

UC San Diego

UC San Diego Previously Published Works

Title

Rational genomic optimization of DNA detection for human papillomavirus type 16 in head and neck squamous cell carcinoma

Permalink

<https://escholarship.org/uc/item/8ng706bx>

Journal

Head & Neck, 42(4)

ISSN

1043-3074

Authors

Saito, Yuki

Favorov, Alexander V

Forman, Michael

et al.

Publication Date

2020-04-01

DOI

10.1002/hed.26041

Peer reviewed



Published in final edited form as:

Head Neck. 2020 April ; 42(4): 688–697. doi:10.1002/hed.26041.

## Rational genomic optimization of DNA detection for human papillomavirus type 16 in head and neck squamous cell carcinoma

Yuki Saito, MD, PhD<sup>1</sup>, Alexander V. Favorov, PhD<sup>2,3</sup>, Michael Forman, PhD<sup>4</sup>, Shuling Ren, MD<sup>1</sup>, Akihiro Sakai, MD, PhD<sup>1</sup>, Takahito Fukusumi, MD, PhD<sup>1</sup>, Chao Liu, MD<sup>1</sup>, Sayed Sadat, BA<sup>1</sup>, Mizuo Ando, MD, PhD<sup>1</sup>, Guorong Xu, PhD<sup>5</sup>, Zubair Khan, MD<sup>6</sup>, John Pang<sup>1</sup>, Alex Valsamakis, MD<sup>4</sup>, Kathleen M. Fisch, PhD<sup>5</sup>, Joseph A. Califano, MD<sup>1,7</sup>

<sup>1</sup>Moore's Cancer Center, University of California San Diego, San Diego, California <sup>2</sup>Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland <sup>3</sup>Laboratory of Systems Biology and Computational Genetics, Vavilov Institute of General Genetics, Moscow, Russia <sup>4</sup>Department of Pathology, Johns Hopkins University, Baltimore, Maryland <sup>5</sup>Center for Computational Biology & Bioinformatics, Department of Medicine, University of California San Diego, San Diego, California <sup>6</sup>Department of Otolaryngology-Head and Neck Surgery, Johns Hopkins Hospital, Baltimore, Maryland <sup>7</sup>Division of Otolaryngology Head and Neck Surgery, Department of Surgery, University of California San Diego, San Diego, California

### Abstract

**Background:** We aimed to use genomic data for optimizing polymerase chain reaction (PCR) primer/probe sets for detection of human papillomavirus (HPV)-16 in body fluids of patients with HPV-related head and neck squamous cell carcinoma (HPV-HNSCC).

**Methods:** We used genomic HPV-HNSCC sequencing data from a single institutional and a TCGA cohort. Optimized primer/probe sets were designed and tested for analytical performance in CaSki HPV-16 genome and confirmed in salivary rinse samples from patients with HPV-HNSCC.

**Results:** The highest read density was observed between E5 and L2 regions. The E1 region contained a region that was universally present. Among candidate PCR primer/probe sets created, six reliably detected 30 HPV-16 copy number. In a CLIA certified laboratory setting, the combination of two novel primer/probe with E7 sets improved performance in salivary rinse samples with a sensitivity of 96% and specificity of 100%.

---

**Correspondence:** Joseph A. Califano, MD, 3855 Health, Sciences Dr., MC 0803 La Jolla, CA, 92037-0803., jcalifano@ucsd.edu.

CONFLICT OF INTEREST

The authors have no potential conflict of interest to declare.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**Conclusions:** PCR-based detection of HPV-16 DNA in HPV-HNSCC can be improved using rational genomic design.

### Keywords

DNA primer; head and neck squamous cell carcinoma; human papillomavirus; polymerase chain reaction; whole-genome sequencing

## 1 | INTRODUCTION

Human papillomavirus (HPV)-associated head and neck squamous cell carcinoma (HPV-HNSCC) has rapidly increased in incidence and is associated with unique epidemiological, molecular, and biological characteristics.<sup>1,2</sup> The presence of primary tumor HPV DNA is a favorable prognostic biomarker in HNSCC, particularly in oropharyngeal carcinoma (OPC).<sup>3</sup>

Post-treatment HPV DNA levels may be measured in a noninvasive manner by obtaining specimens, such as salivary rinse and plasma, which may predict the presence of HPV-OPC and risk of recurrence.<sup>3-5</sup> However, the sensitivity of the detection of HPV DNA in salivary rinses and plasma from HPV oropharynx patients with cancer hovers around 50%. We hypothesized that an optimized quantitative polymerase chain reaction (qPCR) assay with improved performance in HPV-16 DNA detection in body fluids has potential to improve early detection and subsequent management of HPV-OPC.

Current PCR-based methods for detection of high-risk HPV often rely on historical, non-optimized primer/probe sets developed over two decades ago, and therefore, do not build upon the opportunities provided by recent technological advances, such as whole genome sequencing (WGS) or improved assay design, or benefit from analytic validation in a clinical testing environment. The detection of HPV in the fragmented DNA from formalin-fixed, paraffin-embedded samples using historical primer/probe sets is sometimes challenging,<sup>5-7</sup> because of amplicon size. To maximize the sensitivity for HPV DNA detection in body fluids, a rational primer/probe design should ideally incorporate an analysis of the distribution and frequency of HPV genome regions in HPV-HNSCC. This approach allows primers/probes to target regions of HPV DNA that are abundant in tumors as well as regions that are universally or consistently present in a large proportion of tumors related to HPV.

In addition, prior literature describes an assay application in research settings. However, test performance in a clinical environment is defined by U.S. federal regulatory standards, including The Clinical Laboratory Improvement Amendments (CLIA) of 1988 that apply to all clinical laboratory testing performed on humans in the United States.

In the present study, we focused on the HPV-16 subtype, which is associated with approximately 95% of HPV-HNSCC.<sup>8</sup> We analyzed two datasets using WGS of HPV-HNSCC, followed by rational primer/probe design based on read density across the HPV-16 genome as well as the presence of HPV DNA in a high proportion of tumors. The efficacy of candidate primer/probe sets was assessed using qPCR, followed by analytic optimization and validation of the sensitivity of detection of HPV-16. Subsequently, we adapted this assay

to conform to CLIA standards and performed this assay in a CLIA certified clinical laboratory to define assay performance in salivary rinse specimens from patients with HPV-HNSCC.

## 2 | MATERIALS AND METHODS

### 2.1 | Biological samples

Primary tumor tissue samples were obtained from a cohort of 72 patients with HPV-related oropharyngeal squamous cell carcinoma, as previously described.<sup>9</sup> All tissue samples were collected from the Johns Hopkins Tissue Core under an approved IRB protocol (#NA\_00-36 235). Of the 72 primary tumors, 40 were used as the “discovery set” for rational primer/probe design; 32 primary tumors and salivary rinse samples obtained from 28/32 of these patients were used to assess the efficacy of the primers/probes designed for HPV detection and compared to separately collected noncancer control salivary rinses. The discovery set included 40 primary tumor samples from patients with primary OPCs with validated presence of HPV genome.

### 2.2 | Sequence data analysis

DNA was isolated from 0.35-mm frozen tissue sections and purified as previously described.<sup>9</sup> Sequencing was performed with the TruSeq Cluster Kit (Illumina, San Diego, California) using the HiSeq 2500 platform sequencer, which provided approximately 80 million paired reads per sample. An in-house pipeline was developed to extract all unmapped reads from BAM files and re-align them to HPV-16 genomes (accession number: AY686584.1).<sup>8,10</sup> To determine the distribution of HPV-16 genome reads in HPV-HNSCC, we performed realignment of the HPV-16 genomic sequence in a discovery set. WGS data were used to define the read coverage of each nucleotide position in the HPV-16 genome. This in-house pipeline consisted of samtools 1.19,<sup>11</sup> bowtie 1.1,<sup>12</sup> bamtools, and Rsamtools.<sup>13</sup>

The pipeline steps included preparation of the HPV reference genome file, quality control on BAM files, extraction of unmapped read pairs, conversion of unmapped read pairs to FASTQ format, alignment of unmapped read pairs to the HPV reference genomes (accession number: AY686584.1) using Rsamtools, and parsing of binary files with scanBam. The final steps included excluding aligned reads and determination of coverage.

### 2.3 | Validation of TCGA data

TCGA WGS was performed on the Illumina GA2000 and HiSeq platforms; sequencing as well as clinical data are publicly available. A similar in-house pipeline was developed to extract all unmapped reads from BAM files and to re-align these to HPV-16 genomes (accession number: AY686584.1). We analyzed WGS data from HPV-HNSCC (certain cases were not primary OPCs) in 94 samples from the TCGA cohort (TCGA, Provisional version updated 2016, <http://cancergenome.nih.gov/>). Clinical records of HPV positivity were obtained from the Broad GDAC Firebrowse website (<http://firebrowse.org/>). We redefined read coverage for each nucleotide position.

## 2.4 | Primer/probe design

Primer/probe sets were designed using Primer-BLAST on the NCBI website (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). We entered the HPV-16 accession number with focused regions and constructed the PCR product size to be as small as possible (<85 bp). Probes were designed to be <35 bp, with an annealing temperature of <60°C. Previously designed primers and probes that specifically amplify the E6 and E7 regions of HPV-16 were used as positive controls for PCR. We used the E6/E7 primer/probe sets as previously reported.<sup>3</sup>

## 2.5 | Assay design

The ability of primer/probe sets to detect HPV-16 was validated using the HPV-positive cancer cell line CaSki (American Type Culture Collection) and leukocytes obtained from healthy patients. This CaSki cell line reportedly contains 600 copies of HPV-16 DNA per genome.<sup>14</sup> Thresholds used for HPV-16 copy number and PCR efficiency were defined using the CaSki cell line, primary tumor samples obtained from 32 patients with HPV-positive HNSCC, and saliva samples obtained from 28 patients with HPV-positive HNSCC.

PCR was performed on DNA derived from the CaSki cell line using the primer/probe sets designed in combination with 0.2- $\mu$ M JumpStart REDTaq ReadyMix (SIGMA-Aldrich, Missouri) primer. Amplification consisted of a denaturation phase (95°C for 3 minutes, 35 cycles of 95°C for 30 seconds, followed by 55°C for 30 seconds and 72°C for 1 minute). All samples were stored at 4°C. Each PCR product was electrophoresed on a 2.0% agarose gel to evaluate the success of PCR.

Quantitative PCR (qPCR) performed using the Quant Studio 6 Flex Real-Time PCR System (Thermo Fisher Scientific Inc., Massachusetts) was used to examine the threshold in each experiment; the experiment for all samples were conducted in triplicates. HPV-16 viral amplicon copy number estimates were developed using the CaSki cell line. Following amplification of the amplicon by PCR, DNA concentration was calculated by weight NanoDrop (Thermo Fisher Scientific Inc., Massachusetts), and these values were converted to DNA copy number by calculations based on the molecular weight of the double-stranded amplicon. This known-copy-number DNA was serially diluted into 100, 30, and 10 copies among 10-ng normal human DNA. qPCR was performed and regarded as positive based on consistent detection in all reactions of a triplicate experiment.

The efficiency of PCR was calculated from the slope of the standard curve, using HPV-16 known-copy-number DNA for each primer/probe set designed. The lowest concentration that demonstrated consistent amplification in a particular triplicate was defined as the threshold. A combination of the primer/probe sets was regarded as positive when at least one of each primer/probe was positive.

Extracted DNA from primary tumor samples was diluted to 150, 30, 15, 6, 1.2, 0.6, and 0.3 cells' worth of DNA based on the following formula: mass of tumor DNA/mass of normal human leukocyte DNA. This calculation was based on the assumption that one tumor cell has 6.6 pg DNA [=  $3 \times 10^9$  (human DNA genome bp)  $\times$  2 (diploid)  $\times$   $1.67 \times 10^{-24}$  g (weight/Da)]. These DNA levels were achieved by diluting 990 pg, 198 pg, 99 pg, 39.6 pg,

7.9 pg, 3.2 pg, and 1.6 pg of tumor DNA, each, in 10 ng of normal human leukocyte DNA. Amplification thresholds were defined in triplicate for each primer/probe set.

## 2.6 | CLIA test development

For samples analyzed in a CLIA certified environment, total nucleic acid was extracted from saliva samples on the easyMag (bioMerieux, Marcy l'Etoile, France) automated extraction platform using the generic protocol. Saliva input/elution volume was 1 mL/50  $\mu$ L. DNA from salivary rinse samples was diluted to 20 ng. HPV-16 qPCR was performed using TaqMan 2X Universal PCR master mix (Applied Biosystems, Foster City, California). For E7, forward and reverse primers (final concentration 450 nM) were 5'-6-carboxyfluorescein-labeled E7, highest read (HR) E5L 2-4, or BR E1-5 probes containing 3'-6-carboxytetramethylrhodamine quencher (final concentration 100 nM), distilled H<sub>2</sub>O, and template (10  $\mu$ L; total reaction volume, 50  $\mu$ L). qPCR was performed on a 7500 Sequence Detection System (Applied Biosystems, Foster City, CA). Cycling parameters were 50°C for 2 minutes, 95°C for 10 minutes, 45 cycles of 95°C for 15 seconds, and 60°C for 60 seconds. Calibrators for target quantification included subclones of target amplicons prepared in pCR2.1-TOPO plasmid vector (pATCC HPV16; Invitrogen, Carlsbad, California). Calibration curves were determined from serial 10-fold dilutions of plasmids (from 7.0 to 1.0 log<sub>10</sub> copies/reaction). Figure 1 presents the study pipeline.

## 3 | RESULTS

### 3.1 | HPV-16 DNA genomic read coverage HPV-HNSCC

Total read count for the HPV-16 genome was 452 590 608 reads. Average read counts of each nucleotide position were evaluated in all 40 samples. Median read for a given nucleotide position was 1019. There were significant spikes (>25 000 reads) in regions extending from 4189 bp to 4337 bp. This result indicated that the region between the E5 and L2 noncoding areas constituted the region of the highest read density (HR E5L2) throughout the HPV genome (Figure 2A).

From TCGA cohort, total read count for the HPV-16 genome was 166 026 reads. On analyzing average read counts for each nucleotide position of these 94 patients, it was found that the median read at each nucleotide position was 0.81. Significant spikes (>40 reads) were observed in regions between 4190 bp and 4261 bp at this instance as well. These observations were consistent with the findings from our “discovery set” cohort (Figure 2B).

Further, we attempted to define regions where the HPV-16 genome was most likely to appear, across the largest number of tumors, although still at a minimal detectable level in the “discovery set.” Using the 40 tumor samples in the cohort, minimum read counts were plotted for each nucleotide position. Of note, HR E5L2, which had the highest mean read count across the genome, was not universally present in all tumors in our HPV-16-related HNSCC dataset. Nevertheless, we were able to define a region between 1975 bp and 2195 bp in the E1 region “BR E1” for which every tumor analyzed had a read count of >4 (BR E1; Figure 2C). However, we were unable to confirm these observations by TCGA because of low-read coverage data across large parts of the HPV genome in this dataset. As noted,

our discovery cohort coverage (1019 median reads/nucleotide) greatly exceeded that of the TCGA dataset (0.81 median reads/nucleotide; Figure S1).

To identify additional regions with the presence of HPV-16 genome at a minimal detectable level in a large proportion of tumors, we omitted the sample with lowest read coverage (3.92 reads per nucleotide sample) and replotted the minimum read counts for each nucleotide position in the remaining 39 tumor samples. BR E1 remained as the region containing the presence of HPV DNA in all tumors. HR E5L2, along with the entire E6 region (BR E6), was also present in all 39 tumor samples in our original cohort (Figure 2D).

Further, we investigated the following regions for design of the primer/probe sets: BR E6, 172-373 bp; BR E1, 1975-2195 bp; HR E5L2, 4185-4334 bp.

To optimize HPV DNA primer/probe design and validation for HPV-HNSCC using Primer-BLAST, we constructed candidate primer/probe sets targeting the three regions noted above. We created six candidate primer/probe sets (“BR E1-1”-“BR E1-6”) for the BR E1 region (1975-2195 bp; Figure S2A and S2B), four candidate primer/probe sets (“HR E5L2-1”-“HR E5L2-4”) for the HR E5-L2 region (4185-4334 bp; Figure S2C and S2D), and five candidate primer/probe sets (“BR E6-1”-“BR E6-5”) for the BR E6 region (172-373 bp; Figure S2E and S2F). Positive PCR reactions were observed in the CaSki positive control. No additional primer/probe set (other than those used for minimum read coverage area in the area of regions E7, E2, and L1) was designed because these regions were too limited in size to design any additional appropriate primer probe combinations (Figure 2C).

In addition, we performed sensitivity analysis to identify the most sensitive primer/probe set from among the candidate sets. First, we confirmed the sensitivity of historical E6 and E7 primer/probe sets; both sets reacted consistently in triplicates to detect <30 copies of HPV-16 genome diluted in samples of normal genomic DNA (10 ng). However, both reacted inconsistently (one or two times of triplicate) to detect 10 copies of HPV-16 genome. Of the 15 designed primer/probe sets, six sets consistently amplified under 30 copies of HPV-16 genome. These candidate primer/probe sets included “BR E1-4,” “BR E1-5,” “BR E1-6,” “HR E5L2-4,” “BR E6-1,” and “BR E6-3.” Notably, with the “BR E1-5” primer/probe set, there was consistent detection of 10 copies of HPV-16 genome in triplicate.

To validate the function of primer/probe sets, we calculated the slope of delta Rn against cycle threshold for HPV-16 copy number detection from defined HPV copy number DNA derived from CaSki cell line. Of the six designed primer/probe sets, the values for “BR E1-4,” “BR E1-5,” and “HR E5L2-4” were between -3.58 and -3.10 (Table 1).

Furthermore, to determine whether these designed primer/probe sets also have a high sensitivity in a separate cohort, we performed qPCR in a validation cohort of 32 additional primary tumor samples with validated presence of HPV-16. We prepared serial of dilutions of tumor DNA in normal lymphocyte DNA from noncancer individuals and determined the threshold for each primer/probe set and defined amplification thresholds expressed in the number of tumor cells for each sample (Table S1). In the present study, sensitivity for the detection of 100 cells (660 pg DNA) was over 84% for all six designed primer/probe sets and the existing E6/E7 primer/probe sets. In addition, sensitivity for the detection of 10 cells

(66 pg DNA) was between 46% and 78% for all primer/probe sets. Furthermore, sensitivity for the detection of DNA from a single cell (6.6 pg DNA) was the highest (28.1%, 9/32) in “BR E1-5” primer/probe sets, followed by “BR E1-4” (21.9%, 7/32), “BR E1-6” (15.6%, 5/32), and “HR E5L2-4” (15.6%, 5/32) primer/probe sets (Table S2). According to the McNemar test, there was a significant difference between the E7 primer set and “BR E1-5” primer/probe sets with respect to single cell detection sensitivity (Figure 3A). Therefore, the sensitivity of the combination of “BR E1-5” and “HR E5L2-4” was higher than that of E6 and E7 primer/probe sets at each level of tumor dilution (Figure 3B).

### 3.2 | CLIA test development

These primer/probe sets were adapted for development of a CLIA test for HPV16 DNA detection in plasma and salivary rinses (Table 2). Test performance was improved, so that analytical sensitivity in salivary rinse spike in assays ranges from 2.5 to 5.0 DNA copies/mL with 100% specificity for the three primer/probe sets that are now part of the component assay, resulting in a fivefold improvement in analytic sensitivity that is in the range of sensitivity found in sequence-based liquid biopsy methods (Table 2, Figure S3). Additionally, the assays can be used quantitatively, with measuring ranges between 2.0 log<sub>10</sub> and 7.0 log<sub>10</sub> copies/mL. As a final validation, we performed assays on a blinded set of saliva rinse samples from 28 patients with HPV-HNSCC and 15 salivary rinses from non-cancer controls. We performed sensitivity analysis for E7, “BR E1-5,” and “HR E5L2-4” in saliva rinses. The highest sensitivity for the detection of HPV DNA in saliva rinse samples was observed with E7 (23/28; 82%), followed by “BR E1-5” (22/28; 79%) and “HR E5L2-4” (20/28; 71%). The sensitivity increased to 96% when all three sets were used (Table 3). No signal was noted in noncancer control patients, yielding a specificity of 100%.

## 4 | DISCUSSION

In the present study, we constructed optimized PCR primer/probe sets for detection of HPV-16 related HNSCC using a rational genomic-based approach. In addition, we evaluated the performance of these primer/probe sets for low HPV-16 copy number in tumor samples and salivary rinses.

There are several differences between our designed primer/probe sets and previously reported primer/probe sets. First, consensus primer sets “My09/MY11,” “GP5+/GP6+,” and others are not specific to HPV-16 as they amplify other high-risk HPV subtypes,<sup>15</sup> which limits the sensitivity for HPV-16. In addition, their amplicons are over 150 bp<sup>16</sup>; therefore, these are less suitable for screening fragmented DNA in plasma or formalin-fixed specimens because of DNA degradation in these compartments and conditions. Second, the primer/probe sets we have designed have a higher sensitivity than the previously reported HPV16 E6/E7 primer/probe sets.

We focused on the three regions that showed a combination of high copy number and near universal read presence based on WGS data from 39 cases. Kalu et al<sup>17</sup> have reported similar spikes of read coverage from HPV-related tumor cell line analyses, and Nulton et al<sup>18</sup> have also reported the detailed HPV genome status in HNSCC from TCGA database. They categorized the HPV genome status in HPV-HNSCC as episomal-only, integrated, or



hybrid viral-human episomal state. In addition, approximately three quarters of HPV-16-positive HNSCC may potentially replicate in an E1/E2-dependent manner. Their results support the existence of BR E1 location in all HPV-HNSCCs because of the essential role of E1 in replication. In certain saliva rinse samples, elevated high copy numbers of “HR E5L2-4” location were also detected, thereby indicating that relative E5L2 spikes are also found in saliva samples (Figure 1A,B). It remains unclear whether the universal presence of HR E5L2 spike and BR E1 in HPV-16 genome is unique to HPV-HNSCC or is a common feature in cervical carcinoma as well.

We designed the primer/probe “BR E1-5” that can detect 10 copies of HPV-16 DNA diluted in 10-ng DNA derived from normal lymphocytes; this also suggests that the “BR E1-5” primer/probe set is highly sensitive for the validation of tumor samples. The combination of “BR E1-5” and “HR E5L2-4” along with previously reported E7 primer/probe sets provides improved sensitivity compared with prior E6- and E7-based PCR assays. With further refinement and adaptation to a CLIA certified clinical laboratory environment, assay performance was improved such that we could reliably measure two copies of HPV DNA in an mL of plasma or salivary rinse for most primer/probe sets, with excellent quantitation over a large dynamic range, a sensitivity of 96% with 100% specificity.

HPV detection in saliva samples obtained from patients with HPV-HNSCC is a known biomarker for HPV-HNSCC. Wang et al<sup>19</sup> have reported that the single E7 primer/probe PCR method successfully detects HPV in only 40% of saliva samples obtained from patients with HPV-positive OPC. Qureishi et al<sup>20</sup> have reported successful detection of 72.2% HPV genome in their saliva samples obtained from HPV-HNSCC cases. In a study by Chai et al,<sup>21</sup> E6/E7 primer/probe sets showed 92.9% and 60% sensitivity for HPV-16 genome in saliva samples for DNA and RNA, respectively. The data presented indicate that our combination of primer/probe sets would improve the sensitivity of detection of HPV16 DNA in tissues and body fluids and improve assessment of disease burden in patients with HPV-HNSCC.

Digital PCR (dPCR), a relatively new technology, has shown superior performance<sup>22</sup> than qPCR, particularly in the context of low template numbers. The sensitivity of dPCR could be higher than that observed in our QPCR results, and we expect that these primer/probe sets may improve performance in a dPCR context. However, we chose to report performance in a modality consistent with prior reports of PCR-based detection of HPV sequences. In addition, we note that reliable detection of two copies of HPV DNA/mL is similar to that of dPCR and high-throughput sequencing methods.

These data show that we have a PCR-based CLIA certified assay for HPV16 DNA detection that is optimized for HPV-OPC that is economical (estimated cost <\$100), broadly and easily adoptable by clinical testing laboratories, and approaches the performance of sequencing-based liquid biopsy with excellent sensitivity and specificity in salivary rinses that exceeds performance in prior reports.

Of note, we did not perform validation of this assay in plasma samples obtained from patients with HPV-HNSCC and this limits the applicability of our study performance in the

plasma compartment. In addition, it would be helpful to test the performance of these improved assays in well-designed prospective clinical trials to further confirm performance.

In conclusion, using a sequence-based rational approach, we extracted a new optimized primer/probe set for the detection of HPV-16 in HNSCC. The results of this preclinical study are promising and call for further prospective validation.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

The authors thank the colleagues in the Moores Cancer Center at University of California, San Diego. This work was supported by National Cancer Institute [grant number UH2CA211396]. J.A. Califano received this grant. The project described was partially supported by the National Institutes of Health, Grant UL1TR001442 of CTSA.

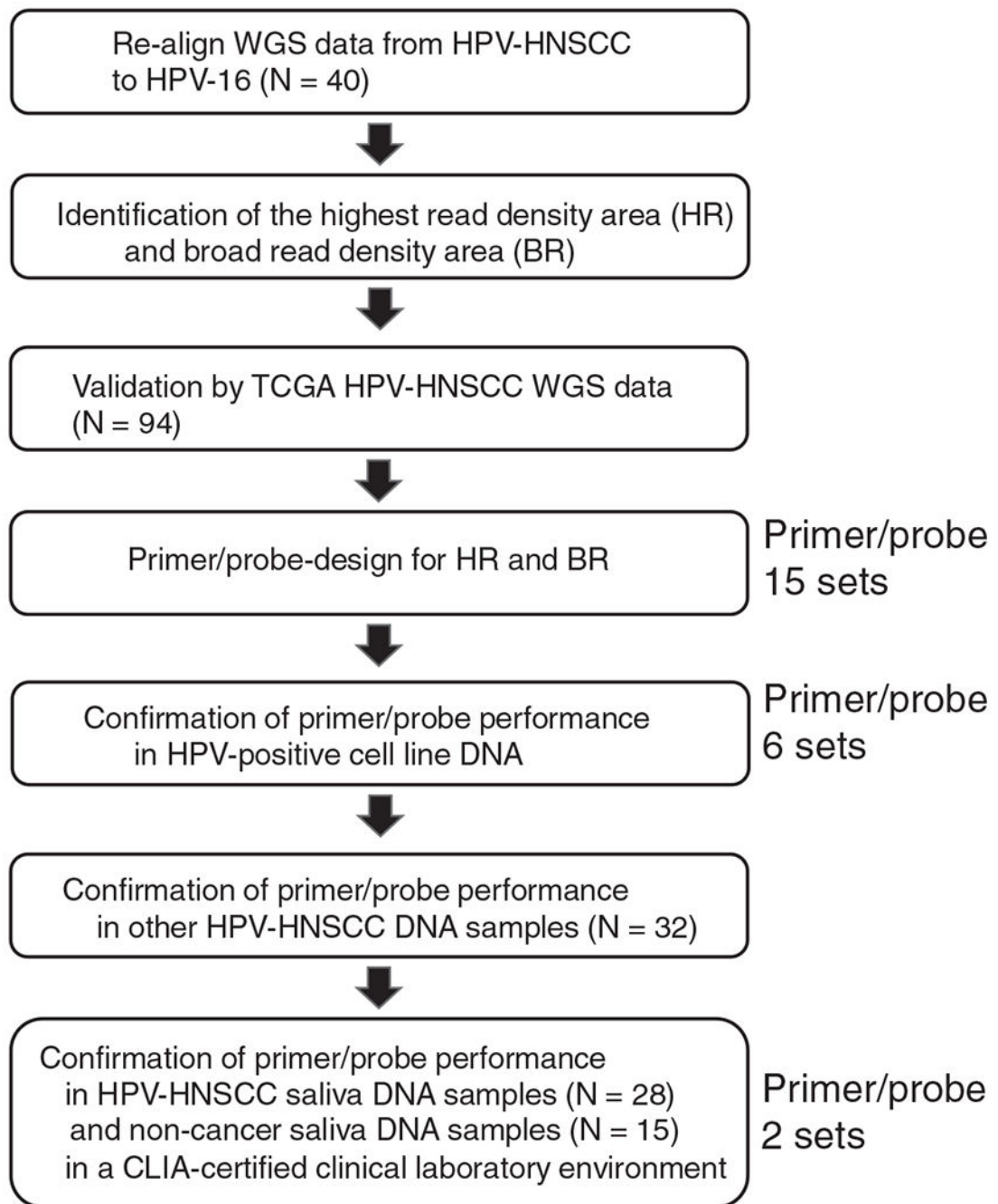
### Funding information

National Cancer Institute, Grant/Award Number: UH2CA211396

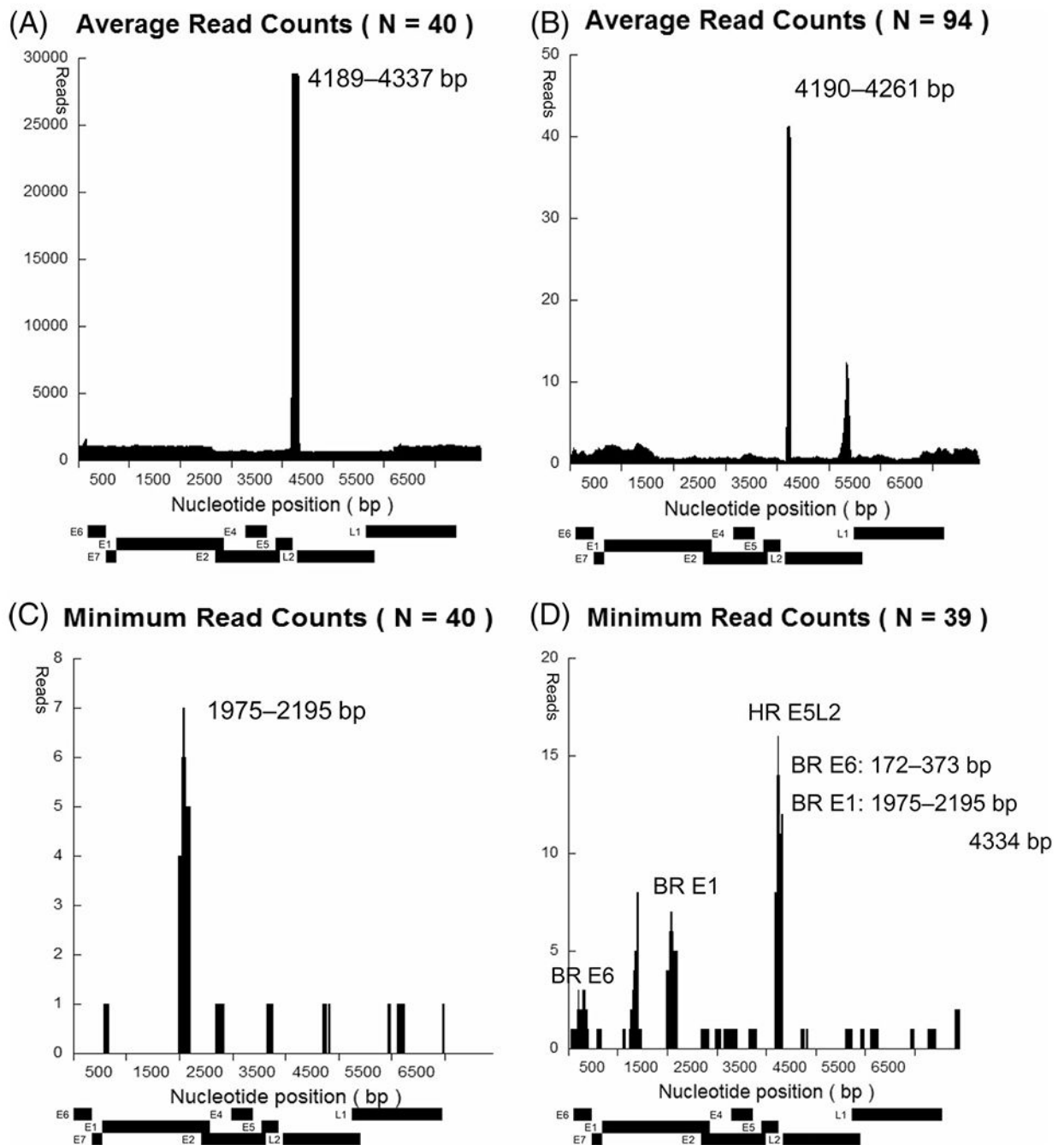
## REFERENCES

- Zandberg DP, Bhargava R, Badin S, Cullen KJ. The role of human papillomavirus in nongenital cancers. *CA Cancer J Clin.* 2013;63:57–81. [PubMed: 23258613]
- Mork J, Lie AK, Glattre E, et al. Human papillomavirus infection as a risk factor for squamous-cell carcinoma of the head and neck. *N Engl J Med.* 2001;344:1125–1131. [PubMed: 11297703]
- Ahn SM, Chan JY, Zhang Z, et al. Saliva and plasma quantitative polymerase chain reaction-based detection and surveillance of human papillomavirus-related head and neck cancer. *JAMA Otolaryngol Head Neck Surg.* 2014;140:846–854. [PubMed: 25078109]
- Hanna GJ, Sridharan V, Margalit DN, et al. Salivary and serum HPV antibody levels before and after definitive treatment in patients with oropharyngeal squamous cell carcinoma. *Cancer Biomark.* 2017;19:129–136. [PubMed: 28387659]
- Wasserman JK, Rourke R, Purgina B, et al. HPV DNA in saliva from patients with SCC of the head and neck is specific for p16-positive oropharyngeal tumours. *J Otolaryngol Head Neck Surg.* 2017;46:3. [PubMed: 28061890]
- Baay MF, Quint WG, Koudstaal J, et al. Comprehensive study of several general and type-specific primer pairs for detection of human papillomavirus DNA by PCR in paraffin-embedded cervical carcinomas. *J Clin Microbiol.* 1996;34:745–747. [PubMed: 8904451]
- Qureishi A, Ali M, Fraser L, Shah KA, Moller H, Winter S. Saliva testing for human papilloma virus in oropharyngeal squamous cell carcinoma: a diagnostic accuracy study. *Clin Otolaryngol.* 2018;43:151–157. [PubMed: 28620984]
- Hahne F, Ivanek R. Visualizing genomic data using Gviz and Bioconductor In: Mathé E, Davis S, eds. *Statistical Genomics: Methods and Protocols.* New York: Springer New York; 2016: 335–351.
- Guo T, Sakai A, Afsari B, et al. A novel functional splice variant of AKT3 defined by analysis of alternative splice expression in HPV-positive oropharyngeal cancers. *Cancer Res.* 2017;77:5248–5258. [PubMed: 28733453]
- Lawrence M, Huber W, Pagès H, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol.* 2013;9:e1003118. [PubMed: 23950696]
- Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25:2078–2079. [PubMed: 19505943]
- Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10:R25. [PubMed: 19261174]

13. Morgan M, Pagès H, Obenchain V, Hayden N. Rsamtools. Binary Alignment (BAM), FASTA, Variant Call (BCF), and Tabix File Import. R Package Version 1.32.0, 2018
14. Chardonnet Y, Lizard G, Chagnol MC, Schmitt D. Analytical methods for evaluation on whole cells of human papillomavirus infection. *Bull Cancer*. 1995;82:107–113. [PubMed: 10846526]
15. Zehbe I, Wilander E. Two consensus primer systems and nested polymerase chain reaction for human papillomavirus detection in cervical biopsies: a study of sensitivity. *Hum Pathol*. 1996;27:812–815. [PubMed: 8760015]
16. Karlsen F, Kalantari M, Jenkins A, et al. Use of multiple PCR primer sets for optimal detection of human papillomavirus. *J Clin Microbiol*. 1996;34:2095–2100. [PubMed: 8862564]
17. Kalu NN, Mazumdar T, Peng S, et al. Genomic characterization of human papillomavirus-positive and -negative human squamous cell cancer cell lines. *Oncotarget*. 2017;8:86369–86383. [PubMed: 29156801]
18. Nulton TJ, Olex AL, Dozmorov M, Morgan IM, Windle B. Analysis of the cancer genome atlas sequencing data reveals novel properties of the human papillomavirus 16 genome in head and neck squamous cell carcinoma. *Oncotarget*. 2017;8:17684–17699. [PubMed: 28187443]
19. Wang Y, Springer S, Mulvey CL, et al. Detection of somatic mutations and HPV in the saliva and plasma of patients with head and neck squamous cell carcinomas. *Sci Transl Med*. 2015;7:293ra104.
20. Qureishi A, Ali M, Fraser L, Shah KA, Moller H, Winter S. Saliva testing for human papilloma virus in oropharyngeal squamous cell carcinoma: a diagnostic accuracy study. *Clin Otolaryngol*. 2018;43:151–157. [PubMed: 28620984]
21. Chai RC, Lim Y, Frazer IH, et al. A pilot study to compare the detection of HPV-16 biomarkers in salivary oral rinses with tumour p16(INK4a) expression in head and neck squamous cell carcinoma patients. *BMC Cancer*. 2016;16:178. [PubMed: 26940728]
22. Biron VL, Kostiuik M, Isaac A, et al. Detection of human papillomavirus type 16 in oropharyngeal squamous cell carcinoma using droplet digital polymerase chain reaction. *Cancer*. 2016;122:1544–1551. [PubMed: 26989832]

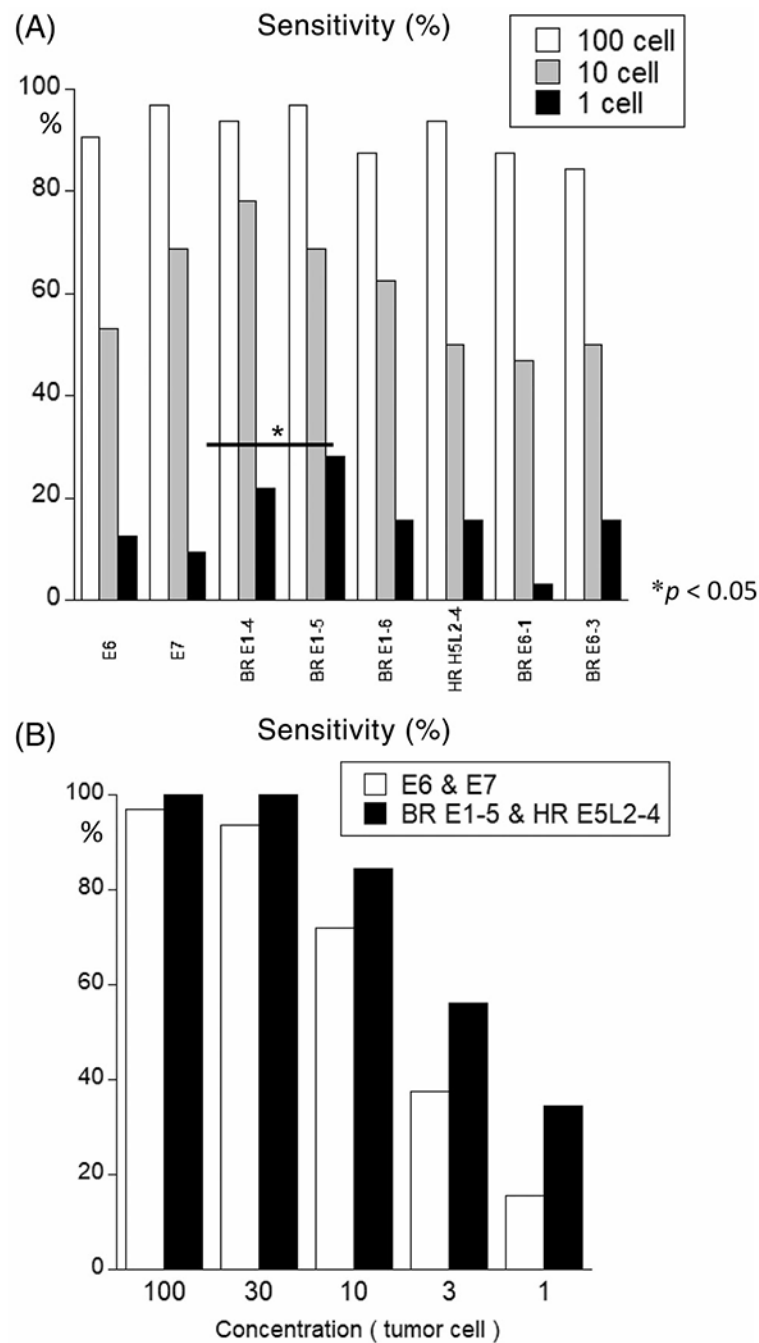
**FIGURE 1.**

The study pipeline. Schematic outline of the primer optimizing approach utilized in this study, which includes the highest read density area (HR) and broad read density area (BR) identified in our cohort, validation by the TCGA cohort, 15 primer/probe sets designed for HR and BR using Primer-BLAST, and PCR confirmation of primer/probe function in HPV-positive cell lines DNA, HPV-HNSCC DNA, and saliva DNA

**FIGURE 2.**

HPV-16 DNA reads in each nucleotide position. Plots are shown for our cohort (A, C, D) and TCGA cohort (B). A, Average read counts at each HPV-16 nucleotide position in 40 samples. The HR density (over 25 000 reads) is observed at the junction between E5 and L2 regions (nucleotide position: 4189-4227). B, Average read counts at each HPV-16 nucleotide position in TCGA samples (N = 94). The HR density (>40 reads) is observed at the junction between E5 and L2 regions (nucleotide position: 4190-4261). C, Minimum read counts at each HPV-16 nucleotide position in all 40 samples. E1 specific region (nucleotide position:

1975-2195) contained the minimum read density across all samples. D, After exclusion of the lowest read sample data, minimum read counts replotted in each HPV-16 nucleotide position. Besides E1 specific region (BR E1; nucleotide position: 1975-2195), E6 specific region (BR E6; nucleotide position: 172-373) and the junction between E5 and L2 regions (HR E5L2; 4185-4334) contained the minimum read density across these 39 samples

**FIGURE 3.**

Sensitivity of each primer/probe sets along with that of E6/E7 primer/probe sets against the 31 clinically known HPV-positive head and neck squamous cell carcinoma samples among DNA derived from healthy individual. A, Sensitivity of samples against each planned primer/probe sets along with E6/E7 primer/probe sets. To calculate the threshold, we prepared 32 clinically known HPV-positive head and neck squamous cell carcinoma anonymized samples that were serially diluted to 150, 30, 15, 6, 1.2, 0.6, and 0.3 cells (corresponding to 990 pg., 198 pg, 99 pg, 39.6 pg, 7.9 pg, 3.2 pg, and 1.6 pg of tumor

weight, respectively) among 10-ng normal human DNA. Each qPCR was studied thrice, and the lowest dilution series showing three consistent reactions was considered as the threshold. Sensitivity was calculated as follows: (a) white, 100 cells (corresponds to 660 pg of tumor weight), (b) gray, 10 cells (corresponds to 66 pg of tumor weight), and (c) black, 1 cell (corresponds to 6.6 pg of tumor weight). The “BR E1-5” primer/probe sets showed the highest sensitivity (28.1%) for one cell among these eight primer/probe sets. B, The sensitivity of the combination of “BR E1-5” and “HR E5L2-10” primer/probe sets was superior to that of E6 & E7 combination at all concentrations (100 cells, 30 cells, 10 cells, 3 cells, 1 cell). Experiments were performed in triplicates



**TABLE 1**

The slope, efficiency, and Y-intercept of the standard curve in each primer/probe sets

	<b>Slope</b>	<b>Efficiency</b>	<b>Y-Intercept</b>
E6	-3.802	83.254	43.787
E7	-3.560	90.948	40.749
BR E1-4	-3.496	93.212	40.332
BR E1-5	-3.274	102.022	39.435
BR E1-6	-3.723	85.6	42.346
HR E5L2-4	-3.560	90.944	48.739
BR E6-1	-3.647	88.03	43.562
BR E6-3	-3.740	85.079	42.147

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Test performance of the second-generation genomic-based assay for HPV16 DNA in HPV-OPC

TABLE 2

Test characteristics	Analyte	Primer-probe region		
		E7	HR E5L 2-4	BR E1-5
Limit of detection—Plasma (cp/mL)	pATCC HPV16	1.5	1.5	1.5
	SiHa	13	20	22
Limit of detection—Salivary rinse (cp/mL)	pATCC HPV16	3	2	3
	SiHa	23	19	12
Measuring range—Plasma (cp/mL)	pATCC HPV16	2.0-8.5 log10	2.0-8.5 log10	2.0-8.5 log10
	SiHa	2.0-5.0 log10	2.0-6.5 log10	2.0-6.5 log10
Measuring range—Salivary rinse(cp/mL)	pATCC HPV16	2.0-8.5 log10	2.0-8.5 log10	2.0-8.5 log10
	SiHa	2.0-6.5 log10	2.0-6.5 log10	
Precision—Plasma (% CV)	pATCC HPV16			
	2.0 E6 cp/mL	11%	28%	7%
	2.0 E4 cp/mL	8%	20%	13%
	2.0 E2 cp/mL	25%	24%	19%
	SiHa	5.0 E4 cp/mL, 10%	2.0 E5 cp/mL, 26%	2.0 E5 cp/mL, 24%
		1.0 E2 cp/mL, 11%	2.0 E3 cp/mL, 55%	2.0 E3 cp/mL, 23%
			2.0 E1 cp/mL, 65%	2.0 E1 cp/mL, 32%
Total precision—Salivary rinse (% CV)	pATCC HPV16			
	8.4 E6 cp/mL	35%	39%	54%
	8.4 E4 cp/mL	50%	64%	50%
	8.4 E2 cp/mL	53%	57%	40%
	SiHa			
	2.0 E5 cp/mL	31%	21%	24%
	2.0 E3 cp/mL	50%	33%	43%
	2.0 E1 cp/mL	62%	42%	65%
Accuracy—Plasma blind panel	Positive % agreement	100	100	100

Test characteristics	Primer-probe region		
	Analyte	E7	HR E5L 2-4 + BR E1-5
Negative % agreement		90	100
Total % agreement		97	100
Accuracy—Salivary rinses from HNSCC cases	N = 28 salivary rinse DNAs tested	<b>The Triple Test (E7 + HR E5L 2-4 + BR E1-5 results) detected HPV16 DNA in 27/28 (96%) of salivary rinses</b>	
Specificity	No positive results were obtained for any of the three primer/probe sets for HIV-1, EBV, Coxsackie A9, echovirus 11, adenovirus, influenza A virus, herpes simplex virus type 1, herpes simplex virus type 2, cytomegalovirus, BK virus, JC virus, human metapneumovirus, respiratory syncytial virus, human rhinovirus, HCV, HBV, parvovirus B19		
Reference range	All positive results are reportable		

**TABLE 3**

Copy numbers detected from 20 ng DNA derived from saliva samples in each primer/probe sets

<b>Saliva samples from HPV-HNSCC patients</b>			
<b>#</b>	<b>E7</b>	<b>BR E1-5</b>	<b>HR E5L2-4</b>
1	3E+05	866 000	322 000
2	50 000	118 000	52 500
3	48 500	47 600	25 800
4	46 500	104 000	33 200
5	14 700	N/A	N/A
6	7320	N/A	N/A
7	4950	9790	9670
8	1570	3590	2510
9	1020	2600	1590
10	981	N/A	N/A
11	521	N/A	N/A
12	155	127	86.7
13	140	290	53.1
14	82.7	114	31.1
15	62.6	160	49.7
16	41.2	121	41.3
17	14.1	67.4	15.8
18	9.69	15.8	2.61
19	8.17	61.1	26.9
20	6.5	30.9	13.4
21	4.02	25.7	6.54
22	2.52	2.68	1.16
23	0.986	1.82	N/A
24	N/A	14.4	N/A
25	N/A	2.76	N/A
26	N/A	N/A	2.15
27	N/A	2.17	1.19
28	N/A	N/A	N/A
<b>Saliva samples from nontumor patients</b>			
<b>#</b>	<b>E7</b>	<b>BR E1-5</b>	<b>HR E5L2-4</b>
1	N/A	N/A	N/A
2	N/A	N/A	N/A
3	N/A	N/A	N/A
4	N/A	N/A	N/A
5	N/A	N/A	N/A
6	N/A	N/A	N/A

<b>Saliva samples from HPV-HNSCC patients</b>			
<b>#</b>	<b>E7</b>	<b>BR E1-5</b>	<b>HR E5L2-4</b>
7	N/A	N/A	N/A
8	N/A	N/A	N/A
9	N/A	N/A	N/A
10	N/A	N/A	N/A
11	N/A	N/A	N/A
12	N/A	N/A	N/A
13	N/A	N/A	N/A
14	N/A	N/A	N/A
15	N/A	N/A	N/A

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript