography. Additionally, data on HPV +ve OSCC is scarce and there is less agreement on HPV being a good prognostic factor in OSCC. Here with 153 OSCC patients, using multiple analytes and assays, we show that a high prevalence (33-58%) of HPV16/18 DNA did not correlate with an equally high prevalence of transcriptionally active viral genomes (E6/E7 RNA prevalence 15%) in tumors. Only 6% of the tumors showed the presence of both HPV DNA and HPV16/18 E6/E7 RNA and none with both p16 and HPV RNA. Most tumors with relatively high-copy HPV DNA and/or HPV E6/E7 mRMA, but not with HPV DNA alone (irrespective of copy number), were wild type for *TP53* and *CASP8* genes. In our study, p16 protein, HPV DNA and E6/E7 RNA, either alone or in combinations, did not correlate with patient survival. Using genome-wide methylation data, 9 HPV-associated genes stratified the HPV +ve from the HPV –ve tumor groups with high confidence ($p<0.008$) when relatively high-copy number of HPV DNA and/or HPV E6/E7 RNA were considered to define HPV positivity and not HPV DNA alone irrespective of their copy number ($p<0.2$). Taken together, we conclude that tests measuring HPV DNA alone without viral load and/or viral RNA may not be a true measures of HPV infection in oral cavity tumors and therefore are not informative.

**Keywords:** HPV, OSCC, HPV DNA, HPV RNA, copy number, survival, somatic mutations, methylation.

## Introduction

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer worldwide with an incidence of 550,000 cases annually [1]. Oral cavity squamous cell carcinoma (OSCC) constitute a majority of HNSCC including tumors of oral/anterior tongue and buccal mucosa [2]. The major known risk factors of OSCC are tobacco products, alcohol, and infection with human papillomavirus (HPV) [3]. Unlike oropharyngeal tumors, where the HPV incidence is reported to be very high, (upto 90%) [4,5], the incidence of HPV in OSCC varies widely (from none to 74%) depending on the detection methodology used and the geography of the patient cohort [6]. Additionally, unlike oropharngeal tumors [7-11], the role of HPV in disease prognosis and response to therapy in patients with OSCC is less equivocal. Among the high risk HPV types, infection with HPV16 or HPV18 has been epidemiologically linked with head and neck malignancies [12]. Presennce of HPV DNA is routinely used as a measure of HPV infection in tumors. However, there is a considerable variation in the sensitivity of DNA-based assays, which leads to differences in reporting HPV DNA prevalence among different cohorts. On the otherhand, the presence of HPV E6/E7 RNA and/or their associated protein products remain as gold standard tests defining HPV positivity.

Large-scale sequencing studies have demonstrated that the majority of HPV-negative tumors harbor mutations in *TP53* and *CASP8* and a large proportion of HPV-positive tumors in *PIK3CA* [13-15]. Additionally, past studies have identified specific mutations in potential drug targets like *FGFR2/3* and lack of *EGFR* abberations in HPV-positive patients [15] and the role of *CASP8* in HPV-negative cell line and patients [13,16]. Despite the wealth of information, which HPV tests are accurate and whether HPV plays an important factor in oral cavity tumor stratification and treatment remain to be answered.

In the present study, we addressed the following questions on HPV in a cohrt of 153 patients with oral cavity tumors. Does sensitivity matter in the detection of HPV DNA in tumors? Does

presence of tumor p16 and HPV DNA correlate with that of HPV E6/E7 RNA? Is high-copy HPV DNA a true refelction of HPV infection in oral cavity tumors? Are p16 protein, HPV DNA, HPV E6/E7 RNA, individually or together, linked with patient survival? And finally, do somatic mutations and gene methylation at 5-Cytosine residues distinguish the HPV +ve from the HPV –ve tumors?

## Materials and Methods

*Patient cohort*

Informed consent was obtained voluntarily from each patient enrolled in the study. Ethical approval (NHH/MEC-CL/2014/197; NHH/MEC-CL-2014/224) was obtained from the Institutional Ethics Committees of the Mazumdar Shaw Medical Centre. Patients diagnosed and treated at the cancer clinic of the Mazumdar Shaw Medical Centre for oral cavity tumors were subjected to a screening procedure before being enrolled in the study. Only those tumors where the histological sections confirmed the presence of squamous cell carcinoma and with at least 70% tumor cells in the section were used in the current study. One hundred and fifty three patients underwent staging according to AJCC criteria, and curative intent treatment as per NCCN guideline involving surgery with or without post-operative adjuvant radiation or chemo-radiation at the Mazumdar Shaw Medical Centre were accrued for the study (Supplementary Table 1). Post-treatment surveillance was carried out by clinical and radiographic examinations as per the NCCN guidelines. All the tissues were frozen immediately in liquid nitrogen and stored at -80$^0$C until further use.

*Cell cultures*

Cell lines (UM-SCC47, from Dr. Thomas Carey, University of Michigan, USA [17]; Hep2, National Centre for Cell Science, Pune, India; UPCI:SCC29B and UPCI:SCC040, from Dr. Susanne Gollin, University of Pittsburg, USA [18]) were maintained in Dulbecco's Modified Eagles' Media (DMEM) supplemented with 10% FBS, 1X MEM non-essential amino acids

97 solution and 1X penicillin/streptomycin mixture (Gibco) and grown in incubators at $37^0$C with 5%
98 $CO_2$.
99
100 *Immunohistochemistry (IHC)*
101 For p16 immunohistochemistry (IHC) staining was carried out using primary antibody from
102 Biogenex, USA (cat: AM540-5M) and using the poly-HRP detection system (Cat: QD400-60KE,
103 BioGenex, Fremont, CA, USA) following manufacturers' instructions. Sections of cervical cancer
104 were used as the positive control.
105
106 *DNA extraction*
107 The genomic DNA from tumor tissues and cell lines were extracted using DNeasy Blood
108 and Tissue extraction kit using the spin column method with an intermediate RNase digestion
109 following manufacturer's instructions (Qiagen, USA). The DNA was quantified using Qubit 2.0
110 fluorometer (Invitrogen, USA) and stored at -$20^0$C until further use. Total 300ng of genomic DNA
111 (unless specified otherwise) from tumors was used for the detection of HPV using PCR, qPCR and
112 digital droplet PCR (ddPCR) methods.
113
114 *Detection of HPV DNA with PCR*
115 PCR primers were designed to amplify either consensus or type-specific regions of the virus
116 (Figure 1A). Five sets of primers, including 4 HPV consensus (GP5-6; CP-I/II; MY09/11 and
117 PGMY09/11) and 1 type-specific (HPV16-L1) primers as described in the literature [19], in the
118 PCR assays. Additionally, we desinged two sets of new type-specific primers (HPV16-E6 and
119 HPV18-L1) to detect HPV DNA. The primer sequences and the amplification conditions used are
120 provided in Supplementary Table 2.
121
122 *Detection of HPV DNA with qPCR*

123 For qPCR, E6 and E7 regions from HPV16 and HPV18 respectively were cloned in pUC19
124 vector. The sequences for the recombinant plasmids were confirmed by Sanger sequencing. The
125 qPCR was carried out using KAPA Probe Fast qPCR master mix Universal (2X) (cat: KK4701,
126 KAPA Biosystems, USA). The primers and probes were designed (Supplementary Table 2) within
127 the cloned regions of the HPV16 and HPV18 plasmids and standard quantitative PCR was
128 performed. The standard curves were generated using serial dilutions (from 10 to 100,000 copies) of
129 HPV16 and HPV18 cloned pUC19 plasmids. Genomic DNA from postive and negative control cell
130 lines wers used as positive and negative controls for HPV18 and HPV16 respectively
131 (Supplementary Figure 1). All amplification reactions were carried out in triplicates, using nuclease
132 free water (cat: AM9932, Ambion, USA) as negative control. The analysis for each sample was
133 done by using the absolute quantification using standard curve generated with serial dilutions of
134 cloned plasmids. Tumors and cell lines samples were counted as positive for those having *Ct* values
135 3 times away from the standard deviation from the negative controls.
136
137 *Detection of HPV DNA using droplet digital PCR (ddPCR)*
138 Digital PCR was performed using the QX100 droplet digital PCR system (Bio-Rad, USA)
139 using primers and probes as provided in Supplementary Table 2. The reaction mix consisted of
140 10 µl of 2× ddPCR Supermix without dUTP (cat: 1863024; Bio-Rad, USA), 450 nM of both
141 forward and reverse primers of HPV16 and 250 nM probe and 300ng of genomic DNA in a final
142 volume of 20 µl. The entire 20 µl reaction was loaded into a droplet cartridge, followed by 70 µl of
143 droplet generation oil and droplets were generated using droplet generator. The droplets were
144 transferred into a 96-well PCR plate (PCR96-FLT-C, Axygen, USA) and sealed using an
145 aluminium foil. Further, the PCR amplification was carried out on in a thermal cycler (C1000, Bio-
146 Rad, USA). After amplification, the plate was transferred to a Bio-Rad droplet reader. The droplets
147 per well were measured and normalized in every reaction using the Quantasoft V.6 software (Bio-
148 Rad, USA). All the samples were processed in triplicates. Tumors and cell lines samples were

149 counted as positive for those having droplet numbers 3 times away from the standard deviation
150 from the negative controls.
151
152 *HPV DNA copy number*
153 We deduced the HPV absolute copy number from the qPCR standard curves using cloned
154 HPV16/18 (Figure 2A,D). We considered a tumor or cell line to have relatively high-copy of HPV
155 DNA when the copy number for HPV16 and HPV18 DNA were more than $3.3 \times 10^2$ and $3.3 \times 10^3$
156 per microgram of tumor DNA respectively. In order to minimize the effect of tumor cellularity,
157 ploidy and heterogeneity, we expressed the HPV copy number as copies per microgram of tumor
158 DNA used in the reaction.
159
160 *Detection of HPV16/18 E6/ E7 RNA*
161 The total RNA was extracted from 25mg of tumor tissues using the RNeasy mini kit from
162 Qiagen (cat: 74104) following manufacturer's instructions. The residual DNA was digested on
163 column during extraction using RNase free DNase set from Qiagen (cat: 79254). Total 500ng of
164 RNA was subjected to cDNA synthesis using Takara's Prime Script first strand cDNA synthesis kit
165 (cat: 6110A). The qPCR was carried out using KAPA SYBR Fast qPCR master mix Universal (cat:
166 KK4601, KAPA Biosystems, USA). The GAPDH was amplified and used as internal control. The
167 primer sequences and the amplification conditions are provided in Supplementary Table 2. All
168 amplification reactions were carried out in triplicates, using nuclease free water (cat: AM9932,
169 Ambion, USA) as no template control and cDNA from UPCI:SCC29B and UPCI:SCC40 cell lines
170 as negative controls. Complementary DNA from cell lines UM-SCC47 and Hep2 were used as
171 positive controls for E6 and E7 amplifications respectively. Data was normalized using relative
172 quantification method ($2^{\wedge -\Delta\Delta Ct}$) [20]. Positive tumors were counted as ones that have *Ct* values 3
173 times from the standard deviation of the negative control.
174
175 *Integration status of HPV genome*

176 We performed HPV16E2 PCR [21,22] to determine whether HPV has integrated into tumor
177 genomes or remained as episomal DNA in tumors those were positive for HPV DNA and RNA.
178 Genomic DNA from the early passage of W12 cell line (gift from Prof. Sharmila Sengupta,
179 NIBMG, Kalyani, India) and p3796 pGEX-16E2 (wt) (Addgene plasmid # 24123) were used as
180 positive controls for E2 PCR. In addition, genomic DNA from UPCI;SCC29B and PCI:SCC040
181 cell lines were used as negative controls. The primer sequences and the amplification conditions are
182 provided in Supplementary Table 2. The PCR amplicons were resolved on 2% agarose gel and
183 visualized with a Gel Doc system (G-box, Syngene, UK). Tumors with no amplification were
184 considered to have HPV integrated into host genomes.
185
186 *Mutation analysis by Sanger sequencing*
187 The mutation data on tumors for *TP53*, *CASP8* and *RASA1* were retrieved from the
188 previously published data [16].
189
190 *Statistical Analysis*
191 The Chi square test was used to analyse significance of different clinical parameters of
192 patients. The relationship between tumor HPV status and survival in patients was examined by
193 Kaplan-Meier analysis and *Cox* proportional hazards regression with overall survival (OS) and
194 disease-free survival (DFS) with logrank test for determining significance ($p$ values <0.05).
195
196 *Whole genome methylation and statistical data analyses*

197 Production of whole genome methylation data using Illumina Infinium Methylation450
198 BeadChip, chip scanning and the data pre-processing are described previously [23]. Three tumors in
199 HPV +ve group (relatively with high-copy HPV DNA and/or E6/E7 RNA) and 18 tumors in HPV -
200 ve group (HPV DNA –ve and HPV RNA –ve) were considered for further analysis. Differential
201 methylation for probes was estimated as $\Delta\beta$ values ($Tumor_{\beta}$ - $Normal_{\beta}$) resulting in values ranging
202 from -1 (hypo-methylation) to 1 (hyper-methylation). Using a supervised clustering approach,

203 differentially methylated probes that discriminated HPV +ve and HPV -ve patients samples were
204 identified. Hyper-methylated probes in the HPV +ve group that were hypo-methylated in the HPV -
205 ve group, and *vice versa*, along with other probes with Δβ values that differed by at least 0.5
206 between the two groups, with a sample frequency of at least 60% were considered further. Finally,
207 the list of genes consisting of average Δβ values across tumors for multiple probes (where present)
208 was drawn. Neighborhood interaction network for these genes was inferred using PCViz from
209 PathwayCommons (http://www.pathwaycommons.org/pcviz/#neighborhood/). The genes and their
210 interacting partners were mapped to the Viral Carcinogenesis pathway from KEGG (hsa05203), and
211 mapped genes were linked back to respective genes found using supervised clustering. For the
212 linked genes, an unpaired *t*-test was performed to determine the significance of difference between
213 the HPV +ve and HPV -ve tumor groups. The same was performed for the HPV-associated genes
214 between the DNA-based (irrespective of the DNA copy number) HPV +ve and HPV -ve tumor
215 groups.

216

217 **Results**

218 *p16 expression*

219 The tumors (n = 56) were first confirmed by H and E staining of the paraffin tissue sections
220 followed by the immunohistochemical staining with the surrogate marker p16. Supplementary
221 Figure 2 shows the representative p16 +ve and –ve staining in oral cavity tumor sections (OSCC)
222 along with the cervical tissue section (+ve control). Fifty one percent (28/55) of the oral cavity
223 tumors studied showed positive p16 staining (Table 1).

224

225 *Incidence of HPV DNA*

226 We used three different molecular techniques (PCR, quantitative PCR and digital PCR) to
227 detect HPV DNA in oral cavity tumors. We tested 5 sets of primers published in the literature and 2
228 newly designed ones in the amplification reactions (Figure 1A). The primers were either consensus
229 ($GP5^{+}6^{+}$, MY09/11, CPI-II and PGMY09/11) or type-specific (HPV16L1, HPV16E6 and

230 HPV18L1). We used genomic DNA from UMSCC-47 and Hep2 cell lines to detect HPV16 and
231 HPV18 respectively at various dilutions to determine relative efficiency, sensitivity and accuracy of
232 amplification (Figure 1B). Primers described in the literature ($GP5^{+}6^{+}$, MY09/11, CPI-II or
233 PGMY09/11, HPV16L1) could only detect HPV DNA when 0.03ng or higher amount of genomic
234 DNA from the cell line UMSCC-47/Hep2 was used (Figure 1B). However, the newly designed type
235 specific primers (HPV16E6 and HPV18L1) were able to detect HPV DNA with as low as 0.0003ng
236 and 0.003ng genomic DNA for HPV16 and HPV18 respectively (Figure 1B). In order to check the
237 efficiency of the newly designed primers in relation to absolute copy numbers of HPV genome, we
238 used serial diluted HPV cloned plasmids DNA. As shown in Supplementary Figure 3, efficient
239 amplification with 100 copies of HPV genome or higher could be detected. Once we optimized the
240 conditions and primers for all the amplification reactions, we performed PCR in oral cavity tumors
241 ($n$ = 70). In our cohort, 58% (41/70) of the tumors were HPV DNA +ve for any of the PCR (Table
242 1, Supplementary Figure 4). Figure 1C shows the efficiency of the consensus and type specific
243 primers in a set of representative oral cavity tumors. Widely used primers from the literature
244 (MY09/11, PGMY09/11, $GP5^{+}6^{+}$ and HPV16L1) yielded either least or moderate (CPI-II)
245 sensitivity of detection while the newly designed HPV16E6 and HPV18L1 primers showed
246 optimum sensitivity of detecton (Figure 1C). We observed inhibition of amplification reactions at
247 high concentration of tumor genomic DNA with positive cell line spike-in experiment
248 (Supplementary Figure 5) and therefore avoided in the reactions. Additionally, an increase in
249 amplification cycles also did not aid in the detection of HPV DNA in PCR as shown in
250 Supplementary Figure 6.
251
252 Next, we used quantitative PCR using Taqman chemistry to detect HPV DNA. We
253 performed serial dilution using HPV16 E6 and HPV18 L1 cloned plasmids to obtain standard
254 curves (Figure 2 A,D). We used increasing amount of genomic DNA from negative control cell
255 lines to demonstrate specific amplification of HPV DNA (Supplementary Figure 1). Genomic DNA
256 from HPV +ve and –ve cell lines were used to plot the baseline. Tumors were counted as HPV

257 DNA positive for those having *Ct* values 3 times away from the standard deviation from the
258 negative controls (Figure 2 B,E). Results from qPCR indicated that 33% (35/106) of tumors were
259 positive for HPV DNA (Table 1). Although, we found high incidence for HPV16 (30%; 32/106)
260 than HPV18 (18%; 19/106) type, the HPV18 positive tumors had very high copy numbers of viral
261 DNA as refelcted in their *Ct* values (Figure 2 C,F). The prevalence of HPV DNA in oral tongue
262 tumors was higher (44%) compared to that of buccal mucosa (4%).
263
264 Finally, we used one of the most sensitive methods, digital PCR, to detect HPV DNA.
265 Digital PCR is recently shown succefully to detect HPV DNA in oropharyngeal tumors in highly
266 specific manner [24]. We performed droplet digital PCR (ddPCR) with serially diluted plasmid
267 HPV16E6 clones to generate the standard curve (Figure 2G). The genomic DNA from positive-
268 (UM-SCC47) and negative control cell lines (Hep2) for HPV16 along with a no template control
269 (NTC) were used to plot the baseline (Figure 2H). Digital PCR results indicated that 43% (59/136)
270 of oral cavity tumors were positive for HPV16 DNA (Table 1). Moreover, HPV16 infection was
271 more prevalent in oral tongue tumors (55%) compared to buccal mucosal ones (17%) (Table 1).
272
273 *HPV E6 and E7 RNA expression*
274 The viral mRNA, E6 and E7 (for both HPV16 and HPV18), were measured by qPCR in oral
275 cavity tumors and from cell lines (UMSCC-47 and Hep2). Compared to the cell lines, tumors
276 showed very low levels of expression of E6 and E7 mRNA (Figure 3). Unlike HPV DNA, only
277 15% of the tumors showed expression of E6 and/or E7 RNA and 6% of the tumors showed the
278 presence of both HPV DNA (with all the 3 assays) and transcriptionally active HPV genomes
279 (Table 1). Similar to DNA, oral tongue tumors had higher number (17%) of HPV E6/E7 RNA
280 compared to the buccal mucosa (9%, Table 1). In our cohort, younger patients (≤40yrs age) were
281 significantly more HPV RNA +ve than the older patients when analyzed by *Chi*-square analysis ($p$
282 $= 0.029$).

When combining the results from all the assays (p16 IHC, HPV DNA and HPV RNA), we found 0-48% of the tumors were positive in various assay combinations with PCR + ddPCR yielding the highest (48%) and p16 IHC + RNA the lowest (0%) numbers (Table 1). We found that 22% (23/106) of the tumors had relatively high-copy of HPV DNA and/or HPV E6/E7 mRNA.

***Integration of HPV DNA in the host tumor genomes***

We performed HPV16 E2 PCR to understand what percent of HPV DNA is integrated into the host tumor genomes and found that HPV is episomal in 18 % (n = 40) of the oral cavity tumors and in the majority of cases, HPV DNA was integrated into the host genome (Table 1). We used a positive control (genomic DNA from W12 cell line) to demonstrate the accuracy of the E2 PCR (Supplementary Figure 7).

*Linking tumor attributes, somatic mutations and HPV with survival*

We performed Kaplan Meir survival analysis with various tumor attributes that revealed significant association between tumor differentiation (*p=0.03*) and clinical stage (*p=0.0003*) with survival (Supplementary Figure 8 A-D). None of the other tumor attributes showed significant association with survival (Supplementary Figure 8 E-J). Expression of p16 in oral cavity tumors did not stratify patients based on both overall and disease-free survival (Supplementary Figure 9 A,B). HPV DNA status alone in any of the DNA-based assays or in combination did not correlate with survival (Supplementary Figure 9 C-L), except when measured with ddPCR for overall survival (Supplementary Figure 9 I, *p*=0.03). We found no correlation between HPV E6/E7 mRNA and with both overall (*p*=0.7) and disease-free (*p=0*.3) survival (Supplementary Figure 9 M,N).

Furhter, we investigated whether somatic mutations in significantly mutated genes in OSCC have a role to play in HPV DNA +ve tumors and patient survival. We looked at 3 genes (*TP53*, *CASP8*, *RASA1*) shown to be significantly mutated in oral cavity tumors [13,16,25] and tested

310 whether the mutation in any of the genes, alone or in combination, with relatively high HPV DNA
311 copy and/or HPV E6/E7 mRNA is linked with survival. As shown in Figure 4A, we did not find any
312 significant association with this group of tumors with both overall- ($p = 0.45$) and disease-free ($p =$
313 $0.68$) survival. Ninty five percent of the HPV +ve tumors in the group were wild-type for *TP53*, and
314 *CASP1* gene and 85% of the HPV +ve tumors for *RASA1*. (Figure 4B). Somatic mutations in a
315 group with relatively high-copy of HPV DNA, either alone or in combinations, were not associated
316 with both overall and disease-free survival (Supplementary Figure 10 A-F).

317 *Linking methylation to HPV*

318 Supervised clustering of the first group of patients (a group defined as one with high-copy
319 number HPV DNA and/or E6/E7 RNA) resulted in a list of 60 genes, out of which 9 (*FERMT3*,
320 *GIT2*, *HK3*, *PRKCZ*, *ZCCHC8*, *IRF5*, *IFFO1*, *ARID3A*, *HOXA2*) were mapped to the HPV pathway
321 (Figure 4C). Methylation of those genes are involved in downstream control of the expression of
322 different target genes. For example, *ZCCHC8* methylation is linked with the expression of *RB1*,
323 *PRKCZ* methylation controls-state-change-of *DLG1*, methylation in *ARID3A*, *IRF5*, *IFFO1* and
324 *HOXA2* are linked with the expression of *TP53*, and *FERMT3*, *HK3* and *GIT2* genes control the
325 expression of AP1 *(JUN)* (Figure 4C). All the genes, except for *HOXA2*, were significantly
326 hypermethylated in the HPV +ve group of tumors compared to the HPV –ve group (Figure 4C). The
327 linked 4 genes, obtained from the 9 significantly methylated genes, were mapped to the pathways
328 involving HPV E6 and E7 proteins (Figure 4D). To test significance, we performed unpaired *t*-tests
329 between the two groups, group 1 that has relatively high-copy HPV DNA and/or HPV RNA and
330 group 2 that were negative for both HPV DNA and HPV RNA. All of the 8 hypermethylated genes
331 showed very significance ($p < 0.00001$) and for *HOXA2* that was hypomethylated the *p*-value was
332 0.007. However, when the patients were grouped based on HPV DNA positivity alone (irrespective
333 of their copy number), most of these 9 HPV-linked genes did not show significant association
334 (Supplementary Table 3).

335

## Discussion

HPV plays an important role in the prognosis of patients with oropharyngeal tumors [26,27]. Unlike oropharynx, HPV incidence and its role in disease prognosis in oral cavity tumors is not well established. Past results on HPV DNA incidence in oral cavity tumors were widely variable (from very low to very high) depending on the assay sensitivity, analyte and patient cohort used [28]. The accuracy of the HPV tests employed is important and HPV positivity need to be answered in order to make treatment decisions in patients with head and neck tumors confidently [29]. There are fewer studies that used multiple analytes (protein, DNA and RNA) and multiple molecular tests (IHC, PCR, qPCR and digital PCR) to establish HPV positivity in oral cavity tumors and that correlated HPV with tumor attributes (including somatic mutations and methylation) and survival. In the present study, we attempted to do this in 153 patients with oral cavity tumors.

While p16 expression, as measured by immunohistochemistry (IHC), is a commonly used proxy for HPV in HNSCC, its expression is not specific in HPV associated tumors [27]. Although, numerous studies correlated p16 expression with HPV [30-32], p16 IHC has shortcomings, especially when correlating its expression with patient survival. Limitations, such as, variation in staining intensities [33,34], nonspecific binding of antibodies and the lack of scoring and interpretive criteria used for p16 staining [35] make the method less reliable. Unlike some of the earlier studies [36,37], we could not find any correlation (either positive or negative) between p16 expression with both HPV DNA and survival (Supplementary Figure 9 A,B). Like previous reports [30,38], we found p16 expression not to be an effective surrogate marker for HPV detection in oral cavity tumors (there was no tumor that expressed both p16 and had detectable HPV RNA) (Table 1). A poor correlation between p16 IHC and HPV *in situ* hybridization (ISH) is reported in the past [39]. In our study, we did not perform HPV ISH, which could have provided additional information on p16 positivity and HPV prevalence.

Unlike antibody-based methods, nucleic acids-based methods detect HPV with high sensitivity and therefore used widely [40]. Meta-analysis of 5478 oral cavity tumors suggested the

363 overall HPV DNA prevalence to be 24.2% with 11% tumors being positive for both HPV DNA and
364 E6/E7 RNA [41]. India has one of the highest incidence rates of oral cavity cancers with a
365 significant difference in the trend of incidence between oropharyngeal and oral cavity cancer [42].
366 Previously, PCR coupled with MassArray is shown to provide high sensitity of detection with low
367 amount of input genomic DNA [5]. From our results, we found 38% and 13% of the tumors to be
368 positive and negative in all 3 DNA-based assays (PCR, qPCR and ddPCR) respectively. Overall,
369 the prevalence (33-58%) of HPV DNA was dependent on the type of assay used with PCR yielding
370 the highest prevalence over the more sensitive methods like qPCR and ddPCR assays (Table 2A).
371 This was possibily due to the use of consensus primers in PCR but not in qPCR and ddPCR, in
372 addition to the type-specific ones, resulting in the detection of non-HPV16/18 subtyps. As expected,
373 digital PCR, being the most-sensitive method out of the three DNA-based assays, resulted in more
374 number of tumors being HPV16-DNA positive resulting in the detection of very low copy viral
375 genomes in tumor samples. We wanted to test whether the presence of HPV DNA alone reflects a
376 biological role of HPV in tumors. Based on several levels of evidence as noted below, we conclude
377 that the presence of low copy HPV DNA alone may not be a reflection of functionally active HPV.
378 First, in the tumors positive in HPV DNA, we found only a fraction (15%) with HPV E6/E7 RNA.
379 Second, only 6% of tumors were positive for both the presence of HPV genome and E6/E7 RNA.
380 Third, almost all the tumors with relatively high-copy HPV genome and/or HPV RNA had wild
381 type *TP53* and *CASP8* genes, which was not the case with tumors with the low-copy HPV DNA.
382 Both *TP53* and *CASP8* are known to be wild-type primarily in HPV positive tumors [13] and in our
383 study too we found this corresponds to the tumors with high-copy HPV genome and/or
384 transcriptionally active genome only (Table 2B) . From all of the evidence above, we believe that
385 the presence of HPV DNA in the tumor tissues might have been a result of highly sensitive assays
386 used in our study and might suggest either the presence of passenger HPV genomes coming from
387 adjacent normal cells, as earlier reported [43,44] and/or be a reflection of inactive or passenger
388 virus in oral cavity tumors. Although very few (n=3), we can not explain why some tumors in our
389 study with HPV E6/E7 RNA did not show the presence of HPV DNA. It is possible that the

390 genomic DNA for those tumors were degraded and therefore, could not serve as ideal templates for
391 DNA-based assays. An additional factor could be the presence of inhibitors for DNA-based assays
392 in those tumors. Although HPV RNA is considered to be gold standard analyte to test for HPV in
393 tumors, RNA is a difficult analyte to handle and is more labile than DNA. Additionally, studying
394 RNA in archival samples may pose additional set of challenges.
395
396 We did not find any significant correlation between p16, HPV-DNA and/or HPV RNA and
397 disease outcome (Supplementary Figure 9). Even the tumors with relatively high-copy HPV
398 genomes and/or E6/E7 RNA did not support the role of HPV in patient survival (Figure 4A). There
399 are several interesting aspects to our results that need further studies and analysis. First, like
400 cervical tumors [41] but unlike oropharyngeal tumors [13 14 15 16], we did not find HPV to be a
401 predictor of survival. It is possible that HPV has a different mechanism of action in oral cavity
402 tumors. An indirect support to linking high HPV genome copy number with HPV positivity in our
403 study comes from the fact that we did not find significance in HPV-linked genes stratifying HPV
404 +ve from HPV –ve group when HPV DNA irrespective of their copy number is taken into account
405 in defining the HPV positivity (Figure 4C-D). It is possible that the presence of high-copy HPV
406 genome in those tumors does not correlate with the presence of biologically active virus. Further
407 studies may help answer this question. Till we have a good answer on high-copy number HPV
408 DNA, HPV RNA may be the best analyte to test for HPV positivity in oral cavity tumors.
409
410 **Acknoweledgements**
411 Research presented in this article is funded by Department of Electronics and Information
412 Technology, Government of India (Ref No:18(4)/2010-E-Infra., 31-03-2010) and Department of IT,
413 BT and ST, Government of Karnataka, India (Ref No:3451-00-090-2-22). We thank Mr. Osama
414 Mansour of BioRad, India in helping us with ddPCR.
415
416 **Author contributions**

417 VP and BP designed the study, performed data analysis amd wrote the manuscript. VP carried out
418 all the molecular assays and cell culture experiments; JB, UC and MP assisted in conducting
419 molecular assays; NMK performed methylation data analysis; AS, GS and BLJ conducted p16 IHC
420 studies, GS assisted in cell culture and maintenance of cell lines; AS, VK and MAK collected
421 patient samples, collected follow up information and provided clinical attributes. BP provided the
422 overall study supervision.
423

424 **Abbreviations**

425 HPV, human papillomavirus; HNSCC, head and neck squamous cell carcinoma; OSCC, oral cavity
426 squamous cell carcinoma; IHC, immunohistochemistry; PCR, polymerase chain reaction; qPCR,
427 quantitative polymerase chain reaction; ddPCR, droplet digital polymerase chain reaction
428

429 **References**


430 1. Torre LA, Bray F, Siegel RL, *et al*. Global cancer statistics, 2012. *CA: a cancer journal for*
431 *clinicians* 2015; **65**: 87-108.
432 2. Bhat SP, Bhat VS, Permi H, *et al*. Oral and oropharyngeal malignancy: A clinicopathological study.
433 *Internet Journal of Pathology and Laboratory Medicine* 2015; **1**.
434 3. Mishra A, Meherotra R. Head and neck cancer: global burden and regional trends in India. *Asian*
435 *Pac J Cancer Prev* 2014; **15**: 537-550.
436 4. Nasman A, Attner P, Hammarstedt L, *et al*. Incidence of human papillomavirus (HPV) positive
437 tonsillar carcinoma in Stockholm, Sweden: an epidemic of viral-induced carcinoma? *Int J Cancer*
438 2009; **125**: 362-366.
439 5. Walline HM, Komarck C, McHugh JB, *et al*. High-risk human papillomavirus detection in
440 oropharyngeal, nasopharyngeal, and oral cavity cancers: comparison of multiple methods. *JAMA*
441 *Otolaryngology–Head & Neck Surgery* 2013; **139**: 1320-1327.
442 6. Isayeva T, Li Y, Maswahu D, *et al*. Human papillomavirus in non-oropharyngeal head and neck
443 cancers: a systematic literature review. *Head Neck Pathol* 2012; **6 Suppl 1**: S104-120.
444 7. Ang KK, Harris J, Wheeler R, *et al*. Human papillomavirus and survival of patients with
445 oropharyngeal cancer. *New England Journal of Medicine* 2010; **363**: 24-35.

446 8. Fakhry C, Westra WH, Li S, *et al*. Improved survival of patients with human papillomavirus-positive
447 head and neck squamous cell carcinoma in a prospective clinical trial. *Journal of the National*
448 *Cancer Institute* 2008; **100**: 261-269.
449 9. Fakhry C, Zhang Q, Nguyen-Tan PF, *et al*. Human papillomavirus and overall survival after
450 progression of oropharyngeal squamous cell carcinoma. *Journal of clinical oncology : official*
451 *journal of the American Society of Clinical Oncology* 2014; **32**: 3365-3373.
452 10. Lassen P, Eriksen JG, Hamilton-Dutoit S, *et al*. Effect of HPV-associated p16INK4A expression on
453 response to radiotherapy and survival in squamous cell carcinoma of the head and neck. *Journal of*
454 *clinical oncology : official journal of the American Society of Clinical Oncology* 2009; **27**: 1992-
455 1998.
456 11. Rischin D, Young RJ, Fisher R, *et al*. Prognostic significance of p16INK4A and human
457 papillomavirus in patients with oropharyngeal cancer treated on TROG 02.02 phase III trial. *Journal*
458 *of clinical oncology : official journal of the American Society of Clinical Oncology* 2010; **28**: 4142-
459 4148.
460 12. Huang C-G, Lee L-A, Tsao K-C, *et al*. Human papillomavirus 16/18 E7 viral loads predict distant
461 metastasis in oral cavity squamous cell carcinoma. *Journal of Clinical Virology* 2014; **61**: 230-236.
462 13. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell
463 carcinomas. *Nature* 2015; **517**: 576-582.
464 14. Lechner M, Frampton GM, Fenton T, *et al*. Targeted next-generation sequencing of head and neck
465 squamous cell carcinoma identifies novel genetic alterations in HPV+ and HPV- tumors. *Genome*
466 *Med* 2013; **5**: 49.
467 15. Seiwert TY, Zuo Z, Keck MK, *et al*. Integrative and comparative genomic analysis of HPV-positive
468 and HPV-negative head and neck squamous cell carcinomas. *Clin Cancer Res* 2015; **21**: 632-641.
469 16. Krishnan N, Gupta S, Palve V, *et al*. Integrated analysis of oral tongue squamous cell carcinoma
470 identifies key variants and pathways linked to risk habits, HPV, clinical parameters and tumor
471 recurrence. *F1000Res* 2015; **4**: 1215.
472 17. Brenner JC, Graham MP, Kumar B, *et al*. Genotyping of 73 UM-SCC head and neck squamous cell
473 carcinoma cell lines. *Head & neck* 2010; **32**: 417-426.

474 18. Martin CL, Reshmi SC, Ried T, *et al*. Chromosomal imbalances in oral squamous cell carcinoma:
475 examination of 31 cell lines and review of the literature. *Oral oncology* 2008; **44**: 369-382.
476 19. Romero-Pastrana F. Detection and typing of human papilloma virus by multiplex PCR with type-
477 specific primers. *ISRN microbiology* 2012; **2012**.
478 20. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. *Nature*
479 *protocols* 2008; **3**: 1101-1108.
480 21. Bhattacharjee B, Sengupta S. CpG methylation of HPV 16 LCR at E2 binding site proximal to P97 is
481 associated with cervical cancer in presence of intact E2. *Virology* 2006; **354**: 280-285.
482 22. Gallo G, Bibbo M, Bagella L, *et al*. Study of viral integration of HPV-16 in young patients with
483 LSIL. *J Clin Pathol* 2003; **56**: 532-536.
484 23. Krishnan NM, Dhas K, Nair J, *et al*. A Minimal DNA Methylation Signature in Oral Tongue
485 Squamous Cell Carcinoma Links Altered Methylation with Tumor Attributes. *Mol Cancer Res* 2016;
486 **14**: 805-819.
487 24. Biron VL, Kostiuk M, Isaac A, *et al*. Detection of human papillomavirus type 16 in oropharyngeal
488 squamous cell carcinoma using droplet digital polymerase chain reaction. *Cancer* 2016; **122**: 1544-
489 1551.
490 25. India Project Team of the International Cancer Genome C. Mutational landscape of gingivo-buccal
491 oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups.
492 *Nature communications* 2013; **4**: 2873.
493 26. Maxwell JH, Grandis JR, Ferris RL. HPV-Associated Head and Neck Cancer: Unique Features of
494 Epidemiology and Clinical Management. *Annu Rev Med* 2016; **67**: 91-101.
495 27. Vokes EE, Agrawal N, Seiwert TY. HPV-Associated Head and Neck Cancer. *Journal of the*
496 *National Cancer Institute* 2015; **107**: djv344.
497 28. Bruni L, Barrionuevo-Rosas L, Serrano B, *et al*. Human papillomavirus and related diseases report.
498 *L'Hospitalet de Llobregat: ICO Information Centre on HPV and Cancer* 2014.
499 29. Seiwert T. Accurate HPV testing: a requirement for precision medicine for head and neck cancer.
500 *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 2013;
501 **24**: 2711-2713.

502 30. Chung CH, Zhang Q, Kong CS, *et al*. p16 protein expression and human papillomavirus status as
503 prognostic biomarkers of nonoropharyngeal head and neck squamous cell carcinoma. *Journal of*
504 *Clinical Oncology* 2014: JCO. 2013.2054. 5228.
505 31. Stephen JK, Divine G, Chen KM, *et al*. Significance of p16 in site-specific HPV positive and HPV
506 negative head and neck squamous cell carcinoma. *Cancer and clinical oncology* 2013; **2**: 51.
507 32. Gröbe A, Hanken H, Kluwe L, *et al*. Immunohistochemical analysis of p16 expression, HPV
508 infection and its prognostic utility in oral squamous cell carcinoma. *Journal of Oral Pathology &*
509 *Medicine* 2013; **42**: 676-681.
510 33. Larsen CG, Gyldenløve M, Jensen D, *et al*. Correlation between human papillomavirus and p16
511 overexpression in oropharyngeal tumours: a systematic review. *British journal of cancer* 2014; **110**:
512 1587-1594.
513 34. Seiwert TY. Ties That Bind: p16 As a Prognostic Biomarker and the Need for High-Accuracy
514 Human Papillomavirus Testing. *Journal of Clinical Oncology* 2014: JCO. 2014.2057. 9268.
515 35. El-Naggar AK, Westra WH. p16 expression as a surrogate marker for HPV-related oropharyngeal
516 carcinoma: A guide for interpretative relevance and consistency. *Head & neck* 2012; **34**: 459-461.
517 36. Alexander RE, Hu Y, Kum JB, *et al*. p16 expression is not associated with human papillomavirus in
518 urinary bladder squamous cell carcinoma. *Modern Pathology* 2012; **25**: 1526-1533.
519 37. Auw-Haedrich C, Martin G, Spelsberg H, *et al*. Expression of p16 in conjunctival intraepithelial
520 neoplasia does not correlate with HPV-infection. *The open ophthalmology journal* 2008; **2**: 48.
521 38. Lingen MW, Xiao W, Schmitt A, *et al*. Low etiologic fraction for high-risk human papillomavirus in
522 oral cavity squamous cell carcinomas. *Oral Oncol* 2013; **49**: 1-8.
523 39. Chung CH, Zhang Q, Kong CS, *et al*. p16 protein expression and human papillomavirus status as
524 prognostic biomarkers of nonoropharyngeal head and neck squamous cell carcinoma. *Journal of*
525 *clinical oncology : official journal of the American Society of Clinical Oncology* 2014; **32**: 3930-
526 3938.
527 40. Gibson JS. Nucleic acid-based assays for the detection of high-risk human papillomavirus: A
528 technical review. *Cancer cytopathology* 2014; **122**: 639-645.
529 41. Ndiaye C, Mena M, Alemany L, *et al*. HPV DNA, E6/E7 mRNA, and p16INK4a detection in head
530 and neck cancers: a systematic review and meta-analysis. *Lancet Oncol* 2014; **15**: 1319-1331.

42. Chaturvedi AK, Anderson WF, Lortet-Tieulent J, *et al*. Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2013; **31**: 4550-4559.

43. Terai M, Hashimoto K, Yoda K, *et al*. High prevalence of human papillomaviruses in the normal oral cavity of adults. *Oral Microbiol Immunol* 1999; **14**: 201-205.

44. Gillison ML, Broutian T, Pickard RK, *et al*. Prevalence of oral HPV infection in the United States, 2009-2010. *JAMA* 2012; **307**: 693-703.

**Table 1:** Summary of HPV prevalence by all assays. DNA PCR was performed with consensus or HPV16/18 type-specific primers; DNA qPCR with Taqman assays for HPV16/18; DNA ddPCR with HPV16 primers; RNA qPCR with SYBR based assay to detect HPV16/18 E6/E7 mRNA. The primer and probe sequences along with the amplification conditions are provided in Supplementary Table 2.

| **Analyte/Assay** | **Detection Method** | **HPV Positivity (%)** | | |
|---|---|---|---|---|
| | | **Oral tongue** | **Buccal mucosa** | **Oral cavity (both subsites)** |
| Protein | Immunohistochemistry | 51% (28/55) | NA | 51% (28/55) |
| | | | | |
| DNA | PCR | 59% (39/66) | 50% (2/4) | 58% (41/70) |
| | qPCR | 44% (34/78) | 4% (1/28) | 33% (35/106) |
| | ddPCR | 55% (52/95) | 17% (7/41) | 43% (59/136) |
| | | | | |
| RNA | qPCR | 17% (05/30) | 9% (1/11) | 15% (06/41) |
| | | | | |
| Combinations | PCR + qPCR | 38% (23/60) | | |
| | PCR + ddPCR | 48%(29/60) | | |
| | qPCR + ddPCR | 27% (27/99) | | |
| | PCR +qPCR + ddPCR (all 3 methods, 3/3 method) | 38% (20/53) | | |
| | p16 + 3/3 Methods | 19% (7/36) | | |
| | p16 + RNA | 0% (0/16) | | |
| | RNA + 3/3 methods | 6% (1/17) | | |
| | | | | |
| Final HPV positivity | Relative high-copy HPV DNA and/or HPV E6/E7 RNA positive | 22% (23/106) | | |
| | | | | |
| Integrated/episomal | HPV16E6 PCR followed with E2-PCR | 18% (7/38) | 0% (0/2) | 18% (07/40) |

540
541

**Table 2:** Summary of the current study on prevalence (A) and tumors with somatic mutations in key genes in tumors with relatively high-copy HPV DNA and/or HPV E6/E7 mRNA. BM: buccal mucosa, OT: oral tongue, ●: detected, ○: not-detected, Y: high-copy present, N: high-copy absent, L: low copy present, -: not done, M: mutant, WT: wild –type.

**A**

| Sample Code | Protein | DNA | | | | | RNA | Integration Status |
|---|---|---|---|---|---|---|---|---|
| | p16 IHC | PCR | qPCR | ddPCR | +ve in 3/3 | DNA high-copy | E6/E7 RNA | E2 PCR |
| BM1 | - | - | ○ | ○ | - | N | - | - |
| BM10 | - | - | - | ○ | - | - | - | - |
| BM11 | - | - | - | ○ | - | - | - | - |
| BM12 | - | - | ○ | ○ | - | N | - | - |
| BM13 | - | - | ○ | ○ | - | N | ○ | - |
| BM14 | - | - | ○ | ○ | - | N | ○ | - |
| BM15 | - | - | - | ○ | - | - | - | - |
| BM16 | - | - | - | ● | - | - | - | - |
| BM17 | - | - | ○ | ○ | - | N | - | - |
| BM18 | - | - | ○ | ● | - | N | - | - |
| BM19 | - | - | ○ | ○ | - | N | - | - |
| BM2 | - | - | - | ● | - | - | - | - |
| BM20 | - | - | ○ | ○ | - | N | - | - |
| BM21 | - | - | ○ | ● | - | N | - | - |
| BM22 | - | - | ○ | ○ | - | N | ○ | - |
| BM23 | - | - | ○ | ○ | - | N | - | - |
| BM24 | - | - | ○ | ○ | - | N | ○ | - |
| BM25 | - | - | ○ | ○ | - | N | - | - |
| BM26 | - | - | ○ | ● | - | N | ○ | - |
| BM27 | - | - | - | ○ | - | - | - | - |
| BM28 | - | - | - | ● | - | - | - | - |
| BM29 | - | ○ | ○ | ○ | ○ | N | - | - |
| BM3 | - | - | ○ | ○ | - | N | - | - |
| BM30 | - | - | ○ | ○ | - | N | - | - |
| BM31 | - | - | - | ○ | - | - | - | - |
| BM32 | - | - | - | ○ | - | - | - | - |
| BM33 | - | - | - | ○ | - | - | - | - |
| BM34 | - | - | - | ○ | - | - | - | - |
| BM35 | - | ● | ○ | ○ | ○ | N | - | ○ |
| BM36 | - | ○ | ○ | ○ | ○ | N | - | - |
| BM37 | - | - | ○ | ○ | - | N | - | - |
| BM38 | - | - | ○ | ○ | - | N | ○ | - |
| BM39 | - | - | ○ | ○ | - | N | ○ | - |
| BM4 | - | - | ○ | ○ | - | N | ○ | - |
| BM40 | - | - | ○ | ○ | - | N | - | - |
| BM41 | - | - | - | ○ | - | - | - | - |
| BM5 | - | ● | ● | ○ | ○ | Y | - | ○ |
| BM6 | - | - | ○ | ○ | - | N | ● | - |
| BM7 | - | - | - | ○ | - | - | - | - |
| BM8 | - | - | ○ | ○ | - | N | ○ | - |
| BM9 | - | - | ○ | ● | - | N | ○ | - |
| OT10 | ● | ● | ○ | ● | ○ | N | - | ○ |
| OT100 | - | ● | ○ | ● | ○ | N | ○ | ○ |
| OT101 | - | - | ○ | ○ | - | N | ● | - |
| OT102 | - | - | - | ● | - | - | - | - |
| OT103 | - | - | ○ | ○ | - | N | - | - |
| OT104 | - | - | - | ● | - | - | - | - |
| OT105 | - | - | - | ○ | - | - | - | - |
| OT106 | - | - | ○ | ○ | - | N | ○ | - |
| OT107 | - | - | - | ● | - | - | - | - |
| OT108 | - | - | - | ○ | - | - | - | - |
| OT109 | - | - | ○ | ○ | - | N | - | - |
| OT11 | ● | ● | - | ○ | - | - | ○ | ○ |
| OT110 | - | - | ○ | ○ | - | N | ● | - |
| OT111 | - | - | - | ○ | - | - | - | - |
| OT112 | - | ● | ○ | ● | ○ | N | - | ○ |
| OT113 | - | - | - | ○ | - | - | - | - |
| OT114 | ● | ● | ● | ● | ● | Y | - | - |
| OT115 | ○ | ● | - | - | - | - | - | ○ |
| OT116 | - | - | ○ | ● | - | N | - | - |
| OT12 | ● | ● | - | - | - | - | - | ○ |
| OT13 | ○ | ● | - | - | - | - | - | ○ |
| OT14 | ○ | ● | ● | ● | ● | Y | ○ | ● |
| OT15 | ○ | - | - | - | - | - | - | - |
| OT16 | ○ | ● | ● | ● | ● | L | - | - |
| OT17 | ○ | ● | ● | - | - | L | ● | ○ |
| OT18 | ○ | ● | ○ | - | - | N | - | ○ |
| OT19 | ○ | ○ | ○ | ○ | ○ | N | - | - |
| OT2 | ● | ○ | ○ | ○ | ○ | N | - | - |
| OT20 | ○ | ● | ● | ○ | ○ | L | - | ○ |
| OT21 | - | ○ | ○ | - | - | N | ○ | - |
| OT22 | - | ○ | - | ○ | - | - | - | - |
| OT23 | ○ | - | - | - | - | - | - | - |
| OT24 | ● | ● | ○ | ○ | ○ | N | ○ | ○ |
| OT25 | ● | ● | - | ● | - | - | ○ | ● |
| OT26 | ○ | ○ | ○ | ○ | ○ | N | - | - |
| OT27 | - | ○ | - | ○ | - | - | ○ | - |
| OT28 | ○ | ○ | ○ | - | - | N | - | - |
| OT29 | ● | ○ | ○ | ● | ○ | N | - | ○ |
| OT3 | ● | ● | ○ | ○ | ○ | N | ○ | ○ |
| OT30 | ● | ● | ● | ● | ● | L | - | ● |
| OT31 | ● | ○ | ● | ● | ○ | - | - | - |
| OT32 | ○ | ○ | ● | ● | ○ | Y | - | ○ |
| OT33 | ● | ○ | ● | ○ | ○ | L | ○ | - |
| OT34 | ● | ○ | - | ● | - | - | - | - |
| OT35 | - | - | - | - | - | - | - | - |
| OT36 | - | ○ | ● | - | - | Y | - | - |
| OT37 | - | ○ | ○ | - | - | N | - | - |
| OT38 | ○ | ○ | ● | ● | ○ | L | - | - |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| OT39 | ● | - | - | - | - | - | - | - |
| OT4 | ○ | ○ | ○ | ● | ○ | N | - | ● |
| OT40 | ● | ○ | - | ● | - | N | ○ | - |
| OT41 | ○ | ○ | ○ | ○ | ○ | N | - | - |
| OT42 | ○ | ● | ● | ● | ● | L | ○ | ○ |
| OT43 | ● | ○ | ○ | ○ | ○ | N | ○ | ○ |
| OT44 | ● | ● | ● | ● | ● | L | ○ | ○ |
| OT45 | - | ○ | - | ○ | - | - | - | - |
| OT46 | - | ● | ● | ● | ● | L | - | ● |
| OT48 | ○ | ● | ● | ● | ● | L | - | - |
| OT5 | ● | - | - | - | - | - | - | - |
| OT50 | - | ○ | ○ | ● | ○ | N | - | - |
| OT51 | ● | ○ | ● | - | - | L | - | - |
| OT52 | ○ | - | - | - | - | - | - | - |
| OT53 | ● | ● | ○ | ○ | ○ | N | - | ○ |
| OT54 | ● | ● | ● | ● | ● | L | - | ○ |
| OT55 | ● | - | ○ | ● | - | N | ○ | ○ |
| OT56 | - | - | ○ | ○ | - | N | - | - |
| OT57 | - | - | - | - | - | - | - | - |
| OT58 | - | - | ● | ○ | - | Y | - | - |
| OT59 | - | - | - | ○ | - | - | - | - |
| OT6 | ● | ○ | ● | ○ | ○ | - | - | - |
| OT60 | - | - | ○ | ○ | - | N | ○ | - |
| OT61 | ○ | ● | ● | ● | ● | Y | ● | ○ |
| OT62 | - | ● | ○ | ● | ○ | N | ○ | ○ |
| OT63 | - | - | - | ○ | - | - | - | - |
| OT64 | - | ● | ● | ● | ● | L | ○ | ○ |
| OT65 | ○ | ● | ● | ● | ● | Y | - | ● |
| OT66 | - | - | ○ | ● | - | N | - | - |
| OT67 | ○ | ● | ● | ● | ● | Y | ○ | ○ |
| OT68 | - | - | ● | ● | - | Y | - | - |
| OT69 | ● | ● | ● | ● | ● | Y | - | - |
| OT7 | ● | ● | ○ | ● | ○ | N | - | - |
| OT70 | - | - | ○ | ○ | - | N | - | ○ |
| OT71 | - | - | ○ | ○ | - | N | - | - |
| OT72 | - | - | ○ | ● | - | N | ● | - |
| OT73 | - | - | - | ● | - | - | - | - |
| OT74 | - | - | - | ○ | - | - | - | - |
| OT75 | - | ● | ● | ● | ● | Y | ○ | ○ |
| OT76 | - | ● | ● | ● | ● | Y | - | ● |
| OT77 | ● | ● | ● | ● | ● | L | ○ | ○ |
| OT78 | ○ | ○ | ○ | ● | ○ | N | ○ | - |
| OT79 | - | - | - | ○ | - | - | - | - |
| OT80 | - | ○ | ● | ● | ○ | Y | ○ | - |
| OT81 | ● | ● | ● | ● | ● | Y | - | - |
| OT82 | ○ | ● | ● | ● | ● | Y | - | ○ |
| OT83 | ○ | ● | ○ | ● | ○ | N | - | ○ |
| OT84 | ○ | - | ○ | ○ | - | N | - | - |
| OT85 | - | ● | ○ | ● | ○ | N | - | ○ |
| OT86 | - | - | ○ | ○ | - | N | - | - |
| OT87 | - | - | ○ | ○ | - | N | ○ | - |
| OT88 | - | ○ | ● | ● | ○ | Y | ○ | ○ |
| OT89 | - | - | - | ○ | - | - | - | - |
| OT9 | ● | ● | ○ | ● | ○ | N | - | - |
| OT90 | - | - | ○ | ○ | - | N | - | - |
| OT91 | ○ | - | ○ | ○ | - | N | - | - |
| OT92 | - | ○ | ● | ● | ○ | L | - | - |
| OT93 | - | - | - | ● | - | - | - | - |
| OT94 | - | - | - | ○ | - | - | - | - |
| OT95 | - | ● | ● | ● | ● | Y | - | - |
| OT96 | - | - | - | ○ | - | - | - | - |
| OT97 | - | - | ○ | ○ | - | N | - | - |
| OT98 | - | - | - | ● | - | - | - | - |
| OT99 | - | - | ○ | ○ | - | N | - | - |

**B**

| Sample Code | DNA | RNA | Somatic mutation | | |
|---|---|---|---|---|---|
| | High-copy | E6/E7 RNA | TP53 | CASP8 | RASA1 |
| BM5 | Y | - | NA | NA | NA |
| BM6 | N | ● | NA | NA | NA |
| OT101 | N | ● | M | NA | WT |
| OT110 | N | ● | WT | M | WT |
| OT114 | Y | - | WT | WT | WT |
| OT14 | Y | ○ | WT | WT | WT |
| OT17 | L | ● | WT | WT | WT |
| OT32 | Y | - | WT | M | WT |
| OT36 | Y | - | WT | WT | M |
| OT58 | Y | - | NA | NA | NA |
| OT61 | Y | ● | WT | WT | WT |
| OT65 | Y | - | WT | WT | WT |
| OT67 | Y | ○ | WT | WT | WT |
| OT68 | Y | - | WT | WT | WT |
| OT69 | Y | - | WT | WT | WT |
| OT72 | N | ● | WT | WT | WT |
| OT75 | Y | ○ | WT | WT | WT |
| OT76 | Y | - | WT | M | WT |
| OT80 | Y | ○ | WT | WT | WT |
| OT81 | Y | - | WT | WT | WT |
| OT82 | Y | - | NA | WT | NA |
| OT88 | Y | ○ | WT | WT | WT |
| OT95 | Y | - | WT | WT | WT |

549 **Figure legends.**

550 Figure 1A: Primer and probes locations used in the study to detect HPV DNA and HPV E6/E7
551 RNA.

552

553 Figure 2: HPV qPCR (A-F) and digital PCR (G-I). Standard curves were obtained using cloned
554 HPV16/18 plasmids (A,D,G), positive and negative cell line DNA (B,E,H) were used to count DNA
555 in tumors (C,F,I).

556

557 Figure 3: HPV16/18 E6/E7 mRNA expression using qPCR.

558

559 Figure 4: Kaplan Meier survival plots linking tumors with high-copy HPV DNA and/or HPV RNA
560 status (A), somatic mutations in 3 frequently mutated genes in tumors (B), clustering of nine
561 methylated genes stratifying HPV positive from negative group of tumors (C) and HPV-associated
562 pathways (D) in HPV positive tumors.

563

564 **Supplementary Figures and Tables legends:**

565

566 Supplementary Table 1: Clinical details of patients (n = 153) used in the study.

567

568 Supplementary Table 2: Primer and probe sequences along with amplicon size and the conditions of
569 amplification reactions used for nucleic acid based detection.

570

571 Supplementary Table 3: Significance (*p* values) from unpaired *t*-tests between HPV positive and
572 negative tumors in two groups based on HPV DNA copy number.

573

574 Supplementary Figure 1: Genomic DNA from cell lines were used as controls for HPV16 (A,B) and
575 HPV18 (C,D) PCR.

576

577 Supplementary Figure 2: Representative images for p16 immunohistochemical staining (IHC) in –
578 ve OSCC, +ve OSCC and cervical cancer (+ve control) samples.

579

580 Supplementary Figure 3: Amplification efficiency as measured by serially diluted HPV16/18 cloned
581 plasmids.

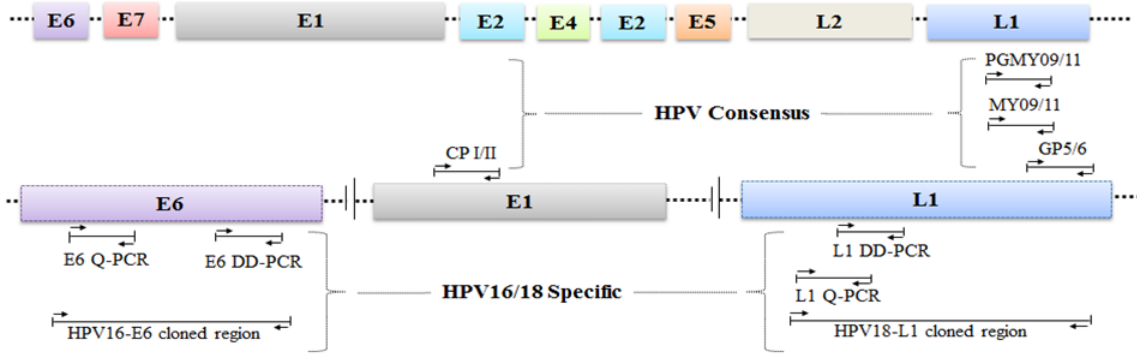
582

583 Supplementary Figure 4: PCR gel images with OSCC tumor DNA.

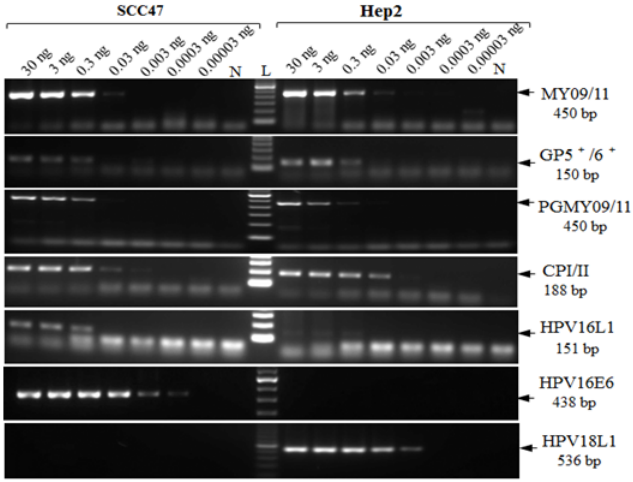
584

585 Supplementary Figure 5: Inhibition of amplification reactions at high concentration of tumor
586 genomic DNA with spike-in experiment.
587
588 Supplementary Figure 6: The effect of amplification cycle on PCR.
589
590 Supplementary Figure 7: HPV E2PCR to test viral integration into host genomes.
591
592 Supplementary Figure 8: KM survival analysis plots with tumor attributes. Tumor differentiation
593 (A), stage (B-D), habits (E-H), age (I), and nodal status (J).
594
595 Supplementary Figure 9: KM survival analysis plots with p16 (A,B), HPV DNA- PCR (C,D), qPCR
596 (E-H), ddPCR(I,J), positive in 3/3 DNA-based method (K,L), and E6/E7 mRNA (M,N).
597
598 Supplementary Figure 10: KM survival analysis plots in various somatic mutational backgrounds
599 (A-F).

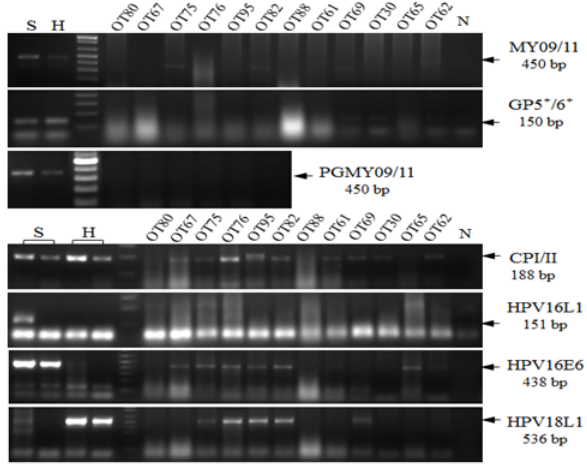
A

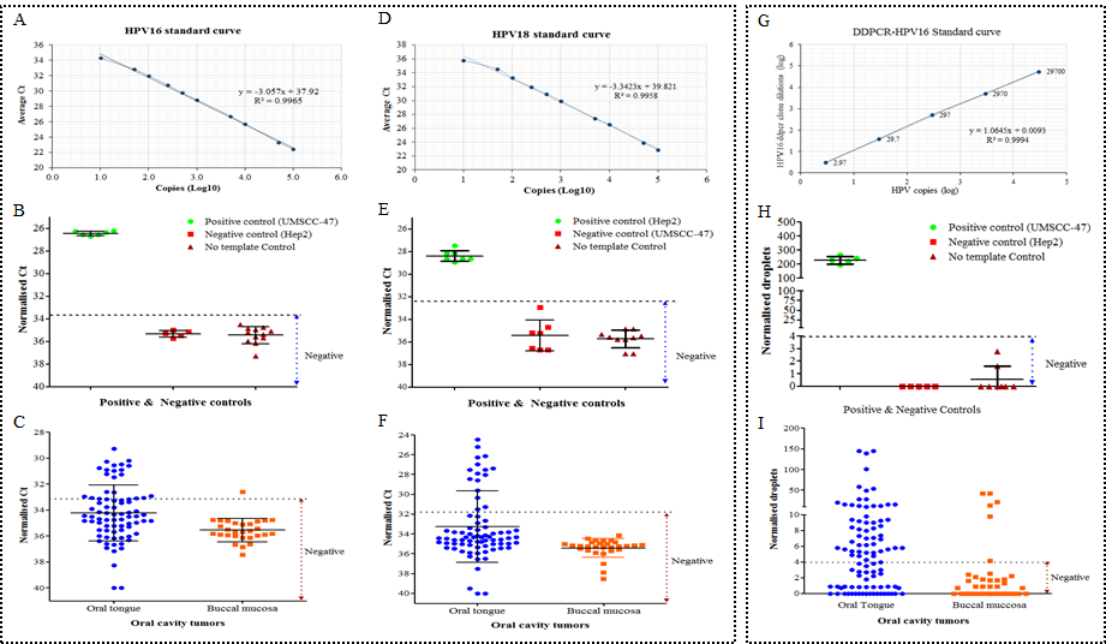

E6
E7
E1
E2
E4
E2
E5
L2
L1
HPV Consensus
PGMY09/11
MY09/11
GP5/6
CP I/II
E6
E1
L1
E6 Q-PCR
E6 DD-PCR
HPV16/18 Specific
L1 DD-PCR
L1 Q-PCR
HPV16-E6 cloned region
HPV18-L1 cloned region

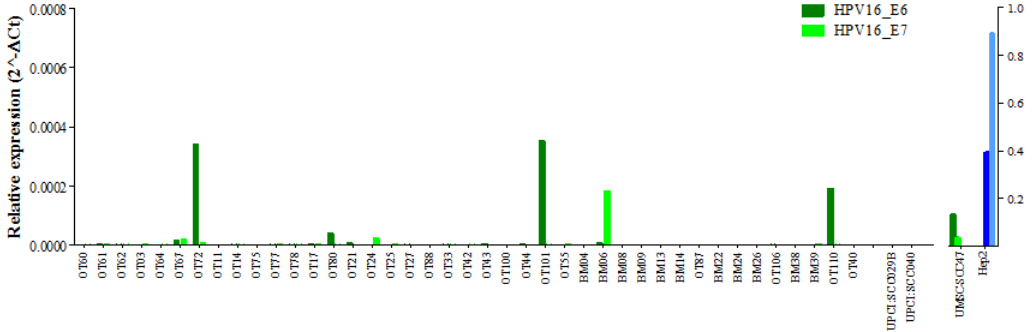
B

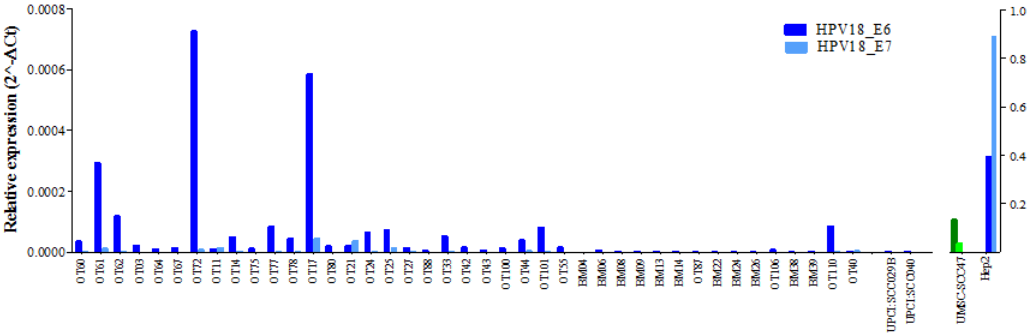

SCC47
Hep2
30 ng
3 ng
0.3 ng
0.03 ng
0.003 ng
0.00003 ng
0.00003 ng
N
L
MY09/11
450 bp
GP5 + /6 +
150 bp
PGMY09/11
450 bp
CPI/II
188 bp
HPV16L1
151 bp
HPV16E6
438 bp
HPV18L1
536 bp

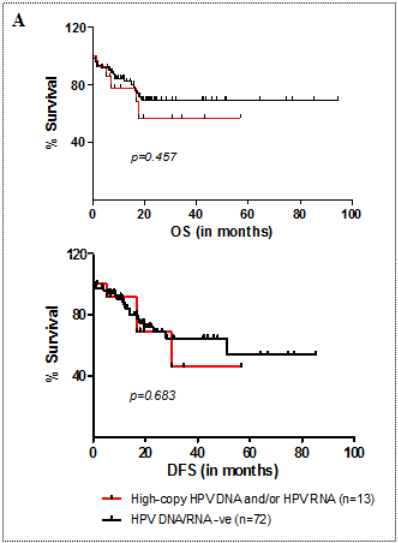
C

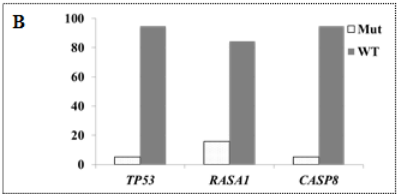

S
H
OT80
OT67
OT75
OT76
OT95
OT82
OT88
OT61
OT69
OT30
OT65
OT62
N
MY09/11
450 bp
GP5+/6+
150 bp
PGMY09/11
450 bp
CPI/II
188 bp
HPV16L1
151 bp
HPV16E6
438 bp
HPV18L1
536 bp

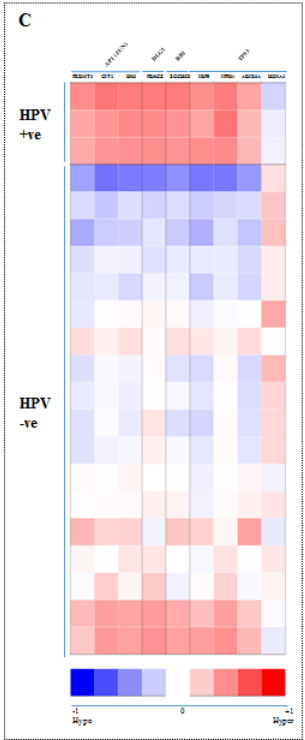

A
HPV16 standard curve
Average Ct
y = -3.057x + 37.92
R² = 0.9965
Copies (Log10)
D
HPV18 standard curve
Average Ct
y = -3.3423x + 39.821
R² = 0.9958
Copies (Log10)
G
DDPCR-HPV16 Standard curve
HPV16 clone dilutions (log)
29700
2970
297
29.7
2.97
y = 1.0645x + 0.0093
R² = 0.9994
HPV copies (log)
B
Positive control (UMSCC-47)
Negative control (Hep2)
No template Control
Normalised Ct
Negative
Positive & Negative controls
E
Positive control (Hep2)
Negative control (UMSCC-47)
No template Control
Normalised Ct
Negative
Positive & Negative controls
H
Positive control (UMSCC-47)
Negative control (Hep2)
No template Control
Normalised droplets
Negative
Positive & Negative Controls
C
Normalised Ct
Negative
Oral tongue
Buccal mucosa
Oral cavity tumors
F
Normalised Ct
Negative
Oral tongue
Buccal mucosa
Oral cavity tumors
I
Normalised droplets
Negative
Oral Tongue
Buccal muccosa
Oral cavity tumors

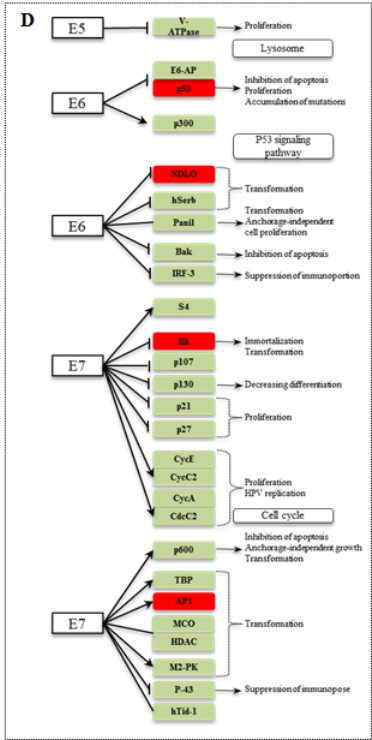

HPV16_E6
HPV16_E7
Relative expression (2^-ACt)
0.0008
0.0006
0.0004
0.0002
0.0000
1.0
0.8
0.6
0.4
0.2
OT60
OT61
OT62
OT03
OT64
OT67
OT72
OT11
OT14
OT75
OT77
OT78
OT17
OT80
OT21
OT24
OT25
OT27
OT88
OT33
OT42
OT43
OT100
OT44
OT101
OT55
BM04
BM06
BM08
BM09
BM13
BM14
OT87
BM22
BM24
BM26
OT106
BM38
BM39
OT110
OT40
UPCI:SCC029B
UPCI:SCC040
UM-SCC-47
Hep2


HPV18_E6
HPV18_E7
Relative expression (2^-ACt)
0.0008
0.0006
0.0004
0.0002
0.0000
1.0
0.8
0.6
0.4
0.2
OT60
OT61
OT62
OT03
OT64
OT67
OT72
OT11
OT14
OT75
OT77
OT78
OT17
OT80
OT21
OT24
OT25
OT27
OT88
OT33
OT42
OT43
OT100
OT44
OT101
OT55
BM04
BM06
BM08
BM09
BM13
BM14
OT87
BM22
BM24
BM26
OT106
BM38
BM39
OT110
OT40
UPCI:SCC029B
UPCI:SCC040
UM-SCC-47
Hep2


A
% Survival
p=0.457
OS (in months)
p=0.683
DFS (in months)
High-copy HPV DNA and/or HPV RNA (n=13)
HPV DNA/RNA -ve (n=72)


B
Mut
WT
TP53
RASA1
CASP8


C
HPV +ve
HPV -ve
Hype
Hyper


D
E5
V-ATPase
Proliferation
Lysosome
E6
E6-AP
p53
p300
Inhibition of apoptosis
Proliferation
Accumulation of mutations
P53 signaling pathway
NDLO
hSerb
Panil
Bak
IRF-3
Transformation
Transformation
Anchorage-independent cell proliferation
Inhibition of apoptosis
Suppression of immunoportion
E7
S4
Rb
p107
p130
p21
p27
CycE
CycC2
CycA
CdcC2
Immortalization
Transformation
Decreasing differentiation
Proliferation
Proliferation
HPV replication
Cell cycle
p600
TBP
AP1
MCO
HDAC
M2-PK
P-43
hTid-1
Inhibition of apoptosis
Anchorage-independent growth
Transformation
Transformation
Suppression of immunopose