

UCSF

UC San Francisco Previously Published Works

Title

Characterization of HPV and host genome interactions in primary head and neck cancers

Permalink

<https://escholarship.org/uc/item/52z6c2wh>

Journal

Proceedings of the National Academy of Sciences of the United States of America,  
111(43)

ISSN

0027-8424

Authors

Parfenov, Michael

Pedamallu, Chandra Sekhar

Gehlenborg, Nils

et al.

Publication Date

2014-10-28

DOI

10.1073/pnas.1416074111

Peer reviewed

# Characterization of HPV and host genome interactions in primary head and neck cancers

Michael Parfenov<sup>a,1</sup>, Chandra Sekhar Pedamallu<sup>b,c,1</sup>, Nils Gehlenborg<sup>c,d</sup>, Samuel S. Freeman<sup>c</sup>, Ludmila Danilova<sup>e</sup>, Christopher A. Bristow<sup>f</sup>, Semin Lee<sup>d</sup>, Angela G. Hadjipanayis<sup>a</sup>, Elena V. Ivanova<sup>b,g</sup>, Matthew D. Wilkerson<sup>h</sup>, Alexei Protopopov<sup>f</sup>, Lixing Yang<sup>d</sup>, Sahil Seth<sup>f</sup>, Xingzhi Song<sup>f</sup>, Jiabin Tang<sup>f</sup>, Xiaojia Ren<sup>a</sup>, Jianhua Zhang<sup>f</sup>, Angeliki Pantazi<sup>a</sup>, Netty Santoso<sup>a</sup>, Andrew W. Xu<sup>d</sup>, Harshad Mahadeshwar<sup>f</sup>, David A. Wheeler<sup>i</sup>, Robert I. Haddad<sup>j</sup>, Joonil Jung<sup>c</sup>, Akinyemi I. Ojesina<sup>b,c</sup>, Natalia Issaeva<sup>k</sup>, Wendell G. Yarbrough<sup>k</sup>, D. Neil Hayes<sup>l</sup>, Jennifer R. Grandis<sup>m,n</sup>, Adel K. El-Naggar<sup>o,p</sup>, Matthew Meyerson<sup>b,c,q</sup>, Peter J. Park<sup>d</sup>, Lynda Chin<sup>c,f</sup>, J. G. Seidman<sup>a,2</sup>, Peter S. Hammerman<sup>b,c,2</sup>, Raju Kucherlapati<sup>r,2</sup>, and the Cancer Genome Atlas Network<sup>3</sup>

Departments of <sup>a</sup>Genetics and <sup>p</sup>Pathology and <sup>d</sup>Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115; <sup>b</sup>Department of Medical Oncology and <sup>e</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215; <sup>c</sup>The Eli and Edythe L. Broad Institute, Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02142; <sup>f</sup>The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD 21287; <sup>i</sup>Institute for Applied Cancer Science, Department of Genomic Medicine and Departments of <sup>o</sup>Pathology and <sup>p</sup>Head and Neck Surgery, University of Texas MD Anderson Cancer Center, Houston, TX 77030; <sup>g</sup>Belfer Institute for Applied Cancer Science, Dana-Farber Cancer Institute, Boston, MA 02115; <sup>h</sup>Lineberger Comprehensive Cancer Center and Department of Genetics and <sup>l</sup>Lineberger Comprehensive Cancer Center and Department of Internal Medicine, Division of Medical Oncology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; <sup>j</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030; <sup>k</sup>Department of Surgery, Division of Otolaryngology, Yale School of Medicine, New Haven, CT 06520-8041; Departments of <sup>m</sup>Otolaryngology and <sup>n</sup>Pharmacology and Chemical Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213; and <sup>r</sup>Harvard Medical School, Partners HealthCare Center for Genetics and Genomics, Boston, MA 02115

Contributed by J. G. Seidman, September 9, 2014 (sent for review March 25, 2014; reviewed by Fadlo R. Khuri and Gregory T. Wolf)

Previous studies have established that a subset of head and neck tumors contains human papillomavirus (HPV) sequences and that HPV-driven head and neck cancers display distinct biological and clinical features. HPV is known to drive cancer by the actions of the E6 and E7 oncoproteins, but the molecular architecture of HPV infection and its interaction with the host genome in head and neck cancers have not been comprehensively described. We profiled a cohort of 279 head and neck cancers with next generation RNA and DNA sequencing and show that 35 (12.5%) tumors displayed evidence of high-risk HPV types 16, 33, or 35. Twenty-five cases had integration of the viral genome into one or more locations in the human genome with statistical enrichment for genic regions. Integrations had a marked impact on the human genome and were associated with alterations in DNA copy number, mRNA transcript abundance and splicing, and both inter- and intrachromosomal rearrangements. Many of these events involved genes with documented roles in cancer. Cancers with integrated vs. nonintegrated HPV displayed different patterns of DNA methylation and both human and viral gene expressions. Together, these data provide insight into the mechanisms by which HPV interacts with the human genome beyond expression of viral oncoproteins and suggest that specific integration events are an integral component of viral oncogenesis.

cancer | head and neck | papilloma virus | genome rearrangement | integration sites

**H**ead and neck cancer (HNC) is a heterogeneous group of tumors characterized by a common anatomic origin, and most such tumors develop from within the mucosa and are classified as head and neck squamous cell carcinomas (HNSCCs) (1). HNSCC, the sixth most common cancer diagnosed worldwide and the eighth most common cause of cancer death (2), is frequently associated with human papillomavirus (HPV) infection (3, 4). Depending on the anatomic site of the tumor, HPV prevalence is estimated at 23–36% (5). HPV-positive HNSCCs form a distinct subset of HNCs that differs from HPV-negative HNSCCs in tumor biology and clinical characteristics, including superior clinical outcomes (6–9).

The molecular pathogenesis of HPV-driven HNSCC also seems distinct from HPV-negative tumors, with previous studies showing a divergent spectrum of alterations in gene expression, mutations, amplifications, and deletions as well as distinct epigenome

alterations (10–15). HPV is known to drive tumorigenesis through the actions of its major oncoproteins E6 and E7, which target numerous cellular pathways, including inactivation of p53 and the retinoblastoma (Rb) protein (16–18). Together with E5, they also play an important role in immune evasion, being involved in both innate and adaptive immunity (19, 20).

Initially after infection, HPV is identified in circular extrachromosomal particles or episomes. A critical step in progression to cancer is the integration of viral DNA into the host cell

## Significance

**A significant proportion of head and neck cancer is driven by human papillomavirus (HPV) infection, and the expression of viral oncogenes is involved in the development of these tumors. However, the role of HPV integration in primary tumors beyond increasing the expression of viral oncoproteins is not understood. Here, we describe how HPV integration impacts the host genome by amplification of oncogenes and disruption of tumor suppressors as well as driving inter- and intrachromosomal rearrangements. Tumors that do and do not have HPV integrants display distinct gene expression profiles and DNA methylation patterns, which further support the view that the mechanisms by which tumors with integrated and nonintegrated HPV arise are distinct.**

Author contributions: M.P., C.S.P., M.D.W., W.G.Y., D.N.H., M.M., J.G.S., P.S.H., R.K., and T.C.G.A.N. designed research; M.P., C.S.P., N.G., L.D., C.A.B., A.G.H., M.D.W., A. Protopopov, L.Y., X.S., J.T., X.R., J.Z., A. Pantazi, N.S., A.W.X., H.M., D.A.W., R.I.H., N.I., D.N.H., J.R.G., A.K.E.-N., M.M., L.C., P.S.H., and R.K. performed research; M.P., S.S.F., S.L., A.G.H., E.V.I., M.D.W., L.Y., S.S., J.Z., D.A.W., R.I.H., N.I., W.G.Y., D.N.H., J.R.G., A.K.E.-N., M.M., P.J.P., L.C., P.S.H., and R.K. contributed new reagents/analytic tools; M.P., C.S.P., N.G., S.S.F., L.D., C.A.B., S.L., A.G.H., E.V.I., M.D.W., A. Protopopov, L.Y., S.S., X.S., J.T., X.R., J.Z., A. Pantazi, N.S., A.W.X., H.M., J.J., A.I.O., N.I., D.N.H., M.M., P.J.P., L.C., J.G.S., P.S.H., and R.K. analyzed data; T.C.G.A.N. collected tumor tissue and normal tissue from more than 200 cancer patients; T.C.G.A.N. provided clinical data on all of these patients; and M.P., C.S.P., N.G., S.S.F., L.D., A.G.H., E.V.I., P.S.H., and R.K. wrote the paper.

Reviewers: F.R.K., Emory University; and G.T.W., University of Michigan.

The authors declare no conflict of interest.

<sup>1</sup>M.P. and C.S.P. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: seidman@genetics.med.harvard.edu, peterh@broadinstitute.org, or rkucherlapati@partners.org.

<sup>3</sup>A complete list of the TCGA Network can be found in *SI Appendix*.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1416074111/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1416074111/-DCSupplemental).

genome (21). Integration breakpoints often occur in the HPV early genes *E1* or *E2*, disrupting their expression, which leads to the deregulation of negative feedback control of *E6* and *E7* oncogene expression by the viral regulatory *E2* protein. Integrant-derived transcripts are more stable than those derived from episomal viral DNA, and HPV integration has been associated with increased proliferative capacity and selective growth advantage for the affected cells (6, 22). Another crucial event frequently associated with HPV integration is genomic instability of the infected cells (23).

The physical state of the virus and its interaction with the host genome in HPV-associated tumors has been well-characterized in cervical cancer (24–30), but less is known about the molecular architecture of HPV infection in HNC (31, 32). Recently, a comprehensive genome-wide analysis of HPV integration in human cancers showed associations between HPV integrants and extensive host genomic amplifications and rearrangements, including deletions, inversions, and chromosomal translocations (12). However, the study primarily used HNSCC cell lines and included only two primary HNCs. There is still a lack of comprehensive analyses of primary HNSCCs using a multiplatform approach, in which integrations are characterized in the context of changes in the DNA, RNA, and epigenome at the site of integration.

Here, by analysis of 35 primary HPV-positive HNCs with whole-genome, transcriptome, and methylation profiling, we describe genomic changes, including massive structural variations, differential DNA methylation and expression patterns, that occur in HPV-positive primary head and neck tumors and are associated with viral integrations.

## Results

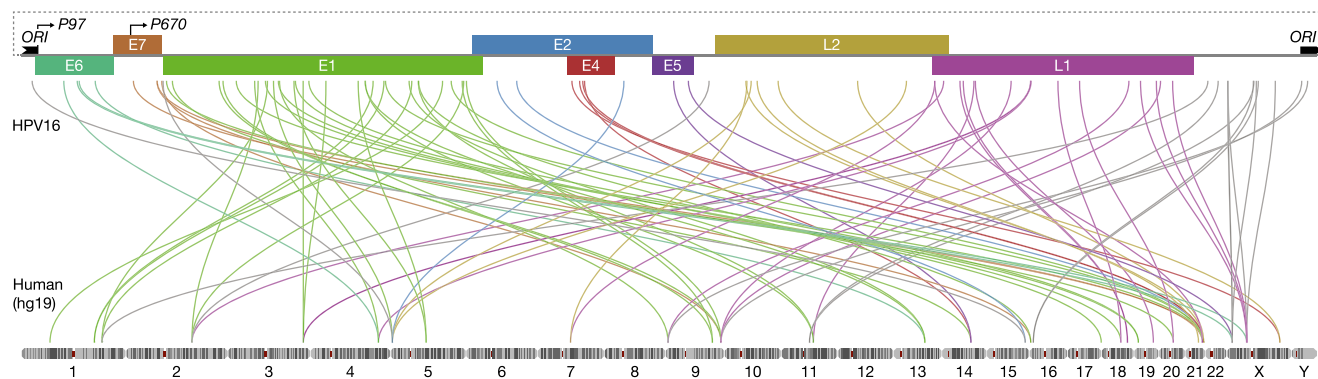
**Characterization of HPV Integrations.** Using the dataset of 279 HNSCCs developed by The Cancer Genome Atlas, we performed computational subtraction (33) and quantitative assessment of HPV *E6* transcripts and identified 35 tumors with evidence of HPV DNA and RNA sequences. HPV16 was the most common type, and it was detected in 29 of 35 tumors, with the remaining cases having either HPV33 or HPV35. Based on single-nucleotide variants present in HPV16 sequences (Dataset S1, Table S1), we assessed the variant lineage and sublineage in each case. The variant lineages and sublineages of 29 HPV16-positive samples were 22 European, 3 African 1, 2 Asian, 1 African 2, and 1 Asian American. Demographic data and clinical characteristics of these patients are presented in Dataset S1, Table S2.

To gain insight into the interaction of HPV with the host genome, we determined if the HPV genome was integrated into the tumor genome using whole-genome sequencing (WGS; 6–8× or 30–60× target coverage) together with RNA sequencing. We

considered the viral genome integrated if there were at least three discordant pairs (in which one end of the paired end read mapped to the viral genome, and its mate pair mapped to human genome) and one split read (in which one end of the paired end read spanned the junction, and its mate pair mapped to either the human or HPV genome), and these seven total reads supported integration at the same locus. Such criteria allow for detection of integrations present in a single copy per cell with 6–8× coverage WGS if the majority of tumor cells in the sample derived from the clone containing integration, although higher coverage may be needed to detect subclonal integration events, the pathologic significance of which is unclear. Of 35 cases examined, 25 cases had integration of the viral genome into 1–16 regions in the human genome, with 103 breakpoints in total (Fig. 1 and Dataset S1, Table S2). The presence of reads spanning the insertion site allowed us to determine the sites of integration with single-nucleotide precision. Of more than one-half of integration breakpoints, 60% occurred in regions of microhomology (1–10 bp) among the viral and human genomes and were nonrandom given that, in all cases but one, integration left at least HPV *E6* or *E7* intact.

Of 103 integration events, we found that, in 56 (54%) cases, the virus integrated into a known gene, and in 19 (17%) cases, integration occurred within 20 kb of a gene, suggesting a selective pressure for viral integration into or near genes ( $P = 0.0029$ ). We also noted a spatial correlation among HPV breakpoints and miRNAs ( $P = 0.0015$ ). Viral integration was associated with increases in somatic DNA copy number of the integrated region; 84 (82%) HPV breakpoints colocalized with somatic copy number variants (CNVs;  $P < 0.0001$ ), and the rest were at least 1 Mb from a CNV. Some colocalized CNVs were amplification events seeming to result from excision of the integrated virus along with human sequences with subsequent circularization, suggestive of maintenance as an episome containing the viral replication origin. Tumors with HPV integration displayed a nonsignificant trend toward a higher mutational burden (Fig. S1B).

In an exploratory analysis, we correlated the presence or absence of HPV integration in the host genome with expression of HPV genes and several clinical variables, including anatomic site of tumor, tumor stage (TNM), age, and smoking status of the subject. No statistically significant correlations among HPV integration and these clinical variables were identified in this relatively small dataset (Fig. S1C). Furthermore, there was no statistical association of HPV integration status with clinical outcome (Fig. S1A). Integration-negative tumors displayed a trend toward higher levels of HPV *E2/E5* expression and lower levels of *E6/E7* expression compared with integration-positive tumors (Fig. S2A) ( $P = 0.002$ ,  $P = 0.05$ ,  $P = 0.002$ , and  $P = 0.004$ , respectively). We did not



**Fig. 1.** Distribution of integration sites across HPV genome and human genome. Integration breakpoints are shown for the HPV16-positive tumors. Two breakpoints involving human GL00220.1 are excluded. Breakpoint colors correspond to HPV genes where integration event occurred. The HPV genome scheme includes early and late genes, the viral origin of replication *ori*, the early viral promoter (*p97*), and the late differentiation-dependent promoter (*p670*).

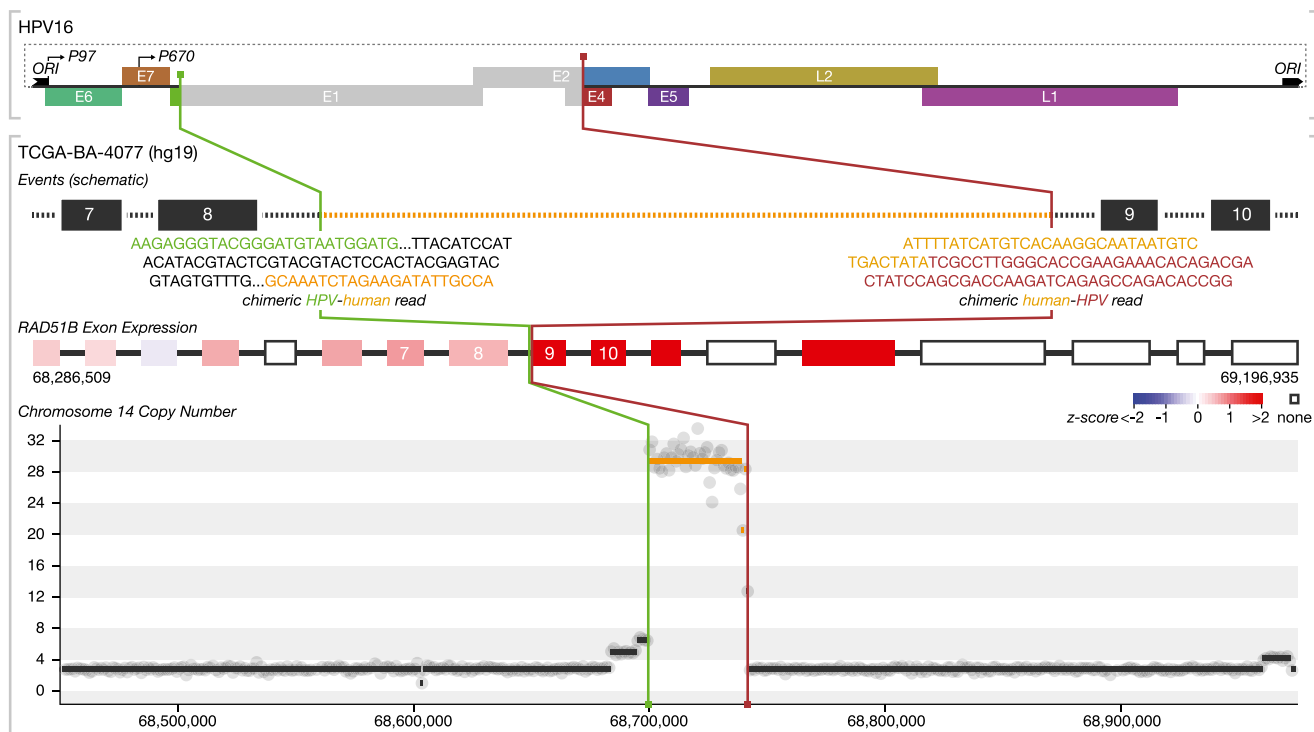
observe a correlation among the presence of integration within specific HPV genes and their expression level. One possible reason for this lack of correlation is the presence of intact copies of HPV genome in the same samples.

**HPV Integrations Are Associated with Somatic Alterations of Key Cancer Genes.** Because the genomic localization of viral integration appeared nonrandom, we hypothesized that the integration events had an impact on cancer development. Detailed analysis of the genes and the insertion sites revealed several distinct possible mechanisms by which the viral insertion may confer a selective advantage. One case involved integration of the viral genome into the DNA repair protein RAD51 homolog 2 (*RAD51B*) gene on chromosome 14 (Fig. 2). In this tumor, we identified three viral integration breakpoints within this gene, disrupting the *E1*, *E4*, or *E5* genes of the virus. All three integration events occurred within intron 8 of the gene. After integration, a 42-kb segment of intron 8 and the viral genome amplified to 28-fold, consistent with the possibility that this amplified chimeric segment is episomal. Extrachromosomal location of the amplified *RAD51B* region was supported by FISH analysis (Fig. S3). Examination of RNA expression data revealed that transcripts corresponding to exons 9–11 and 13 of *RAD51B* were significantly elevated, generating alternative transcripts that were unlikely to produce a functional RAD51B protein. RAD51B is an important component of the DNA double-strand break repair pathway, and loss-of-function variants in this gene have been reported in uterine leiomyoma and breast cancer (34).

A second mechanism by which the viral insertion may be involved in tumor development is illustrated by the insertion of HPV16 into the v-ets avian erythroblastosis virus E26 oncogene homolog 2 (*ETS2*) gene on chromosome 21 (Fig. S4A). In this

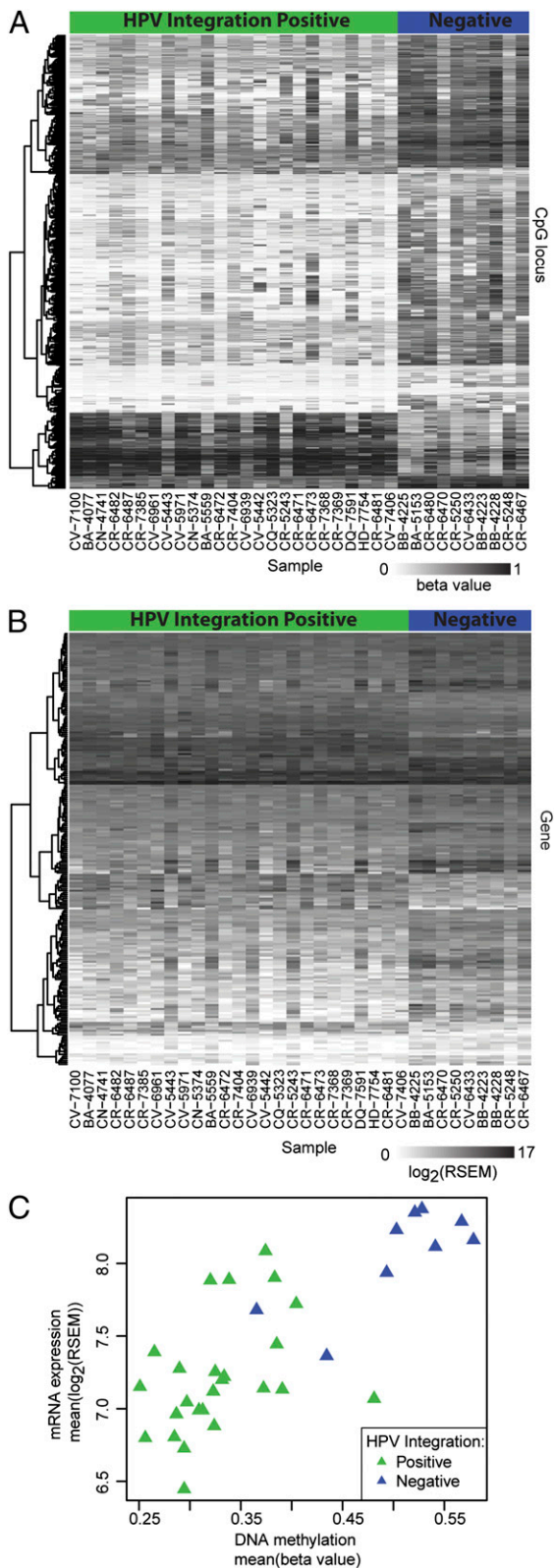
case, the integration resulted in the replacement of exons 7 and 8 of the *ETS2* gene by HPV 16. The second copy of the *ETS2* gene appeared intact. RNA expression data revealed that the overall level of expression of *ETS2* was unchanged in this tumor, although the levels of exons 7 and 8 were markedly reduced, suggesting that exon skipping occurs as a result of viral integration. *ETS2* is a tumor suppressor gene (35), and previous reports have shown that truncated ETS family proteins may act in a dominant negative fashion (36). Our findings suggest that insertion of viral sequences into one of two copies could result in the loss of function of this gene. A similar effect of HPV integration was seen in the case of the *CD274* [programmed death ligand 1 (*PDL1*)] gene on chromosome 9 (Fig. S5). In this case, the virus integrated into an intron flanked by exons 4 and 5, and expression levels of exons 5–7 were attenuated, suggesting that viral integration into this locus resulted in an altered form of *PDL1* that is expressed at high levels. Alternative *PDL1* transcripts have been reported (37, 38) and associated with poor prognosis in renal cell carcinoma (37).

A third mechanism was illustrated by integration of HPV 30 kb upstream of the nuclear receptor subfamily 4, group A, member 2 (*NR4A2*) gene (Fig. S4B). The insertion was followed by 248-fold amplification of a 75-kb genomic region including the entire *NR4A2* gene. RNA-sequencing (RNA-seq) analysis revealed that the level of expression of the *NR4A2* gene was significantly elevated (Fig. S4B). *NR4A2* is a member of the nuclear transcription factor gene family and has been reported as an oncogene (39, 40). Our results suggest that the HPV integration in this sample and subsequent amplification resulted in overexpression of the *NR4A2* oncogene. RNA-seq analysis of HPV expression showed low levels of E6 and E7 transcripts in this tumor (Fig. S2A), suggesting that this integration event may



**Fig. 2.** Integrated analysis of HPV integration events. Breakpoint locations and joining patterns are shown with regard to schematic representations of the HPV and human genomes, regions of copy number alteration, and exon expression of genes involved and/or located near integration sites. Colored regions on the genome schemes represent sequences included in the integrated structure. Gray regions represent sequences that were replaced or lost as a result of integration. *RAD51B* integration leads to 28-fold amplification of the intronic region and overexpression of exons located downstream of the integrated virus. The integrated HPV genome retained oncogenes E6 and E7 (both promoters and *ori*).





**Fig. 3.** DNA methylation and gene expression analysis. (A and B) Heat map of DNA methylation levels for significantly differentially methylated CpG loci and mRNA expression of significantly differentially expressed genes (rows) for all HPV-positive samples (columns). The columns of the heat maps are organized by HPV integration status: samples with and without virus integration are marked at the top of the heat maps in green and blue, respectively. (C) Scatter plot of averaged significantly differential methylation

be of greater importance in this tumor than the impact of the HPV oncoproteins.

A fourth example is represented by HPV insertion in the region of an interchromosomal translocation between chromosomes 3 and 13. HPV was integrated in a nongenic region, but the rearranged region contained the tumor protein p63 regulated 1 (*TPRG1*) and tumor protein p63 (*TP63*) loci on chromosome 3 and the Kruppel-like factor 5 (*KLF5*) gene on chromosome 13 (Fig. S4C). The integration occurred near the *KLF5* promoter and about 150 bp downstream of the translocation junction. Genomic regions involved in this complex event were amplified. RNA-seq data showed increased levels of expression of *KLF5*, *TP63*, and *TPRG1*. *KLF5* (41) and *TP63* are transcription factors known to regulate cellular proliferation and tumor formation, and *TP63* has a documented role in the pathogenesis of squamous cell carcinomas (11, 42). Our results suggest that viral integration was associated with a structural rearrangement and that these events resulted in the overexpression of these oncogenes.

#### HPV Integration Is Associated with a Specific Methylation Signature.

Previous work has shown that HPV-positive tumors have a distinct methylation profile compared with HPV-negative tumors (43, 44). To study the influence of HPV integration events on the host cell, we compared the patterns of DNA methylation and gene expression of HNSCCs with and without integration. Using a linear model of differential methylation and expression for each cytosine-phosphate-guanine (CpG) locus and mRNA, respectively, we found subsets of genes with significantly distinct patterns of methylation and expression among tumors with and without integrated HPV (Fig. 3 and Dataset S1, Tables S3 and S4). Unsupervised clustering of these subsets (Fig. S6) for all studied tumors and adjacent normal tissue controls revealed that the patterns of DNA methylation and gene expression profiles of integration-positive tumors are similar to those of HPV-negative tumors and normal tissue, whereas integration-negative tumors differed from the other three groups. There were no associations of distinct methylation patterns with clinical variables, with the exception of weak correlations with tumor site ( $P = 0.038$ ) and E7 expression ( $P = 0.045$ ) (Fig. S7B).

Four genes among those found in both differentially methylated and expressed subsets are of particular interest. Suppressors of tumor progression *BARX2*, which encodes a homeobox transcription factor, and *IRX4*, a homeodomain-containing transcription factor of the Iroquois family (45, 46), were hypermethylated and underexpressed in the integration-negative tumors (Fig. S8 A and B). Conversely, a single-minded family bHLH transcription factor 2 (*SIM2*) and an intracellular proteinase *CTSE*, both known to be associated with tumorigenesis and cancer progression (47, 48), displayed hypomethylation and increased expression in integration-negative tumors (Fig. S8 C and D). Recent studies have reported hypermethylation and inversely correlated expression of *IRX4* in HPV-driven oropharyngeal tumors compared with non-HPV-driven tumors, and the hypermethylation signature correlated with better clinical outcomes (49). Using Gene Set Enrichment Analysis, we observed differential methylation of genes in the cAMP response element-binding protein (CREB) pathway and differential expression of genes involved in regulation of apoptosis and NF- $\kappa$ B signaling. These data suggest that the biology of HPV-integrated vs. nonintegrated HNCs may be distinct and that HPV may drive changes in cellular gene expression when HPV presents in the episomal form.

and expression values. One HPV(+)/integration(–) sample did not have RNA-seq data and is not represented on the expression heat map and scatter plot.

## Discussion

Here, we show by WGS, transcriptome, and DNA methylation analyses of 35 HPV-associated HNCs that HPV is present in both integrated and nonintegrated forms in these cancers. We show that, in addition to viral E6 and E7 expression, HPV is likely to drive tumorigenesis by alteration of the host genome at sites of integration. Mechanisms by which integration affects host genes include generation of altered transcripts, disruption of tumor suppressor genes, high-level DNA amplifications, and interchromosomal rearrangements. Such mechanisms revealed for primary tumors are in agreement with those observed by Akagi et al. (12) in their study of HNC cell lines.

It is noteworthy that not all tumors with integrated HPV showed enhanced expression of viral E6 and E7 oncoproteins. In fact, at least seven samples in which viral integration occurred showed an HPV expression profile similar to the integration-negative tumors. Although elevated levels of E6 and E7 proteins are considered essential for development of the HPV-associated malignant phenotype (12, 50), our observations raise the question of whether, in some cases, HPV infection and subsequent integration can drive tumorigenesis because of HPV-induced perturbations of the host genome independent of the activities of E6 and E7. Such a possibility could be illustrated by the tumor that showed significant amplification and overexpression of the *NR4A2* oncogene without evidence of elevated E6 or E7 expression. Our studies suggest that the actions of HPV in the process of tumorigenesis likely extend beyond the expression of viral oncoproteins (at least in a subset of HNSCCs) and suggest a more nuanced role for the virus than the conventional view of E6/E7-mediated transformation by disruption of predominantly the *TP53* and *CDKN2A/RB1* axes. It may also be the case that the high prevalence of HPV infection worldwide compared with HPV-driven cancers is explained by a requirement for additional genomic events downstream of HPV infection to drive tumor formation. Although many host genome alterations, most notably PI3K pathway mutations, have been shown to be enriched in HPV-positive HNSCCs and cooperate in oncogenesis (14, 51), our study suggests that additional interactions of HPV itself with the host genome are likely to be relevant in the oncogenic process. Although the use of preventative HPV vaccines has gained traction worldwide, we believe that attention should be also focused on the development of therapeutic HPV vaccines for individuals with HPV-driven HNSCCs, which is currently being explored by the Stand Up to Cancer Initiative ([www.standup2cancer.org](http://www.standup2cancer.org)) and other investigators.

We observed the occurrence of integration breakpoints across the entire viral genome; only in E1 gene, however, there were more breakpoints than would have been expected by chance (Fig. S2B). Sites of integration in the tumor genome were non-random, suggesting that there is likely to be selection for particular clones within a tumor based on sites of HPV integration that drive a survival and/or proliferation advantage, and we observed that the molecular architectures of tumors with integrated and nonintegrated HPV are likely distinct. Comparison with a recent study in cervical cancer (52) revealed some similarities, with HPV integration events targeting *RAD51B*, *NR4A2*, and *TP63*. In addition, several HPV integration sites seem to be

unique to each dataset, but these differences may result from the fact that the cervical cancer data were mainly derived from transcriptome sequencing as opposed to the WGS data in this study. It is likely that the size of our cohort is limiting in this analysis, and we believe that a much larger cohort of HPV-driven HNSCCs will need to be studied to fully account for patterns of integration, address recurrence of integration in key genes, and understand implications of integration on the biology and pathogenesis of HPV-driven HNCs and the association with clinical outcomes.

Distinct host genome methylation and expression patterns in HPV integration-positive and -negative HNSCCs are additional findings that may shed some new light on the possible mechanisms by which HPV infection could facilitate or drive carcinogenesis independent of integration. Here, we present data describing a small number of differentially methylated and expressed genes in tumors with integrated and nonintegrated HPV, including *IRX4* and *SIM2*, genes that have shown roles in cancer biology (53). However, this hypothesis requires additional exploration to establish the roles of these genes and others that display distinct methylation and expression profiles in HPV-positive tumors with and without HPV integration.

Our study, using DNA sequencing, RNA-seq, and DNA methylation data, highlights previously undescribed roles of HPV in primary HNCs beyond production of the E6 and E7 viral oncoproteins and suggests that HPV is likely to drive HNCs through a variety of mechanisms, including expression of E6 and E7, integration in key cancer-related genes, changes in global DNA methylation and gene expression, and facilitation of genomic instability. Our study also provides evidence for increased sensitivity of next generation sequencing methods for HPV detection compared with p16 immunohistochemistry or HPV FISH and shows evidence of HPV in tumors with atypical clinical features, such as larynx origin or a background of heavy tobacco use. Given that HPV status is used for risk stratification for patients undergoing therapy for HNSCC, we feel that it is important to continue to refine methods for HPV detection to both better enable preclinical studies of the role of HPV in HNSCCs and provide accurate information to clinicians who are treating patients with HNSCCs.

## Materials and Methods

The presence of HPV and its integration into tumor genome was detected using WGS (30–60×,  $n = 28$  and 6–8×,  $n = 122$ ), RNA-seq, whole-exome sequencing, and molecular approaches (a PCR-based MassArray Assay, p16 staining, and HPV in situ hybridization). A subset of samples ( $n = 129$ ), which lacked evidence of HPV on the basis of RNA-seq, was not analyzed by WGS. Both human and HPV gene expressions were assessed using the Illumina TrueSeq protocol and Illumina HiSeq sequencers, with a target coverage of 60 million read pairs. Whole-genome methylation was assayed using the Infinium HM4501 array. CNV was analyzed using the Bayesian Information Criterion-sequencing algorithm.

Full methods and any associated references are available in *SI Materials and Methods*.

**ACKNOWLEDGMENTS.** This study was funded, in part, by National Cancer Institute Grants U24 CA143867 (to M.M.), K08 CA163677 (to P.S.H.), and U24 CA144025 (to R.K.). L.C. is a Cancer Prevention and Research Institute of Texas Scholar in Cancer Research.

- Joseph AW, D'Souza G (2012) Epidemiology of human papillomavirus-related head and neck cancer. *Otolaryngol Clin North Am* 45(4):739–764.
- O'Rorke MA, et al. (2012) Human papillomavirus related head and neck cancer survival: A systematic review and meta-analysis. *Oral Oncol* 48(12):1191–1201.
- D'Souza G, et al. (2007) Case-control study of human papillomavirus and oropharyngeal cancer. *N Engl J Med* 356(19):1944–1956.
- Gillison ML, et al. (2000) Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J Natl Cancer Inst* 92(9):709–720.
- Goon PK, et al. (2009) HPV & head and neck cancer: A descriptive update. *Head Neck Oncol* 1:36.
- Rautava J, Syrjänen S (2012) Biology of human papillomavirus infections in head and neck carcinogenesis. *Head Neck Pathol* 6(Suppl 1):S3–S15.
- Koch WM, Lango M, Sewell D, Zahurak M, Sidransky D (1999) Head and neck cancer in nonsmokers: A distinct clinical and molecular entity. *Laryngoscope* 109(10):1544–1551.
- Chaturvedi AK, et al. (2011) Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol* 29(32):4294–4301.
- Gillison ML, et al. (2008) Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. *J Natl Cancer Inst* 100(6):407–420.
- Agrawal N, et al. (2011) Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* 333(6046):1154–1157.

11. Stransky N, et al. (2011) The mutational landscape of head and neck squamous cell carcinoma. *Science* 333(6046):1157–1160.
12. Akagi K, et al. (2014) Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res* 24(2):185–199.
13. Walter V, et al. (2013) Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes. *PLoS ONE* 8(2):e56823.
14. Lechner M, et al. (2013) Targeted next-generation sequencing of head and neck squamous cell carcinoma identifies novel genetic alterations in HPV+ and HPV- tumors. *Genome Med* 5(5):49.
15. Wilson GA, et al. (2013) Integrated virus-host methylome analysis in head and neck squamous cell carcinoma. *Epigenetics* 8(9):953–961.
16. Munger K, Phelps WC, Bubb V, Howley PM, Schlegel R (1989) The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. *J Virol* 63(10):4417–4421.
17. de Freitas AC, Coimbra EC, Leitao MdaC (2014) Molecular targets of HPV oncoproteins: Potential biomarkers for cervical carcinogenesis. *Biochim Biophys Acta* 1845(2): 91–103.
18. Tommasino M (2014) The human papillomavirus family and its role in carcinogenesis. *Semin Cancer Biol* 26:13–21.
19. Grabowska AK, Riemer AB (2012) The invisible enemy - how human papillomaviruses avoid recognition and clearance by the host immune system. *Open Virol J* 6:249–256.
20. DiMaio D, Petti LM (2013) The E5 proteins. *Virology* 445(1-2):99–114.
21. Wentzensen N, Vinokurova S, von Knebel Doeberitz M (2004) Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. *Cancer Res* 64(11):3878–3884.
22. Moody CA, Laimins LA (2010) Human papillomavirus oncoproteins: Pathways to transformation. *Nat Rev Cancer* 10(8):550–560.
23. Korzeniewski N, Spardy N, Duensing A, Duensing S (2011) Genomic instability and cancer: Lessons learned from human papillomaviruses. *Cancer Lett* 305(2):113–122.
24. Pett MR, et al. (2004) Acquisition of high-level chromosomal instability is associated with integration of human papillomavirus type 16 in cervical keratinocytes. *Cancer Res* 64(4):1359–1368.
25. Peter M, et al. (2010) Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma. *J Pathol* 221(3):320–330.
26. Peter M, et al. (2006) MYC activation associated with the integration of HPV DNA at the MYC locus in genital tumors. *Oncogene* 25(44):5985–5993.
27. Schmitz M, et al. (2012) Loss of gene function as a consequence of human papillomavirus DNA integration. *Int J Cancer* 131(5):E593–E602.
28. Kraus I, et al. (2008) The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes. *Cancer Res* 68(7): 2514–2522.
29. Xu B, et al. (2013) Multiplex Identification of Human Papillomavirus 16 DNA Integration Sites in Cervical Carcinomas. *PLoS ONE* 8(6):e66693.
30. Durst M, Croce CM, Gissmann L, Schwarz E, Huebner K (1987) Papillomavirus sequences integrate near cellular oncogenes in some cervical carcinomas. *Proc Natl Acad Sci USA* 84(4):1070–1074.
31. Kim SH, et al. (2007) HPV integration begins in the tonsillar crypt and leads to the alteration of p16, EGFR and c-myc during tumor formation. *Int J Cancer* 120(7):1418–1425.
32. Hafkamp HC, et al. (2003) A subset of head and neck squamous cell carcinomas exhibits integration of HPV 16/18 DNA and overexpression of p16INK4A and p53 in the absence of mutations in p53 exons 5-8. *Int J Cancer* 107(3):394–400.
33. Kostic AD, et al. (2011) PathSeq: Software to identify or discover microbes by deep sequencing of human tissue. *Nat Biotechnol* 29(5):393–396.
34. Suwaki N, Klare K, Tarsounas M (2011) RAD51 paralogs: Roles in DNA damage signalling, recombinational repair and tumorigenesis. *Semin Cell Dev Biol* 22(8):898–905.
35. Kabbout M, et al. (2013) ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer. *Clin Cancer Res* 19(13):3383–3395.
36. Kumar-Sinha C, Tomlins SA, Chinnaiyan AM (2008) Recurrent gene fusions in prostate cancer. *Nat Rev Cancer* 8(7):497–511.
37. Frigola X, et al. (2011) Identification of a soluble form of B7-H1 that retains immunosuppressive activity and is associated with aggressive renal cell carcinoma. *Clin Cancer Res* 17(7):1915–1923.
38. Steidl C, et al. (2011) MHC class II transactivator CIITA is a recurrent gene fusion partner in lymphoid cancers. *Nature* 471(7338):377–381.
39. Imielinski M, et al. (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 150(6):1107–1120.
40. Ke N, et al. (2004) Nuclear hormone receptor NR4A2 is involved in cell transformation and apoptosis. *Cancer Res* 64(22):8208–8212.
41. Nandan MO, et al. (2008) Kruppel-like factor 5 mediates cellular transformation during oncogenic KRAS-induced intestinal tumorigenesis. *Gastroenterology* 134(1): 120–130.
42. Cancer Genome Atlas Research Network (2012) Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 489(7417):519–525.
43. Lechner M, et al. (2013) Identification and functional validation of HPV-mediated hypermethylation in head and neck squamous cell carcinoma. *Genome Med* 5(2):15.
44. Lleras RA, et al. (2013) Unique DNA methylation loci distinguish anatomic site and HPV status in head and neck squamous cell carcinoma. *Clin Cancer Res* 19(19): 5444–5455.
45. Sellar GC, et al. (2001) BARX2 induces cadherin 6 expression and is a functional suppressor of ovarian cancer progression. *Cancer Res* 61(19):6977–6981.
46. Nguyen HH, et al. (2012) IRX4 at 5p15 suppresses prostate cancer growth through the interaction with vitamin D receptor, conferring prostate cancer susceptibility. *Hum Mol Genet* 21(9):2076–2085.
47. Lu B, Asara JM, Sanda MG, Arredouani MS (2011) The role of the transcription factor SIM2 in prostate cancer. *PLoS ONE* 6(12):e28837.
48. Konno-Shimizu M, et al. (2013) Cathepsin E is a marker of gastric differentiation and signet-ring cell carcinoma of stomach: A novel suggestion on gastric tumorigenesis. *PLoS ONE* 8(2):e56766.
49. Kostareli E, et al. (2013) HPV-related methylation signature predicts survival in oropharyngeal squamous cell carcinomas. *J Clin Invest* 123(6):2488–2501.
50. Pett M, Coleman N (2007) Integration of high-risk human papillomavirus: A key event in cervical carcinogenesis? *J Pathol* 212(4):356–367.
51. Lui VW, et al. (2013) Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. *Cancer Discov* 3(7):761–769.
52. Ojesina AI, et al. (2014) Landscape of genomic alterations in cervical carcinomas. *Nature* 506(7488):371–375.
53. Su Y, et al. (2014) MiR-200a impairs glioma cell growth, migration, and invasion by targeting SIM2-s. *Neuroreport* 25(1):12–17.