

Variations in HPV function are associated with survival in squamous cell carcinoma

Frederico O. Gleber-Netto,¹ Xiayu Rao,² Theresa Guo,³ Yuanxin Xi,² Meng Gao,¹ Li Shen,² Kelly Erikson,¹ Nene N. Kalu,⁴ Shuling Ren,⁵ Guorong Xu,⁵ Kathleen M. Fisch,⁶ Keiko Akagi,⁴ Tanguy Seiwert,⁷ Maura Gillison,⁴ Mitchell J. Frederick,⁸ Faye M. Johnson,^{4,9} Jing Wang,^{2,9} Jeffrey N. Myers,^{1,9} Joseph Califano,⁵ Heath D. Skinner,¹⁰ and Curtis R. Pickering^{1,9}

¹Department of Head and Neck Surgery and ²Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA. ³Department of Otolaryngology – Head and Neck Surgery, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA. ⁴Department of Thoracic/Head & Neck Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA. ⁵Division of Otolaryngology – Head and Neck Surgery, Department of Surgery, Moores Cancer Center, University of California, San Diego, La Jolla, California, USA. ⁶Center for Computational Biology & Bioinformatics, Department of Medicine, University of California, San Diego, La Jolla, California, USA. ⁷Department of Medicine and Comprehensive Cancer Center, University of Chicago, Chicago, Illinois, USA. ⁸Department of Otolaryngology, Baylor College of Medicine, Houston, Texas, USA. ⁹The University of Texas Graduate School of Biomedical Sciences, Houston, Texas, USA. ¹⁰Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA.

Incidence of HPV⁺ oropharyngeal squamous cell carcinoma (OPSCC) has been increasing dramatically. Although long-term survival rates for these patients are high, they often suffer from permanent radiotherapy-related morbidity. This has prompted the development of de-escalation clinical protocols to reduce morbidity. However, a subset of patients do not respond even to standard therapy and have poor outcomes. It is unclear how to properly identify and treat the high- and low-risk HPV⁺ OPSCC patients. Since HPV positivity drives radiotherapy sensitivity, we hypothesized that variations in HPV biology may cause differences in treatment response and outcome. By analyzing gene expression data, we identified variations in HPV-related molecules among HPV⁺ OPSCC. A subset of tumors presented a molecular profile distinct from that of typical HPV⁺ tumors and exhibited poor treatment response, indicating molecular and clinical similarities with HPV⁻ tumors. These molecular changes were also observed in vitro and correlated with radiation sensitivity. Finally, we developed a prognostic biomarker signature for identification of this subgroup of HPV⁺ OPSCC and validated it in independent cohorts of oropharyngeal and cervical carcinomas. These findings could translate to improved patient stratification for treatment deintensification and new therapeutic approaches for treatment-resistant HPV-related cancer.

Introduction

HPV is an etiological factor for human tumors in multiple squamous tissue sites. Cervical squamous cell carcinoma (CESCC) is the most prevalent HPV-related tumor among woman, but incidence of HPV-related oropharyngeal squamous cell carcinoma (OPSCC) has increased in recent decades and is now the most prevalent HPV-related cancer in the United States (1, 2). HPV⁺ head and neck squamous cell carcinoma (HNSCC) is considered a separate oncological entity from its HPV⁻ counterpart that is largely tobacco driven. Its distinct etiology results in unique clinical characteristics, including affecting younger nonsmokers (3). The most remarkable clinical difference is a higher rate of response to radiation-based treatment among HPV⁺ HNSCCs (4, 5). Nearly all OPSCC patients present with locally advanced disease and are treated with a combination of chemotherapy and radiotherapy. While response and survival rates are very high, the symptom burden after treatment with these modalities can be significant and permanent and can include diminished swallowing function, xerostomia, altered appearance, and trismus (6). Therefore, there has been an emphasis on reducing radiation exposure for low-risk patients without impairing survival outcomes. However, given that current therapy fails in 15% of HPV⁺ OPSCCs with

Conflict of interest: The authors have declared that no conflict of interest exists.

License: Copyright 2019, American Society for Clinical Investigation.

Submitted: September 7, 2018

Accepted: November 27, 2018

Published: January 10, 2019

Reference information:

JCI Insight. 2019;4(1):e124762.

<https://doi.org/10.1172/jci.insight.124762>.

insight.124762.

locoregional recurrence or distant metastatic disease, treatment deintensification is not appropriate for all patients (7). Clinical predictors of outcome are crude and include tumor stage and smoking history. Currently, there is an unmet need for reliable clinical tools or biomarkers for risk stratification in HPV⁺ OPSCC that could be used to improve patient outcomes.

Although extensive research efforts have addressed the molecular mechanisms involved in HPV-mediated carcinogenesis (8), little is known about the role of HPV in modulating tumor clinical parameters, such as progression and treatment response. Cancer-related HPV gene expression and mechanistic studies have focused mainly on the E6 and E7 oncogenes. However, the HPV genome contains significant complexity in other genes, which may provide important biological information for understanding clinical variations among HPV-mediated tumors (9–11). In this study, our objective was to explore the association between HPV function and molecular and clinical phenotypes. We hypothesized that changes in HPV function would have an important role in tumor progression and response to treatment. As a measure of HPV function, we identified human genes correlated with HPV gene expression.

In this study, we determined that expression of HPV-correlated tumor genes characterized groups of patients with distinct tumor biology and clinical outcome. Dysfunctional HPV activity was associated with worse survival in patients and with lower sensitivity to radiation *in vitro*. These genes also showed promising prognostic power and were successfully validated in independent cohorts of HPV⁺ squamous cell carcinomas from oropharynx and cervix.

Results

A measure of HPV function stratifies OPSCC tumors into groups with distinct HPV biology and survival. Our first goal was to develop a measure of HPV function from tumor RNA expression data. To accomplish, this we considered HPV status as a continuous variable based on total expression level of HPV RNA and identified human genes correlated with the range of viral expression. We performed Pearson correlation analysis ($r \leq -0.6$ or ≥ 0.6 ; adjusted $P < 0.05$) in all 80 OPSCC tumor tissues from The Cancer Genome Atlas (TCGA) (12) and identified 582 HPV-correlated human transcripts (Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.124762DS1>) that were also highly differentially expressed between HPV⁺ and HPV⁻ OPSCC cases (Supplemental Table 1 and Supplemental Figure 1). Ingenuity Pathway Analysis (IPA) indicated that positively correlated genes (those with expression levels that vary but follow the same trend as the expression levels of HPV genes) enriched cellular functions and canonical pathways, which are highly associated with the HPV functions, including cell proliferation and survival and DNA replication and repair (Supplemental Tables 2 and 3). Negatively correlated genes were associated with cell movement, survival, and cell-to-cell signaling (Supplemental Tables 4 and 5).

The expression pattern of these 582 HPV-correlated genes, evaluated by hierarchical clustering, revealed 3 distinct sample clusters, which we termed OPSCC C1, OPSCC C2, and OPSCC C3 (Figure 1A). OPSCC C1 and C2 comprised primarily HPV⁺ cases, and OPSCC C3 comprised exclusively HPV⁻ cases. OPSCC C2 and OPSCC C3 groups showed clearly distinct expression profiles, while differences in expression between OPSCC C1 and the other groups were less dramatic, suggesting an intermediate gene expression state. Analysis of the relative expression levels of these genes in OPSCC C1 compared with the other 2 groups found that 50% of the genes had similar expression to that of OPSCC C2 (Figure 1B). However, 44% of the genes showed expression levels between those of OPSCC C2 and OPSCC C3. Differences among these groups were further confirmed by principal component analysis (PCA) (Figure 1C).

We next tested whether there was any difference in survival between these groups. As expected, since OPSCC C3 comprised HPV⁻ cases, it had the lowest 5-year overall survival (OS) rate of the 3 clusters (13). However, a significant difference in 5-year OS was also observed between the OPSCC C1 and C2 clusters ($P < 0.001$, Figure 1D). This difference was observed even after removing the 2 HPV⁻ cases from OPSCC C1 ($P < 0.001$).

To further explore the differences between OPSCC C1 and OPSCC C2, we investigated if there were HPV-correlated genes differentially expressed specifically between these groups, while excluding the HPV⁻ cases; this analysis identified 38 HPV-correlated genes with highly differential expression ($FDR < 0.05$ and \log_2 fold change > 1.0) (Supplemental Table 6). Hierarchical clustering demonstrated a clear differential expression of these genes between 2 clusters that we named HPV⁺ C1 and HPV⁺ C2 (Figure 1E). Groups formed by this set of genes were nearly identical to those defined by the 582 genes (i.e., HPV⁺ C1 is similar to OPSCC C1 and HPV⁺ C2 is similar to OPSCC C2 group) and exhibited a similarly significant difference in survival ($P < 0.001$; Figure 1F). PCA (Figure 1G) and relative gene expression based on these 38

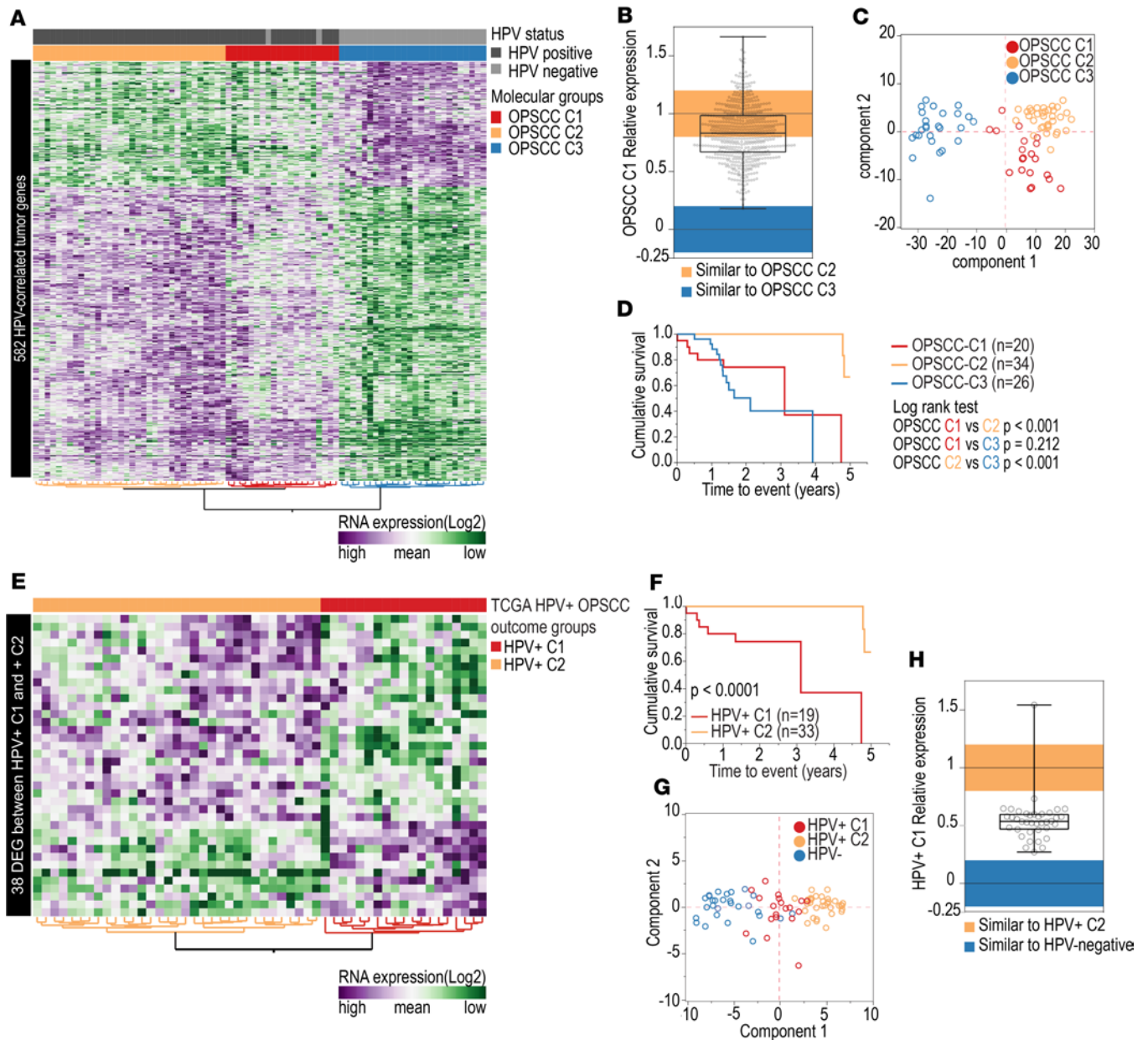


Figure 1. Expression of HPV-correlated human tumor genes among OPSCC samples. (A) Expression of 582 HPV-correlated genes (vertical axis) among 80 OPSCC patients (horizontal axis) from TCGA. Three groups with distinct gene expression were identified: OPSCC C1 ($n = 20$, 2 HPV⁻ and 18 HPV⁺), C2 ($n = 34$, all HPV⁺), and C3 ($n = 26$, all HPV⁺). (B) Expression of 582 HPV-correlated genes in OPSCC C1 ($n = 20$) group relative to OPSCC C2 ($n = 34$) and to OPSCC C3 ($n = 26$) groups. Vertical axis represents the product of fold change (\log_2) between OPSCC C1 and OPSCC C3 and between OPSCC C2 and C3. Values close to 0 (-0.2 to 0.2, blue) indicate OPSCC C1 gene expression similar to OPSCC C3 mean. Values close to 1 (0.8 to 1.2, orange) indicate OPSCC C1 gene expression similar to OPSCC C2 mean. Gene expression in OPSCC C1 group was similar to OPSCC C2 in 50% of cases ($n = 291$) and was intermediate between OPSCC C2 and C3 in 44% ($n = 256$) of cases. (C) PCA confirmed the different expression profiles among the groups of OPSCC ($n = 80$). (D) Survival curves indicate that the OPSCC C1 (7 deaths) and OPSCC C3 (13 deaths) groups had significantly lower survival rates in the 5-year follow-up period compared with OPSCC C2 group (2 deaths). (E) Expression of 38 HPV-correlated genes (differentially expressed between OPSCC C1 and OPSCC C2) (vertical axis) among TCGA patients with HPV⁺ OPSCC ($n = 52$) shows 2 groups of patients, HPV⁺ C1 ($n = 19$) and HPV⁺ C2 ($n = 33$). (F) Survival curves show that the HPV⁺ C1 group had a significantly lower 5-year survival rate than the HPV⁺ C2 group (log-rank test). (G) PCA based on expression levels of 38 HPV-correlated genes indicates HPV⁺ C1 ($n = 19$) samples have an intermediate expression profile between HPV⁻ ($n = 28$) and HPV⁺ C2 ($n = 33$) samples. (H) Expression of 38 HPV-correlated genes in HPV⁺ C1 ($n = 19$) group relative to HPV⁺ C2 ($n = 33$) and HPV⁻ ($n = 19$) groups. Vertical axis represents the product of fold change (\log_2) between HPV⁺ C1 and HPV⁻ and between HPV⁺ C2 and HPV⁻. Values close to 0 (-0.2 to 0.2, blue), indicate HPV⁺ C1 gene expression similar to HPV⁻ cases. Values close to 1 (0.8 to 1.2, orange), indicate HPV⁺ C1 gene expression similar to HPV⁺ C2 mean. Expression levels of 97% ($n = 37$) of these genes in HPV⁺ C1 group were intermediate between HPV⁺ C2 and HPV⁻ groups.

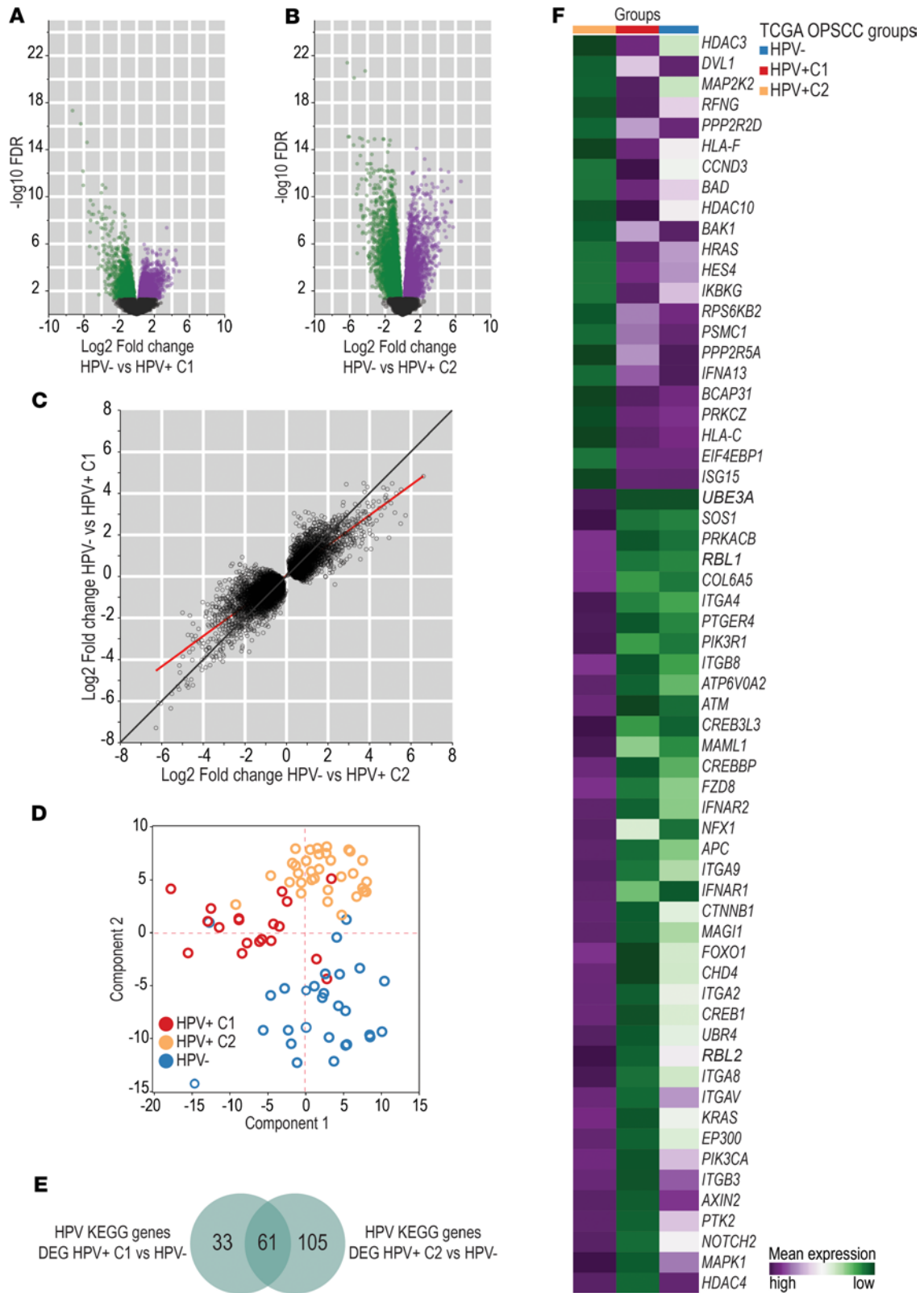


Figure 2. Human transcriptome differences between HPV⁻ C1 and HPV⁻ C2 groups. Volcano plot showing \log_2 fold-change (horizontal axis) and $-\log_{10}$ FDR (vertical axis) values from the whole transcriptome analysis between HPV⁻ ($n = 28$) and HPV⁺ C1 ($n = 19$) OPSCC (A) and between HPV⁻ ($n = 28$) and HPV⁺ C2 ($n = 33$) OPSCC groups (B). Genes were considered differentially expressed at a significance level of 0.05 ($n = 4366$ in A and $n = 8724$ in B). Green dots represent differentially expressed genes with lower expression in HPV⁻ cases than in HPV⁺ C1 cases ($n = 2537$ in A and $n = 5268$ in B). Purple dots represent differentially expressed genes with higher expression in HPV⁻ cases than in HPV⁺ C1 cases ($n = 1829$ in A and $n = 3456$ in B). Black dots represent genes whose expression was not significantly different between groups. (C) 7760 differentially expressed genes between HPV⁺ ($n = 52$) and HPV⁻ ($n = 28$) OPSCC cases and their corresponding \log_2 fold-change values when their expression was compared between HPV⁻ ($n = 28$) and HPV⁺ C1 cases ($n = 19$) (vertical axis) and between HPV⁻ ($n = 28$) and HPV⁺ C2 cases ($n = 33$) (horizontal axis). The red line represents the best-fit calculated by linear regression. The black line represents the 45° line. The line of best-fit slope was lower than 1 (0.72 ± 0.004), indicating greater \log_2 fold-change values in the horizontal axis (HPV⁻ vs. HPV⁺ C2) than in the vertical axis (HPV⁻ vs. HPV⁺ C1). This suggests that differences in expression levels of HPV-related genes were smaller between HPV⁺ C1 and HPV⁻ cases than between HPV⁺ C2 and HPV⁻ cases. (D) PCA using expression of 319 genes from the HPV-KEGG pathway database among all 80 TCGA OPSCC cases. All 3 groups (HPV⁺C1, $n = 19$; HPV⁺C2, $n = 33$; HPV⁻, $n = 28$) showed distinct expression profiles of HPV-related genes. (E) Venn diagram representing HPV-KEGG pathway genes differentially expressed between HPV⁺ C1 ($n = 19$) and HPV⁻ cases ($n = 28$) and between HPV⁺ C2 ($n = 33$) and HPV⁻ ($n = 28$) cases. A greater number of HPV-related genes was differentially expressed between HPV⁺ C2 and HPV⁻ cases. This suggests that the HPV⁺ C1 group exhibited expression of some HPV-related genes similar to that in HPV⁻ cases. (F) Mean expression of 61 HPV-KEGG pathway genes among HPV⁺ C1 ($n = 19$), HPV⁺ C2 ($n = 33$), and HPV⁻ ($n = 28$) cases. These 61 genes were differentially expressed between HPV⁺ C1 and HPV⁺ C2 but not between HPV⁺ C1 and HPV⁻ cases, suggesting that HPV⁺C1 cases lost part of their HPV characteristics.

HPV-correlated genes (Figure 1H) showed that the group with worse survival (HPV⁺ C1) presented an intermediate gene expression profile between HPV⁻ and HPV⁺ tumors with better survival (HPV⁺ C2). These results indicate HPV⁺ C1 tumors present a distinct, and intermediate, expression pattern of genes potentially modulated by HPV function.

Deregulation of HPV-regulated pathways between groups. We further explored this distinct gene expression of HPV⁺ C1 tumors by testing whether the global gene expression profile of HPV⁺ C1 and HPV⁺ C2 groups was different from that of HPV⁻ OPSCC TCGA tumors. HPV⁺ C1 tumors were significantly ($P < 0.0001$) less distinct from HPV⁻ tumors (21.3% of the transcriptome was differentially expressed, FDR < 0.05; Figure 2A) than HPV⁺ C2 tumors (42.5% of the transcriptome was differentially expressed, FDR < 0.05; Figure 2B). In addition, considering the differentially expressed genes between HPV⁻ and all HPV⁺ cases together (including HPV⁺ C1 and C2 as a unique group), the fold-change differences between HPV⁻ and HPV⁺ C1 cases tended to be smaller than those between HPV⁻ and HPV⁺ C2 cases (Figure 2C). Only 45.6% ($n = 3496$) of the differences in expression between HPV⁻ and all HPV⁺ cases were also significant when considering HPV⁺ C1 cases only, while 88.5% ($n = 6792$) were significantly differentially expressed when considering the HPV⁺ C2 group (Supplemental Table 7). These results indicate that the HPV⁺ C1 group was less distinct from HPV⁻ cases than the HPV⁺ C2 group.

We then investigated whether genes known to be directly linked to HPV biology were differentially expressed between these groups. The expression of 319 genes associated with HPV biology in the KEGG pathway database (14) was analyzed among all OPSCC samples by PCA analysis. We found that HPV⁻, HPV⁺ C1, and HPV⁺ C2 samples showed distinct expression profiles among HPV-related genes (Figure 2D). Compared with HPV⁻ cases, about 52% ($n = 166$) of these genes were differentially expressed in HPV⁺ C2 tumors, while only 29.5% ($n = 94$) were differentially expressed in HPV⁺ C1 cases, a significant difference ($P < 0.001$; Supplemental Table 8 and Figure 2E). Beyond that, among the 89 HPV KEGG genes differentially expressed between HPV⁺ C1 and HPV⁺ C2 tumors, 61 (69%) showed similar or not statistically different expression levels between HPV⁺ C1 and HPV⁻ cases (Figure 2F), including protein targets of HPV E7 (*RBL1* and *RBL2*), E6 (*E6AP*, *EP300*, and *CREBBP*), and E5 (*BCAP31* and *ATP6V0A2*), suggesting differences in HPV protein functions between HPV⁺ C1 and HPV⁺ C2 groups.

We further explored differences between HPV⁺ C1 and C2 groups by pathway analysis of the genes differentially expressed between these groups (Supplemental Table 9 and Supplemental Figure 2). By setting a cutoff point of FDR < 0.01, 3368 differentially expressed genes were analyzed through IPA. The most significantly enriched pathways were involved in RNA and protein metabolism, DNA replication and damage repair, cell proliferation, cell metabolism, and cell signaling (Supplemental Tables 10 and 11). Many of the enriched molecular functions are related to important mechanisms modulated by HPV (Supplemental Table 11). We also assessed all 5482 differentially expressed genes (FDR < 0.05) by gene set enrichment analysis (GSEA) (15, 16), which confirmed the enrichment of HPV biology-related pathways (Supplemental Tables 12 and 13). Together, these findings support the hypothesis that a subset of HPV⁺ OPSCC (HPV⁺ C1) have altered HPV function that manifests as clinical behavior and gene expression that is similar to that of HPV⁻ OPSCC.

HPV gene expression levels are different among tumor groups. Since our findings suggest a differential function of HPV between HPV⁺ C1 and C2 groups, we examined differences in expression of HPV genes. We used TCGA HPV16⁺ cases only, since HPV16 was the most prevalent HPV type (Supplemental Table 14), and direct comparison across different HPV types is challenging because of issues with proper alignment of alternatively spliced transcripts. No difference in total HPV gene expression level or HPV integration status was observed between groups (Supplemental Table 14). Given the complex gene splicing and overlapping gene structure of the HPV genome, 3 analytic approaches were employed to examine differences at the nucleotide, gene, and splicing variants levels.

First, expression levels of individual nucleotides were assessed by calculating the mean expression values in the HPV⁺ C1 and C2 groups (Figure 3A). The expression levels between the E1 and E2 exons were lower among HPV⁺ C1, specially between nucleotides 925 and 1161, 1340 and 1368, and 1743 and 2130 (FDR < 0.1).

HPV expression was also measured based on the number of transcripts aligned to each known HPV16 gene irrespective of its isoforms (Supplemental Table 15). Transcription of the E5, E6, and E7 genes was similar between groups (Figure 3, B–D), and only E1 expression was significantly lower within HPV⁺ C1 patients when compared with the HPV⁺ C2 group ($P = 0.036$, FDR = 0.290; Figure 3E). Expression of E2, E4, E5, L1, and L2 trended lower among HPV⁺ C1 samples, but the differences were not statistically significant.

We evaluated the expression levels of splicing isoforms by measuring reads crossing splicing junctions (Supplemental Table 15). We observed a strong difference in E1[^]E4 isoform expression, which was significantly higher among HPV⁺ C2 samples (Figure 3F). Since HPV16 E1[^]E4 shares nt 3358 with the E2[^]E8 isoform, we also evaluated the expression of its splicing junction. No expression of E2[^]E8 was observed. Expression across splice junctions within the E1 gene (nt 880:881 and 1302:1303) was lower in OPSCC C1 as well (Figure 3G), but the magnitude of expression difference was not as strong as for the E1[^]E4 splicing junction. E1[^]E4 expression level was not associated with HPV integration status ($P = 0.129$).

These results indicate significant changes in HPV transcription profile between the groups of tumors with distinct outcomes. Differences were not related to E6, E7, or E5 HPV oncogenes but were characterized by lower expression of E1 and E2 genomic regions. The most striking distinction was the lower expression of E1[^]E4 splicing isoform, suggesting it may have a role in the different molecular and clinical phenotypes observed between these groups.

HPV E1[^]E4 expression associates with altered HPV function and radiation sensitivity in vitro. In order to test whether lower transcription of E1[^]E4 is associated with the phenotypic and molecular changes observed in HPV⁺ C1, we assessed its expression among HPV-infected cell lines. Among 10 HPV16⁺ cell lines from head and neck and cervical sites (17), E1[^]E4 expression was detected in 6 lines while 4 showed no expression. Expression was determined by RNA sequencing (RNASeq) and confirmed by quantitative PCR (qPCR), and we found a high correlation between these detection methods (Supplemental Figure 3).

To test whether the phenotypic differences observed in tumor samples were associated with E1[^]E4 in vitro, we used cell line radiation sensitivity as a surrogate marker of outcome. Radiotherapy response is the main determinant of curability among patients with HPV⁺ OPSCC, and it is probable that HPV⁺ C1 patients presented a worse prognosis due to increased radiation resistance. To test whether E1[^]E4 expression is associated with cell response to ionizing radiation, we irradiated 6 HPV16⁺ cell lines with a 2-Gy dose (4 other HPV16⁺ lines do not readily form colonies in this assay). By correlating the cell line surviving fraction (SF2) (18) with their E1[^]E4 expression, we observed a significant inverse correlation between these two parameters (Pearson $r = -0.88$, $P = 0.007$; Figure 4A). We also evaluated the correlation of the expression of the entire E4 ORF with SF2 values and obtained similar results (Pearson $r = -0.85$, $P = 0.014$). These findings indicate that decreased E1[^]E4 and E4 expression is associated with resistance to ionizing radiation and corroborates the hypothesis that lower E1[^]E4 expression is associated with worse outcome in patients.

We then used E1[^]E4 expression levels as a surrogate marker of cell line aggressiveness and evaluated whether lines with no E1[^]E4 expression have similar molecular characteristics to HPV⁺ C1 tumors. First, we evaluated differences in expression of HPV KEGG pathway genes between E1[^]E4⁺ and E1[^]E4⁻ cell lines (Supplemental Table 8). For many of these genes the expression level in E1[^]E4⁻ lines was similar to that in HPV⁻ squamous cell carcinoma lines (Supplemental Table 16), indicating a disruption of HPV function among E1[^]E4⁻ cell lines. Some genes, such as *RBL2* and *UBE3A* (Figure 4B), direct targets of E6 and E7, were downregulated in E1[^]E4⁻ lines, as also observed among HPV⁺ C1 tumors. These results suggests that loss or decrease in E1[^]E4 expression is associated with distinct HPV molecular characteristics in tumors and in cell lines.

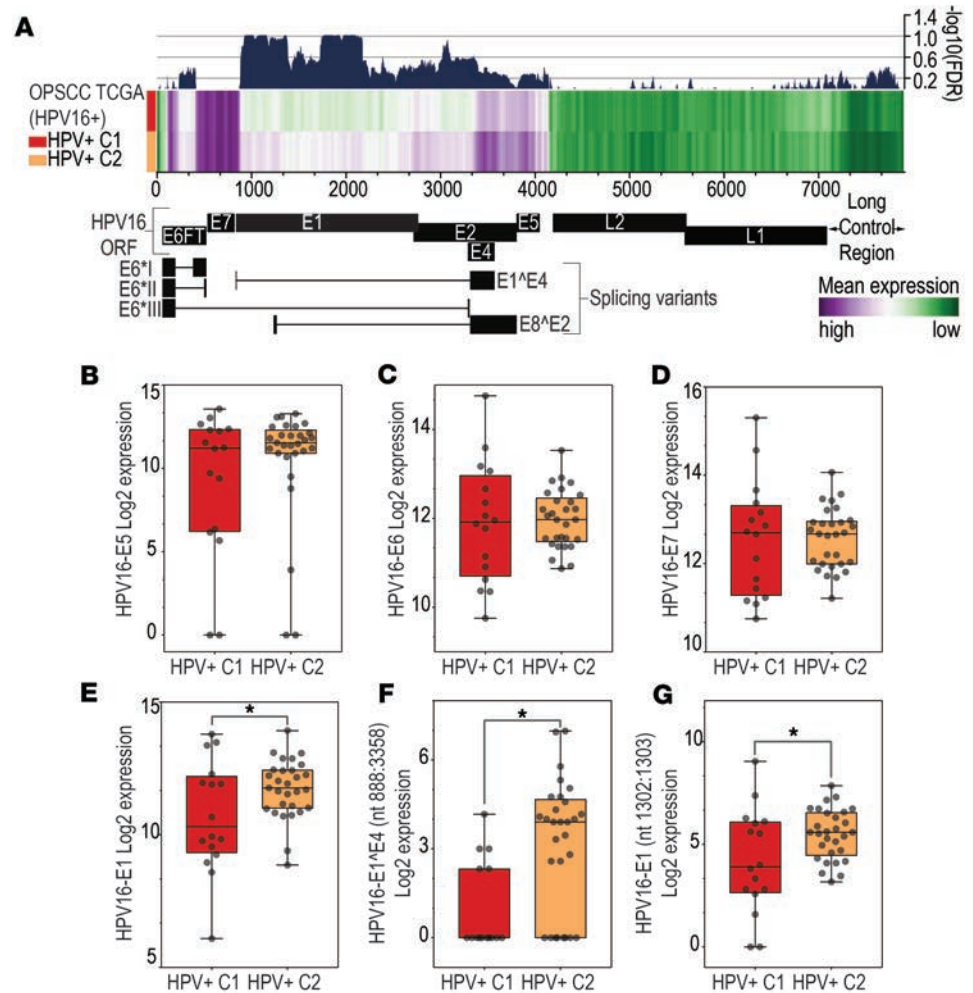


Figure 3. HPV transcriptome differences between HPV⁺ C1 and HPV⁺ C2 groups. (A) Heatmap representing the mean expression levels of each HPV16 nucleotide (horizontal axis) among the HPV⁺ C1 ($n = 19$) and HPV⁺ C2 ($n = 33$) groups. Dark blue indicates the $-\log_{10}$ FDR (right vertical axis) determined by comparing differences in expression between the 2 groups. Expression levels of E5 ($P > 0.05$) (B), E6 ($P > 0.05$) (C), E7 ($P > 0.05$) (D), and E1 ($*P = 0.03$, FDR = 0.29) (E) HPV16 exons in HPV⁺ C1 ($n = 19$) and HPV⁺ C2 ($n = 33$) groups. Expression of E1 (nt1302:1303) ($*P = 0.02$, FDR = 0.09) (F) and E1[^]E4 (nt880:3358) ($*P < 0.001$, FDR = 0.006) (G) splicing junctions was significantly lower among HPV⁺ C1 ($n = 19$) cases.

To further confirm that HPV molecular characteristics are altered in E1[^]E4⁺ cell lines as well as in HPV⁺ C1 tumors, we analyzed RNASeq data from another 62 HPV⁻ cell lines (60 HNSCC and 2 CESCC) and 7 cell lines (1 HNSCC and 6 CESCC) positive for HPV types other than HPV16 from our panel (Supplemental Table 16). These additional cell lines were used to determine transcriptome differences between HPV⁻ and HPV⁺ cells in vitro (Supplemental Table 17). Among the 925 differentially expressed genes between HPV⁺ and HPV⁻ cell lines, 73.4% ($n = 679$) were also differentially expressed between E1[^]E4⁺ and HPV⁻ cases, but only 43.1% ($n = 399$) were differentially expressed between E1[^]E4⁺ and HPV⁻ cell lines (Supplemental Table 17 and Figure 4C). The number of differentially expressed genes between the E1[^]E4⁺ and HPV⁻ cell lines was significantly lower ($P < 0.0001$) when compared with the number of differentially expressed genes between E1[^]E4⁺ and HPV⁻ cell lines. This is similar to the relationship observed between HPV⁺ C1, C2, and HPV⁻ tumors.

Next, we identified the genes differentially expressed between E1[^]E4⁺ and E1[^]E4⁻ cell lines ($n = 1708$; considering $P < 0.05$ and FDR < 0.83 ; uncorrected P value was used in this analysis because of the small number of samples in each group) (Supplemental Table 18) and performed IPA (Supplemental Table 19). Among the 36 enriched canonical pathways, 19 were also found to be enriched between HPV⁺ C1 and C2 (Table 1). Molecular functions related to gene expression, cell death, DNA repair, cell cycle, and protein synthesis were also enriched by genes differentially expressed between cell lines with different E1[^]E4 expression

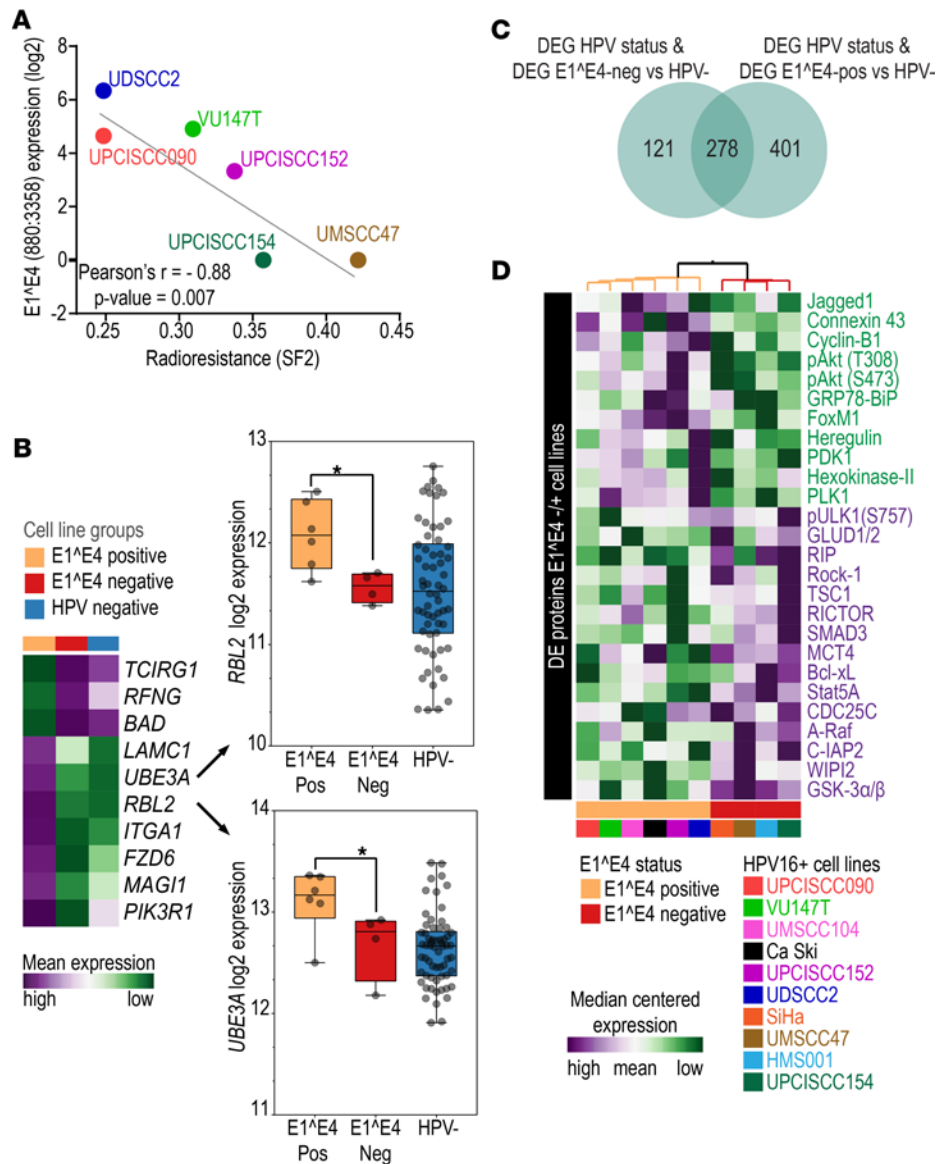


Figure 4. Molecular phenotypic characteristics of HPV16⁺ squamous cell carcinoma cell lines. (A) Pearson's correlation analysis between radioresistance (SF2) (horizontal axis) and expression of E1^{E4} splicing junction (vertical axis) among 6 cell lines indicates a significant inverse correlation. Cell lines with higher radioresistance (higher SF2 values) tended to express lower levels of E1^{E4} splicing junction transcripts. (B) Mean expression levels of 10 HPV-KEGG genes among E1^{E4}⁺ ($n = 6$), E1^{E4}⁻ ($n = 4$), and HPV⁻ ($n = 62$) cell lines. These genes were differentially expressed between E1^{E4}⁻ and E1^{E4}⁺ HPV16⁺ cell lines and between HPV⁺ C1 ($n = 19$) and HPC+ C2 ($n = 33$) tumors and showed similar expression between E1^{E4}⁻ and HPV⁻ cells and between HPV⁺ C1 and HPV⁻ ($n = 28$) tumors. This suggests altered HPV function among E1^{E4}⁻ cells and HPV⁺ C1 tumors. (C) Venn diagram representing the numbers of differentially expressed genes (DEG) between HPV⁺ ($n = 17$) and HPV⁻ ($n = 62$) cell lines ($n = 925$) that were also differentially expressed between E1^{E4}⁻ ($n = 4$) and HPV⁻ cell lines ($n = 399$ genes) and between E1^{E4}⁺ ($n = 6$) and HPV⁻ cell lines ($n = 679$ genes). E1^{E4}⁺ cell lines showed a significantly ($P < 0.0001$, 2-tailed Fisher's exact test) higher number of differentially expressed genes than E1^{E4}⁻ cell lines. (D) Expression profile of the most differentially expressed (DE) proteins ($P \leq 0.1$) ($n = 26$) (vertical axis) among HPV⁺ cell lines ($n = 10$) (horizontal axis) with distinct E1^{E4} expression status (E1^{E4}⁺, $n = 6$; E1^{E4}⁻, $n = 4$).

status. These results indicate that gene expression differences in cell lines with distinct E1^{E4} expression resemble the gene expression differences observed among HPV⁺ C1 and C2 OPSCC tumors and support our hypothesis that E1^{E4} expression is associated with the molecular differences among tumor samples.

We also tested whether molecular differences between cell line groups are also evident at the protein level, as detected by reverse-phase protein array (RPPA) (17) (Supplemental Table 20 and Figure 4D). E1^{E4}⁺ cell lines exhibited activation of the PIK3/AKT pathway, as indicated by elevated levels of heregulin, Jagged-1, PDK1,

PLK1, and phosphorylated AKT (T308 and S473). This was further indicated by upregulation of FoxM1 and cyclin B1 and downregulation GSK3a/b, TSC1, Smad3, and Rock-1 relative to E1⁺E4⁻ cell lines. Additionally, metabolism-related proteins hexokinase II and connexin 43 were upregulated in E1⁺E4⁻ cell lines, while glutamate dehydrogenase and MCT4 were downregulated. E1⁺E4⁻ cell lines also showed downregulation of inhibitors of autophagy (WIPI2, ULK1) and apoptosis (IAP2, BCL-XL, STAT5a). These results indicate a decrease in PI3K/AKT activity and activation of glutamine metabolism and resistance to autophagy and apoptosis among E1⁺E4⁻ cell lines. These changes in protein levels support the transcriptome findings (Table 1) and indicate dysfunctional HPV activity at the protein level in cell lines with no E1⁺E4 expression.

These in vitro findings indicate that HPV and tumor gene expression changes are directly associated and suggest that E1⁺E4 may have a functional role. E1⁺E4⁻ cell lines were more resistant to ionizing radiation and exhibited changes in HPV-mediated functions, such as metabolic, autophagy, and antiapoptotic pathways. Importantly, molecular and phenotypic findings in HPV⁺ cell lines confirm those observed in clinical samples.

Biomarker evaluation and validation. Considering the strong association between gene expression and phenotypic changes in HPV⁺ tumors and cell lines, we sought to determine the value of the 38 HPV-correlated genes as a prognostic biomarker.

First, we evaluated whether differences in survival between HPV⁺ C1 and HPV⁺ C2 groups could be influenced by other clinical or biological factors. Analysis for differences in clinical features (Supplemental Table 21), mutation patterns (Supplemental Figure 4), and HPV infection characteristics (Supplemental Table 14) between HPV⁺ C1 and C2 identified a higher number of nonsmokers in the HPV⁺ C2 group as well as a lower number of packs smoked per year. However, neither amount of smoking (Figure 5A) nor specific HPV⁺ OPSCC risk score or clinical stage (performed retrospectively with AJCC 8th ed.) (13, 19) (Figure 5, B and C) was significantly associated with survival, despite of a trend toward lower survival among heavy smokers and advanced clinical stages. Patient categorization by expression profile showed higher prognostic power than any of these clinical variables. These other factors were also assessed through multivariate analysis, which indicated that the gene expression signature was significantly associated with survival ($P < 0.01$), even after adjusting for other clinical variables (Supplemental Table 22).

The prognostic potential of the 38-gene set was then tested in an independent validation cohort of HPV16⁺ OPSCC (Johns Hopkins University [JHU] cohort) (20). Two groups of patients with distinct patterns of gene expression were observed (Figure 5D), which we identified as OPSCC Val. C1 and Val. C2. The 5-year OS rate was significantly different between OPSCC Val. C1 and Val. C2 (Figure 5E). The number of patients categorized in the group with poorer prognosis (OPSCC Val. C1) was significantly smaller than in the TCGA cohort (35.5% in the TCGA cohort and 12.8% in the JHU cohort, $P = 0.010$). However, the JHU cohort also presented a significantly smaller number of patients with stage III OPSCC ($P = 0.041$), which may explain the low number of patients in the OPSCC Val. C1 group (Supplemental Table 21). Influence of other clinical and biological factors was also verified in this validation cohort by multivariate analysis, and the gene expression signature was significantly associated with survival ($P < 0.001$), even after adjusting for other variables (Supplemental Table 22).

A second validation cohort composed of HPV16⁺ CESCC (21) was used to test whether this signature was associated with survival independently of tumor site. Expression of signature genes indicated two distinct tumor groups in this cohort as well (Figure 5F), which exhibited significantly different survival rates (Figure 5G).

To test the similarity in gene expression pattern among the cohorts, we calculated the expression difference (\log_2 fold change) for each of the 38 genes between the poorer (C1) and the better (C2) prognosis groups from each cohort. The direction of expression of 36 genes was the same in the poorer prognosis groups in all cohorts (Figure 5H) (Supplemental Table 23), confirming that this gene expression signature is consistent across the cohorts and indeed is associated with a poorer prognosis phenotype in HPV⁺ squamous cell carcinomas. A similar result was found for the full set of 582 HPV-correlated genes (Supplemental Figure 5).

We also tested this molecular prognostic signature among patients primarily treated by chemoradiation in the JHU cohort ($n = 16$) (Figure 5I). Patients classified as OPSCC Val. C2 treated with chemoradiation had better survival than patients classified as OPSCC Val. C1 receiving the same treatment ($P = 0.008$) (Figure 5J). This suggests that the signature should be further evaluated as a potential predictive biomarker for chemoradiation.

Pursuing an even smaller list of prognostic genes, we performed univariate Cox regression on the 38 genes in order to evaluate their prognostic power as individual markers. Among HPV⁺ OPSCCs from TCGA, 16 individual genes showed significant prognostic power (5-year OS, Supplemental Table 24). Four of these genes were also significantly prognostic in the JHU OPSCC cohort (IKAROS family zinc finger 3 [IKZF3],

Table 1. Canonical pathways significantly enriched by genes differentially expressed between cell lines and tumor groups

Canonical pathways (IPA)	HPV ⁺ C1 vs. HPV ⁺ C2		E1 ⁺ E4 ⁻ vs. E1 ⁻ E4 ⁺ cell lines	
	Rank	P value (-log ₁₀)	Rank	P value (-log ₁₀)
EIF2 signaling	4	13.4	1	4.01
PTEN signaling	16	5.02	2	3.7
PI3K/AKT signaling	131	1.91	4	2.97
Chronic myeloid leukemia signaling	97	2.33	5	2.67
ATM signaling	23	4.4	8	2.46
Regulation of eIF4 and p70S6K signaling	6	9.39	10	2.07
Telomerase signaling	50	3.33	11	2.06
IL-6 signaling	152	1.54	12	2.01
NF-κB signaling	144	1.7	13	1.9
NF-κB activation by viruses	148	1.62	14	1.88
PEDF signaling	106	2.22	15	1.88
Glucocorticoid receptor signaling	20	4.67	17	1.63
Prostate cancer signaling	29	4.07	18	1.6
Huntington's disease signaling	9	6.14	19	1.6
Lymphotoxin β receptor signaling	85	2.57	23	1.59
Adipogenesis pathway	171	1.31	27	1.5
Mitochondrial dysfunction	3	14	29	1.45
Ceramide signaling	72	2.84	35	1.31
Hereditary breast cancer signaling	8	6.99	36	1.31

cytochrome B5 reductase like [*CYB5RL*], *ARHGAP26*, and calcium voltage-gated channel subunit $\alpha 1 D$ [*CACNA1D*]) and three were prognostic in the JHU and CESCC cohorts (*IKZF3*, *ARHGAP26*, and *CACNA1D*) (Table 2). Lower expression of these genes (*IKZF3*, *ARHGAP26*, and *CACNA1D*) was significantly (likelihood ratio test and Walt test, $P < 0.05$) associated with shorter survival in all cohorts. These results indicate tumor genes highly associated with HPV biology may represent promising prognostic biomarkers for HPV⁺ SCC.

Since *IKZF3* is a transcription factor involved with lymphocyte development, we hypothesized that its expression might be determined by inflammatory cell infiltration in the tumors. However, *IKZF3* RNA was detected in our HPV⁺ cell lines, suggesting that its expression could instead be driven by the tumor cells. Interestingly, *IKZF3* was significantly downregulated in E1⁺E4⁻ cell lines (Figure 5K), supporting an association between this biomarker and our molecular findings. To confirm that *IKZF3* is expressed by tumor cells, we performed immunohistochemistry on HPV⁺ OPSCC cases from our tissue bank. We observed a variable protein expression pattern among the samples. The majority of tumors presented strong nuclear and moderate-to-weak cytoplasmic *IKZF3* expression in tumor cells (Figure 5L), while a few cases showed weak or no expression in tumor cells (Figure 5M). Immune cell infiltrate was strongly positive for *IKZF3* irrespective of tumor cell expression pattern (Figure 5, L and M).

Discussion

By transforming HPV status into a continuous variable, based on expression of HPV genes, rather than as a discrete variable (positive vs. negative), we gained relevant insight into the function of HPV in cancer. We identified potential molecular mechanisms driving poor prognosis in HPV⁺ OPSCC, characterized by changes in key HPV-related pathways and viral transcription. The high-risk group presented an altered or intermediate gene expression profile between HPV⁻ and HPV⁺ OPSCC tumors. Although some HPV molecular characteristics were maintained, many other crucial functions of the HPV life cycle appeared significantly changed.

This analysis of HPV function also identified a dramatic prognostic biomarker for HPV⁺ tumors that is valid across cohorts and tissue sites. This biomarker has the potential to affect one of the most important unmet clinical needs in HPV⁺ OPSCC, a risk classification system more effective than the currently available clinical methods (13, 19). This could be used for selection of treatment de-escalation in low-risk patients and selection of more intense treatment options for high-risk patients. While these biomarkers need to be optimized for clinical use and validated in large prospective cohorts, it may soon be possible to incorporate molecular markers into the treatment of HPV⁺ OPSCC.

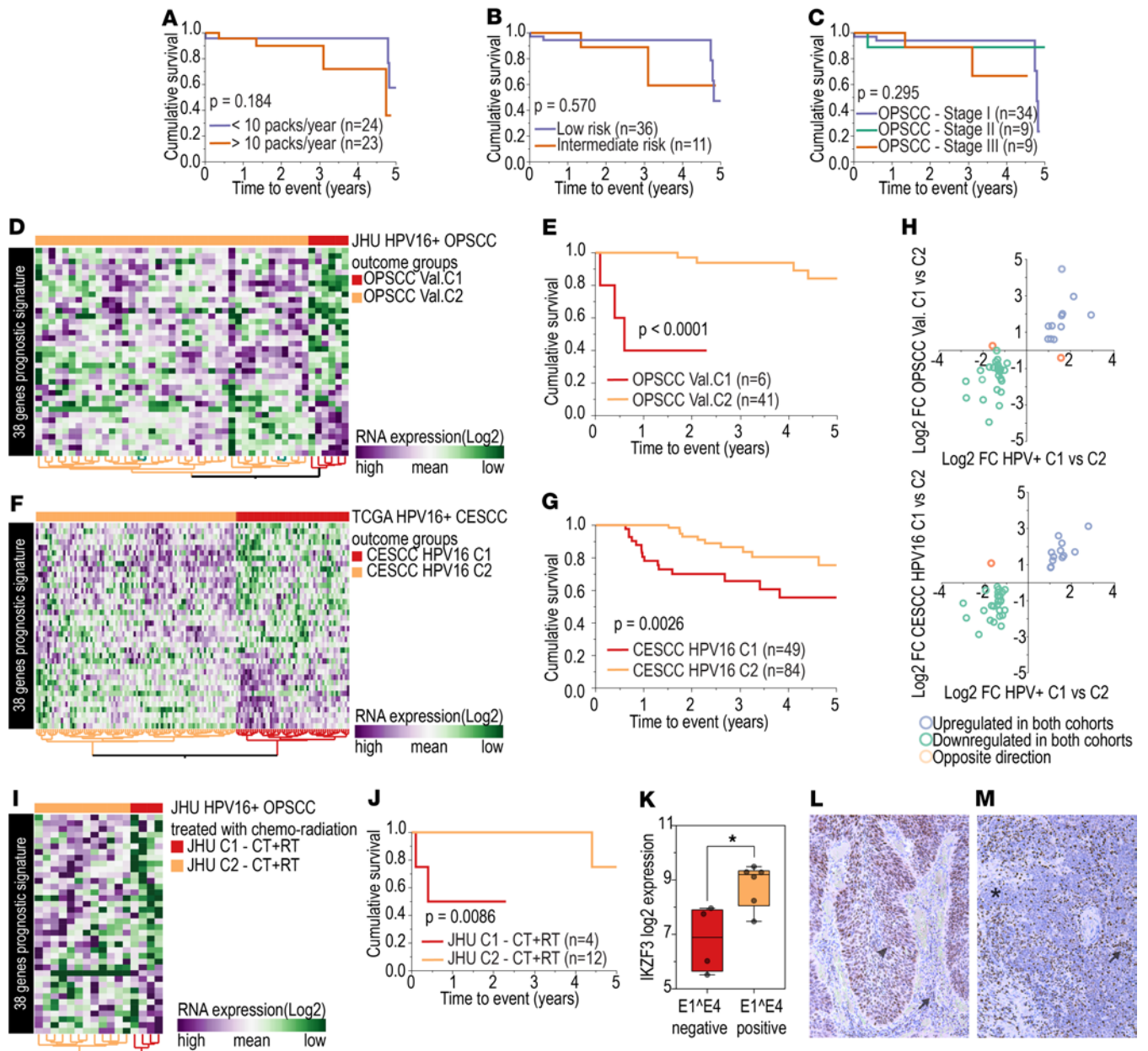


Figure 5. Prognostic biomarker assessment. Log-rank test shows no significant differences in survival among patients with HPV⁺ OPSCC grouped according to (A) number of cigarette packs smoked per year ($n = 47$); (B) low or intermediate risk according to the risk stratification system proposed by Ang et al. (13) ($n = 47$; low risk: ≤ 10 packs smoked per year and N0-N2a; intermediate risk: > 10 packs smoked per year and N2b-N3); and (C) AJCC staging system for OPSCC (8th edition) ($n = 52$). (D) Expression profile of 38 HPV-correlated genes (vertical axis) among HPV16⁺ OPSCC ($n = 47$). JHU cohort shows 2 groups of tumors, OPSCC Val. C1 ($n = 6$) and OPSCC Val. C2 ($n = 41$). (E) OPSCC Val.C1 patients had a significantly lower 5-year survival rate (3 deaths) than those in the OPSCC Val.C2 group (4 deaths) (log-rank test). (F) Expression profile of 38 HPV-correlated genes (vertical axis) among HPV16⁺ CESC patients ($n = 138$) from TCGA shows 2 groups of tumors, CESC HPV16 C1 ($n = 50$) and CESC HPV16 C2 ($n = 88$). (G) CESC HPV16 C1 patients (15 deaths) had a significantly lower 5-year survival rate than the CESC HPV16 C2 group (10 deaths) (log-rank test). (H) Of the 38 genes in the signature, 36 showed similar patterns of expression (upregulation or downregulation) among the groups with lower survival rates (C1) from the TCGA ($n = 19$) (horizontal axis) and JHU ($n = 6$) (vertical axis) cohorts compared with the groups with higher survival rates (C2) (TCGA, $n = 33$; JHU $n = 41$). Only *INHBA* and *DPF1* genes showed the opposite expression direction. Comparing TCGA HPV⁺ OPSCC ($n = 52$) and CESC ($n = 138$) cohorts, all but one gene (*INHBA*) showed the same pattern of expression in the poorer prognosis group from both cohorts (TCGA HPV⁺ OPSCC, $n = 19$; TCGA HPV⁺ CESC, $n = 50$). (I) Expression profile of 38 selected genes among samples from the JHU cohort treated primarily with chemotherapy and/or radiotherapy ($n = 16$) revealed 2 distinct groups (JHU C1 - CT+RT, $n = 4$, and JHU C2 CT+RT, $n = 12$). (J) JHU C1 - CT+RT group showed significantly shorter survival (log-rank test). (K) Expression of *IKZF3* was significantly lower among E1⁺E4⁻ cell lines ($n = 4$) than among E1⁺E4⁺ cell lines ($n = 6$) ($*P = 0.013$, t test). Low *IKZF3* expression was observed among HPV⁺ C1 tumors, and its low expression was associated with shorter survival. (L) HPV⁺ OPSCC showing *IKZF3* protein expression in tumor (arrowhead) and inflammatory cells (arrow) detected by immunohistochemistry. (M) HPV⁺ OPSCC showing no *IKZF3* expression in tumor cells (asterisk) and strong expression in inflammatory cells (arrow) (original magnification, $\times 40$).

Table 2. Candidate prognostic biomarkers for HPV⁺ squamous cell carcinoma determined by univariate Cox regression analysis

		HPV ⁺ OPSCC TCGA				HPV ⁺ OPSCC JHU				HPV16 ⁺ CESC TCGA			
Gene ID	Gene Symbol	HR	CI (95%)	Wald test (P value)	Likelihood test (P value)	HR	CI (95%)	Wald test (P value)	Likelihood test (P value)	HR	CI (95%)	Wald test (P value)	Likelihood test (P value)
22806	IKZF3	0.38	0.17–0.67	0.0036	0.0005	0.29	0.10–0.71	0.0106	0.0065	0.77	0.62–0.96	0.0198	0.0208
606495	CYB5RL	0.19	0.04–0.56	0.0071	0.0018	0.19	0.04–0.73	0.0180	0.0159	0.98	0.62–1.59	0.9497	0.9497
23092	ARHGAP26	0.45	0.19–0.91	0.0376	0.0268	0.33	0.10–0.88	0.0350	0.0265	0.58	0.40–0.85	0.0047	0.0058
776	CACNA1D	0.62	0.38–0.98	0.0474	0.0426	0.4	0.18–0.77	0.0127	0.0046	0.81	0.68–0.98	0.0289	0.0315

HR, hazard ratio; CI, confidence interval.

The biomarker currently includes 38 genes, but our results suggest that this has the potential to be reduced to as few as a single gene. At least 3 of the HPV-correlated genes have prognostic power when assessed alone among HPV⁺ OPSCC and CESC tumors: *IKZF3*, *ARHGAP26*, and *CACNA1D*. *IKZF3* is a lymphocyte transcriptional factor that regulates apoptosis and differentiation (22). This gene has been described as a target of HPV integration (23), although it was not observed in our cases. While this gene is highly expressed in lymphocytes and could indicate a difference in immune infiltration, its expression was not associated with other immune signature markers in our analysis and variable expression was found on tumor cells. *CYB5RL* belongs to the family of NADH-cytochrome b5 reductases, which are involved in the mitochondrial electron transport system and suggests that metabolic changes have an important role in tumor treatment susceptibility (24). The *CACNA1D* gene encodes the voltage-dependent calcium channel CaV1.3 and is also associated with metabolism. Its protein has been linked to regulation of cell migration and invasion in cancer (25). It is encouraging that these genes are mechanistically linked to the differential biology between the groups. They should be further studied both mechanistically and as biomarkers.

An intriguing question is whether the measured expression differences in signature genes are due to changes in tumor cells or stromal/immune cells. Overall, the majority of the signature genes have not been strongly associated with immune cells, and most of the changes are likely to be driven by the tumor cells. All genes demonstrated detectable levels of expression in HPV⁺ cell lines. According to The Human Protein Atlas database (<http://www.proteinatlas.org/>), only 3 genes (*IKZF3*, *CDA*, and *LRMP*) have higher expression in immune-related tissues (26, 27). Intriguingly, *IKZF3* and *CDA* exhibited high RNA expression in HPV⁺ lines, suggesting a possible role for tumor-intrinsic expression. We investigated *IKZF3* protein expression in a small cohort of HPV⁺ OPSCC cases. *IKZF3* was highly expressed in inflammatory cells in all investigated cases, as expected; however, *IKZF3* also exhibited variable expression in the nucleus and cytoplasm of tumor cells (Figure 5, L and M). This supports our hypothesis that most of these biomarkers are driven by tumor cell-intrinsic expression. Additionally, in general, immune-related signatures are not differentially expressed between C1 and C2.

To promote its life cycle, HPV hijacks host-cellular machineries and deregulates functions related to gene expression, cell cycle control, DNA repair, and genome maintenance (8). All of these key molecular processes linked to HPV appeared dramatically transformed among patients with poor prognosis in our data set and suggest that these tumors have a unique biology distinct from the majority of HPV-mediated tumors. A critical cancer-related effect of HPV is the deregulation of the p53 and Rb pathways through the function of E6 and E7 (8). Our findings demonstrate that the expression levels of many of HPV E6 and E7 direct target genes, including *RBL1*, *RBL2*, *UBE3A*, *EP300*, and *CREBBP*, in the high-risk patients (HPV⁺ C1) are similar to those in HPV⁻ OPSCC cases. It is plausible to speculate that high-risk tumors progressed to become less dependent on HPV control.

Although E6 and E7 functions appear reduced in HPV⁺ C1, there was no difference in their expression between the groups. In contrast, decreased E1 and E1^ΔE4 mRNA expression was observed in the HPV⁺ C1 tumors. E1^ΔE4 protein has been linked to inhibition of E6 and E7 protein activity by promoting their direct turnover (28) and, indirectly, by preventing E2 degradation (29), which is a modulator of E6 and E7 activity (30, 31). It is possible that changes in the expression of E1^ΔE4 could be responsible for the observed alterations in HPV function. These findings may indicate an important role for underappreciated HPV genes in cancer and provide new insights into the role of these other viral proteins in cancer progression.

In addition to modulation of E6 and E7, the E1^ΔE4 protein has been linked to mitochondrial dysfunction by displacing mitochondria toward the nucleus and causing a drop in the mitochondrial membrane potential (32). The decrease in expression of mitochondria-related genes among samples with higher E1^ΔE4 suggests this may have occurred in the studied samples. Mitochondria are one of the main targets of ionizing radiation, and inhibition of their activity has been tested as a radiosensitizer (33–39). Downregulation of mitochondrial genes among the HPV⁺ C2 group may suggest some level of mitochondrial impairment and explain the better treatment response among this group. If this assumption is true, treatment-resistant HPV⁺ OPSCC tumors may benefit from mitochondrial targeting for radiosensitization.

The association between E1^ΔE4 expression and the molecular and clinical phenotypes was tested and validated *in vitro*. Deregulation of HPV biology and metabolic imbalances are further supported by the results of the RPPA analysis. We identified differences in the AKT pathway between E1^ΔE4⁺ and E1^ΔE4⁻ cells. AKT activation has been associated with a higher dependency on glucose availability and more sensitivity to oxidative stress (40, 41). Interestingly, E1^ΔE4⁻ cells also exhibited high expression of inhibitors of autophagy and apoptosis that promote survival and are known to be inhibited by phospho-AKT under oxidative stress (41). PI3K-AKT activation has also been attributed to HPV infection, specifically to E7 protein function, which leads to impairment of autophagy in HPV-infected cell lines (42, 43). This suggests that changes in metabolism, apoptosis, and autophagy in E1^ΔE4⁻ cells may be directly related to imbalances in HPV functions and that E1^ΔE4-expressing cells may be more radiosensitive because of a reduced ability to survive under oxidative stress caused by radiation (44).

Furthermore, some of the most enriched pathways identified in our analysis, such as eIF2a, oxidative phosphorylation and mitochondrial dysfunction, protein ubiquitination, regulation of eIF4 and p70S6k, and sirtuin signaling, may have a synergistic effect by being directly involved with the integrated stress response (45–47), a cytoprotective signaling cascade that can be activated by cellular stress. It has been linked to mechanisms of treatment resistance, which may explain the poorer survival in this group (48). Additionally, eIF2a phosphorylation is druggable and may represent an alternative treatment approach (49).

During this analysis, we have come to appreciate that our understanding of HPV function during carcinogenesis is incomplete. There is not even a consensus definition of HPV function that could be applied during our analysis. Therefore, we used panels of genes linked to HPV in our assessment of HPV function, but this is an imperfect measure. Depending on the gene set used in the analysis, the poor prognosis group could be defined as having either reduced HPV function (Figure 1, G and H) or reduced and different HPV function (Figure 1, B and C). These distinct interpretations could have important implications for how to target this group and whether there will be treatments that work only for them. In addition, it is unclear whether molecular and phenotypic changes are mediated by E1^ΔE4 protein or whether E1^ΔE4 RNA expression is a surrogate indicator of other HPV functional changes. Additionally, it is unclear why E1 and E1^ΔE4 expression varied among the tumors and groups. While HPV integration can disrupt the E4 region, we found only a modest statistical association between E1^ΔE4 expression and integration status and no association between E1 expression and integration status. Other possibilities for regulation of expression are host factors, such the local immune response, or variations in the HPV genome. Variations in the HPV genome may be the result of different HPV sublineages or *de novo* mutations in HPV (50, 51). Future studies will be necessary to determine the precise mechanisms regulating E1 and E1^ΔE4 and improve our understanding the HPV function and its role in disease progression.

Additionally, these data demonstrate that there are a range of phenotypes within HPV⁺ tumors. In the future, we believe it will be important to better define key HPV functions during carcinogenesis and develop assays to evaluate them *in vitro* and *in vivo*. This will improve our mechanistic understanding of HPV function, but it could also help identify the most useful molecular targets in each patient based on their individual HPV functional state. Our biomarker should just be a first step in this process.

In conclusion, by analyzing HPV as a continuous variable we have revealed relevant variations in HPV function among HPV⁺ patients. Since HPV promotes treatment response in HPV⁺ OPSCC, it should not be surprising that variations in HPV function could alter this endpoint. Our findings have led to the development of a prognostic and potentially predictive biomarker, associated with HPV function. This could lead to safer implementation of deintensified treatment protocols and facilitate testing of new treatment approaches for patients with unresponsive tumors, both important unmet clinical needs. Biologically, we have found that these variations in function are not driven by E6 or E7 RNA expression, and the mechanism may require a more complete understanding of the functions of other HPV genes, such as E1[^]E4. E1[^]E4 expression is associated with key pathways in patients and cell lines, and it may drive metabolic and stress signaling that regulates the tumor radiation treatment response. We have identified cell lines, stratified by E1[^]E4 status, that are similar to the good and poor outcome patient groups. These represent models to study both good and poor treatment response in this disease with few available models, and they may facilitate the development of treatment options. These findings may eventually lead to better treatment outcomes for patients with HPV⁺ tumors.

Methods

Study design. This study began with in silico analysis of transcriptome data on 80 OPSCCs obtained from TCGA. The results of this initial exploratory analysis provided a molecular phenotype associated with outcome in patients with HPV⁺ OPSCC as well as a set of candidate prognostic markers. The molecular phenotype associated with patient outcome was validated in vitro by using squamous cell carcinoma cell lines. The candidate prognostic biomarkers were validated in silico using transcriptome data from 2 cohorts of HPV⁺ squamous cell carcinoma (OPSCC HPV16⁺, $n = 47$, and CESCC HPV16⁺, $n = 138$).

Data collection. Clinical samples used in this study were separated into discovery and validation cohorts. The discovery cohort was composed of the 80 OPSCCs from TCGA (52 HPV⁺ and 28 HPV⁻) (12). The validation cohorts comprised the 47 HPV16⁺ OPSCCs from JHU (20) and the 138 HPV16⁺ CESCCs from TCGA (21). TCGA RNASeq data was obtained from National Cancer Institute Genomic Data Commons (<https://portal.gdc.cancer.gov/>) (12). Mutation and clinical data were obtained from FireBrowse (<http://firebrowse.org/>). The JHU validation cohort was provided in-house, and details on the cohort have been previously published (20). RNASeq data is available at the Gene Expression Omnibus (GEO) database (accession GSE112027).

HPV detection. Detection of HPV infection and HPV integration and measurement of HPV gene expression in TCGA data sets were performed by applying the VirusSeq pipeline (52). Tumor HPV status was established by considering an empirical cutoff of 1000 read counts mapped against the viral genome, as was done by TCGA (12). HPV reference genomes were obtained from the Papillomavirus Episteme database (<https://pave.niaid.nih.gov/>). HPV status for JHU samples was determined by in situ hybridization or p16 immunohistochemistry (20).

Cell lines. A panel of 10 HPV16⁺ cell lines was used in this study; the panel included 8 HNSCC lines (UDSCC2, UMSCC47, UMSCC104, UPCISCC090, UPCISCC152, UPCISCC154, 93VU147T, and HMS001) and two CESCC lines (CaSki and SiHa). Another 69 squamous cell carcinoma cell lines (62 HPV⁻ and 7 infected by a HPV type other than HPV16) were used for further comparison (Supplemental Table 16). Their short tandem repeat and genomic profiles have been previously published as well as their whole-exome, transcriptome, and RPPA profiles (17). The gene expression profile of cell lines is available through GEO (accession GSE122512).

qPCR. Total RNA was extracted from cell lines using the RNeasy Mini Kit (catalog 74104, Qiagen). RNA was treated with DNase (catalog 18068-015, Life Technologies) and reverse transcribed using the RevertAid First Strand cDNA Synthesis Kit (catalog K1622, Life Technologies) according to the manufacturer's instructions using random hexamer primers provided in the kit. The resulting cDNA was subjected to qPCR using SYBR Green qPCR Master Mix (catalog 4364344, Life Technologies). HPV16 E1[^]E4 was detected with PCR primers (forward, 5'-GATCCTGCACAGCAACGAAG-3'; reverse, 5'-CGACG-GCTTTGGTATGGGT-3') that covered the splicing junction between HPV16 nucleotides 880 and 3358. GAPDH expression was used as internal control (forward, 5'-CTCTGCTCCTCTGTTTCGAC-3'; reverse, 5'-TTGACTCCGACCTTCACCTT-3').

Radioresistance assay. Radioresponse was assessed via standard colony formation assay (18).

Pathway analysis. Pathway analysis was performed using IPA Software (Qiagen) and by GSEA (16). GSEA was performed using MSigDB gene set collection "C2" and subcollection "CP," representing curated gene sets from pathway databases.

Immunohistochemistry. Formalin-fixed, paraffin-embedded tissue from 10 deidentified HPV⁺ oropharyngeal squamous cell carcinoma cases was retrieved from The University of Texas MD Anderson Cancer Center tissue bank for evaluation of IKZF3 protein expression. Immunohistochemical staining was performed in an automated staining system using the polyclonal rabbit antibody anti-IKZF3 [1:1000] (HPA024377, MilliporeSigma). Morphological distribution of IKZF3 expression was assessed in order to determine the most likely cell source of expression in tumor samples.

Statistics. Statistical analyses were performed using JMP Pro 12.0.1 software (SAS Institute Inc.). Correlation between human and HPV gene expression was assessed by Pearson correlation analysis between log₂-normalized counts from each human gene and total HPV gene expression among TCGA OPSCC samples. Differences in gene expression between groups were considered significant when the FDR *P* value (calculated using the Benjamini-Hochberg technique) was lower than 0.05, except for the HPV16⁺ cell line analysis, where we used uncorrected *P* values lower than 0.05 because of the small sample size (4 and 6 in each group), which impaired the multiple testing correction. Hierarchical clustering analysis was performed using the Ward minimum variance method for defining distances between clusters. Cluster distances were computed on standardized expression values. Associations between categorical variables and survival were determined by the log-rank test, considering a 5-year follow-up interval. For continuous variables, univariate and multivariate Cox regression analysis was employed, again considering a 5-year follow-up interval, and significance was established by the Wald and likelihood ratio tests. Associations between categorical variables were determined by the 2-tailed Fisher exact test, and a *P* value lower than 0.05 was considered significant. Data represented in box plots shows lower, median, and upper quartiles, and error bars indicate minimum to maximum values. Associations between categorical and continuous variables were determined by 2-tailed *t* test, and a *P* value lower than 0.05 was considered significant.

Study approval. JHU cohort patients were recruited under protocol NA_00-36235, which was approved by the IRB of JHU. Patients from The University of Texas MD Anderson Cancer Center were recruited under protocol PA15-0138/PA14-0964, which was approved by The University of Texas MD Anderson Cancer Center IRB. All patients provided informed consent prior to participation in the study or a waiver of consent was granted by the IRB.

Author contributions

FOGN and CRP conceived and designed the study. FOGN, XR, YX, LS, CRP, and JW developed the methodology. FOGN, XR, TG, YX, LS, KE, NNK, SR, GX, KMF, JW, JC, HDS, CRP, TS, KA, and MG acquired data. FOGN, XR, TG, YX, LS, KE, MJF, FMJ, JW, JNM, HDS, and CRP analyzed and interpreted data. FOGN, MG, JNM, JC, HDS, and CRP wrote, reviewed, and/or revised the manuscript. MG, KMF, JW, JC, HDS, and CRP provided administrative, technical, or material support. CRP supervised the study.

Acknowledgments

The authors thank the patients for their courage and generosity. We also thank the members of the Pickering, Skinner, and Myers laboratories for moral support, technical support, and helpful discussions and Kathryn L. Hale in The University of Texas MD Anderson Cancer Center's Department of Scientific Publications for editing the manuscript. This work was supported by The University of Texas MD Anderson Cancer Center's HPV-Related Cancers Moon Shot Program, Institutional Startup Funds for CRP, and the National Institutes of Health Clinical and Translational Science Awards Program (UL1TR001442). Bioinformatics analyses and use of the Characterized Cell Line Core Facility, RPPA Core Facility, Research Histology Core Laboratory, and Sequencing and Microarray Facility were supported by the National Cancer Institute through The University of Texas MD Anderson Cancer Center Support Grant (P30CA016672).

Address correspondence to: Curtis R. Pickering, Department of Head and Neck Surgery, Unit T5.3900, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030 USA. Phone: 1.713.794.4512; email: crpickering@mdanderson.org.

KE's present address is: Department of Neuroscience, Baylor College of Medicine, Houston, Texas, USA.

NNK's present address is: Lonza Viral Therapy, Houston, Texas, USA.

HSD's present address is: Department of Radiation Oncology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA.

1. Chaturvedi AK, et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol*. 2011;29(32):4294–4301.
2. Cancers associated with human papillomavirus, United States - 2010 - 2014. Centers for Disease Control and Prevention. <https://www.cdc.gov/cancer/uscs/about/data-briefs/no1-hpv-assoc-cancers-UnitedStates-2010-2014.htm>. Accessed December 26, 2018.
3. Marur S, D'Souza G, Westra WH, Forastiere AA. HPV-associated head and neck cancer: a virus-related cancer epidemic. *Lancet Oncol*. 2010;11(8):781–789.
4. Lindel K, Beer KT, Laissue J, Greiner RH, Aebersold DM. Human papillomavirus positive squamous cell carcinoma of the oropharynx: a radiosensitive subgroup of head and neck carcinoma. *Cancer*. 2001;92(4):805–813.
5. Rischin D, et al. Prognostic significance of p16INK4A and human papillomavirus in patients with oropharyngeal cancer treated on TROG 02.02 phase III trial. *J Clin Oncol*. 2010;28(27):4142–4148.
6. Høxbroe Michaelsen S, Grønhoj C, Høxbroe Michaelsen J, Friberg J, von Buchwald C. Quality of life in survivors of oropharyngeal cancer: A systematic review and meta-analysis of 1366 patients. *Eur J Cancer*. 2017;78:91–102.
7. Masterson L, et al. De-escalation treatment protocols for human papillomavirus-associated oropharyngeal squamous cell carcinoma: a systematic review and meta-analysis of current clinical trials. *Eur J Cancer*. 2014;50(15):2636–2648.
8. Moody CA, Laimins LA. Human papillomavirus oncoproteins: pathways to transformation. *Nat Rev Cancer*. 2010;10(8):550–560.
9. Chen J, et al. Mapping of HPV transcripts in four human cervical lesions using RNAseq suggests quantitative rearrangements during carcinogenic progression. *Virology*. 2014;462-463:14–24.
10. Schmitt M, Dalstein V, Waterboer T, Clavel C, Gissmann L, Pawlita M. Diagnosing cervical cancer and high-grade precursors by HPV16 transcription patterns. *Cancer Res*. 2010;70(1):249–256.
11. Coupe VM, et al. Transcriptional analysis of human papillomavirus type 16 in histological sections of cervical dysplasia by in situ hybridisation. *J Clin Pathol*. 2012;65(2):164–170.
12. Broad Institute TCGA Genome Data Analysis Center. Analysis-ready standardized TCGA data from Broad GDAC Firehose 2016_01_28 run. Broad Institute of MIT and Harvard Web site. <https://doi.org/10.7908/C11G0KM9>. Accessed December 7, 2018.
13. Ang KK, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med*. 2010;363(1):24–35.
14. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
15. Mootha VK, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34(3):267–273.
16. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102(43):15545–15550.
17. Kalu NN, et al. Genomic characterization of human papillomavirus-positive and -negative human squamous cell cancer cell lines. *Oncotarget*. 2017;8(49):86369–86383.
18. Liu X, et al. BAP1 is a novel target in HPV-negative head and neck cancer. *Clin Cancer Res*. 2018;24(3):600–607.
19. Amin, MB, et al. *AJCC Cancer Staging Manual*. 8th ed. New York, NY: Springer International Publishing; 2017.
20. Guo T, et al. Characterization of functionally active gene fusions in human papillomavirus related oropharyngeal squamous cell carcinoma. *Int J Cancer*. 2016;139(2):373–382.
21. Campbell JD, et al. Genomic, pathway network, and immunologic features distinguishing squamous carcinomas. *Cell Rep*. 2018;23(1):194–212.e6.
22. Duhamel M, Navarro P, Cario-Andre M, Billot K, Arrous I, Rebollo A. Differential epigenetic regulation of Aiolos expression in human tumoral cell lines and primary cells. *FEBS Lett*. 2008;582(3):457–467.
23. Rusan M, Li YY, Hammerman PS. Genomic landscape of human papillomavirus-associated cancers. *Clin Cancer Res*. 2015;21(9):2009–2019.
24. Schenkman JB, Jansson I. The many roles of cytochrome b5. *Pharmacol Ther*. 2003;97(2):139–152.
25. Fourbon Y, et al. Ca²⁺ protein alpha 1D of CaV1.3 regulates intracellular calcium concentration and migration of colon cancer cells through a non-canonical activity. *Sci Rep*. 2017;7(1):14199.
26. Uhlén M, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419.
27. Uhlen M, et al. A pathology atlas of the human cancer transcriptome. *Science*. 2017;357(6352):eaan2507.
28. Kajitani N, Satsuka A, Yoshida S, Sakai H. HPV18 E1⁺E4 is assembled into aggresome-like compartment and involved in sequestration of viral oncoproteins. *Front Microbiol*. 2013;4:251.
29. Davy C, et al. A novel interaction between the human papillomavirus type 16 E2 and E1–E4 proteins leads to stabilization of E2. *Virology*. 2009;394(2):266–275.
30. Gammoh N, Grm HS, Massimi P, Banks L. Regulation of human papillomavirus type 16 E7 activity through direct protein interaction with the E2 transcriptional activator. *J Virol*. 2006;80(4):1787–1797.
31. Dowhanick JJ, McBride AA, Howley PM. Suppression of cellular proliferation by the papillomavirus E2 protein. *J Virol*. 1995;69(12):7791–7799.
32. Raj K, Berguerand S, Southern S, Doorbar J, Beard P. E1 empty set E4 protein of human papillomavirus type 16 associates with mitochondria. *J Virol*. 2004;78(13):7199–7207.
33. Anand SK, Tikoo SK. Viruses as modulators of mitochondrial functions. *Adv Virol*. 2013;2013:738794.
34. Richardson RB, Harper ME. Mitochondrial stress controls the radiosensitivity of the oxygen effect: Implications for radiotherapy. *Oncotarget*. 2016;7(16):21469–21483.
35. Cullen KJ, Yang Z, Schumaker L, Guo Z. Mitochondria as a critical target of the chemotherapeutic agent cisplatin in head and neck cancer. *J Bioenerg Biomembr*. 2007;39(1):43–50.

36. Thomas RJ, et al. HPV/E7 induces chemotherapy-mediated tumor suppression by ceramide-dependent mitophagy. *EMBO Mol Med.* 2017;9(8):1030–1051.
37. Cheng G, et al. Mitochondria-targeted analogues of metformin exhibit enhanced antiproliferative and radiosensitizing effects in pancreatic cancer cells. *Cancer Res.* 2016;76(13):3904–3915.
38. Liu J, et al. Enhanced cytotoxic effect of low doses of metformin combined with ionizing radiation on hepatoma cells via ATP deprivation and inhibition of DNA repair. *Oncol Rep.* 2012;28(4):1406–1412.
39. Anoopkumar-Dukie S, Conere T, Sisk GD, Allshire A. Mitochondrial modulation of oxygen-dependent radiosensitivity in some human tumour cell lines. *Br J Radiol.* 2009;82(982):847–854.
40. Nogueira V, et al. Akt determines replicative senescence and oxidative or oncogenic premature senescence and sensitizes cells to oxidative apoptosis. *Cancer Cell.* 2008;14(6):458–470.
41. Zhao Y, et al. ROS signaling under metabolic stress: cross-talk between AMPK and AKT pathway. *Mol Cancer.* 2017;16(1):79.
42. Surviladze Z, Sterk RT, DeHaro SA, Ozbun MA. Cellular entry of human papillomavirus type 16 involves activation of the phosphatidylinositol 3-kinase/Akt/mTOR pathway and inhibition of autophagy. *J Virol.* 2013;87(5):2508–2517.
43. Menges CW, Baglia LA, Lapoint R, McCance DJ. Human papillomavirus type 16 E7 up-regulates AKT activity through the retinoblastoma protein. *Cancer Res.* 2006;66(11):5555–5559.
44. Azzam EI, Jay-Gerin JP, Pain D. Ionizing radiation-induced metabolic oxidative stress and prolonged cell injury. *Cancer Lett.* 2012;327(1-2):48–60.
45. Rajendran R, Garva R, Krstic-Demonacos M, Demonacos C. Sirtuins: molecular traffic lights in the crossroad of oxidative stress, chromatin remodeling, and transcription. *J Biomed Biotechnol.* 2011;2011:368276.
46. Tang BL. Sirt1 and the Mitochondria. *Mol Cells.* 2016;39(2):87–95.
47. Ghosh HS, Reizis B, Robbins PD. SIRT1 associates with eIF2-alpha and regulates the cellular stress response. *Sci Rep.* 2011;1:150.
48. Palam LR, Gore J, Craven KE, Wilson JL, Korc M. Integrated stress response is critical for gemcitabine resistance in pancreatic ductal adenocarcinoma. *Cell Death Dis.* 2015;6:e1913.
49. Sidrauski C, et al. Pharmacological brake-release of mRNA translation enhances cognitive memory. *Elife.* 2013;2:e00498.
50. Mirabello L, et al. HPV16 Sublineage Associations With Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women. *J Natl Cancer Inst.* 2016;108(9):djwt100.
51. Mirabello L, et al. HPV16 E7 genetic conservation is critical to carcinogenesis. *Cell.* 2017;170(6):1164–1174.e6.
52. Chen Y, Yao H, Thompson EJ, Tannir NM, Weinstein JN, Su X. VirusSeq: software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. *Bioinformatics.* 2013;29(2):266–267.