



Published in final edited form as:

Mol Cancer Res. 2016 October ; 14(10): 941–952. doi:10.1158/1541-7786.MCR-16-0105.

Genomic Integration of High-Risk HPV Alters Gene Expression in Oropharyngeal Squamous Cell Carcinoma

Heather M. Walline¹, Christine M. Komarck², Jonathan B. McHugh³, Emily L. Bellile⁴, J. Chad Brenner², Mark E. Prince², Erin L. McKean², Douglas B. Chepeha², Gregory T. Wolf², Francis P. Worden⁵, Carol R. Bradford², and Thomas E. Carey²

¹Cancer Biology Program, Program in the Biomedical Sciences, Rackham Graduate School, University of Michigan Ann Arbor MI

²Department of Otolaryngology/Head and Neck Surgery, University of Michigan, Ann Arbor, MI

³Department of Pathology, University of Michigan, Ann Arbor, MI

⁴Biostatistics, University of Michigan Cancer Center, Ann Arbor, MI

⁵Department of Internal Medicine, Division of Clinical Oncology, University of Michigan, Ann Arbor, MI

Abstract

High-risk HPV (hrHPV) is the leading etiologic factor in oropharyngeal cancer. HPV-positive oropharyngeal tumors generally respond well to therapy, with complete recovery in approximately 80% of patients. However, it remains unclear why some patients are non-responsive to treatment, with 20% of patients recurring within 5 years. In this study, viral factors were examined for possible clues to differences in tumor behavior. Oropharynx tumors that responded well to therapy were compared to those that persisted and recurred. Viral oncogene alternate transcripts were assessed and cellular sites of viral integration were mapped and sequenced. Effects of integration on gene expression were assessed by transcript analysis at the integration sites. All of the tumors demonstrated active viral oncogenesis, indicated by expression of HPV E6 and E7 oncogenes and alternate E6 splicing. In the responsive tumors, HPV integration occurred exclusively in intergenic chromosome regions, except for one tumor with viral integration into TP63. Each recurrent tumor exhibited complex HPV integration patterns into cancer-associated genes, including: TNFRSF13B, SCN2A, SH2B1, UBE2V2, SMOC1, NFIA, and SEMA6D. Disrupted cellular transcripts were identified in the region of integration in four of the seven affected genes.

Implications—Integration of transcriptionally active hrHPV into cellular intergenic regions associates with tumor behavior by altering gene expression.

Keywords

HPV; Integration; HNSCC; Cancer; Oropharynx

Corresponding author: Thomas E. Carey, 1150 W. Medical Center Dr., 5311 Med Sci I, Ann Arbor, MI. 48109-5616. careyte@med.umich.edu; Phone: 734-764-4371; Fax: 734-464-0014.

The authors disclose no potential conflicts of interest.

INTRODUCTION

High-risk human papillomaviruses (hrHPV) are known factors in the etiology of head and neck squamous cell carcinoma, particularly in association with the increasing incidence of oropharynx cancers. Conventional treatment for these patients includes high dose radiation therapy often combined with concurrent chemotherapy. These treatments are associated with significant acute and long-term morbidity. In oropharyngeal tumors, hrHPV is associated with better prognosis, suggesting that hrHPV-positive tumors may be responsive to alternate therapies that are more tolerable than those currently used (1–5). However, a reduction in treatment intensity is hampered by our current inability to distinguish the most responsive HPV-positive oropharynx tumors from those that would fail if given reduced-intensity treatment or that fail to respond to current therapies.

Carcinogenesis in hrHPV-induced tumors is driven by sustained expression of viral E6 and E7 oncogenes, which is secondary to disruption of the viral E2 gene that regulates E6–E7 expression (6). The HPV16 E6 gene contains multiple splice sites, generating alternate E6*I-E7 and E6*II-E7 transcripts that have been linked to increased expression of E7, considered to be the more potent oncoprotein, at the expense of full length E6 (7, 8). The E6*I and E6*II alternate transcripts result from a single donor site at nucleotide (nt) 226 of the viral genome and two acceptor sites at nt 407 (E6*I) and at nt 526 (E6*II) (Figure 1A). Secondary carcinogenic mechanisms of viral integration could include disruption of tumor suppressor genes or upregulation of genes that promote cell-cycle progression. Integration of hrHPV into the host cellular genome has been reported to be associated with high E6 and E7 transcription and carcinogenic progression from cervical intraepithelial neoplasia (CIN) to invasive disease in many cervical cancer studies (6, 9, 10). Cellular sites of viral integration in cervical cancer are primarily into gene poor regions or chromosome common fragile sites (11–13), but there are studies that report viral integration into known genes in cervical cancer (14–17). Similarly, in head and neck cancers, viral integration into the host genome and into cellular genes is now being appreciated. A study by Partenov et al. that analyzed TCGA data from 35 HPV positive head and neck tumors noted frequent integration of HPV into regions of microhomology between the viral and host genomes. Furthermore, they observed that more than half of these integration events occurred into a known gene, and another subset (19%) occurred within 20 kb of a gene. They also found that the integration of HPV often altered the expression of cellular genes and was associated with focal amplifications; however, there was no significant association of HPV integration status with clinical outcome (18).

We and others have previously demonstrated transcriptionally active hrHPV integration into known cancer-related genes (those that are known to be involved in cancer pathways or that have been reported to have altered expression in one or more types of cancer) in HPV16-positive HNSCC cell lines (19–21). Partenov et al. (18) postulate that integration events affecting expression and function of cellular genes may be a secondary driver of HPV-positive head and neck tumors. Vojtechova, et al. (21) reported HPV integration, extrachromosomal or mixed integrated and extrachromosomal HPV in 14 fresh and 186 archival tumors and found HPV positivity, tumor size and lymph node positivity were associated with survival, but integration status was not. In this study we investigate viral

copy number, viral oncogene transcript production, sites of viral integration, and effects of viral integration on cellular gene transcripts across the integration sites in 10 patients who differed in tumor resolution after therapy.

MATERIALS AND METHODS

Tumor specimens

Oropharyngeal tumors were obtained from the Head and Neck Cancer SPORE Biorepository. We evaluated ten HPV-positive oropharyngeal tumor pre-treatment biopsies from patients who had also provided fresh frozen tumor tissue and written informed consent to investigate their tissue under a study approved by the Institutional Review Board for the University of Michigan medical school. Tumor information, patient gender, age, smoking status, year of diagnosis, treatment, and outcome are listed in Table 1. Areas in the paraffin tissue block enriched for tumor cells were identified and marked by a pathologist on a freshly cut H & E stained slide prior to tumor sampling. Genomic DNA was extracted from formalin-fixed, paraffin-embedded (FFPE) tumor cores using the DNeasy Blood and Tissue Kit (Qiagen) or from fresh-frozen tumor sections using a standard phenol extraction. Tumor tissue was microdissected for RNA from fresh-frozen tumor sections immediately following histological evaluation. Total RNA was isolated using the RNeasy Mini Kit with QIAzol (Qiagen), followed by on-column DNase treatment.

HPV genotyping and copy number analysis

hrHPV genotyping was performed on DNA from all tumors using the HPV PCR-MassArray assay (22, 23). Type-specific TaqMan quantitative PCR was used to determine HPV copies per cell, assessing both E6 and E7 amplicons, with a GAPDH assay as an endogenous two copy/cell endogenous reference control.

HPV E6 and E7 transcript analysis

HPV16 E6 and E7 transcripts were evaluated by reverse-transcription PCR (RT-PCR) with gel electrophoresis and TaqMan quantitative RT-PCR. To analyze the expression of HPV16 E6 and E7, transcript-specific assays were used that exclusively amplify each product: the intact, non-spliced, full-length E6–E7 transcript, the spliced E6*I-E7 transcript, and the spliced E6*II-E7 transcript, as illustrated in Figure 1A (Primer sets are listed in Supplemental Table s1). An assay for human endogenous GAPDH was included to verify the absence of contaminating genomic DNA. Quantitative RT-PCR was performed using similar transcript-specific TaqMan assays that individually interrogate each HPV E6 and E7 transcript: non-spliced full length E6, spliced E6*I, spliced E6*II, and E7 (Figure 1B) (Primer sequences are listed in Supplemental Table s2). A TaqMan quantitative assay for GAPDH was included as an endogenous control to calculate relative viral gene expression.

Detection of Integrated Papillomavirus Sequences-Polymerase Chain Reaction (DIPS-PCR)

Viral integration was evaluated using an adaptation of the DIPS-PCR method previously published (13). Genomic DNA from each tumor was subjected to Taq^α1 restriction enzyme digestion, producing fragmented DNA. There are approximately 1.5 million Taq^α1 restriction sites within the human cellular genome, but only one in the non-variant HPV16

genome, located in the E6 open reading frame (ORF) at nucleotide 505. Additional HPV16 Taq^α1 restriction sites have been described in HPV16 variants at positions 311 and 2608. Following restriction digest, a ligation reaction attached a double-strand adapter oligo (5'-CGCAACGTGTAAGTCTG-NH₂-3' annealed to 5'-GGGCCATCAGTCAGCAGTCGTAGCCGGATCCAGACTTACACGTTG-3') to the overhanging ends of each fragment. Linear amplification of the ligated fragments was performed using 11 viral-specific primers, generating amplicons that originate in the viral genome, extend into adjacent cellular sequence, and terminate at the end of the adapter. This was followed by a second, logarithmic, PCR using 11 nested viral primers with a reverse adapter-specific primer (Primers are listed in Supplemental Table s3). Thermocycling conditions used for linear and exponential PCR included 3 minute extension cycles, allowing limitation of amplicon size to 3kb or less, therefore excluding production of any of large (>3kb), episome-only fragments. PCR products were separated by gel electrophoresis.

Sequence analysis of HPV16 integration products

Viral-cellular fragments were distinguished from episomal virus fragments based on predicted viral-only amplicon sizes of 2750bp or larger (Supplemental Table s3). DIPS-PCR amplicons of approximately 2500bp or smaller were identified, the corresponding bands were excised, and the amplicons were purified and sequenced. Integration events into known cellular genes were confirmed by direct PCR and sequencing of the original tumor genomic DNA, using primers designed for each viral and cellular region.

Integration site transcript analysis

RT-PCR assays were designed to amplify viral-cellular fusion transcripts and cellular transcripts from tumor RNA in cases expected to be altered by confirmed viral integration into known cellular genes. Assays included virus-cellular fusion transcripts (although expected only in the single case where the integration into the cellular gene followed the same orientation as the virus) from HPV ORFs into cellular gene exons, cellular gene exon-exon transcripts spanning the integration site, and exon-exon or within-exon transcripts outside of the integration site region. All successfully amplified transcripts were sequenced for verification.

RESULTS

Patient material

Of 227 HPV16-positive oropharynx tumors, there were 19 with sufficient FFPE and fresh frozen tumor tissue for our study. Of these, tumor DNA and RNA of sufficient quantity and quality and patient follow-up of at least 2 years, were obtained from 10 patients, 5 with recurrent and 5 with responsive tumors. Patient information, tobacco and alcohol use, HPV copy number, year of diagnosis, tumor staging, treatment, and outcome are listed by patient in Table 1. Tumors were from patients treated with 1) surgery alone (1 patient), 2) surgery and chemotherapy with weekly cisplatin (40 mg/m²)(1 patient), or 3) concurrent chemotherapy with weekly carboplatin (AUC 1) and paclitaxel (30 mg/m²) and intensity-modulated radiation therapy (IMRT) to defined tumor targets to maximize tumor dose and minimize dose to normal tissue (8 patients). All patients had advanced stage III or IV

disease. Patients with recurrent tumors all had advanced stage disease (IVA or IVB) and large (T3 and T4) tumors. Of patients who remained free of disease, three had stage III, one had IVA and one had IVB disease, two with T1, two with T2, and one with T3 tumors. At the time of publication, all of the patients with responsive tumors were free of disease (38–56 (mean 48.6) months following diagnosis) and four of the five patients with recurrent tumors have died with survival time from diagnosis ranging from 11–75 (mean 29.5) months. The single surviving patient in the recurrent group is alive with disease at 109 months.

HPV genotyping and copy number analysis

All 10 tumors were positive for HPV16 and negative for all other 13 hrHPV types included in the PCR-MassArray assay. HPV16 copy number for the responsive tumors ranged from 16 to over 500 copies/cell with a mean viral copy number of 242.8. The recurrent tumors had overall lower copy number, ranging from 6 to 279 HPV16 copies per cell with a mean of 92.6 (Table 1). Although the mean copy number was lower in the nonresponsive tumors, the difference in the mean copy number values for responsive and recurrent tumors was not significant ($p=0.26$).

HPV E6 and E7 transcript analysis

HPV16 E6 and E7 transcripts were expressed in all ten tumors. The alternate E6*I transcript was the most highly expressed transcript in eight of the ten tumors. Only in responsive tumor 1971 and recurrent tumor 0732, was the E6*I less abundant than the full length E6 transcript. The E7 transcript was expressed at levels equivalent to the E6 (full length or E6*I) transcript in all responsive tumors except one (2148). In all of the non-responsive tumors, the E7 transcript was expressed at higher levels than any of the E6 transcripts (Figures 1 and 2).

Detection of Integrated Papillomavirus Sequences-Polymerase Chain Reaction (DIPS-PCR)

All ten HPV16-positive tumor specimens demonstrated viral integration. (Representative DIPS-PCR gels for responsive and recurrent tumors are shown in Supplemental Figures s1 and s2, respectively). A total of 207 hybrid viral-cellular amplicons were isolated and sequenced, 99 amplicons generated from the responsive tumors and 108 amplicons from the recurrent tumors. The numbers of viral-cellular amplicons generated and sequenced from each tumor are listed in Supplemental Table s4. All amplicons were analyzed and viral-host DNA fusions were identified by sequence and BLAST analysis. The sequence reads mapped to viral-only sequence, viral-cellular hybrids, or were unmapped due to poor sequence resolution. All identifiable integrations are reported, multiple amplicons from each tumor were of the same integration.

Analysis of integration events

Each integration into a cellular gene was confirmed on the undigested tumor genomic DNA by direct PCR and sequencing. Representations of the viral integration events are depicted in Figure 3; and summarized in Table 2, indicating the chromosome locus, cellular gene, and the region of integration. Based on the integration results from DIPS-PCR, transcript

analysis was used to investigate expression of the affected gene. Electrophoretic gel images of transcript amplicons are shown in Supplemental Figure s3, cellular gene transcript RT-PCR and sequencing results are summarized in Figure 4 and summarized in Supplemental Table s5.

Integration Events in Responsive tumors

Eleven of the twelve HPV integration events identified in the responsive tumors involved intergenic chromosome regions (Figure 3B–F). Tumor 1733 had an HPV E2 integration into a known chromosome fragile site in 2p16 (24), tumor 1979 had 3 intergenic integration events, HPV E2 into 9q21, HPV L1 into 16q11.2, and another L1 into 4q27; four intergenic integration events were identified in tumor 1804, HPV E1 into 6q16, HPV L2 into 10p11.1, HPV E5 into 16q11.2, and HPV E2 into 16q11.2; and tumor 2148 had a single integration of HPV L2 into 7p22 a known chromosome fragile site (24).

Only Responsive tumor 1971 had integration into a cellular gene. Of the three integrations in tumor 1971, two were intergenic, HPV E1 into chromosome fragile site 7p22.3 (24), L2 into 4p16.3, also a known chromosome fragile site (24), and one, HPV L1 into intron 4 of *TP63* in 3q28 (Table 2 and Figure 3E). This integration site is located within the region that codes for the DNA binding domain of this tumor suppressor protein. Upon transcript analysis of this integration event, we found that a fusion transcript between HPV L1 and *TP63* exon 4 was not produced (Figure 4A; Table s6). The transcript across *TP63* exons 4 and 5, spanning the viral integration site in intron 4, was produced, and the sequence was in-frame. Additionally, a transcript across *TP63* exons 5 and 6 (outside of the integration region) was generated and was spliced in-frame.

HPV integration into recurrent tumors

The majority of the HPV integrations in the recurrent tumors were viral integrations into cellular genes. Tumor 2049 had two integration events into cellular genes, the first involving a rearrangement of HPV E1 (a duplicated region of E1 was inserted into E1 upstream of the integration) into 8q11.21, at intron 1 of *UBE2V2*, which codes for ubiquitin-conjugating enzyme E2 variant 2 (Table 2; Figure 3G). A fusion transcript was generated between HPV E1 and *UBE2V2*. Sequence analysis of this fusion transcript revealed the entire *UBE2V2* exon1 fused to a portion of HPV L1 reading into nonsense sequence then into HPV E1 followed by a segment of chromosome 17q11.2 and the distal end of the transcript amplicon included the expected region of HPV E1 (Figure 4B; Table s6). However, a *UBE2V2* transcript across exons 1 and 2, spanning the integration site in intron 1, as well as the transcript outside of the integration region across exons 2 and 3 were produced and the sequences were spliced in frame, suggesting that these came from a different chromosome or that the transcript was incomplete. The protein product of *UBE2V2* mediates transcriptional activation of target genes, regulates cell cycle progression and cellular differentiation, and is involved in DNA repair and cell survival after DNA damage. Deregulation of *UBE2V2* expression has been reported to be associated with gastric cancer (25), and in ER-positive/HER2-negative breast cancer, *UBE2V2* was linked to poor prognosis (26).

The second integration event identified in tumor 2049 was HPV E1 into 14q24.1, at intron 1 of *SMOCl*, the gene for SPARC-related modular calcium binding 1 (Table 2; Figure 3G). The fusion transcript between HPV E1 and *SMOCl* was sequenced and contained *SMOCl* exon 1 linked to chromosome 3p23, followed by nonsense sequence (Figure 4B, Table s6). There were other transcripts generated across *SMOCl* exons 1 and 2 (spanning the intron 1 integration) and exons 3 and 4 (outside of the integration region), but the transcript sequences did not contain any homology to *SMOCl*, and were determined to be nonsense sequence. As there was no intact *SMOCl* transcript, it appears that SMOC1 was inactivated by the viral integration. *SMOCl* codes for a secreted protein localized to the basement membrane that is involved in cellular differentiation, and has been associated with brain cancer (27).

DIPS-PCR and sequencing revealed an HPV early gene rearrangement in recurrent tumor 0843, where the distal half of E6 was duplicated and joined within the E2 ORF (Figure 3H). Viral integration in tumor 0843 was identified from HPV L2 into 2q24.3 at intron 16 of *SCN2A* which codes for the voltage-gated type II sodium channel α subunit (Table 2). Transcript analysis of the HPV L2 integration that mapped to intron 16 of the cellular gene *SCN2A* demonstrated that no fusion transcript was created in tumor 0843 between HPV L2 and cellular *SCN2A* exon 17 (Figure 4C, Table s6). Transcript primers in exon 16 and exon 17 of *SCN2A* amplified a cDNA transcript generated across the integration site in intron 16, but the sequence analysis identified a portion of HPV L1 flanked on one side by the cellular gene for the ATP-binding cassette, sub-family A, member 12 (*ABCA12*) located on chromosome 2q34, and on the other side by an intergenic region of chromosome 1q32, indicating a very complex chromosome rearrangement involving multiple chromosomes. Furthermore, there was no transcript generated when *SCN2A* was queried downstream from the HPV L1 integration event, across exons 18 and 19, indicating that *SCN2A* was fully disrupted by this integration event. This integration takes place in the second helical transmembrane S6 region of the protein, which participates in a complex for action potential initiation and propagation in excitable cells, as well as proliferation, migration, and adhesion in non-excitable cells (28). It has been reported that differential expression of voltage-gated sodium channels is associated with the metastatic activity of multiple malignancies such as leukemia and prostate, breast, and lung cancer, and these ion channels are currently being investigated as targets for cancer therapies (28).

Two integration events were identified in tumor 2238, the first was comprised of a rearrangement within HPV, where the L2/L1 overlapping region was inserted into the E1 ORF and inserted into 1p31.3, at intron 9 of *NFIA*, which codes for nuclear factor I/A (Table 2, Figure 3I). No fusion transcript was generated between HPV L1 and *NFIA* (Figure 4D, Table s6). There was no transcript generated across *NFIA* exons 9 and 10, spanning the intron 9 integration, suggesting that the integration disrupted the normal *NFIA* transcript, but the transcript across exons 10 and 11, outside of the integration site, was produced and that sequence was in-frame. The NFIA protein product is a sequence-specific transcription factor that regulates numerous adenoviral and cellular genes, and is independently proficient in activating cellular transcription and replication. It was recently reported that an investigation of acute erythroid leukemia containing t(1;16)(p31;q24) uncovered a gene fusion between *NFIA/CBFA2T3*(29).

The second integration in tumor 2238 was HPV E2 into 15q21.1, at intron 4 of *SEMA6D*, the gene for semaphorin 6D (Table 2, Figure 3I). No fusion transcript was generated between HPV E2 and *SEMA6D*, or across *SEMA6D* exons 4 and 5, spanning the intron 4 integration site (Figure 4D, Table s6). The *SEMA6D* transcript across exons 5 and 6, outside of the integration site, was generated, but was found to be nonsense upon sequence analysis, which is consistent with disruption of *SEMA6D* gene expression. The product of *SEMA6D* is a transmembrane protein historically characterized as an axon guidance molecule, but has more recently been shown to participate in differentiation, organogenesis, and angiogenesis, mediated by Plexin-A1 as the major Sema6D-binding receptor (30, 31). Furthermore, it has been reported that the Sema6D/Plexin-A1 complex binds VEGFR-2 to mediate survival and anchorage-independent growth of tumor cells (31, 32).

Tumor 0732 had integration from HPV E2 into an intergenic region at 10p11.1, as well as HPV L2 into 17p11.2, inserting at intron 3 of *TNFRSF13B*, the gene coding for a member of the tumor necrosis factor receptor superfamily (Table 2, Figure 3J). Transcript analysis of this integration event revealed that no fusion transcript was generated between HPV L2 and *TNFRSF* exon 3 (Figure 4E, Table s6). The transcript across *TNFRSF* exons 3 and 4, spanning the viral integration site in intron 3, as well as the *TNFRSF* transcript across exons 4 and 5, outside of the integration site, was generated and the both sequences were in-frame. This viral integration occurs within the region that produces the extracellular topological domain of the receptor protein, which participates in immunity by interacting with a TNF ligand. *TNFRSF13B* induces B-cell maturation and differentiation and activates multiple transcription factors, including NFAT, AP1, and NF- κ B. It has been reported that hematological malignancies are induced by B-cell survival and aberrant proliferation caused by dysregulated signaling by TNFRSF family members (33), but how this pathway may be involved in HNSCC is not clear.

Tumor 1040 had integration of HPV L2 into an intergenic region of 10p11.1, as well as HPV L1 into 16p11.2, at intron 3 of *SH2B1*, the gene for Src Homology 2B (SH2B) adapter protein 1 (Table 2, Figure 3K). In recurrent tumor 1040, there was no fusion transcript generated between HPV L1 and cellular *SH2B* exon 3 (Figure 4F, Supplemental Table s6). There were, however, transcripts generated across *SH2B* exons 3 and 4 (spanning the intron 3 integration site), and within exon 5 (outside of the integration region), and both sequences were in-frame. This is a mediator protein for tyrosine kinase receptors, and is involved in Janus kinase (JAK) and receptor tyrosine kinase signaling pathways.

HPV integration into the intergenic chromosome region 16q11.2 was identified 3 times among the responsive tumors examined; once in tumor 1769 and in two different events in tumor 1804. A second intergenic region was involved in 3 integration events including responsive tumor 1804 and recurrent tumors 0732 and 1040, which all exhibited viral integration into chromosome 10p11.1. These parallels suggest that there may be sequence or structural similarities that increase the probability of viral integration into these intergenic regions.

To further assess the importance of these genes to HNSCC, we assessed the mutation, copy number and gene fusion status of each gene harboring an integrated copy of HPV using the

Oncomine database to assess all publicly available head and neck next generation sequencing data, including from the Head and Neck Cancer Genome Atlas project (HNSCC-TCGA). Importantly, each of the identified genes harbored a genomic alteration in at least one tissue sample (Supplemental Tables S6–9). For example, the most frequently altered gene amongst this set, *TP63*, was mutated in 12/380 (3.2%) (Detailed in Supplemental Table S7), amplified in 16/390 (4.1%) (Detailed in Supplemental Table S8), and rearranged in 1/302 (0.3%) (Detailed in Supplemental Table S9) of all head and neck samples (Summarized in Supplemental Table S6). Genomic amplification of the *TP63* gene correlated with increased RNA expression as it did for *UBE2V2*, but not *TNFRSF13B* where almost no reads supported the presence of *TNFRSF13B* expression (Supplemental Figure S4). Interestingly, as observed from the sample ID numbers, the genes were altered in a mutually exclusive manner, suggesting that they may individually function as unique cancer drivers or suppressors.

DISCUSSION

The incidence of HPV-positive oropharyngeal cancer is rising, and there remains a lack of understanding around factors that determine or influence tumor response to treatment (1–5, 34, 35). There is significant interest in reducing treatment intensity for patients with HPV-positive tumors, but a decrease would risk the possibility of under-treating some patients who are cured by intensive concurrent chemoRT regimens (3, 36). In addition, our studies and others have found a subset of patients with HPV-positive tumors non-responsive to concurrent chemoRT (1, 37). Thus, it is necessary to identify the differences between 1) tumors that fail current intensive multimodality treatments and require alternate therapies, 2) those that respond to current therapies but will not respond to reduced-intensity treatment, and 3) tumors that are highly likely to respond to reduced-intensity therapy with lower treatment morbidity. Based on our previous work and what is known from HPV in cervical cancer, we examined viral copy number, transcriptional activity, and viral integration of hrHPV in responsive and recurrent tumors to determine whether these factors might be useful as clinically relevant factors to predict response.

The ten tumors studied were positive for HPV16 and negative for all other high-risk HPV types assessed. HPV copy number was established for each tumor; the ranges of viral load values were similar for responsive and recurrent tumors (16–539 copies/cell for responsive tumors, 6–298 copies/cell for recurrent tumors). It is important to note that the values obtained for viral copy number may not be exact because the tumor DNA was extracted from tissue cores that may have contained normal cells. Nevertheless, the average viral copy number for the responsive tumors (242.8 copies/cell) was more than twice that of the recurrent tumors (92.6). While the small number of tumors and wide ranges of copy number values limit our ability to draw conclusions from this result, it does agree with our hypothesis based on earlier work (4) that less advanced tumors contain higher numbers HPV copies, perhaps representing multiple episomal copies of HPV, whereas more advanced cancers have fewer copies due to loss of episomes unless the extra copies represent integrated or extrachromosomal concatenated copies of the viral genome that may or may not contribute to cancer cell proliferation (38, 39).

All of the tumors demonstrated expression of the E6 and E7 oncogenes, indicating that both the responsive and recurrent tumors are HPV-driven, and that the virus is not an incidental passenger to an alternate carcinogenic mechanism. Four of the five tumors in each group (responsive and recurrent) exhibited the alternate E6*I as the most abundant E6 transcript. Only one tumor from each group had the full-length E6 transcript as the most abundant oncogene transcript. E7 transcripts were also very abundant in 9/10 tumors, only one had lower E7 expression than the E6*I transcript. The E6–E7 transcripts (Full length E6–E7, E6*I-E7, and E6*II-E7) are produced as polycistronic mRNAs derived from the first p97 promoter. The E6 oncoprotein is translated from the full length E6–E7 transcript, and E7 is translated from the spliced E6*I–E7 transcript (7, 8, 40). This suggests that the tumors with more abundant full length E6 transcripts would have higher levels of the E6 oncoprotein, while the tumors with more abundant E6*I transcripts would produce higher levels of the E7 oncoprotein (7, 8, 40).

All of the tumors evaluated exhibited HPV16 integration into the cellular genome. We suspect that in most cases integration occurs into intragenic regions that account for greater than 90% of the genome. While the number of viral integrations that we detected and identified varied among the tumors it is important to acknowledge that the DIPS-PCR method may not detect all integration events, and so the number of integration events detected by this method cannot be used as a prognostic measure as suggested in Liu, et al. (38). Integration events in both responsive and recurrent tumors demonstrated forward and reverse orientations into the cellular genome. This is consistent with the “looping” model of viral integration described recently by Akagi, et al. (41) or other rearrangement mechanisms. In each of the responsive tumors, at least one viral integration event was identified into intragenic regions known to be chromosome fragile sites (2p16, 7p22, and 4p16 in tumors 1733, 1971, and 2148) or into intergenic regions that were common to more than one tumor (16q11.2 and 10p11.1 in tumors 1769 and 1804). These results suggest that such integrations are not always entirely random and that viral integration is more likely to occur in gene-poor regions of the cellular genome that are already unstable or into regions with sequence or structural characteristics that favor integration. Zhang et al. (42) analyzed 14 publications and concluded that in cervical cancers HPV integration showed a preference for intragenic areas and transcriptionally active regions of the human chromosomes.

We postulate that the effect of intergenic viral integrations seen in both the responsive and recurrent tumors relates to the primary mechanism of HPV-driven carcinogenesis through disruption of the E6 and E7 transcriptional repressor E2, leading to upregulated and unopposed expression of the E6 and E7 oncogenes. Consistent with this idea, Partenov et al. (18) showed that the most common integration events among the HPV positive TCGA head and neck tumors occurred within the HPV E1 gene disrupting the polycistronic early region transcripts of E1 and E2. Viral integration into intergenic and chromosome fragile sites occurs frequently in cervical cancer (11–15), resulting in disruption of E2 and enhanced expression of E6 and E7 (9, 43, 44).

One of the responsive tumors also had integration into *TP63*. Interestingly, we and others (19, 20) observed HPV integration of HPV E2 into distal *TP63* in the HNSCC cell line UM-SCC-47. Viral integration into *TP63* has also been reported in cervical cancer, and

susceptibility for integration into this gene may be due to short segments of homologous sequence shared by HPV E1 and chromosome 3q28 within the *TP63* gene (14). *TP63* belongs to the p53 family of tumor suppressor genes, and is a sequence-specific DNA binding transcriptional repressor and activator. The p63 protein participates in TGF β and WNT signal transduction as well as differentiation and cell-cycle regulation (45), and as such, HPV integration into the *TP63* gene could cause disruption of these processes and may result in increased proliferation. In contrast to the case with HPV16 integration into UM-SCC-47, in which fusion transcripts were produced, we found no fusion transcript of reverse L1 into *TP63* in tumor 1971. In fact, we did find an intact *TP63* exon4-exon5 transcript. Thus, *TP63* may not have been disrupted by the HPV insertion in tumor 1971.

The integration analysis of the recurrent tumors revealed viral integrations into cellular genes in each case. Alterations of cellular genes as a consequence of viral integration may provide a second mechanism of oncogenesis in HNSCC. Cellular gene disruption caused by viral integration has been reported in rare cases of malignant transformation by low-risk HPV types that lack E6 and E7 oncogenic activity (46, 47). Thus, viral integration events into genic regions cause disruptions that are likely to alter cellular gene expression and mediate additional carcinogenic mechanisms, resulting in a more aggressive tumor phenotype. Partenov et al. (18) noted that the majority of the HPV breakpoints they analyzed colocalized with somatic copy number variants. In the current study, not only was integration into a cellular gene identified in every recurrent tumor, each of the genes disrupted by viral integration (*TNFRSF13B*, *UBE2V2*, *SCN2A*, *SH2B1*, *SMOC1*, *NFIA*, and *SEMA6D*) is involved in a pathway or mechanism that is related to cancer, or is differentially expressed in some cancers (sTables 6–8).

Transcription analysis across the integration sites showed that the effects of integration vary from tumor to tumor and in some cases are associated with very complex genomic rearrangements. In the five recurrent tumors, there were seven integration events into cellular genes. In three of the seven events (*TNFRSF13B*, *UBE2V2*, and *SH2B*), intact transcripts were detected both across the integration site and elsewhere in the gene. In all of these cases, the integration was intronic, and it is possible that the gene was spliced in-frame across the integration, eliminating the virus from the intron without disrupting the local transcript. However it is also possible in these cases that the intact transcripts come from a different copy of the gene. Whether this results in reduced gene dosage or if the full transcript is intact is not known. It is important to note that while intact *UBE2V2* transcripts were identified both across the integration site and elsewhere in the gene, a fusion transcript between HPV E1 and *UBE2V2* was generated. Sequence analysis of this fusion transcript demonstrates the severity of chromosome disorder in the cellular genome, as well as within the viral genome, with rearrangements resulting in production of a transcript containing exon 1 of *UBE2V2* (located on chromosome 8), HPV E1, nonsense sequence, HPV L1, and an intergenic region of chromosome 17q11.2.

In the remaining four recurrent tumors, viral integrations affecting *SCN2A*, *SMOC1*, *NFIA*, and *SEMA6D* resulted in disruption of gene transcription. In two of these cases, genomic instability was again demonstrated by complex chromosome rearrangements involving the viral integration site. In tumor 0843, a transcript spanning the integration site in *SCN2A* was

found to involve HPV L1, a portion of chromosome 2q34 (including part of the *ABCA12* gene), and an intergenic region of chromosome 1q32. In tumor 2049, a fusion transcript generated between HPV E1 and *SMOC1* included both exon 1 of *SMOC1* (located on chromosome 14) and a region of chromosome 3p23.

Viral integration into a gene does not inevitably cause loss of gene expression; we have shown that transcription of some or all of the gene can persist, possibly from additional, unaltered copies of the gene, or by splice removal of intron-integrated virus. Likewise, detection of portions of gene transcripts does not necessarily indicate that full-length transcripts are intact and that the appropriate protein is produced. In recurrent tumor 0843 a complex fusion transcript across the HPV insertion in exon 16 of *SCN2A* was produced, but an upstream transcript across exons 18–19 was not. We cannot eliminate the possibility that the gene transcripts that were found were incomplete, inactive, or otherwise defective. Full transcript analysis will help to provide a better understanding of the effect of viral integration on the gene expression and function.

Upregulation of cellular genes is a possible consequence of viral integration as well, either through disruption of transcriptional repression, generation of fusion transcripts, or other mechanisms. Viral integration can both result from genomic instability and contribute to genomic instability. Oncogenic activities of E6 and E7 promote instability through unregulated cellular proliferation and aberrant progression through the cell cycle bypassing important checkpoints for genomic integrity. These effects of HPV E6 and E7 allow the unstable cell to incur viral integration, resulting in increased viral oncoprotein expression (through disruption of E2). Furthermore, prolonged suppression of p53 function and associated aberrant checkpoint function can cause further chromosomal damage (48). The process of HPV integration into the cellular genome appears to be highly clastogenic, in some cases leading to additional dsDNA breaks, resulting in further rearrangement of the viral and cellular genomes. A recent publication supports this postulate (41). As tumor cells progressively acquire chromosome rearrangements from oncogenic processes, the genome becomes more disorganized and abnormal (48–50).

The limitations of the DIPS-PCR method restrict detection of cellular integration sites to those that have a restriction site in relatively close proximity. Viral rearrangement or convoluted integrations (multiple concatenated copies, alternate orientations) can reduce the sensitivity of the method and increase the complexity of analyzing the results as described in the rolling loop integration model by Akagi et al. (19). Nevertheless, identification of cellular genes affected by viral integration in all five recurrent tumors, together with detection of rearranged chromosomes, demonstrates the validity of the approach and the extent of cellular disorder present in the non-responsive and recurrent tumor cells.

Our evaluation of hrHPV transcriptional activity and integration in these tumors provides support for the hypothesis that viral integration analysis may identify additional virus-induced gene disruptions. Clearly much more work is needed before we fully understand the impact of integration sites on tumor behavior. Eventually, locating cellular genes with viral integration and assessing subsequent alterations in cellular expression may be included in a schema of factors for assessing patients for reduced, alternate or targeted therapy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial Support for each author: Supported by NIH-NCI Head and Neck SPORE P50 CA097248, NIH-NIDCR R01 DE019126, NIH-NCI P30 Cancer Center Support Grant P30 CA46592, and NIH-NIDCD P30 DC05188. HMW supported by the Cancer Biology Training Grant NIH-NCI T32 CA09676, and the Eleanor Lewis Scholarship through the University of Michigan Rackham Graduate School.

The authors acknowledge the assistance of the University of Michigan DNA core, especially Dr. Robert Lyons and Ellen Pedersen. We also thank the members of the University of Michigan Head and Neck SPORE team, as well as our nurses, clinical colleagues, including Drs. Vasu Divi, Lance Oxford, Theodoros Teknos, who are now at other institutions. Most of all we thank our patients who so graciously allowed us to study their tissues and clinical information.

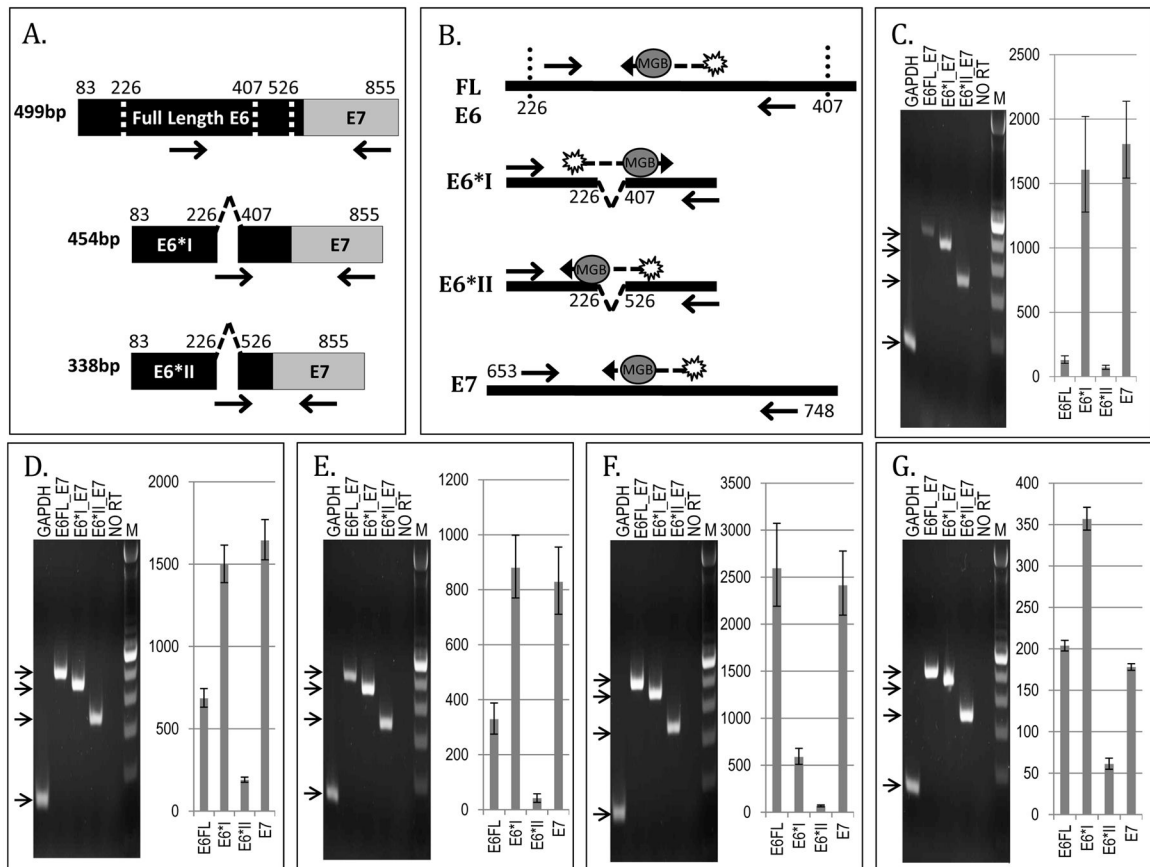
References

- Fakhry C, Gillison ML. Clinical implications of human papillomavirus in head and neck cancers. *J Clin Oncol.* 2006; 24(17):2606–11. Epub 2006/06/10 24/17/2606 [pii]. DOI: 10.1200/JCO.2006.06.1291 [PubMed: 16763272]
- Kumar B, Cordell KG, Lee JS, Worden FP, Prince ME, Tran HH, et al. EGFR, p16, HPV Titer, Bcl-xL and p53, sex, and smoking as indicators of response to therapy and survival in oropharyngeal cancer. *J Clin Oncol.* 2008; 26(19):3128–37. Epub 2008/05/14 JCO.2007.12.7662 [pii]. DOI: 10.1200/JCO.2007.12.7662 [PubMed: 18474878]
- Maxwell JH, Kumar B, Feng FY, Worden FP, Lee JS, Eisbruch A, et al. Tobacco use in human papillomavirus-positive advanced oropharynx cancer patients related to increased risk of distant metastases and tumor recurrence. *Clin Cancer Res.* 2010; 16(4):1226–35. Epub 2010/02/11 1078-0432.CCR-09-2350 [pii]. DOI: 10.1158/1078-0432.CCR-09-2350 [PubMed: 20145161]
- Worden FP, Kumar B, Lee JS, Wolf GT, Cordell KG, Taylor JM, et al. Chemoselection as a strategy for organ preservation in advanced oropharynx cancer: response and survival positively associated with HPV16 copy number. *J Clin Oncol.* 2008; 26(19):3138–46. Epub 2008/05/14 JCO.2007.12.7597 [pii]. DOI: 10.1200/JCO.2007.12.7597 [PubMed: 18474879]
- Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tan PF, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med.* 2010; 363(1):24–35. Epub 2010/06/10 NEJMoa0912217 [pii]. DOI: 10.1056/NEJMoa0912217 [PubMed: 20530316]
- Ho CM, Lee BH, Chang SF, Chien TY, Huang SH, Yan CC, et al. Integration of human papillomavirus correlates with high levels of viral oncogene transcripts in cervical carcinogenesis. *Virus Res.* 2011; 161(2):124–30. Epub 2011/07/23 S0168-1702(11)00235-8 [pii]. DOI: 10.1016/j.virusres.2011.06.012 [PubMed: 21777630]
- Tang S, Tao M, McCoy JP Jr, Zheng ZM. The E7 oncoprotein is translated from spliced E6*I transcripts in high-risk human papillomavirus type 16- or type 18-positive cervical cancer cell lines via translation reinitiation. *J Virol.* 2006; 80(9):4249–63. Epub 2006/04/14 80/9/4249 [pii]. DOI: 10.1128/JVI.80.9.4249-4263.2006 [PubMed: 16611884]
- Belaguli NS, Pater MM, Pater A. Splice sites of human papillomavirus type 16 E6 gene or heterologous gene required for transformation by E7 and accumulation of E7 RNA. *J Med Virol.* 1995; 47(4):445–53. [PubMed: 8636718]
- Cricca M, Morselli-Labate AM, Venturoli S, Ambretti S, Gentilomi GA, Gallinella G, et al. Viral DNA load, physical status and E2/E6 ratio as markers to grade HPV16 positive women for high-grade cervical lesions. *Gynecol Oncol.* 2007; 106(3):549–57. Epub 2007/06/15 S0090-8258(07)00339-3 [pii]. DOI: 10.1016/j.ygyno.2007.05.004 [PubMed: 17568661]
- Li W, Wang W, Si M, Han L, Gao Q, Luo A, et al. The physical state of HPV16 infection and its clinical significance in cancer precursor lesion and cervical carcinoma. *J Cancer Res Clin Oncol.* 2008; 134(12):1355–61. Epub 2008/05/15. DOI: 10.1007/s00432-008-0413-3 [PubMed: 18478264]

11. Ragin CC, Reshmi SC, Gollin SM. Mapping and analysis of HPV16 integration sites in a head and neck cancer cell line. *Int J Cancer*. 2004; 110(5):701–9. Epub 2004/05/18. DOI: 10.1002/ijc.20193 [PubMed: 15146560]
12. Wentzensen N, Vinokurova S, von Knebel Doeberitz M. Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. *Cancer Res*. 2004; 64(11):3878–84. Epub 2004/06/03. DOI: 10.1158/0008-5472.CAN-04-000964/11/3878 [PubMed: 15172997]
13. Luft F, Klaes R, Nees M, Durst M, Heilmann V, Melsheimer P, et al. Detection of integrated papillomavirus sequences by ligation-mediated PCR (DIPS-PCR) and molecular characterization in cervical cancer cells. *Int J Cancer*. 2001; 92(1):9–17. Epub 2001/04/03. DOI: 10.1002/1097-0215(20010401)92:1<9::AID-IJC1144>3.0.CO;2-8 [pii] [PubMed: 11279600]
14. Schmitz M, Driesch C, Jansen L, Runnebaum IB, Durst M. Non-random integration of the HPV genome in cervical cancer. *PLoS One*. 2012; 7(6):e39632. Epub 2012/07/05. PONE-D-12-09523 [pii]. doi: 10.1371/journal.pone.0039632 [PubMed: 22761851]
15. Klimov E, Vinokourova S, Mojsjak E, Rakhmanaliev E, Kobseva V, Laimins L, et al. Human papilloma viruses and cervical tumours: mapping of integration sites and analysis of adjacent cellular sequences. *BMC Cancer*. 2002; 2:24. Epub 2002/10/16. [PubMed: 12377107]
16. Xu B, Chotewutmontri S, Wolf S, Klos U, Schmitz M, Dürst M, et al. Multiplex Identification of Human Papillomavirus 16 DNA Integration Sites in Cervical Carcinomas. *PLoS ONE*. 2013; 8(6):e66693. doi: 10.1371/journal.pone.0066693 [PubMed: 23824673]
17. Ojesina AI, Lichtenstein L, Freeman SS, Peadarallu CS, Imaz-Rosshandler I, Pugh TJ, et al. Landscape of genomic alterations in cervical carcinomas. *Nature*. 2014; 506(7488):371–5. DOI: 10.1038/nature12881 [PubMed: 24390348]
18. Parfenov M, Peadarallu CS, Gehlenborg N, Freeman SS, Danilova L, Bristow CA, et al. Characterization of HPV and host genome interactions in primary head and neck cancers. *P Natl Acad Sci USA*. 2014; 111(43):15544–9.
19. Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, et al. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res*. 2014; 24(2):185–99. DOI: 10.1101/gr.164806.113 [PubMed: 24201445]
20. Olthof NC, Huebbers CU, Kolligs J, Henfling M, Ramaekers FC, Cornet I, et al. Viral load, gene expression and mapping of viral integration sites in HPV16-associated HNSCC cell lines. *Int J Cancer*. 2015; 136(5):E207–18. DOI: 10.1002/ijc.29112 [PubMed: 25082736]
21. Vojtechova Z, Sabol I, Salakova M, Turek L, Grega M, Smahelova J, et al. Analysis of the integration of human papillomaviruses in head and neck tumours in relation to patients' prognosis. *Int J Cancer*. 2016; 138(2):386–95. DOI: 10.1002/ijc.29712 [PubMed: 26239888]
22. Tang AL, Hauff SJ, Owen JH, Graham MP, Czerwinski MJ, Park JJ, et al. UM-SCC-104: a new human papillomavirus-16-positive cancer stem cell-containing head and neck squamous cell carcinoma cell line. *Head Neck*. 2012; 34(10):1480–91. Epub 2011/12/14. DOI: 10.1002/hed.21962 [PubMed: 22162267]
23. Walline HM, Komarck CM, McHugh JB, Byrd SA, Spector ME, Hauff SJ, et al. High-risk human papillomavirus detection in oropharyngeal, nasopharyngeal, and, oral cavity cancers: Comparison of multiple methods. *JAMA Otolaryngology*. 2013
24. Durkin SG, Glover TW. Chromosome fragile sites. *Annual review of genetics*. 2007; 41:169–92. DOI: 10.1146/annurev.genet.41.042007.165900
25. Cheng L, Wang P, Yang S, Yang Y, Zhang Q, Zhang W, et al. Identification of genes with a correlation between copy number and expression in gastric cancer. *BMC medical genomics*. 2012; 5:14. doi: 10.1186/1755-8794-5-14 [PubMed: 22559327]
26. Santarpia L, Iwamoto T, Di Leo A, Hayashi N, Bottai G, Stampfer M, et al. DNA Repair Gene Patterns as Prognostic and Predictive Factors in Molecular Breast Cancer Subtypes. *The oncologist*. 2013; doi: 10.1634/theoncologist.2013-0163
27. Brellier F, Ruggiero S, Zwolanek D, Martina E, Hess D, Brown-Luedi M, et al. SMOC1 is a tenascin-C interacting protein over-expressed in brain tumors. *Matrix biology : journal of the International Society for Matrix Biology*. 2011; 30(3):225–33. DOI: 10.1016/j.matbio.2011.02.001 [PubMed: 21349332]

28. Li M, Xiong ZG. Ion channels as targets for cancer therapy. *International journal of physiology, pathophysiology and pharmacology*. 2011; 3(2):156–66.
29. Micci F, Thorsen J, Panagopoulos I, Nyquist KB, Zeller B, Tierens A, et al. High-throughput sequencing identifies an NFIA/CBFA2T3 fusion gene in acute erythroid leukemia with t(1;16)(p31;q24). *Leukemia*. 2013; 27(4):980–2. DOI: 10.1038/leu.2012.266 [PubMed: 23032695]
30. Toyofuku T, Zhang H, Kumanogoh A, Takegahara N, Suto F, Kamei J, et al. Dual roles of *Sema6D* in cardiac morphogenesis through region-specific association of its receptor, *Plexin-A1*, with off-track and vascular endothelial growth factor receptor type 2. *Genes Dev*. 2004; 18(4):435–47. DOI: 10.1101/gad.1167304 [PubMed: 14977921]
31. Gu C, Giraud E. The role of semaphorins and their receptors in vascular development and cancer. *Exp Cell Res*. 2013; 319(9):1306–16. DOI: 10.1016/j.yexcr.2013.02.003 [PubMed: 23422037]
32. Catalano A, Lazzarini R, Di Nuzzo S, Orciari S, Procopio A. The *plexin-A1* receptor activates vascular endothelial growth factor-receptor 2 and nuclear factor- κ B to mediate survival and anchorage-independent growth of malignant mesothelioma cells. *Cancer Res*. 2009; 69(4):1485–93. DOI: 10.1158/0008-5472.CAN-08-3659 [PubMed: 19176370]
33. Rickert RC, Jellusova J, Miletic AV. Signaling by the tumor necrosis factor receptor superfamily in B-cell biology and disease. *Immunological reviews*. 2011; 244(1):115–33. DOI: 10.1111/j.1600-065X.2011.01067.x [PubMed: 22017435]
34. Kumar B, Cordell KG, Lee JS, Prince ME, Tran HH, Wolf GT, et al. Response to therapy and outcomes in oropharyngeal cancer are associated with biomarkers including human papillomavirus, epidermal growth factor receptor, gender, and smoking. *Int J Radiat Oncol Biol Phys*. 2007; 69(2 Suppl):S109–11. Epub 2007/10/18 S0360-3016(07)00965-0 [pii]. DOI: 10.1016/j.ijrobp.2007.05.072 [PubMed: 17848274]
35. Smith EM, Wang D, Kim Y, Rubenstein LM, Lee JH, Haugen TH, et al. P16INK4a expression, human papillomavirus, and survival in head and neck cancer. *Oral Oncol*. 2008; 44(2):133–42. Epub 2007/03/16 S1368-8375(07)00031-0 [pii]. DOI: 10.1016/j.oraloncology.2007.01.010 [PubMed: 17360226]
36. Bonner JA, Harari PM, Giralt J, Cohen RB, Jones CU, Sur RK, et al. Radiotherapy plus cetuximab for locoregionally advanced head and neck cancer: 5-year survival data from a phase 3 randomised trial, and relation between cetuximab-induced rash and survival. *Lancet Oncol*. 2010; 11(1):21–8. Epub 2009/11/10. S1470-2045(09)70311-0 [pii]. DOI: 10.1016/S1470-2045(09)70311-0 [PubMed: 19897418]
37. Gillison ML, D'Souza G, Westra W, Sugar E, Xiao W, Begum S, et al. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. *J Natl Cancer Inst*. 2008; 100(6):407–20. Epub 2008/03/13 djn025 [pii]. DOI: 10.1093/jnci/djn025 [PubMed: 18334711]
38. Liu Y, Lu Z, Xu R, Ke Y. Comprehensive mapping of the human papillomavirus (HPV) DNA integration sites in cervical carcinomas by HPV capture technology. *Oncotarget*. 2015; doi: 10.18632/oncotarget.6809
39. Orav M, Henno L, Isok-Paas H, Geimanen J, Ustav M, Ustav E. Recombination-dependent oligomerization of human papillomavirus genomes upon transient DNA replication. *J Virol*. 2013; 87(22):12051–68. DOI: 10.1128/JVI.01798-13 [PubMed: 23986589]
40. Zheng ZM, Tao M, Yamanegi K, Bodaghi S, Xiao W. Splicing of a cap-proximal human Papillomavirus 16 E6E7 intron promotes E7 expression, but can be restrained by distance of the intron from its RNA 5' cap. *J Mol Biol*. 2004; 337(5):1091–108. DOI: 10.1016/j.jmb.2004.02.023 [PubMed: 15046980]
41. Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, et al. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res*. 2013; Epub 2013/11/10 gr.164806.113 [pii]. doi: 10.1101/gr.164806.113
42. Zhang R, Shen C, Zhao L, Wang J, McCrae M, Chen X, et al. Dysregulation of host cellular genes targeted by human papillomavirus (HPV) integration contributes to HPV-related cervical carcinogenesis. *Int J Cancer*. 2016; 138(5):1163–74. DOI: 10.1002/ijc.29872 [PubMed: 26417997]
43. Cricca M, Venturoli S, Leo E, Costa S, Musiani M, Zerbini M. Molecular analysis of HPV 16 E6I/E6II spliced mRNAs and correlation with the viral physical state and the grade of the cervical

- lesion. *J Med Virol*. 2009; 81(7):1276–82. Epub 2009/05/29. DOI: 10.1002/jmv.21496 [PubMed: 19475620]
44. Romanczuk H, Howley PM. Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. *Proc Natl Acad Sci U S A*. 1992; 89(7):3159–63. Epub 1992/04/01. [PubMed: 1313584]
45. Pozzi S, Zambelli F, Merico D, Pavesi G, Robert A, Maltere P, et al. Transcriptional network of p63 in human keratinocytes. *PLoS One*. 2009; 4(3):e5008. Epub 2009/04/25. doi: 10.1371/journal.pone.0005008 [PubMed: 19390658]
46. Huebbers CU, Preuss SF, Kolligs J, Vent J, Stenner M, Wieland U, et al. Integration of HPV6 and downregulation of AKR1C3 expression mark malignant transformation in a patient with juvenile-onset laryngeal papillomatosis. *PLoS One*. 2013; 8(2):e57207. Epub 2013/02/26 PONE-D-12-19430 [pii]. doi: 10.1371/journal.pone.0057207 [PubMed: 23437342]
47. Kahn T, Turazza E, Ojeda R, Bercovich A, Stremlau A, Lichter P, et al. Integration of human papillomavirus type 6a DNA in a tonsillar carcinoma: chromosomal localization and nucleotide sequence of the genomic target region. *Cancer Res*. 1994; 54(5):1305–12. Epub 1994/03/01. [PubMed: 8118820]
48. Pett MR, Alazawi WO, Roberts I, Downen S, Smith DI, Stanley MA, et al. Acquisition of high-level chromosomal instability is associated with integration of human papillomavirus type 16 in cervical keratinocytes. *Cancer Res*. 2004; 64(4):1359–68. [PubMed: 14973079]
49. Korzeniewski N, Spardy N, Duensing A, Duensing S. Genomic instability and cancer: lessons learned from human papillomaviruses. *Cancer Lett*. 2011; 305(2):113–22. DOI: 10.1016/j.canlet.2010.10.013 [PubMed: 21075512]
50. Boccardo E. HPV-mediated genome instability: at the roots of cervical carcinogenesis. *Cytogenetic and genome research*. 2010; 128(1–3):57–65. DOI: 10.1159/000290657 [PubMed: 20389033]

**Figure 1.**

HPV Oncogene Transcript-Specific Quantitative RT-PCR and E6-E7 RT-PCR in HPV16-Positive Responsive Tumors. Bar graphs represent TaqMan quantitative PCR relative expression, and electrophoretic gel images represent E6-E7 RT-PCR. Panel A. RT-PCR Strategy for Transcript-Specific E6-E7 Oncogene Evaluation. Panel B. TaqMan Quantitative RT-PCR Strategy for Transcript-Specific E6 Oncogene Evaluation. Primers within the splice region or across splice junctions allow for exclusive amplification of full length E6 or alternate E6 transcripts. Panel C. 1733, Panel D. 1769, Panel E. 1804, Panel F. 1971, Panel G. 2148. Arrows indicate sizes of expected amplicon bands: HPVE6 FullLength_E7=499bp, HPVE6*I_E7= 454bp, and HPVE6*II_E7= 338bp. NO RT=no reverse transcriptase negative control. M= 100bp ladder.

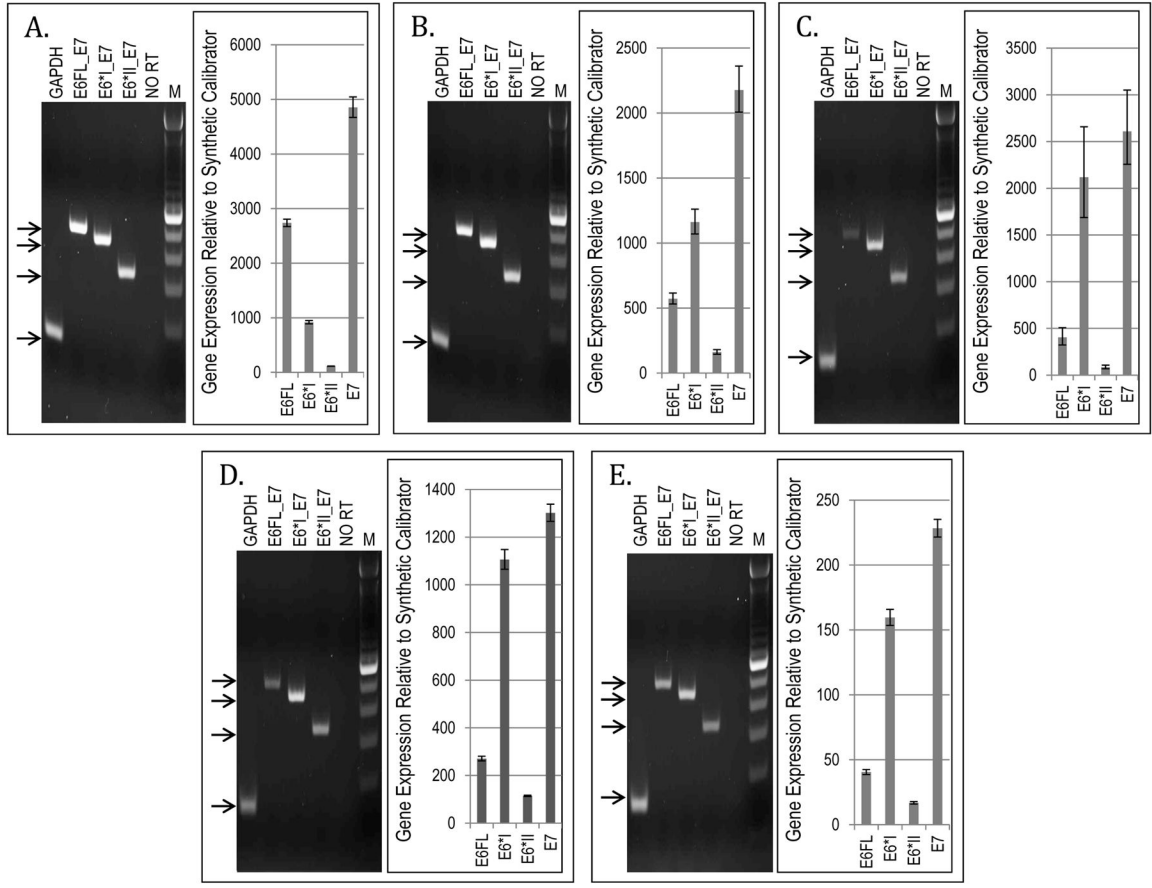


Figure 2. HPV Oncogene Transcript-Specific Quantitative RT-PCR and E6–E7 RT-PCR in HPV16-Positive Recurrent Tumors. Bar graphs represent TaqMan quantitative PCR relative expression, and electrophoretic gel images represent E6–E7 RT-PCR. Panel A. 0732, Panel B. 0843, Panel C. 1040, Panel D. 2049, Panel E. 2238. Arrows indicate sizes of expected amplicon bands: HPVE6 FullLength_E7=499bp, HPVE6*I_E7= 454bp, and HPVE6*II_E7= 338bp. NO RT=no reverse transcriptase negative control. M= 100bp ladder.

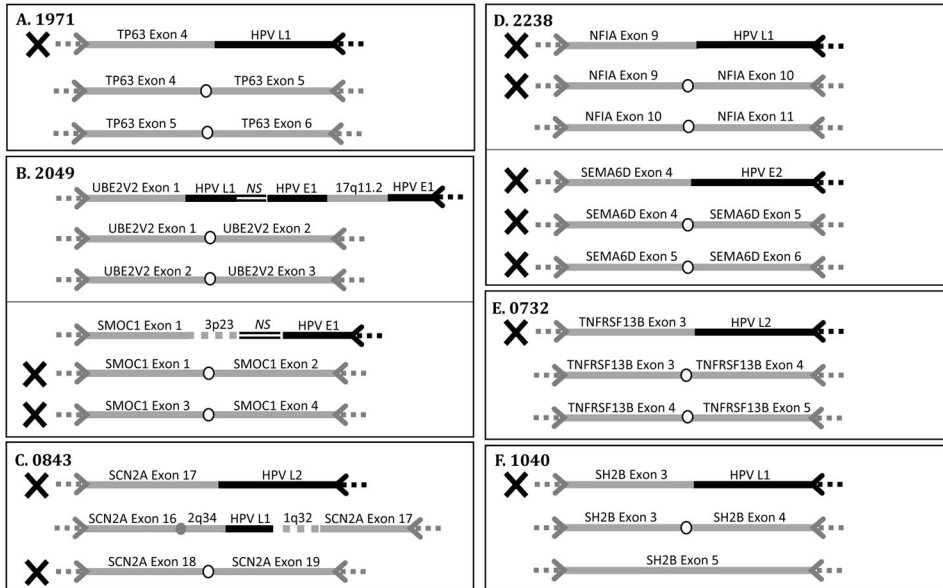


Figure 4. Schematic Representation of Transcript Analysis of Integration Events in HPV16-Positive Responsive and Recurrent Tumors. Panel A. Tumor 1971, Panel B. Tumor 2049, Panel C. 0843, Panel D. 2238, Panel E. 0732, Panel F. 1040. Solid black lines represent HPV, Solid gray lines represent cellular regions, Double black lines represent nonsense (*NS*) sequence, Dashed black arrows represent viral-specific primers, Dashed gray arrows represent the cellular exon-specific primers, and White circles represent exon-exon boundaries. X= no transcript was produced across the indicated sequences.

Table 1

Responsive and recurrent tumors, patient clinical information.

Tumor ID	Patient Gender	Age at Dx	Smoking Status	Smoking History	Alcohol Use	HPV Copy Number	Year of Dx	Status	Stage	TNM Classification	Tx
1733	Female	42	Former Smoker	5py × 20y	?	22.5	2009	Alive with NED 52 months from diagnosis	III	T2 N1 M0	Chemo-RT ^C
1769	Female	60	Non-Smoker	NA	None	475.6	2009	Alive with NED 49 months from diagnosis	III	T3 N1 M0	Surgery
1804	Male	58	Non-Smoker	NA	Social	161.1	2009	Alive with NED 48 months from diagnosis	IVA	T2 N2 M0	Chemo-RT ^D
1971	Male	51	Non-Smoker	NA	None	538.9	2010	Alive with NED 40 months from diagnosis	III	T1 N1 M0	Chemo-RT ^D
2148	Male	67	Current Smoker	Daily Cigar	Social	16.1	2010	Alive with NED 34 months from diagnosis	IVB	T1 N3 M0	Chemo-RT ^D
0732	Male	78	Former Smoker ^A	80py	2–3 glasses wine/wk	110.7	2005	AWD 105 months from diagnosis	IVA	T4 N2b M0	Chemo-RT ^D
0843	Male	50	Current Smoker	20 py quit 27y ago	Social	34.1	2005	DWD 75 months from diagnosis	IVA	T4 N1 M0	Chemo-RT ^D
1040	Male	72	Former Smoker ^B	light quit 35 yr ago	none for 25y	297.9	2006	DWD 11 months from diagnosis	IVB	T4 N3 M0	Chemo-RT ^D
2049	Male	56	Non-Smoker	NA	Social	6.3	2010	DWD 18 months from diagnosis	IVA	T4 N2 M0	Chemo-RT ^D
2238	Male	51	Current Smoker	25py	?	14	2011	DWD 14 months from diagnosis	IVA	T3 N1 M0	Surgery + Chemo ^E

NED= no evidence of disease, AWD= alive with disease, DWD= died with disease,

^A non-smoker for 27 years,

^B non-smoker for 35 years,

^C chemotherapy and radiation received at an outside institution, no additional information,

^D chemotherapy/radiation consisting of carboplatin/paclitaxel and intensity-mediated radiation therapy (IMRT),

^E chemotherapy with weekly cisplatin.

Summary of Integration Events in HPV16-Positive Responsive and Recurrent Tumors. Locus of integration into the cellular genome are shown on the right side of the table. Responsive tumors (top five tumors) are 1733, 1769, 1804, 1971, and 2148. Recurrent tumors (bottom five) are 0732, 0843, 1040, 2049 and 2238, indicated in bold text. (F), (R) = Forward or Reverse viral orientation in relation to the cellular gene.

Table 2

Tumor	HPV Site	Cellular				
		Locus	Gene	Gene Name	Region	Domain
1733	(F)E2	2p16	Intergenic			
	(F)E2	9q21	Intergenic			
1769	(R)L1	16q11.2	Intergenic			
	(R)L1	4q27	Intergenic			
1804	(R)E1	6q16	Intergenic			
	(F)L2	10p11.1	Intergenic			
	(F)E5	16q11.2	Intergenic			
	(F)E2	16q11.2	Intergenic			
1971	(F)E1	7p22.3	Intergenic			
	(R)L2	4p16.3	Intergenic			
	(R)L1	3q28	<i>TP63</i>	Tumor protein p63	Intron 4	DNA Binding domain
2148	(F)L2	7p22	Intergenic			
2049	(R)E1	8q11.21	<i>UBE2V2</i>	Ubiquitin-conjugating enzyme E2	Intron 1	Ubiquitin-Conjugating domain
	(R)E1	14q24.1	<i>SMO1</i>	SPARC related modular calcium binding 1	Intron 1	Kazal-like domain
0843	(F)L2	2q24.3	<i>SCN2A</i>	Sodium channel, voltage-gated, type II, α subunit	Intron 16	Helical Transmembrane S6 of repeat II region
	(R)L1	1p31.3-p31.2	<i>NF1A</i>	Nuclear factor I/A	Intron 9	DNA Binding domain
2238	(R)E2	15q21.1	<i>SEMA6D</i>	Semaphorin 6D	Intron 4	Extracellular Topological Sema domain
	(R)E2	10p11.1	Intergenic			
0732	(R)L2	17p11.2	<i>TNFRSF13B</i>	Tumor necrosis factor receptor superfamily, 13B	Intron 3	Extracellular Topological domain
	(R)L2	10p11.1	Intergenic			
1040	(R)L1	16p11.2	<i>SH2B1</i>	SH2B Adapter Protein 1	Intron 3	Plekstrin Homology domain