Human Papillomavirus Integration Strictly Correlates with Global Genome Instability in Head and Neck Cancer

Brandon Labarge[1,2], Max Hennessy[1,2], Lijun Zhang[2], David Goldrich[1,2], Scott Chartrand[3], Carson Purnell[3], Sage Wright[3], David Goldenberg[1], and James R. Broach[2,3]

ABSTRACT

Human papillomavirus (HPV)-positive head and neck cancers, predominantly oropharyngeal squamous cell carcinoma (OPSCC), exhibit epidemiologic, clinical, and molecular characteristics distinct from those OPSCCs lacking HPV. We applied a combination of whole-genome sequencing and optical genome mapping to interrogate the genome structure of HPV-positive OPSCCs. We found that the virus had integrated in the host genome in two thirds of the tumors examined but resided solely extrachromosomally in the other third. Integration of the virus occurred at essentially random sites within the genome. Focal amplification of the virus and the genomic sequences surrounding it often occurred subsequent to integration, with the number of tandem repeats in the chromosome accounting for the increased copy number of the genome sequences flanking the site of integration. In all cases, viral integration correlated with pervasive genome-wide somatic alterations at sites distinct from that of viral integration and comprised multiple insertions, deletions, translocations, inversions, and point mutations. Few or no somatic mutations were present in tumors with only episomal HPV. Our data could be interpreted by positing that episomal HPV is captured in the host genome following an episode of global genome instability during tumor development. Viral integration correlated with higher grade tumors, which may be explained by the associated extensive mutation of the genome and suggests that HPV integration status may inform prognosis.

Implications: Our results indicate that HPV integration in head and neck cancer correlates with extensive pangenomic structural variation, which may have prognostic implications.

Introduction

Head and neck cancer, an aggressive malignancy with high morbidity and mortality, is the seventh most common cancer worldwide, with 890,000 new cases and 450,000 deaths worldwide and 51,540 new cases and 10,030 deaths in the United States in 2018 (1–3). Greater than 90% of head and neck cancers are squamous cell carcinomas arising from the mucosal surfaces of the oral cavity, oropharynx, and larynx (4, 5). While the classic risk factors are tobacco and alcohol, human papillomavirus (HPV) has emerged in the past few decades as a growing risk factor for these cancers, especially for oropharyngeal squamous cell carcinoma (OPSCC), defining a new subtype of tumor that is distinct from HPV-negative tumors. As a consequence, OPSCC is one of the few cancers with rapidly increasing incidence in recent years, driven predominantly by HPV-positive cases (6, 7).

Following initial infection, HPV persists in the nucleus of its host cell as an extrachromosomal episome, but can subsequently integrate into the host genome (8, 9). The reported proportion of HPV-positive tumors in which the virus integrates into the genome varies by study, but analysis of The Cancer Genome Atlas (TCGA) data indicates that

HPV is found integrated in approximately 71% in virus-positive head and neck cancer and 83% in cervical cancer (10, 11). These integration events occur essentially randomly throughout the genome, although a few loci have been identified as recurrent (12, 13). The recurrent sites are often associated with common fragile genomic locations, transcriptionally active regions, and near regions of microhomology (1–10 bp) between the viral and human genomes (11, 14–16). This suggests that DNA double strand breaks drive vial integration, a conclusion supported by the observation that DNA damage promotes viral integration (17, 18). Integration occurs at essentially random sites within the viral genome as well. Most often only a fragment of the viral genome is retained following integration, spanning E6, E7 and a random amount of the adjacent viral genome often lacking an intact E2 (11, 13). The retention of E6 and E7 and the loss of E2 is likely the consequence of selection during tumorigenesis, as the elimination of E2 results in increased expression of the E6/E7 viral oncogenes, which drive tumorigenesis. The E6 oncoprotein of the oncogenic strains of HPV, primarily 16, 18, and 33, inactivates the p53 pathway by promoting degradation of p53, resulting in abrogation of genome integrity surveillance (19, 20). In addition, the second HPV oncoprotein, E7, binds to and inactivates the cell-cycle inhibitor, Rb, leading to uncontrolled cell-cycle progression. Accordingly, tumorigenesis results predominantly from loss of cell-cycle regulation elicited by E7 and abrogation of DNA damage checkpoint control caused by E6.

Several previous studies have documented an association of HPV integration with structural alteration of the host genome at the site of viral integration in OPSCC (11, 21–24). An elegant study by Akagi and colleagues (21), primarily focused on head and neck cancer cell lines, demonstrated that viral integration was linked to local genome instability, including inversions, duplications and deletions, often leading to amplification of host sequences adjacent to the site of integration. Using a combination of PCR amplification, chromosome walking and Sanger sequencing they determined detailed structures of the focal rearrangements surrounding the sites of integration and

[1]Department of Otolaryngology, Penn State College of Medicine, Hershey, Pennsylvania. [2]Institute for Personalized Medicine, Penn State College of Medicine, Hershey, Pennsylvania. [3]Department of Biochemistry and Molecular Biology, Penn State College of Medicine, Hershey, Pennsylvania.

Corresponding Author: James R. Broach, Department of Biochemistry, Penn State College of Medicine, Hershey, PA 17033. Phone: 717-531-8586; E-mail: jbroach@pennstatehealth.psu.edu

proposed a rolling circle replication and looping model to account for the organization of the genome over the region surrounding viral integration. A separate study proposed amplification of excised hybrid viral-human DNA segments as extrachromosomal circles often followed by reintegration into the host genome (22). Consistent with the observations underlying both models, several groups examined sequence data from HPV-positive head and neck cancer and noted amplification of host sequences surrounding the sites of viral integration (21, 23).

Earlier studies, particularly with cervical cancer, indicate that HPV is associated not only with focal disruptions near the site of integration but also with genome wide genomic instability, most notably aneuploidy (25). Several lines of evidence suggested that HPV16 E7 disrupts genome integrity by directly interfering with centriole duplication control (26, 27). Nonetheless, HPV-positive tonsillar squamous cell carcinoma exhibit a lower frequency of aneuploidy than HPV-negative tumors (28), so whether HPV promotes aneuploidy in oral cancers remains an open question. An additional global effect of HPV on genome integrity is that HPV E6 and E7 appear to enhance mutation frequency in primary human keratinocytes (29). In addition, HPV infection and HPV E6 and E7 oncoproteins alone activate the DNA damage ATM and ATR pathways (30). Our data described below provides a compelling argument that HPV integration, rather than HPV per se, correlates with genome-wide mutation and genomic instability.

We applied a combination of whole-genome sequencing (WGS) and optical genome mapping (OGM) of several HPV-positive OPSCCs in order to identify the viral integration state as well as somatic alterations throughout the tumor genome. OGM using a Bionano Genomics microfluidic instrument, such as the Saphyr, interrogates individual large (>250 kb) genomic DNA fragments rendered strictly linear in nanofluidic channels following fluorescent barcoding targeting specific DNA sequences (31, 32). The aggregated images of these molecules allow de novo assembly of the tumor genome without reference to a scaffold. Once assembled, the resultant genome can be compared with a generic reference map or, in our case, to the normal germline genome from the patient's peripheral blood (33). Our results confirmed focal amplifications and rearrangements surrounding the sites of integration but also documented a high level of genome-wide somatic structural variants (SV) in HPV-positive tumors, but only in those tumors in which HPV was integrated into the host genome. Tumors with comparable amounts of episomal HPV exhibited essentially no somatic structural variation. On the basis of our analysis of the various rearrangements and integration events, we speculate that viral integration occurs following a genome wide structural catastrophe. The causes and consequences of such widespread genome instability remain to be resolved but our data and some previous reports suggest that head and neck tumors with integrated HPV may be more aggressive, resulting in poorer patient outcomes (34–36).

Materials and Methods

Patient sample

Tissue and blood samples were obtained following surgical resection for twelve p16-positive oropharyngeal tumors under protocol PRAMS00040532 approved by the Penn State Health Institutional Review Board. Patients' demographics are provided in Supplementary Table S1. We retained both tumor and resected lateral neck lymph nodes adjacent to the tumor, which we analyzed if they were metastasis-positive. Tissue samples were flash frozen and stored at –80°C.

Blood samples were obtained from all patients and stored at –80°C until use. DNA samples isolated from tumor tissue using the DNeasy Blood & Tissue Kit (Qiagen) were tested by HPV-specific PCR to confirm the presence of virus in the tumor.

OGM

Ultrahigh molecular weight DNA was extracted from tumor tissue and associated blood samples and fluorescently labeled as described previously (33). Samples were analyzed on Saphyr chips (Bionano Genomics, USA), targeting approximately 200X human genome coverage (Supplementary Table S2).

WGS

Previously isolated high molecular weight DNA was sheared into 400 base pair fragments using a Covaris Sonicator, followed by size selection using SparQ PureMag magnetic beads at 70x and 55x bead concentration. Library preparation was performed using the KAPA Hyper Prep Kit with dual-indexed, unique NEXTflex DNA Barcode library adapters. Samples were pooled, applied to a S1 flowcell of an Illumina NovaSeq 6000 Sequencer, from which we obtained an average coverage of 40X (Supplementary Table S2).

Bionano data analysis

Whole-genome imaging data was analyzed using Bionano Access 1.6 Pipeline. Individual consensus genome maps were assembled de novo and compared with GRCh38 reference. On the basis of the de novo assembly results, we further ran dual variant annotation pipeline (Bionano Solve 3.6) for each cancer genome to filter germline SV present in matched blood sample from the same patient. We also removed SVs present in the control reference genome to remove common SVs. Filtered SV counts are shown in **Table 1**.

WGS data analysis and virus detection

WGS reads were mapped to human GRCh38 and combined HPV database, which includes multiple HPV reference genomes, using BWA-mem (version 0.7.17; ref. 37). WGS somatic mutations were identified with DRAGEN pipeline and only variants that passed filter were counted and annotated with Funcotator from Genome Analysis Toolkit (version 4.1.6.0). Somatic SNV counts are listed in **Table 1b**. Copy-number variants (CNV) were determined by Control-FREEC (version 11.5; ref. 38). Mutated cancer associated genes were called in MutSigSV (version 1.41; ref. 39). SAMtools depth was used to obtain the depth of sequence coverage across the ~7.9kb HPV genome based on alignment files in mapped BAM format (40).

Data availability

Raw and aligned next generation sequencing files have been submitted to the European Genome-Phenome Archive (https://www.ebi.ac.uk/ega) within study accession EGAS00001005163. Bionano variant calls and mapped reads for our samples can be downloaded from https://www.datacommons.psu.edu/commonswizard/MetadataDisplay.aspx?Dataset=6286.

Results

Determination of HPV integration

We performed WGS and OGM on OPSCCs or associated lymph node metastases, if available, as well as corresponding whole blood from twelve patients. We mapped WGS reads to the human reference hg38 and, for tumor samples, to either HPV16 (eleven of the twelve tumors) or HPV33 (one tumor) genomes. We noted those sequence

Table 1. Structural somatic mutation counts in tumors with integrated versus episomal HPV.

Sample	HPV status	Insertions	Deletions	Inversion	Duplications	Translocations	Total
20T	I	3	6	0	2	6	17
3718LN	I	9	3	1	6	5	24
3922LN	I	5	17	2	5	11	40
5785T	I	6	12	0	14	23	55
5954LN	I	2	5	2	5	5	19
7122LN	I	5	30	1	18	34	88
7331LN	I	14	19	2	10	9	54
7387T	I	8	13	2	5	43	71
3726T	E	1	3	0	4	1	9
3943T	E	1	0	0	3	0	4
7309LN	E	0	0	0	0	0	0
7313LN	E	0	0	0	0	0	0

reads carrying both human and HPV sequences, including those in which one of the paired ends mapped to the human genome and the other mate pair mapped to the HPV genome (discordant reads) and those in which one of the paired ends carried HPV sequences immediately abutting human sequences (softclip reads). We considered those samples with multiple consistent discordant and softclip reads as likely candidates for containing integrated HPV, with the boundaries of the human sequences in softclip reads marking the likely sites of integration. We then examined the OGM data for evidence of integration at the sites specified by the softclip reads and designated a sample as containing integrated HPV only if the OGM data confirmed an integration event at those sites. In most cases, softclip reads indicated one or two sites of possible integration that were confirmed by OGM. However, all samples had additional low-level softclip reads that were not confirmed by OGM and represented either sequencing artifacts, minor extrachromosomal hybrid molecules or integration events in a small subclone of cells below the 5% limit of detection of OGM in this study. In addition, WGS of sample 3943 returned a large number of consistent softclip reads linking viral sequences to a single chromosome 10 sequence. However, OGM failed to identify an integration event at that genome position and thus we concluded that sample 3943 carried an extrachromosomal hybrid viral genome. In sum, we determined that eight of the twelve tumors carried integrated HPV genomes, while the others carried exclusively extrachromosomal viral genomes.

We mapped the WGS reads to the human and HPV genome, from which we calculated the copy number of the virus as a function of position along its genome (**Fig. 1**; Supplementary Table S3). As evident from these data, all the tumors with only episomal copies of HPV contained sequences covering the entire genome, indicating the presence of an intact HPV genome within the tumor, with copy numbers ranging from 0.6 to 16. Three of these four tumors contained additional viral sequences mapping to only a portion of the genome, indicating the presence of a truncated extrachromosal HPV species as well. However, with the exception of sample 3943, WGS demonstrated that the viral sequences at the two boundaries of the gaps in each of these genomes are linked to each other, demonstrating that the gap results from an internal deletion of the virus rather than an integration event. Contrary to a previous report (23), in no case was the calculated copy level of the truncated viral genome equal to that of the intact genome.

All of the tumors with integrated HPV genomes contained sequences that mapped to only a portion of the genome and the boundaries of the partial HPV genome often correlated with the viral boundaries of the integration event. This indicates that, consistent with earlier studies, only a portion of the HPV genome integrated into the tumor genome such that integration maintained integrity of the E6 and E7 loci but inactivated E2. Three of the tumors carried two distinct partial genomes at equal copy level. Some of the tumors contained sequences mapping to the entire viral genome, indicating the presence of a complete viral genome in the tumors in addition to the partial integrated genomes. Our results do not distinguish between the persistence of an extrachromosomal HPV viral genome in the tumor versus integration of the intact viral genome, perhaps in tandem with the partial genome. The fact that the copy number of the intact genome is roughly equal to that of the partial genome suggests the latter, with the integrating virus consisting of one copy of the intact genome and one copy of the truncated genome. In several cases, the sizes of the integrated virus as determined by OGM are consistent with that interpretation.

From the genome sequences of the viruses, we could extract single nucleotide polymorphisms that allowed us to identify the relatedness among the viruses in the different tumors. Each had a unique SNP profile but with overlapping patterns, yielding a similarity profile as indicated in **Fig. 1**. Of particular significance, those viruses that were exclusively extrachromosomal did not constitute a cluster distinct from those that had integrated into the genome. Thus, the difference between integrating and non-integrating viruses in our cohort does not appear to be an intrinsic feature of the virus itself.

HPV can integrate at multiple sites and induce focal amplification

The combination of WGS and OGM data allowed us to determine the site and structure of the tumor genome spanning the viral integration site in almost all cases (**Fig. 2**; Supplementary Table S4). As has been observed previously for both head and neck and cervical cancers, the sites of integration were recurrent in neither the host genome nor the viral genome. Moreover, we observed focal amplification of the host sequences surrounding the site of insertion in almost every case of integration. These amplifications involved anywhere from 22 kb to 400 Kb of flanking sequences and range from a simple tandem duplication of the virus and surrounding sequences (tumor 7387) to a complex rearrangement spanning 2.5 Mb and at least 26 copies of the virus and various regions of the bordering host genome (tumor 7122; **Fig. 2**). In four cases, we found the virus inserted at two different sites in the genome, located on two different

Figure 1.
Copy number and polymorphisms of viral genomes in OPSCCs. Shown underneath a map of the HPV16 genome are copy number values as a function of genome position of virus from eleven of the twelve OPSCCs examined in this study. Copy number was determined from the total WGS read counts at each position of the virus, normalized to the average read count over unique human genome sequences in the same sample. Positions of single nucleotide polymorphisms relative to reference HPV16 (NC_001526.4) in each virus are designated, color coded to indicate the nucleotide substitution (T, red; C, blue; A, green; G, orange). Tumors with integrated virus are shown in teal and those with only extrachromosomal virus are shown in violet.

chromosomes. In another case, the virus spans the breakpoint in an interchromosomal translocation involving three separate chromatids, and in three other cases, it spans the breakpoint of an intrachromosomal translocation. In all but one case, at least one of the virus insertions lay inside, or within 10 kb, of a protein-coding region, consistent with previous observations (11, 13, 41, 42). Detailed descriptions of the rearrangements at the sites of each integration are provided in Supplementary Material.

Genome instability strictly correlates with HPV integration

By analyzing the genomes of the tumor samples by OGM, we were able to identify SVs not only at the site of viral integration but also across the entire genome. As shown in **Fig. 3**, every one of the tumors in which HPV had integrated had also undergone extensive genome rearrangement, including insertions, deletions, inversions, and translocations. Some of these resulted in substantial segmental aneuploidies (Supplementary Fig. S1) and impinge on a variety of cancer associated genes, including some that have been previously implicated in OPSCC (Supplementary Table S5). In contrast, the tumors in which HPV remained episomal showed no comparable genome instability: the genomes of these tumors were essentially identical to those of the corresponding germline genomes of the patients. The numbers of each type of SV in each of the tumors, listed in **Table 1A**, confirm this visual impression and document that the differences in both individual and total SV load between tumors with integrated versus episomal HPV is statistically significant ($P = 0.008$, Wilcoxon rank sum test, for total SVs).

We also examined the single nucleotide frequency in the tumors as well as the overall mutational load and the specific cancer genes mutated in each tumor. Previous sequence analyses of head and neck tumors demonstrated that HPV-positive head and neck tumors had a mutational landscape quite distinct from that of HPV-negative tumors (43–48). In particular, almost all HPV negative tumors contained mutations in *TP53* and a significant fraction carried mutations of *CDKN2A* or amplification of *CCND1*, whereas HPV-positive tumors contained almost none of these mutations. This is consistent with viral E6 and E7 oncogenes driving tumorigenesis in HPV-positive tumors through inactivation of Rb and p53, eliminating the selective pressure for host mutations in genes comprising these pathways. Previous results also demonstrated that a significant fraction of both HPV-positive and negative tumors contained activating mutations or amplifications of *PIK3A* (44, 45).

The total point mutational burden across all of the tumors in our study spanned a range comparable with that previously reported for head and neck cancers (Supplementary Fig. S2). Moreover, the mutational landscape of our tumors was consistent with that of HPV-positive tumors described in the TCGA cohort. None of our cohort carried *TP53* mutations or amplification of *CCND1*, while a significant fraction carried an activating mutation or amplification of PIK3CA. Nonetheless, the mutational landscape of tumors with episomal HPV was clearly distinct from that of tumors with integrated HPV (**Fig. 4**). The average mutational burden was substantially less in the former group than the latter, albeit not reaching statistical significance. Moreover, the spectrum of mutations was quite distinct. Seven of eight of the tumors with integrated HPV carried a mutation in or amplification of *PIK3CA*. On the other hand, only one of the episomal HPV tumors had a mutation or amplification of *PIK3CA*. These results suggest that HPV E6 and E7 drive tumorigenesis in both episomal and integrated HPV tumors. However, the mutational landscape is distinct between the two classes and the overall mutational frequency is substantially higher in those with integrated (Supplementary Fig. S2; **Table 2**). Whether the different mutational spectrum simply
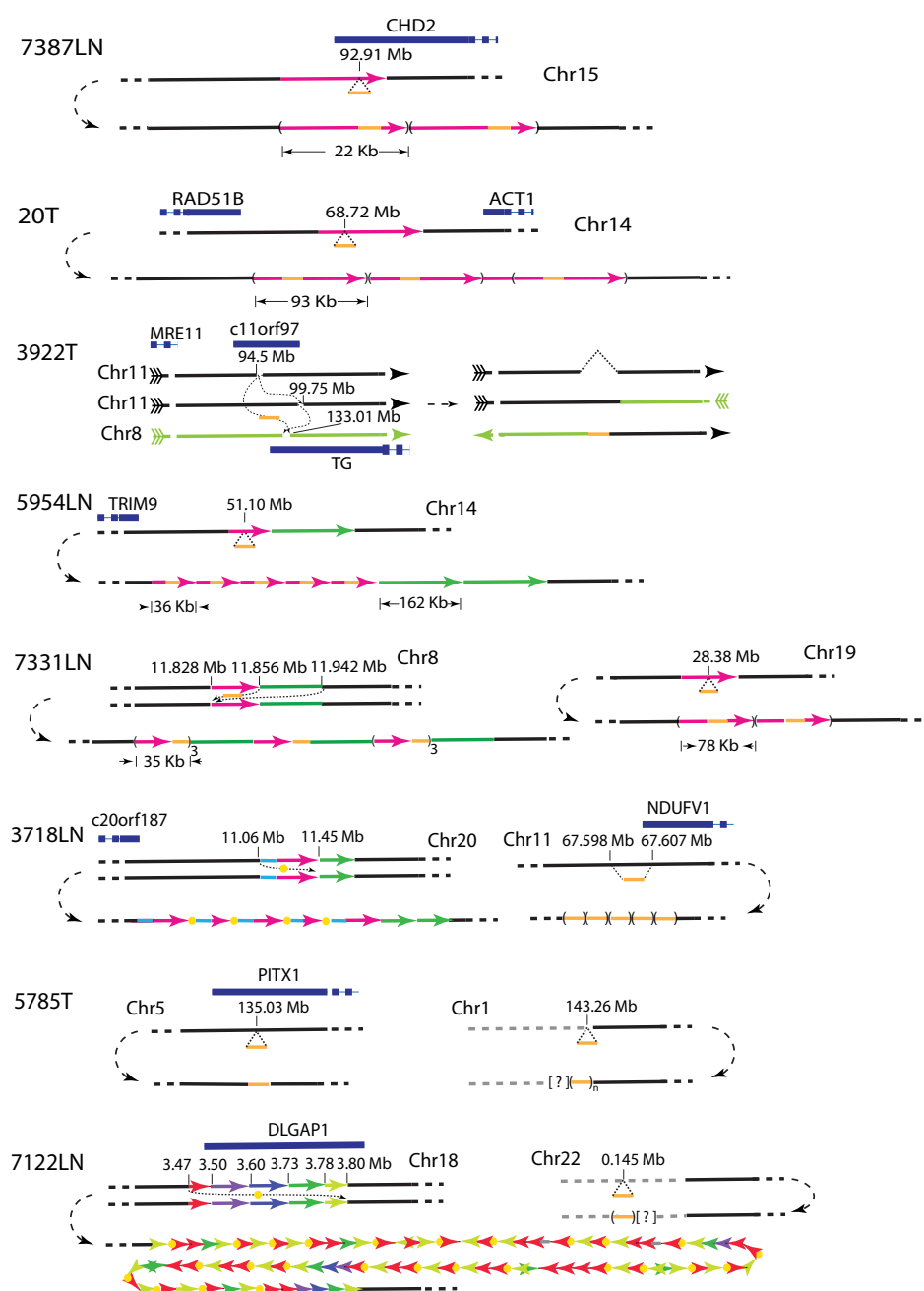
Figure 2.
Genome structure surrounding of the sites of viral integration. Shown are diagrams of the regions surrounding the sites of viral integration in eight OPSCCs. The upper four contained only a single integration site while the lower four contained two separate sites. Virus is shown as a bar or dot in yellow and regions of the genome that become duplicated following integration are shown in color. The upper (or leftward for 3922T) portion of each diagram indicates the location of integration, with nearby genes shown above, while the lower (or rightward) portion represents the local structure following integration and focal amplification. Gray segments indicate regions unmappable by OGM. Diagrams are not to scale.

reflects the difference in mutational burden or a distinction in the growth properties of the two classes requires a more extensive analysis.

Viral integration is associated with tumors that are more aggressive

While the number of patients examined in this study is too small to draw statistically significant conclusions, the available data suggests that the tumors with integrated HPV are more aggressive than those with episomal HPV (Supplementary Table S1). Specifically, the mean size of tumors with integrated HPV was 3.5 cm while those with episomal HPV was 2.5 cm. Moreover, half of the tumors with integrated HPV exhibited perineural invasion while none of the tumors with episomal HPV exhibited this feature. Finally, tumor staging indicated that the HPV integrated tumors were more advanced with 3 of the 8 patients with integrated HPV tumors were stage 4 while all of the tumors with episomal HPV were only stage 1 or 2. These observations beg the question of cause and effect but suggest that further evaluation of features and outcomes of HPV integrated versus episomal tumors is warranted.

Discussion

We applied a combination of WGS and whole-genome imaging to interrogate the genome structure of HPV-positive OPSCCs. These complementary techniques allowed us to determine unequivocally the integration state of HPV in the tumors, to define the genome
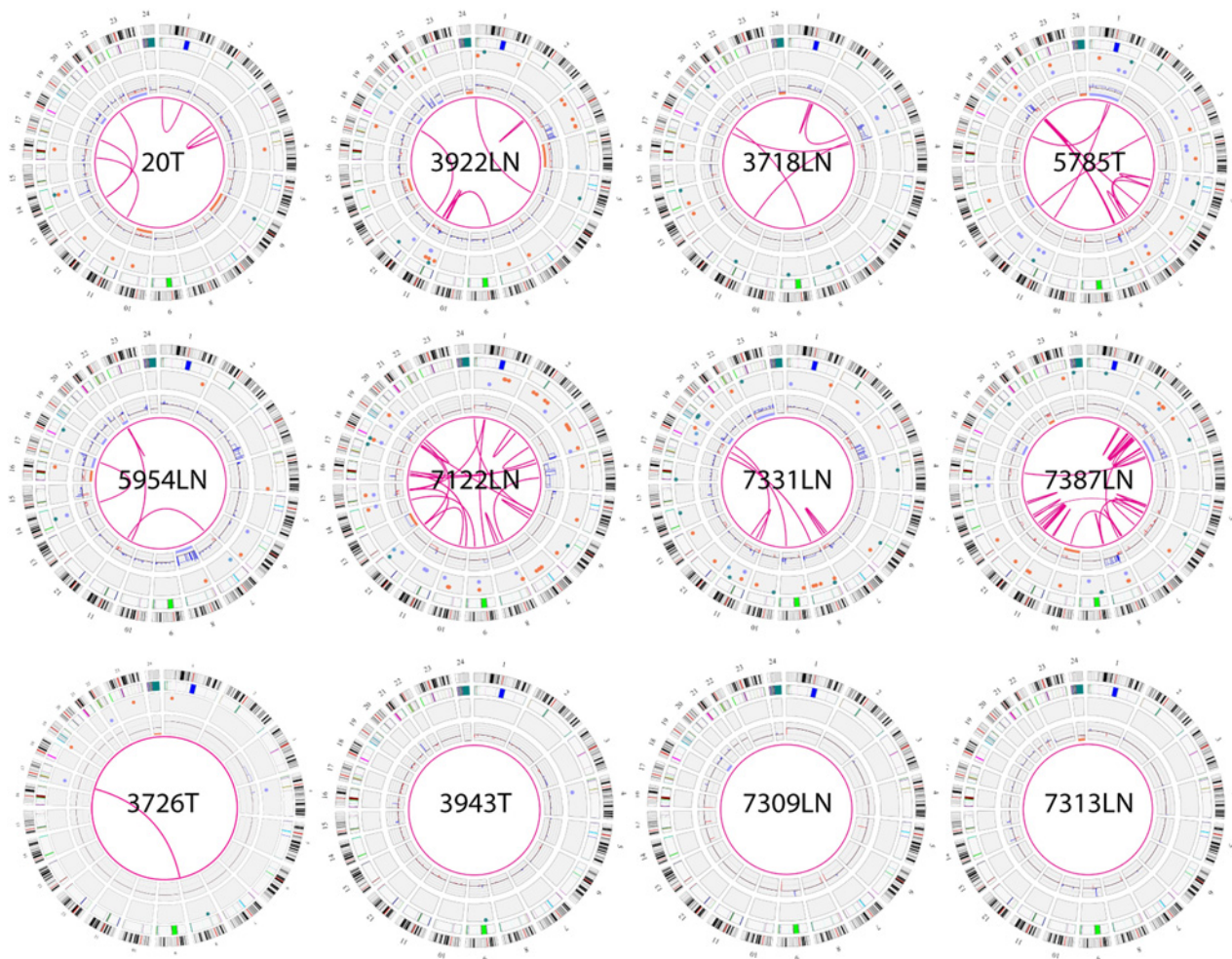
Figure 3.
Global genome structural variation accompanies viral integration. Circos plot diagrams of somatic SVs in all the OPSCC genomes, relative to the patients' normal genome, showing translocations and inversions in the center, copy number on the inner ring and insertions (green), deletions (orange) and duplications (light blue) on the third most outer ring. Chromosomes are ordered sequentially in the outer ring on which are indicated cytologic banding patterns and the centromere (red bar).

organization at the sites of integration and to correlate genome wide mutational patterns with the integration state of the virus. The results yield the striking conclusion that viral integration is tightly linked to global genome instability and increased mutation frequency. Moreover, while the number of cases examined to date is limited, clinical features indicate a trend toward more aggressive tumors associated with viral integration. Using our methodology, we were unable to rule out that some HPV genomes remained extrachromosomal in tumors in which HPV had integrated. However, regardless of whether some viral genomes remained extrachromosomal, the presence of an integrated copy was sufficient and necessary to link it to pervasive genomic instability.

Our criteria for determining viral integration required both the presence of multiple consistent WGS read in which viral sequences abut human sequences and OGM documentation of a structural alteration at the site predicted by the WGS reads. This is more rigorous than relying on WGS or mate pair sequencing alone, which do not distinguish between integration and extrachromosomal human-viral hybrid episomes, and more reliable than more indirect methods, such as RNA sequencing, inverse PCR, FISH, or E2/E6

copy levels (23, 49–51). Moreover, the method allowed us to extract the structure of the region immediately surrounding the site of integration directly from the primary data. These analyses showed that integration can occur at multiple sites within one tumor, that integration is often followed by focal amplification of the virus and the genomics sequences surrounding it and that integration often occurs at the junction between inter- or intra-chromosomal translocations. In all our cases, the local amplification of adjacent host sequences accounts for the increased copy number of those sequences, as has previously reported (13, 21, 23), without positing the formation and persistence of extrachromosomal hybrid molecules.

Our results document a strict correlation between viral integration and whole-genome instability. Tumors in which HPV had integrated contained a significantly greater number of SVs of all types—deletions, insertions, translocations and inversions—and an increase in single nucleotide and frameshift mutations over those found in tumors with only extrachromosomal HPV. Tumors without integrated HPV contained essentially no SVs and relatively few point mutations. While previous analysis of integrated versus episomal head and neck tumors noted significant genomic
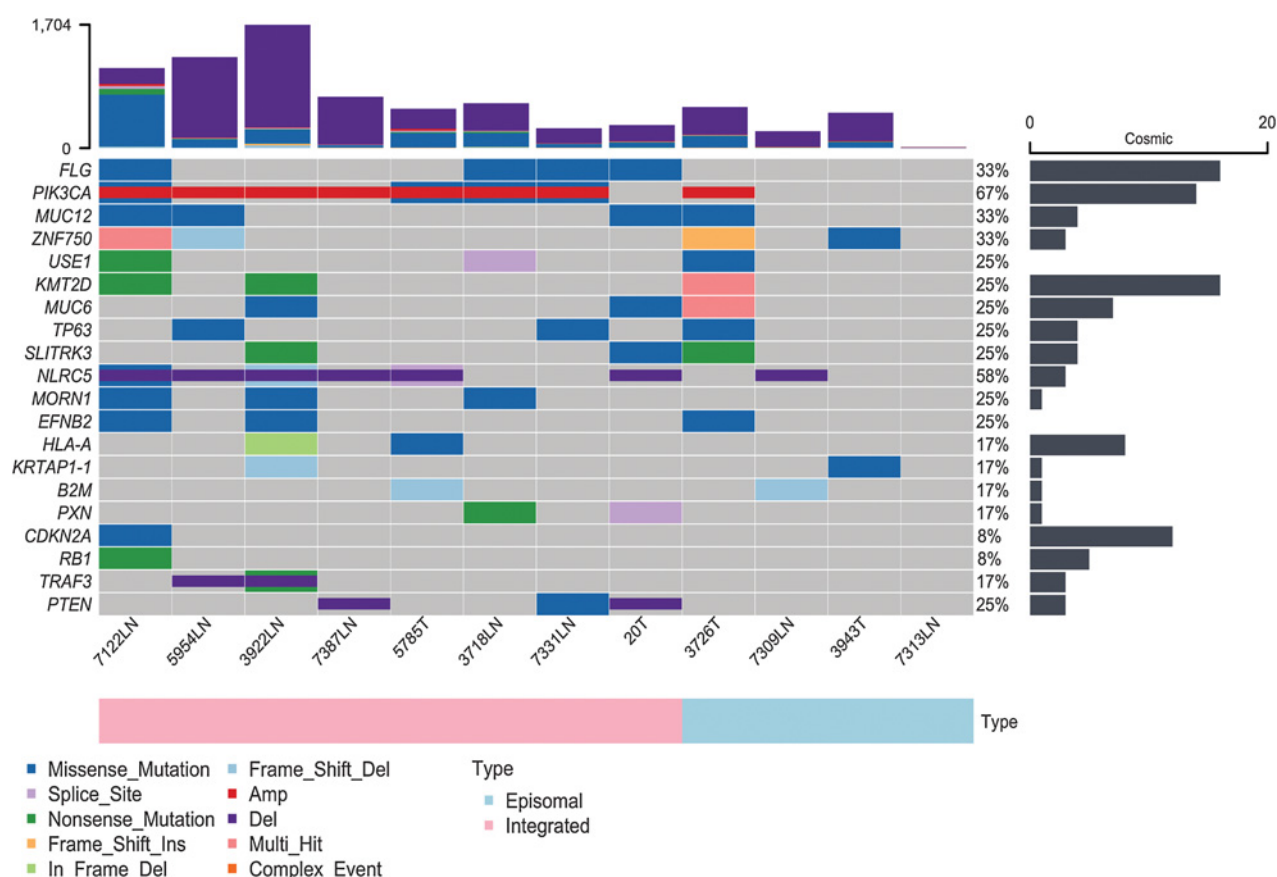
Figure 4.
Significantly mutated genes in integrated versus episomal tumors. Frequently mutated genes in OPSCCs with integrated HPV (left) and only episomal HPV (right) are shown on the heat map for each of the tumor samples. The mutation classes are indicated by color. Total number and type of exonic mutations in each sample are shown in the graph above the heat map. Mutation percentage of each gene in the cohort is shown immediately to the right of the heat map. Graph on the far right shows mutation percentage of the gene in COSMIC (upper aerodigestive tract, head and neck, squamous cell carcinoma).

alterations at the site of viral integration, few examined the overall structural variation in those tumors.

This correlation raises the question of causality. Does integration induce genome instability, perhaps as a consequence of increased expression of E7 following inactivation of E2 and attendant reduction

of p53 activity (25)? Or, does genome instability arise from a plethora of DNA damage events, yielding double strand breaks and activation of nonhomologous end joining repair through which the extrachromosomal HPV genomes could become attached to a chromosomal site (52)? Certainly, previous reports noted in the introduction suggest

Table 2. Single nucleotide somatic mutation counts in tumors with integrated versus episomal HPV.

Sample	HPV status	Missense	Nonsense	In Frame Del	In Frame Ins	Frameshift	Total
20T	I	72	6	2	1	1	82
3718LN	I	188	20	4	0	12	224
3922LN	I	190	9	5	0	58	262
5785T	I	198	17	0	0	5	220
5954LN	I	112	9	1	0	5	127
7122LN	I	715	77	5	0	18	815
7331LN	I	52	2	1	0	3	53
7387T	I	36	2	0	1	4	43
3726T	E	159	10	1	0	4	174
3943T	E	79	3	1	0	4	87
7309LN	E	7	1	2	0	1	11
7313LN	E	6	0	0	0	0	6

that enhanced E6 and E7 expression attendant upon integration could increase mutation and aneuploidy (26, 27, 29, 30). However, in half the tumors with integrated HPV, the viral segment is present at two different sites in the genome. This is more readily explained by a concurrent capture of the virus at two separate sites during a genome catastrophe rather than sequential integration events driven by selection. In addition, in several cases, we found the virus integrated at the junction of an intra- or inter-chromosome translocation. This would require the interaction of two separate chromatids, or in the case of one tumor, three separate chromatids, all broken concurrently and healed in concert. Finally, given the fact that some tumors with only extra-chromosomal HPV carried no obvious chromosomal driver mutations, we conclude that episomal HPV alone is likely sufficient for promoting tumor formation. Thus, subsequent selection for integration of the genome was not required for tumorigenesis. In sum, these observations are consistent with the hypothesis that viral integration results from some genome catastrophe subsequent to HPV infection rather than selection for viral integration to drive tumorigenesis. Such a genome catastrophe could occur in a short period during tumor development through a breakage-fusion-bridge cycle initiated by an initial cell division error (53), accounting for simultaneous viral integration and widespread genomic rearrangements.

Substantial clinical data has documented that HPV-positive OPSCCs are more responsive to radiation treatment or chemotherapy than HPV negative carcinomas and such patients have more favorable outcomes (54–56). However, those studies have not distinguished between integrated versus episomal HPV-positive tumors. Clinical characteristics of the tumors in the dozen cases examined here suggest that those with integrated HPV are more aggressive, although examination of more cases will be necessary to rigorously test that correlation. This differs from previous reports indicating that integration had no effect on outcomes (49) or indicated a more favorable prognosis (28). However, our suggestion that tumors with integrated virus are more aggressive is consistent with the recent observation that high copy-number variation in HPV-positive OPSCCs is strongly associated with worse recurrence-free survival (57) and our observation that copy-number variation is substantially elevated in integrated versus episomal cases (Supplementary Fig. S1). We do not have sufficient outcomes data to determine whether integration is associated with increased recurrence or reduced overall survival. However, our results raise the possibility that such distinction may have prognostic value that could inform treatment options.

Authors' Disclosures

Authors' Contributions

B. Labarge: Investigation, methodology, writing–original draft. **M. Hennessy:** Formal analysis, investigation, methodology. **L. Zhang:** Data curation, software, formal analysis, visualization. **D. Goldrich:** Investigation, methodology. **S. Chartrand:** Investigation. **C. Purnell:** Software, formal analysis, visualization. **S. Wright:** Data curation, formal analysis. **D. Goldenberg:** Conceptualization, supervision, funding acquisition, project administration. **J.R. Broach:** Conceptualization, supervision, writing–original draft, project administration, writing–review and editing.

Acknowledgments

Note

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424.
2. Jou A, Hess J. Epidemiology and molecular biology of head and neck cancer. Oncol Res Treat 2017;40:328–32.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018;68:7–30.
4. Blot WJ, McLaughlin JK, Winn DM, Austin DF, Greenberg RS, Preston-Martin S, et al. Smoking and drinking in relation to oral and pharyngeal cancer. Cancer Res 1988;48:3282–7.
5. Vigneswaran N, Williams MD. Epidemiologic trends in head and neck cancer and aids in diagnosis. Oral Maxillofac Surg Clin North Am 2014;26:123–41.
6. Elrefaey S, Massaro MA, Chiocca S, Chiesa F, Ansarin M. HPV in oropharyngeal cancer: the basics to know in clinical practice. Acta Otorhinolaryngol Ital 2014;34:299–309.
7. You EL, Henry M, Zeitouni AG. Human papillomavirus–associated oropharyngeal cancer: review of current evidence and management. Curr Oncol 2019;26:119–23.
8. Morgan IM, DiNardo LJ, Windle B. Integration of human papillomavirus genomes in head and neck cancer: is it time to consider a paradigm shift? Viruses 2017;9:208.
9. Thierry F. Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. Virology 2009;384:375–9.
10. Cancer Genome Atlas Research N, Albert Einstein College of M, Analytical Biological S, Barretos Cancer H, Baylor College of M, Beckman Research Institute of City of H, et alet al. Integrated genomic and molecular characterization of cervical cancer. Nature 2017;543:378–84.
11. Parfenov M, Pedamallu CS, Gehlenborg N, Freeman SS, Danilova L, Bristow CA, et al. Characterization of HPV and host genome interactions in primary head and neck cancers. Proc Natl Acad Sci USA 2014;111:15544–9.
12. McBride AA, Warburton A. The role of integration in oncogenic progression of HPV-associated cancers. PLoS Pathog 2017;13:e1006211.
13. Symer DE, Akagi K, Geiger HM, Song Y, Li G, Emde AK, et al. Diverse tumorigenic consequences of human papillomavirus integration in primary oropharyngeal cancers. Genome Res 2022;32:55–70.
14. Bodelon C, Untereiner ME, Machiela MJ, Vinokurova S, Wentzensen N. Genomic characterization of viral integration sites in HPV-related cancers. Int J Cancer 2016;139:2001–11.
15. Christiansen IK, Sandve GK, Schmitz M, Durst M, Hovig E. Transcriptionally active regions are the preferred targets for chromosomal HPV integration in cervical carcinogenesis. PLoS One 2015;10:e0119566.
16. Thorland EC, Myers SL, Persing DH, Sarkar G, McGovern RM, Gostout BS, et al. Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites. Cancer Res 2000;60:5916–21.
17. Chen Y, Williams V, Filippova M, Filippov V, Duerksen-Hughes P. Viral carcinogenesis: factors inducing DNA damage and virus integration. Cancers 2014;6:2155–86.
18. Katerji M, Duerksen-Hughes PJ. DNA damage in cancer development: special implications in viral oncogenesis. Am J Cancer Res 2021;11:3956–79.
19. Taberna M, Mena M, Pavon MA, Alemany L, Gillison ML, Mesia R. Human papillomavirus–related oropharyngeal cancer. Ann Oncol 2017;28:2386–98.
20. Leemans CR, Snijders PJF, Brakenhoff RH. The molecular landscape of head and neck cancer. Nat Rev Cancer 2018;18:269–82.

21. 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;24:185–99.

22. Deshpande V, Luebeck J, Nguyen ND, Bakhtiari M, Turner KM, Schwab R, et al. Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. Nat Commun 2019;10:392.

23. 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;8:17684–99.

24. Peter M, Stransky N, Couturier J, Hupe P, Barillot E, de Cremoux P, et al. Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma. J Pathol 2010;221:320–30.

25. Korzeniewski N, Spardy N, Duensing A, Duensing S. Genomic instability and cancer: lessons learned from human papillomaviruses. Cancer Lett 2011;305: 113–22.

26. Duensing A, Liu Y, Perdreau SA, Kleylein-Sohn J, Nigg EA, Duensing S. Centriole overduplication through the concurrent formation of multiple daughter centrioles at single maternal templates. Oncogene 2007;26:6280–8.

27. Duensing S, Duensing A, Crum CP, Munger K. Human papillomavirus type 16 E7 oncoprotein-induced abnormal centrosome synthesis is an early event in the evolving malignant phenotype. Cancer Res 2001;61:2356–60.

28. Mooren JJ, Kremer B, Claessen SM, Voogd AC, Bot FJ, Peter Klussmann J, et al. Chromosome stability in tonsillar squamous cell carcinoma is associated with HPV16 integration and indicates a favorable prognosis. Int J Cancer 2013;132: 1781–9.

29. Liu X, Han S, Baluda MA, Park NH. HPV-16 oncogenes E6 and E7 are mutagenic in normal human oral keratinocytes. Oncogene 1997;14:2347–53.

30. Albert E, Laimins L. Regulation of the human papillomavirus life cycle by DNA damage repair pathways and epigenetic factors. Viruses 2020;12:744.

31. Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol 2012;30:771–6.

32. Mak AC, Lai YY, Lam ET, Kwok TP, Leung AK, Poon A, et al. Genome-wide structural variation detection by genome mapping on nanochannel arrays. Genetics 2016;202:351–62.

33. Goldrich DY, LaBarge B, Chartrand S, Zhang L, Sadowski HB, Zhang Y, et al. Identification of somatic structural variants in solid tumors by optical genome mapping. J Pers Med 2021;11:142.

34. Koneva LA, Zhang Y, Virani S, Hall PB, McHugh JB, Chepeha DB, et al. HPV integration in HNSCC correlates with survival outcomes, immune response signatures, and candidate drivers. Mol Cancer Res 2018;16:90–102.

35. Nulton TJ, Kim NK, DiNardo LJ, Morgan IM, Windle B. Patients with integrated HPV16 in head and neck cancer show poor survival. Oral Oncol 2018;80:52–5.

36. Veitia D, Liuzzi J, Avila M, Rodriguez I, Toro F, Correnti M. Association of viral load and physical status of HPV-16 with survival of patients with head and neck cancer. Ecancermedicalscience 2020;14:1082.

37. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 2010;26:589–95.

38. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 2012;28:423–5.

39. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 2013;499:214–8.

40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25:2078–9.

41. Olthof NC, Speel EJ, Kolligs J, Haesevoets A, Henfling M, Ramaekers FC, et al. Comprehensive analysis of HPV16 integration in OSCC reveals no significant impact of physical status on viral oncogene and virally disrupted human gene expression. PLoS One 2014;9:e88718.

42. Speel EJM. HPV integration in head and neck squamous cell carcinomas: cause and consequence. In: Golusiński W, Leemans CR, Dietz A, editors. HPV infection in head and neck cancer. Cham: Springer International Publishing; 2017. p57–72.

43. Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science 2011;333:1154–7.

44. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 2015;517:576–82.

45. Gillison ML, Akagi K, Xiao W, Jiang B, Pickard RKL, Li J, et al. Human papillomavirus and the landscape of secondary genetic alterations in oral cancers. Genome Res 2019;29:1–17.

46. Haft S, Ren S, Xu G, Mark A, Fisch K, Guo TW, et al. Mutation of chromatin regulators and focal hotspot alterations characterize human papillomavirus–positive oropharyngeal squamous cell carcinoma. Cancer 2019;125:2423–34.

47. Lui VW, Hedberg ML, Li H, Vangara BS, Pendleton K, Zeng Y, et al. Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. Cancer Discov 2013;3:761–9.

48. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The mutational landscape of head and neck squamous cell carcinoma. Science 2011;333:1157–60.

49. Lim MY, Dahlstrom KR, Sturgis EM, Li G. Human papillomavirus integration pattern and demographic, clinical, and survival characteristics of patients with oropharyngeal squamous cell carcinoma. Head Neck 2016;38:1139–44.

50. Mellin H, Dahlgren L, Munck-Wikland E, Lindholm J, Rabbani H, Kalantari M, et al. Human papillomavirus type 16 is episomal and a high viral load may be correlated to better prognosis in tonsillar cancer. Int J Cancer 2002;102:152–8.

51. 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 tumors in relation to patients' prognosis. Int J Cancer 2016;138:386–95.

52. Toledo L, Neelsen KJ, Lukas J. Replication catastrophe: when a checkpoint fails because of exhaustion. Mol Cell 2017;66:735–49.

53. Umbreit NT, Zhang CZ, Lynch LD, Blaine LJ, Cheng AM, Tourdot R, et al. Mechanisms generating cancer genome complexity from a single-cell division error. Science 2020;368:eaba0712.

54. 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;363:24–35.

55. O'Sullivan B, Huang SH, Su J, Garden AS, Sturgis EM, Dahlstrom K, et al. Development and validation of a staging system for HPV-related oropharyngeal cancer by the international collaboration on oropharyngeal cancer network for staging (ICON-S): a multicenter cohort study. Lancet Oncol 2016;17:440–51.

56. Rietbergen MM, Brakenhoff RH, Bloemena E, Witte BI, Snijders PJ, Heideman DA, et al. Human papillomavirus detection and comorbidity: critical issues in selection of patients with oropharyngeal cancer for treatment De-escalation trials. Ann Oncol 2013;24:2740–5.

57. Schrank TP, Lenze N, Landess LP, Hoyle A, Parker J, Lal A, et al. Genomic heterogeneity and copy-number variant burden are associated with poor recurrence-free survival and 11q loss in human papillomavirus–positive squamous cell carcinoma of the oropharynx. Cancer 2021;127:2788–800.