Supplementary figures:

Fig. S1. Human genes highly correlated with HPV transcript levels are the most differentially expressed between HPV-positive and HPV-negative tumors.

Fig. S2. Transcriptome differences between HPV+ C1 and HPV+ C2 groups.

Fig. S3. E1^E4 expression levels detected by RNASeq and qPCR are highly correlated.

Fig. S4. Mutation profiles of HPV+ C1 and C2 tumors are not significantly distinct.

Fig. S5. Expression of 582 HPV-correlated genes in JHU OPSCC and TCGA CESCC cohorts.

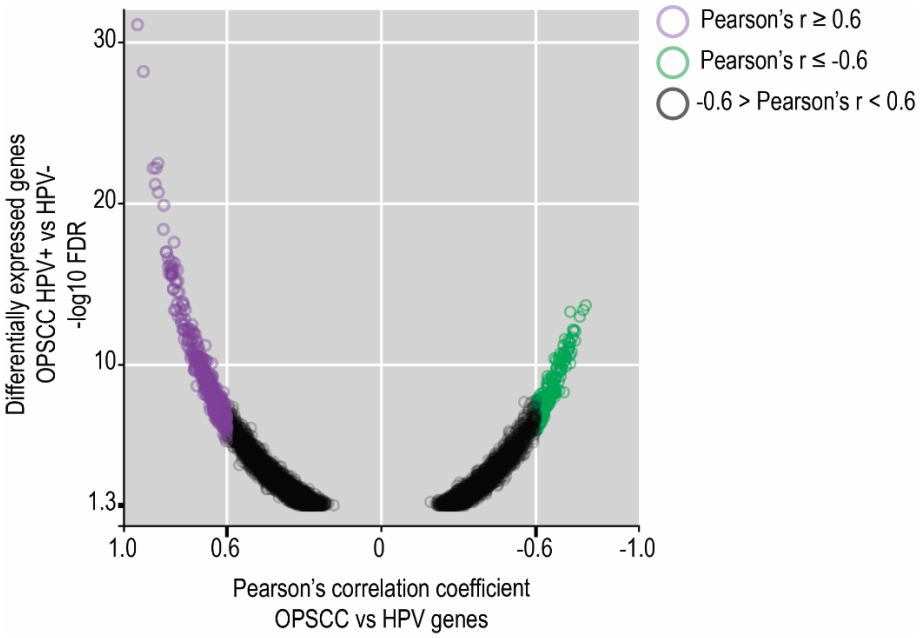

Fig. S1. Human genes highly correlated with HPV transcript levels are the most differentially expressed between HPV-positive and HPV-negative tumors. Distribution of differentially expressed genes between HPV-positive (n=52) and -negative (n=28) OPSCC cases from TCGA according to their FDR (-log10) values (vertical-axis) and their corresponding Pearson correlation coefficient (horizontal-axis) obtained from their correlation with HPV transcripts. Human genes highly correlated with HPV transcription (n=582) (red and green circles) were also highly differentially expressed between HPV-positive and -negative tumors.

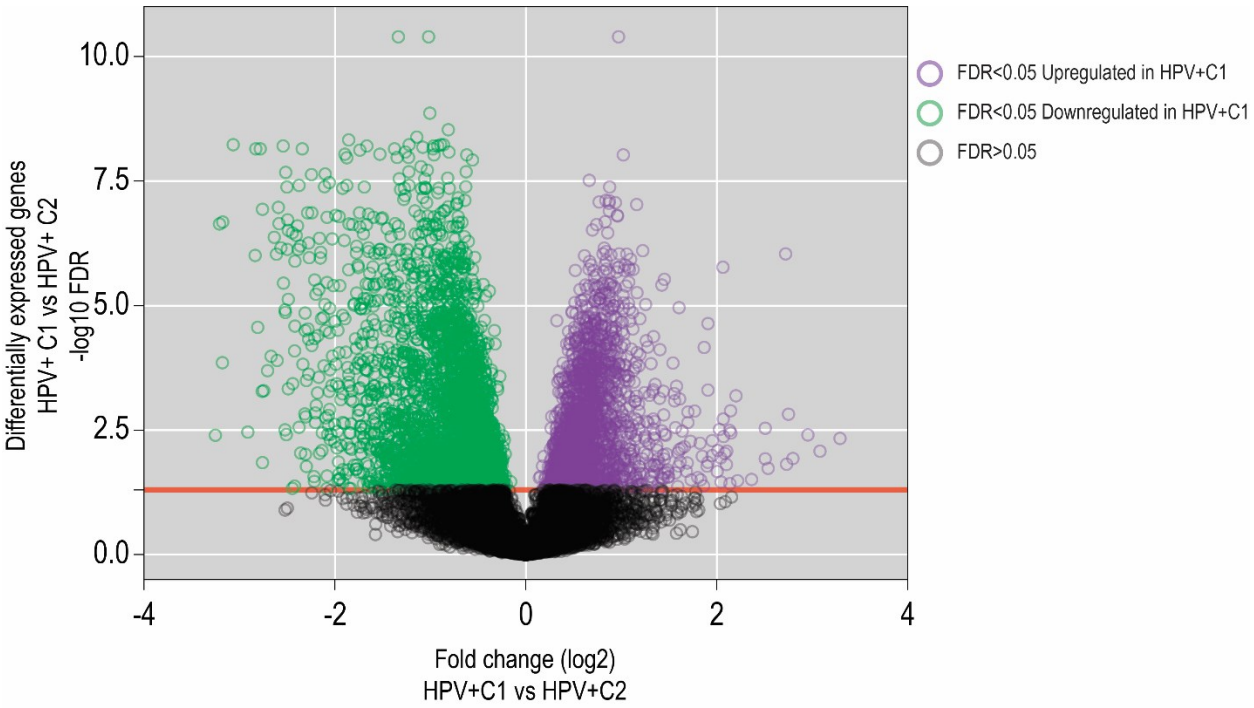

Fig. S2. **Transcriptome differences between HPV+ C1 and HPV+ C2 groups.** Volcano plot representing log2FC (horizontal-axis) and –log10FDR (vertical-axis) from 20081 genes whose expression was compared between TCGA HPV+ C1 (n=19) and HPV+ C2 (n=33) groups. The red line indicates the significance cutoff at 1.30104 –log10FDR (FDR=0.05). Among the 5482 differentially expressed genes, 3037 were downregulated and 2445 upregulated in HPV+ C1 cases.

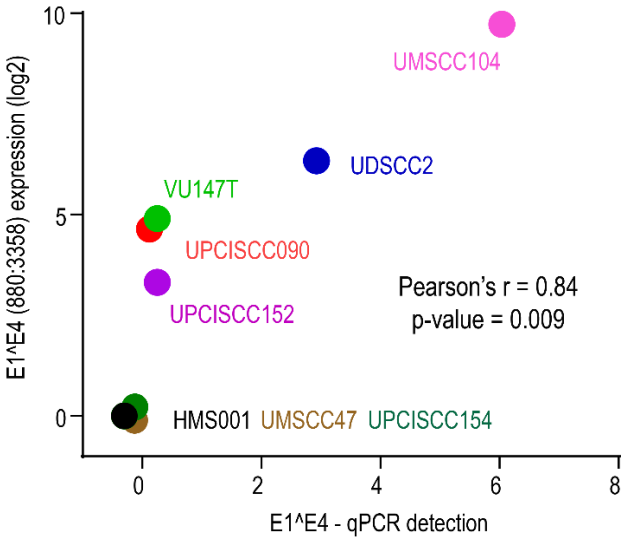

Fig. S3. Levels of E1^E4 expression detected by RNASeq and qPCR are highly correlated. Correlation between levels of E1^E4 detected by RNASeq and by qPCR among HPV-positive head and neck cancer cell lines (n=8). Log2 expression values determined by RNASeq analysis (vertical-axis) were compared to relative expression of E1^E4 determined by qPCR (horizontal-axis). Absence of E1^E4 expression in HMS001, UMSCC47, and UPCISCC154 cells was confirmed by both methods. A highly positive correlation (Pearson's correlation coefficient = 0.84) was detected for the two approaches, providing a technical validation of *in silico* findings.

Top 15 mutated genes (# mutations)	HPV+ C1 mutant cases (%)	HPV+ C2 mutant cases (%)	P-value	Corrected p-value
PIK3CA (11)	3 (16.7)	8 (25)	0.488	0.665
ATAD5 (10)	1 (5.5)	2 (6.2)	0.921	0.921
ANK3 (9)	0	2 (6.2)	0.176	0.463
LRP1B (9)	5 (27.8)	4 (12.5)	0.185	0.463
ASH1L (7)	0	2 (6.2)	0.176	0.463
BBX (7)	1 (5.5)	0	0.149	0.463
CYLD (7)	2 (11.1)	5 (15.6)	0.654	0.818
MACF1 (7)	2 (11.1)	4 (12.5)	0.884	0.921
MLL2 (7)	3 (16.6)	3 (9.4)	0.454	0.665
ASXL3 (6)	0	5 (15.6)	0.029	0.434
CASZ1 (6)	1 (5.5)	5 (15.6)	0.267	0.573
EP300 (6)	2 (11.1)	4 (12.5)	0.884	0.921
FGFR3 (6)	1 (5.5)	4 (12.5)	0.413	0.665
KALRN (6)	4 (22.2)	2 (6.2)	0.103	0.463
SYNE1 (6)	1 (5.5)	4 (12.5)	0.413	0.665

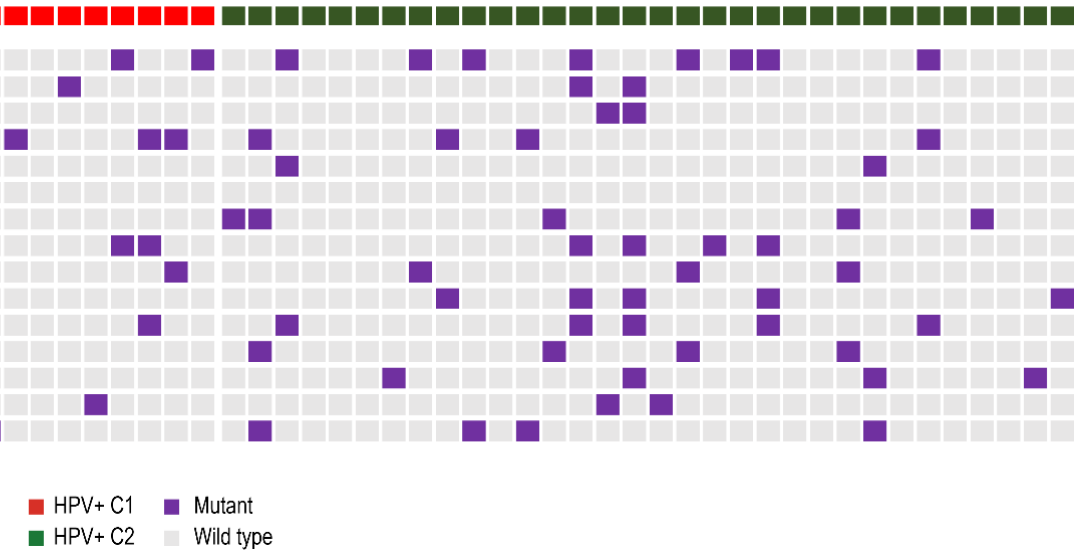

Fig. S4. Mutation profiles of HPV+ C1 and HPV+ C2 tumors are not significantly distinct. The mutation map shows the most common 15 mutated genes and their status among HPV+ C1 (n=19) and HPV+ C2 (n=33) groups. No significant mutation enrichment was found in these groups. The *ASXL3* gene was mutated only among patients in the HPV+ C2 group, but this difference was not confirmed as significant after correction for multiple testing (Fisher's exact test p=0.029, corrected p=0.434). The *TTN*, *MUC16*, *CSMD3*, *DNAH6,* and *DNAH9* genes were highly mutated, but these mutations were excluded since reliability of genetic variation in these genes is uncertain.

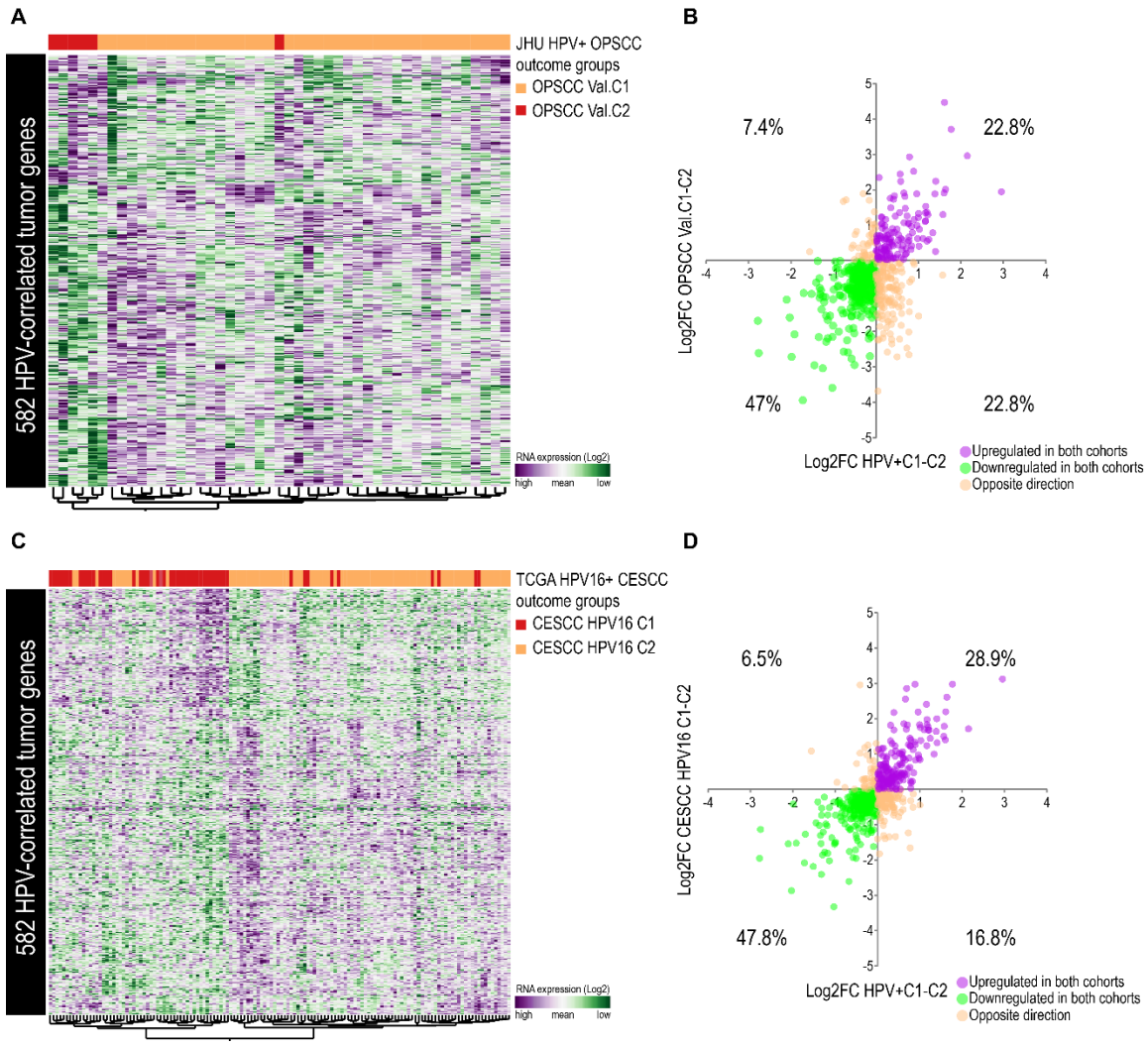

Fig. S5. Expression of 582 HPV-correlated genes in JHU OPSCC and TCGA CESCC cohorts. (A) Expression profile of 582 HPV-correlated genes (vertical-axis) among 47 HPV-positive OPSCC cases from the JHU cohort (horizontal-axis). A tendency toward samples from OPSCC Val.C1 (n=6) clustering together is observed even using the whole set of HPV-correlated genes. **(B)** The expression differences (log2FC) between HPV+ C1 (n=19) and HPV+ C2 (n=33) groups and between OPSCC Val.C1 (n=6) and Val.C2 (n=33) groups indicate that about 70% of the 582 genes presented the same expression pattern in the poorer prognosis group from each cohort. **(C)** Expression profile of 582 HPV-correlated genes (vertical-axis) among 138 HPV-positive CESCC cases from TCGA (horizontal-axis). A tendency toward CESCC HPV16 C1 (n=50) samples clustering together is observed even using the whole set of HPV-correlated genes. **(D)** About 77% of the 582 genes showed a similar pattern of expression among the poorer prognosis groups from the TCGA OPSCC (n=80) and CESCC (n=138) cohorts.