Research Paper

HPV shapes tumor transcriptome by globally modifying the pool of RNA binding protein-binding motif

Yingcheng Wu[1, *], Hao Chen[2, *], Yuyan Chen[3], Lishuai Qu[2], Erhao Zhang[1], Zhou Wang[4], Yuanyuan Wu[1], Riyun Yang[1] , Renfang Mao[5], Cuihua Lu[2], Yihui Fan[1, 6]

[1]Laboratory of Medical Science, School of Medicine, Nantong University, Jiangsu 226001, China
[2]Department of Gastroenterology, Affiliated Hospital of Nantong University, Jiangsu 226001, China
[3]Department of Surgery, Affiliated Hospital of Nantong University, Jiangsu 226001, China
[4]School of Life Sciences, Nantong University, Jiangsu 226001, China
[5]Department of Pathophysiology, School of Medicine, Nantong University, Jiangsu 226001, China
[6]Department of Immunology, School of Medicine, Nantong University, Jiangsu 226001, China
[*] Equal contribution

Correspondence to: Cuihua Lu, Yihui Fan ; email: lch670608@sina.com, fanyihui@ntu.edu.cn

ABSTRACT

Human papillomavirus (HPV) positive head and neck cancer displayed specific transcription landscape but the underlying molecular mechanisms are not fully determined. Here, we interestingly found that HPV infection could globally elongate the 3'-untranslated regions (3'UTRs) in the majority of alternative polyadenylation (APA)-containing genes. Counterintuitively, the 3'UTR elongation does not affect their resident gene expression. Rather, they significantly increase the number of binding sites for RNA-binding proteins (RBPs) and subsequently upregulate a group of oncogenic genes by absorbing RBPs. A significant fraction of HPV affected genes are regulated through such mechanism that is 3'UTR-mediated recruitment of RBPs. As an example, we observed that HPV infection increases the length of 3'UTR of RBM25 transcript and hence recruits much more RNA binding protein including FUS and DGCR8. Consequently, in the absence of FUS and DGCR8 regulation, PD-1 was rescued and up-regulated after HPV infection. Taken together, our findings not only suggest a novel paradigm of how oncogenic viruses shape tumor transcriptome by modifying the 3'UTR, but also present a previously unrecognized layer of APA–RBP interplay in this molecular hierarchy. Modification of the pool of RBP-binding motif might expand our understandings into virus-associated carcinogenesis.

INTRODUCTION

Cancer is a group of diseases involving abnormal and aggressive cell growth accompanied by loss of tissue-specific function. It has already become one of the most challenging diseases that affect health globally, but its etiology is still not fully understood. Several known risk factors such as viral infection, tobacco use, obesity, and excessive drinking of alcohol can significantly increase the incidence of a subtype of cancer [1-4]. In particular, the oncogenic viruses are thought to cause ~20% of all human cancers [5]. Yet, the underlying mechanism of how viruses induce cancers remains poorly defined.

In general, there are seven known human oncogenic viruses, including human papillomavirus (HPV), hepatitis B virus (HBV), hepatitis C virus (HCV), human T- lymphotropic virus 1 (HTLV-1), Epstein–Barr virus (EBV), Kaposi sarcoma- associated herpes-virus (KSHV), and Merkel cell polyomavirus (MCPyV)

[6]. HPV is a small DNA virus and infects mucosal epithelia [7]. Its infection is frequently asso-ciated with cervical carcinoma as well as head and neck squamous carcinomas (HNSCs) [8-9]. Throughout the chronic infection period, oncogenic viruses using oncoproteins to shape cellular processes for replication, immune evasion, and cellular transformation. Onco-proteins directly interact and modify the function of key cellular proteins such as tumor suppressors (pRB and p53), PI3K–AKT–mTOR signaling, MAPK signaling and NF-κB signaling [10-13]. High-throughput sequencing studies revealed that HPV induced profound transcriptional alterations in HNSCs [14]. It suggests that besides direct modification of protein function, oncoproteins also profoundly shape transcriptome. However, the underlying mechanisms about how virus shape cellular transcriptome are not fully defined.

Alternative polyadenylation (APA) was shown to regulate the inclusion of sequences into the 3'-untranslated regions (3'UTRs) and thus involved in multiple biological processes by affecting the stability, localization, and half-life of the mRNAs [15]. Evidence from pan-cancer analysis revealed that a majority of APA-associated genes show shorter 3'UTR compared with normal samples [16]. The shortened 3'UTRs make their resident genes resistance to be attacked by miRNAs [17, 18], which validates the link between 3'UTR and gene expression and perfectly explains how proto-oncogenes were activated. We hence hypo-thesized that alternative 3'UTR length will also possibly change the binding sites for RNA-binding proteins (RBPs) and thus disrupt the RNA output in the post-transcriptional level. Due to the critical role of APA in the mRNA output, we questioned whether oncogenic virus such as HPV could control the APA process to achieve immune evasion and cellular transformation.

To date, freshly published computational tools and datasets [16] make it possible to explore those unanswered questions. Here, by using high-throughput profiling analyses over large patient cohorts, we interestingly found that HPV infection could globally elongate the 3'UTRs in the majority of APA-containing genes. Counterintuitively, the 3'UTR elongation does not affect their resident gene expression, but strikingly, recruit RBPs and reshape cellular transcriptome for cell proliferation and immune evasion.

RESULTS

HPV infection globally elongates 3'UTRs

It remains an open question of how HPV impacts host cells and induces oncogenic changes. Recent findings have suggested the biological significance of miRNAs

in HPV positive tumors [19, 20], where miRNAs bind to 3'UTRs of key molecules such as IL-1α. These discoveries support the importance of 3'UTR and hence raise the question of whether HPV infection affects 3'UTR directly.

To test our hypothesis, we first compute the PDUI for APAome in both HPV+ and HPV- HNSC samples. PDUI is a widely-accepted quantifying unit for evaluating APA level by using RNA-seq data (Methods and Supplementary Table 1) [16, 21]. After generating the PDUI profiles, we perform the differential APA analysis between HPV positive and negative samples (Methods). Strikingly, among the whole APAome within 8831 events, 4824 events show differential APA. With a close look, we find that significant 3'UTR lengthening (3'UL) events predominate (1855 of 3402 longer 3'UTR (54.53%), 1492 of 3402 unchanged 3'UTR (43.86%), 55 of 3402 shorter 3'UTR (1.62%) cut-off |logFC| was set as 0.2). The top 100 differential APA events are shown in Figure 1A. Among the APAome, HPV mediates the globally longer 3'UTR length (Figure 1A and 1B). Next, we conduct the principal component analysis and found that 3'UTR length differentiates the HPV positive and negative tumors clearly (Figure 1C). The 3'UTR of several hallmark genes such as Cyclin-dependent kinase 2 (CDK2), Cyclin-dependent kinase 16 (CDK16), MAPK1 and NFKBIL1 are significantly longer in HPV positive cancer compared with HPV negative cancer (Figure 1D). Together, our results clearly demonstrate that HPV infection could globally increase the length of 3'UTR.

The APA process is orchestrated by multi-protein machinery consisting of a list of key components [21]. To understand the underlying molecular mechanism, we further analyze the expression of individual key components [21] in HPV positive and negative cancers (Supplementary Figure 1A and Figure 1E). Interes-tingly, almost all essential APA regulators show the trend of upregulation. In particular, PABPN1 shows significant higher expression in HPV positive cancers (Figure 1E). The previous report reveals that PABPN1 is a master regulator which governs the APA processes [22], where higher expression of PABPN1 marks the suppression of alternative cleavage and polyadenylation [21]. Consistent with the known function of PABPN1, the upregulation of PABPN1 in HPV infected cells might explain the universal lengthening of 3'UTR. Further analysis clearly indicates a positive correlation between PABPN1 level and 3'UTR lengthening (Figure 1F), while other components do not show a similar trend as PABPN1 (Supplementary Figure 1B). In summary, HPV infection profoundly affects the length of 3'UTR probably through upregulation of PABPN1, the key component of 3′-end processing machinery.
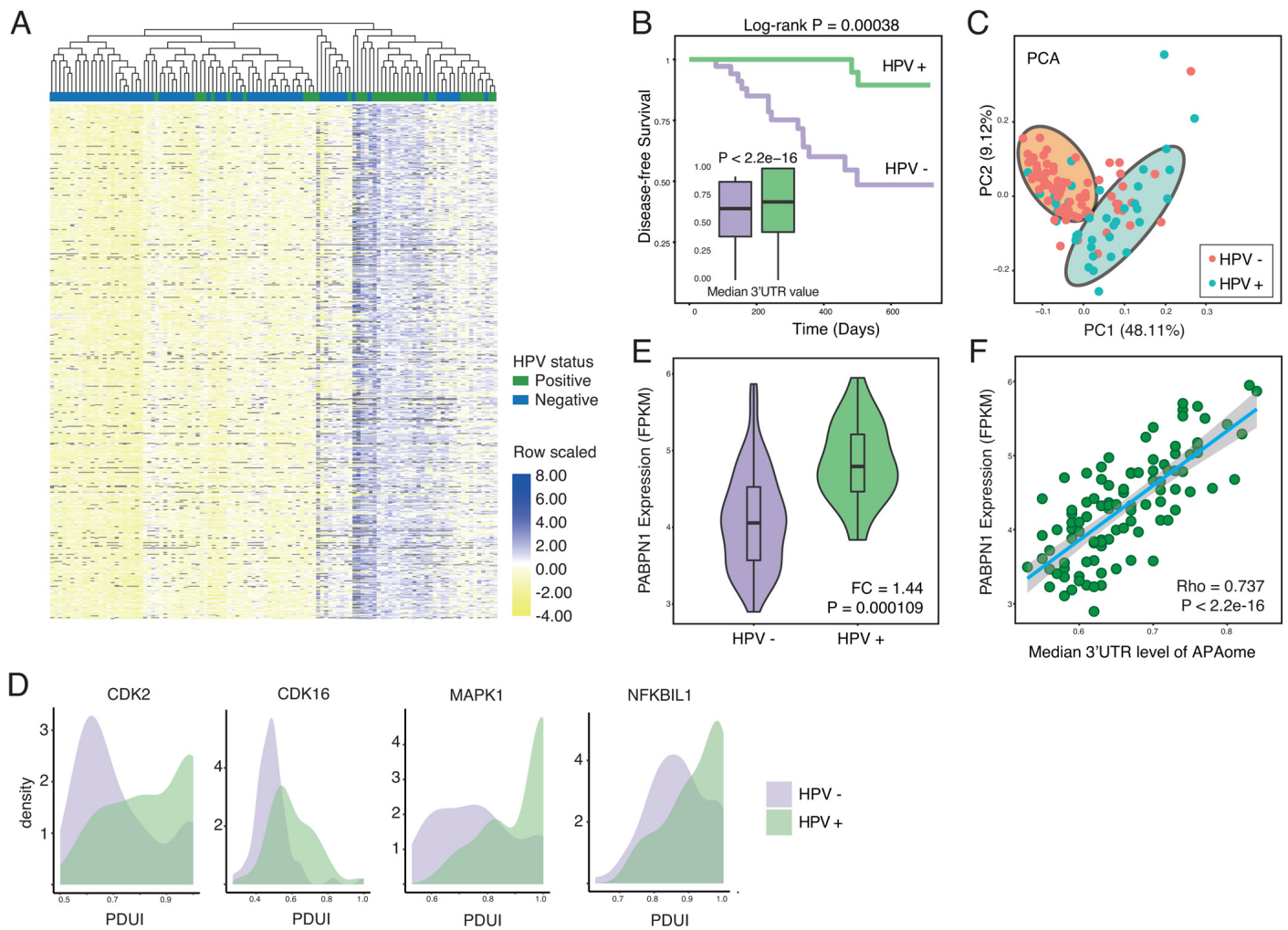
Figure 1. HPV infection causes globally longer 3'UTR in HNSC. (**A**) Among all APA events, 54.53% of genes show significant longer 3'UTR length (1855 of 3402, Wilcoxon test, excluding terms with computation uncertainties). HPV+ samples harbor significant longer 3'UTR. The heat map shows the top 100 APA events ranked by P-value (Wilcoxon test). Values are scaled by rows. (**B**) HPV infection status associates the prognosis of HNSC patients. In general, all APA events shows universally longer 3'UTR in HPV+ HNSC samples compared to HPV- samples in the level of APAome. (**C**) The principle component analysis of all head and neck cancer patients based on all APA events, showing the different HPV infection status. Each dot corresponds to each patient sample, colored according to HPV infection status. (**D**) Key hallmark cancer genes show different 3'UTR profiles between HPV+ and HPV- samples. (**E**) The length of 3'UTR can be used to distinguish the status of HPV infection. Principle component analysis was conducted based on all APA events excluding null values. PC1 and PC2 were showed. (**F**) PABPN1, the master regulator of APA, is significantly upregulated in HPV+ samples. Fold change and P-value were computed through EdgeR.

3'UTR lengthening does not affect their resident gene expression

3'UTR plays a critical role in the regulation of mRNA level. In general, short 3'UTR upregulates gene expression [23]. Longer 3'UTR may expose the RNA molecule in a more attackable manner. Therefore, we ask whether the 3'UTR lengthening after HPV infection will affect their own mRNA level. Among all 3'UL events, we compute the interplay between the APAome and transcriptome. Interestingly, our results suggest that only 6.89% of 3'UL genes shows direct down-regulated expression (Figure 2A). We also verify that differential APA do not fully explain the differential expression (Supplementary Figure 2A). It suggests that 3'UTR lengthening in HPV infected cells has minimal effect on their own gene expression. This is consistent with the previous report to show that only ~2% — 12% of APA events are negatively correlated with their gene expression [21].
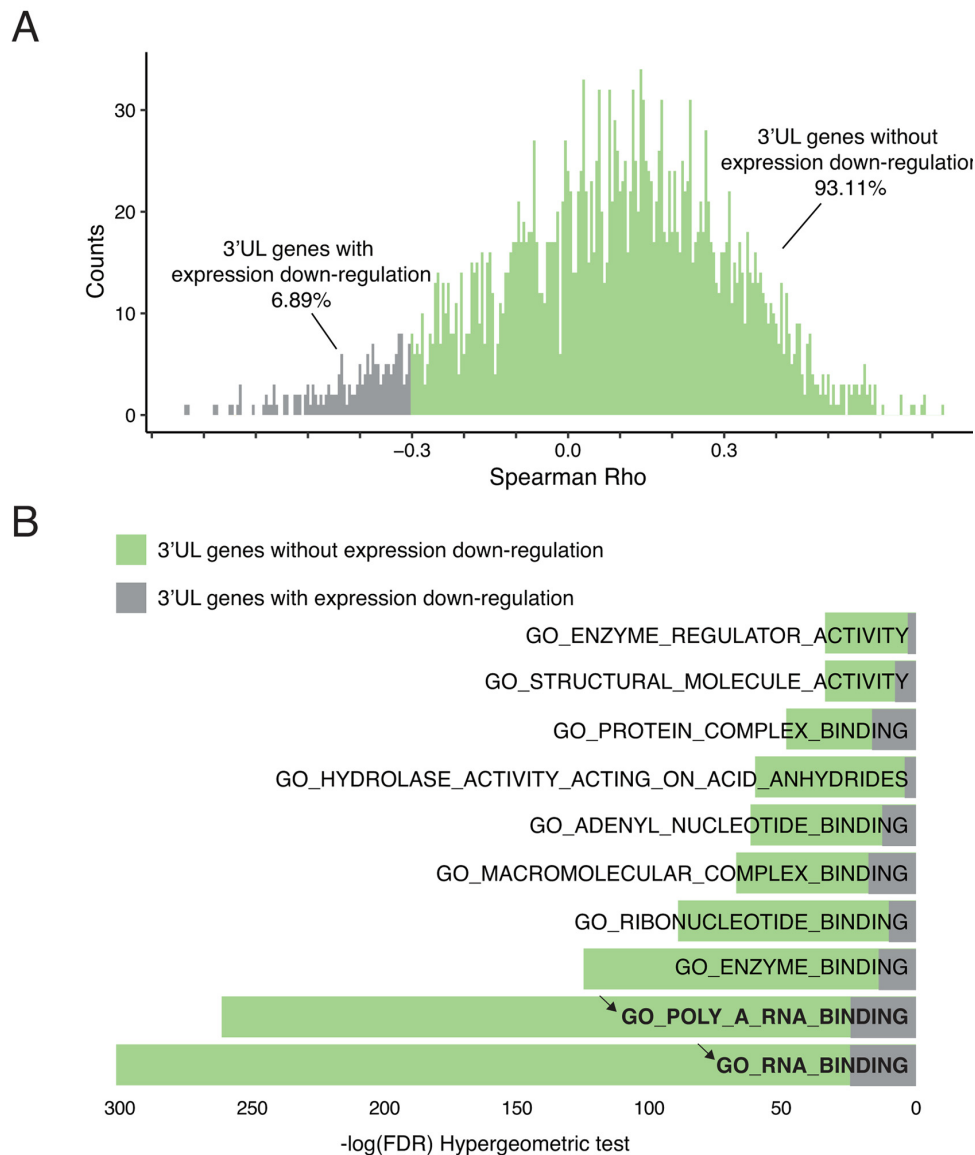
Figure 2. 3'UTR-longer genes do not affect their own expressions. (**A**) 3'UL genes without direct downregulation expression predominate. Only 6.89% of 3'UL genes show direct down-regulated downstream expression. The downregulation was defined as Spearman Rho < -0.3 and P-value < 0.05. (**B**) 3'UL genes without down-regulated gene expression control more aspects of molecular functions (hypergeometric test, retrieved from Molecular Signature Database). RNA-binding ranks highest in the function enrichment computation.

3'UTR lengthening fails to influence their own gene expression, suggesting that these 3'UTR lengthening genes might not be the direct targets for HPV. This is supported by the following functional enrichment analysis. Based on the Molecular Signature Database, among the top 10 overlapped gene ontology enrichment terms, we do not observe any cancer-development-related terms (Figure 2B). In contrast, those genes are enriched in RNA binding, which indicates the potential regulatory roles of these 3'UL genes. In order to verify

that 3'UL genes without down-regulation are not directly associated with biological functions, we examine whether they are involved in immune functions, tumor progression, and HPV infection. We apply six well-established molecular signatures and no significant results were observed (Supplementary Figure 2B). Taken together, our results suggest that 3'UL genes are unlikely the direct targets of HPV infection to transform cells. Instead, these 3'UL genes might regulate other genes by using their RNA binding affinity.

Longer 3'UTR indirectly impacts genes by recruiting RNA-binding proteins

To understand how 3'UL impacts downstream functions and the potential function of RBPs, we fetch the RBP-mRNA interactome data from RAID (July 2018) [24]. RAID is a comprehensive resource for evaluating the RNA-protein interaction based on CLIP-seq data. Under normal conditions, RBPs competitively bind to target RNAs and regulate their targets expression [25]. However, when genes with longer 3'UTRs are able to sequester more RBPs, the other targets of RBPs may get

repressed or activated due to fewer interferences of RBPs. The analyses reveal that, in accordance with our hypothesis, differential APA genes are more capable of being bound by RBPs (Figure 3A). The degree of differential APA also tends to be correlated with the counts of RBPs binding to each gene. Next, we calculate the counts of paired genes controlled by 3'UL via RBPs. The paired genes are defined as the targets of 3'UTR associated RBPs (3'UTR-RBP-targets). Due to the large interactome dataset, the 3'UTR—RBP—gene pairs are randomly selected for 1000 times. Surprisingly, differential-APA genes profoundly impact
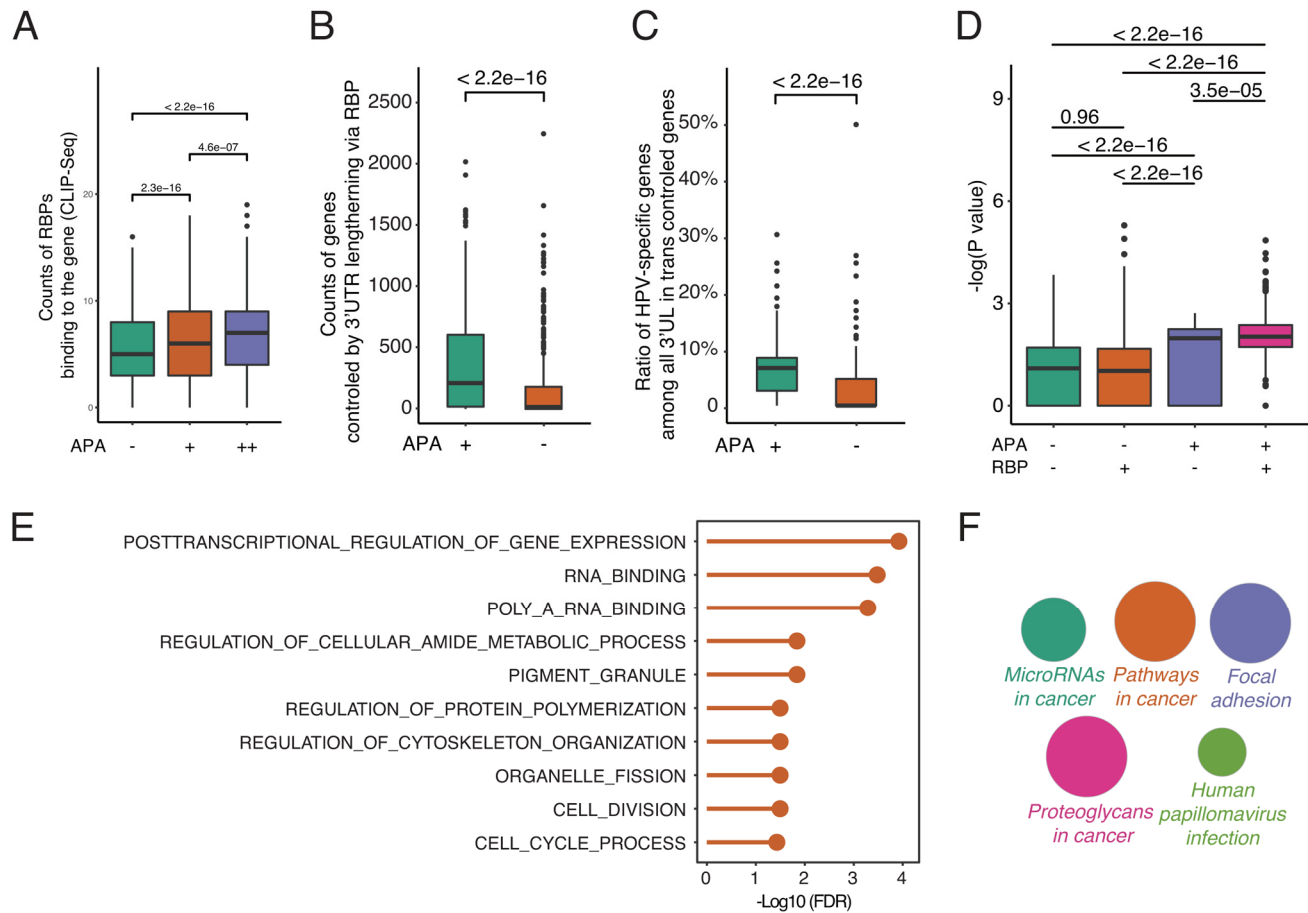

Figure 3. By releasing intermediate RNA-binding proteins (RBPs), 3'UL genes indirectly impact paired genes in trans. (**A**) Differential-APA genes are easier to be attacked by RBPs (Wilcoxon test, data from CLIP-seq). APA+ group and APA++ group represent different differential APA level (P < 0.05 and P < 0.001, respectively). (**B**) 3'UL controls more paired genes in trans compared with non-3'UL APA events. This computation was generated by randomly selecting 1,000 3'UL-gene pairs. In order to retrieve high-confidence interactions, we limit the number of intermediate RBPs > 3. (**C**) 3'UL-in-trans-paired genes show significantly higher ratio of HPV-specific. This computation was generated by randomly selecting 1,000 3'UL-gene pairs with the number of intermediate RBPs > 3. (**D**) Through intermediate RBPs, 3'UL controls paired genes in trans. APA+RBP+ group shows the most significant HPV-specific trend. The bigger the Y-axis value is, the group harbors more HPV-specific trend. The Y-axis marks the -log (P-value) of differential expression between HPV+ and HPV- samples through EdgeR. The P-value between 4 groups was computed through the Wilcoxon test. This computation was generated by randomly selecting 1,000 3'UL-gene pairs. RBP+/RBP- was defined as higher/lower RBP-binding confidence level (intermediate RBPs > 3 and ≤ 3, respectively). In the RBP+ group, 3'UL impacts in-trans-paired genes with higher confidence. (**E**) Biological process enrichment analysis of HUB-3'UL in-trans-paired genes (hypergeometric test, retrieved from Molecular Signature Database). The cell cycle is the most top ranking for enrichment term. HUB-3'UL was defined as the 3'UL genes with interacting RBPs ≥ 15. (**F**) Pathway enrichment of HUB-3'UL in-trans-paired genes. CLUEGO package was used based on KEGG database. The size of each bubble represents the enrichment level of each term.

their paired genes (P < 2.2e-16, Wilcoxon test, Figure 3B). In order to have a closer view of the paired genes, we compute the ratio of HPV-specific genes among them (randomly selected 1000 times). Consequently, the pairs of differential-APA genes are more enriched in HPV-specific molecules (Figure 3C). These data collectively suggest that the pairs of longer 3'UTR genes are enriched in HPV-specific genes. HPV regulates oncogenic genes by modifying the 3'UTR of a group of genes and subsequently affect the function of RBPs.

To further validate the APA—RBP—paired gene axis, we next explore whether the paired genes are in an RBP-dependent manner. We divide the interactome into two groups: RBP+ and RBP- groups, which are defined as higher or lower RBP-binding confidence level (intermediate RBPs > 3 and ≤ 3, respectively). The RBP+ group means APA regulates paired genes via more intermediate RBPs in a high confidence level. In contrast, RBP- group represents a weaker bridge between 3'UTR and paired genes. The Y-axis marks the -log (P-value) of differential expression between HPV+ and HPV- samples. The bigger the Y-axis value is, the group harbors a more HPV-specific trend. Firstly, among the APA negative groups, no significant difference of HPV-specific-trend is observed between RBP+ and RBP- sets (P = 0.96, Wilcoxon test). On the contrary, in the APA+RBP+ group, the -log (P-values) of differential expression are upregulated significantly compared with the other three groups (Figure 3D). This result strongly supports the hypothesis that APA regulates paired genes dependent on RBPs.

Next, to better quantify the functional impact of the global APA-changed paired genes, we monitor the associated biological processes. With a focus on the hub 3'UL genes, we limit the cutoff number of intermediate RBPs to ≥ 15. Interestingly, the cell cycle is the top ranking term, which is strongly in consistency with our previous understandings into HPV+ tumor [26, 27] (Figure 3E). Further, pathway enrichment analysis was computed through CLUEGO. We set the most global network specificity to generate the most credible enrichment results. Strikingly, the hub-3'UL paired genes are strongly associated with HPV infection (Figure 3F). Thus, we provide convincing evidence to show that HPV shapes host cells by modifying the length of 3UTR, subsequently affect the function of RBPs and finally regulate their target genes for survival and transformation.

3'UTR lengthening of RBM25 upregulates PD-1 via FUS and DGCR8

To further confirm the identified paradigm that HPV shapes host cells by using APA—RBP—paired gene

axis, we next focus on individual key molecules to examine the paradigm (Figure 4A). A recent study reports HPV induces the dysfunction of CD8 T cells [30]. Yet, very little is known about the underlying mechanism. Therefore, we asked whether HPV infection affects PD-1, which is a key inhibitory checkpoint of CD8 T cells. Interestingly, in accordance with previous reports [28], we confirm that HPV upregulates the expression of PD-1 (Figure 4B). Next, the upstream 3'UL genes were computed based on CLIP-seq. We rank the Spearman Rho between 3'UL genes and PD-1 and found that RBM25 and GTF3C3 are two potential regulators of PD-1 (Figure 4C). Next, we mine the interactome of RBPs and RNAs and observe two potential intermediate RBPs including FUS and DGCR8 (Figure 4C). In order to validate our computation based on experimental results, we use another independent CLIP-seq to confirm our findings. In consistency with the computational data, we confirm that FUS and DGCR8 bind to PD-1 and the RBM25 3'UTR by combining CLIP-seq and RNA-seq data (Figure 4C-4D).

To further determine whether HPV regulates PD-1 in the 3'UL dependent manner, we focus our analysis on the RNA-seq data. HPV infection can robustly increase the APA level of RBM25 and GTF3C3 (Figure 4E and Supplementary Figure 3A). Without 3'UL, HPV infection alone fails to significantly upregulate PD-1 expression (Figure 4F and Supplementary Figure 3B). In the presence of 3'UL, HPV infection dramatically increases the level of PD-1 (Figure 4F and Supplementary Figure 3B). Our analysis supports that HPV infection increases PD-1 expression by elongating the 3'UTR of RBM25 and GTF3C3. As expected, correlations between 3'UTR scores (PDUI) (RBM25 and GTF3C3) and PDCD1 expression (FPKM) also demonstrate the interplay between 3'UL and PDCD1 (Figure 4G). We next found that PABPN1, the core component of APA machinery, correlates with 3'UTR of RBM25 (Supplementary Figure 3D). We further examine whether HPV affects FUS or DRGR8 directly and find that HPV infection made insignificant changes to RNA expression of those two genes (Supplementary Figure 3E). Together, these results substantiate our hypotheses that 3'UTR lengthening contributes to upregulation of PD-1 via RBPs.

3'UTR lengthening upregulates cell cycle gene E2F1

To further prove our hypothesis that 3'UTR lengthening affects gene expression via RBP, we focus on the cell cycle process of cancer cells which is the hallmark of HPV infection [26, 29]. Though lots of layers of this process were reported such as miRNAs and lncRNAs [30-32], the upper regulator is not fully understood.

E2F1 is a key cell cycle regulator [33-35] and has long been regarded as an oncogene [35-38].

To further gain insights into associations between HPV and cell cycle regulator E2F1, we re-check the differential APA profile. Surprisingly, among the APAome, E2F1 undergo 3'UTR lengthening (Supplementary Figure 4A) indicating a higher likelihood of being downregulated by miRNAs and RBPs. In contrast with the 3'UTR lengthening, E2F1 is paradoxically upregulated in HPV+ samples (Fold change = 4.92, P = 1.45E-37; Supplementary Figure 4A). These opposite results suggest there exists another robust regulator of E2F1 that powerfully controls its expression. The APA—RBP—paired gene axis might be involved in the regulation of E2F1. Thus, after mining the interactome based on CLIP-seq data between RBPs and RNAs as is

similarly conducted in Figure 4, we observe a regulatory network between E2F1 and 3'UL genes EDC3 and AGGF1 and demonstrate that interplay between 3'UL and E2F1 is RBP dependent (Supplementary Figure 4B). HPV infection could significantly upregulate the APA level of AGGF1 and EDC3 (Supplementary Figure 4C and Supplementary Figure 4F). Without 3'UL, HPV infection fails to upregulate E2F1 (Supplementary Figure 4D and Supplementary Figure 4G). However, in the presence of 3'UL, HPV significantly upregulates E3F1 (Supplementary Figure 4D and Supplementary Figure 4G). The APA score of AGGF1 and EDC3 highly correlates with the expression of E2F1 (Supplementary Figure 4E and 5H). Collectively, these results provide new evidence of how HPV alters cell cycles through changing 3'UTR.
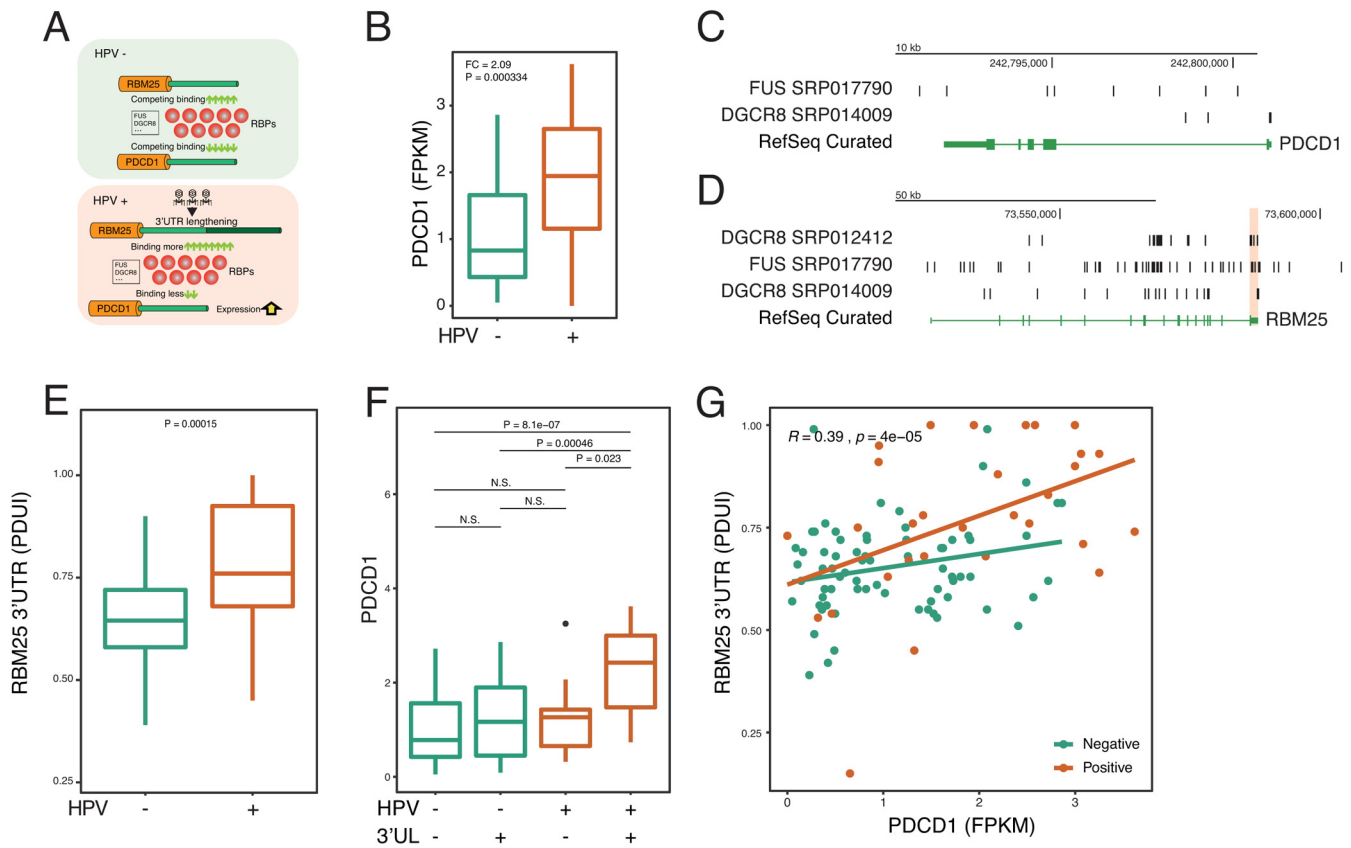

Figure 4. HPV-mediated longer 3'UTR of RBM25 and GTF3C3 upregulates PD-1 via impacting RBPs of FUS and DGCR8. (**A**) In the HPV+ state, 3'UL mediates the imbalance of competing for binding of RBPs. Longer 3'UTR of RBM25 and GTF3C3 drives more RBPs to bind on them. Conversely, less RBPs will bind on PDCD1, which upregulates the expression of PDCD1. (**B**) PDCD1 is upregulated in the HPV+ samples. Fold change and P-value were computed through EdgeR. (**C**) RBPs of FUS and DGCR8 can bind to the 3'UTR of PDCD1 (CLIP-seq). CLIP-seq peaks were combined within all samples. (**D**) RBPs of FUS and DGCR8 can bind to the 3'UTR of RBM25. CLIP-seq peaks were combined within all samples. (**E**) 3'UTR level of RBM25 is significantly longer in HPV+ samples (Wilcoxon test). (**F**) 3'UL+ and HPV+ group shows the highest PDCD1 expression (FPKM). (**G**) 3'UTR length of RBM25 is positively correlated with PDCD1 expression. Spearman Rho value and P value was analyzed in HPV positive and negative samples.

DISCUSSION

Viral infection greatly contributes to the global cancer burden, but how oncogenic virus transforms normal cells to cancerous cells is still poorly understood. Viral infection widely affects gene expression and several common virus-targeted pathways have been identified. For example, virus profoundly altered the miRNA profile and activate oncogenic processes [30]. However, we are convinced that those mechanisms may not be the whole picture. The molecular mechanism of how viral infection extensively influences gene expression is still not fully addressed. Here, we found a new routine which virus utilized to modify gene expression. In our analysis, HPV infection could significantly elongate 3' UTR from many genes. These elongated 3'UTRs significantly affect the pool of RBP binding motif and thus inhibit RBP function. As a result, the other RBP binding targets were rescued and increased expression significantly. Through this mechanism, virus modifies 3'UTR to control the expression of a group of genes via the intermediate RBPs. Thus, our results might provide novel insights to understand how virus shape cell by modifying the length of 3' UTRs. The model providing here might also be suitable for other virus and might be a useful paradigm to understand the interaction between virus and host.

The 3'UTRs of eukaryotic mRNA is generated through endonucleolytic cleavage at poly(A) site and subsequently polyadenylation by the poly(A) polymerases [39]. The choice of different poly(A) site determines the length of 3'UTRs and this process is orchestrated by a multi-protein machinery composed of three main complexes: 1) Cleavage and Polyadenylation Stimulatory Factor (CPSF), 2) Cleavage Stimulatory Factor (CSTF), and 3) Cleavage Factors Im and IIm (CFIm, CFIIm) [40, 41]. After cleavage at poly(A) site, poly(A) tails are added by a poly(A) polymerase (PAP) and recruit poly(A)-binding proteins (PABPs) including the cytosolic PABPC and the nuclear PABPN1. PABPN1 directly interacts with PAS regions and suppresses the usage of weak proximal PASs [22]. In accordance with this mechanism, our data confirmed the loss of PABPN1 in human cells resulted in extensive 3' UTR shortening. We observed significantly higher expression of PABPN1 in HPV infected cells. It suggests that HPV might regulate the key components such as PABPN1 in the 3′-end processing machinery to modify the length of 3'UTR.

Increasing evidence indicates that APA is extensively used and is an important way to regulate gene expression [42-44]. The 3'UTR influences many aspects of mRNA metabolism including but not limited to transcription termination by RNAP II, mRNA stability,

mRNA export to the cytoplasm, and the efficiency of translation. The 3' UTRs often harbor regulatory elements such as AU-rich elements (AREs) or other binding motifs that are easily attacked by micro-RNA (miRNA) [45, 46]. Previous studies show that transcripts with shorter 3' UTRs are more likely to produce higher levels of protein [47, 48]. Interestingly, in our analysis, the status of 3'UTRs do not affect their own RNA level. Instead, they significantly influence their binding RBP's downstream targets. However, it is still possible that the changed 3'UTRs might affect their protein translation even though they not affect mRNA level directly. While this possibility was not directly excluded due to lack of corresponding sequencing data at the protein level. But this possibility was not supported by alternative approaches. For example, functional annotation of RBP's downstream targets but not 3'UTR resident genes was enriched in the cell cycle, which is highly related to virus infection and replication. Therefore, our finding provides a new insight to reveal how virus shapes host transcriptional profile for their survival and replication.

Pan-cancer analysis across 7 cancers including 358 tumor/normal pairs reveals that most APA genes (91%) in tumors have shorter 3′ UTRs. Besides these global features of APA events in cancer, several specific APA regulated genes in different cancer have also been investigated. For example, in hepatocellular carcinoma, downregulation of CFIm25 subunit of cleavage factor results in high expression of PSMB2 and CXXC5 by increasing the usage of the proximal polyadenylation site [49]. In pancreatic ductal adenocarcinoma, the increased ZEB1 protein expression is due to its shorted 3'UTR [50]. In breast cancer, the cellular proliferation marker Ki-67 is increased due to shortening of the Ki-67 3'UTR [51]. To our surprise, HPV infection is significantly associated with elongated 3'UTR. It suggests that the HPV-induced cancer might use a different mechanism to drive cancer. It remains an interesting question to further test the difference between virus-induced cancers and mutation-driven cancer. This is critically important for us to prevent and treat cancer based on their driving factors.

In a nutshell, by comparing the APA events between HPV infected or non-infected cancers, we found HPV-induced features of APA. Through elongating the 3'UTRs, HPV indirectly changes the RBP targets for their survival and cell transformation. Our work not only presents a novel paradigm of how oncogenic viruses shape tumor transcriptome by modifying the 3'UTR, but also highlight a previously unrecognized layer of APA—RBP interplay in this molecular hierarchy. Modification of the pool of RBP-binding

motif might provide new insight to understand virus-associated carcinogenesis.

MATERIALS AND METHODS

Data sets and data availability

All the RNA-seq data were fetched from The Cancer Genome Atlas. We retrieved the level 3 transcriptome raw counts and FPKM profile of head and neck squamous carcinoma samples. The HPV positive/negative statuses were defined by the p16 testing results. Samples without HPV status and their RNA-seq data were excluded for computation. Non-tumor samples were excluded. In all, 111 patient samples were included in the analysis, including 73 HPV negative samples and 38 HPV positive samples. The detailed patient clinical information can be seen in Supplementary Table 1.

Alternative polyadenylation events quantification

The Cancer 3'UTR Atlas yields the characterization of APAome based on RNA-seq by using DaPars [16, 52] (https://github.com/ZhengXia/DaPars). APAome refers to all APA events detected by DaPars. To infer the proximal polyadenylation sites, 3'UTR annotation was generated based on UCSC hg19 reference (python DaPars_Extract_Anno.py -b gene.bed.txt -s symbol.txt -o 3UTR.bed). The Percentage of Distal polyA site Usage Indexes (PDUIs) was computed for all transcripts (python DaPars_main.py my_configure_file.txt). The PDUI score ranges from 0 to 1, which indicate from full shortening to full lengthening of 3'UTR. The larger PDUI indicates the longer 3'UTR that each transcript has. We used the Wilcoxon test for differential APA analysis which has been used before [21]. APA events without enough PDUI scores were excluded for analysis. We define $P < 0.05$ as a significant differential APA event. We use pheatmap package for visualizing the heat map of APA profiles.

Cross-linking immunoprecipitation sequencing analysis

In order to assess the impact of 3'UTR changing, we fetched the RBP-RNA interactome data from RAID [24] (July 2018). In order to retrieve high-confidence interactions between 3'UL genes and paired genes, we regard the pairs with the number of intermediate RBPs > 3 as high-confidence interactions. HUB-3'UL was defined as the 3'UL genes with interacting RBPs > 15. StarBase was used to search the interacting loci between RBP and transcript [53]. The cross-linking immuno-precipitation sequencing (CLIP-seq) data was fetched

from NCBI Sequence Read Archive (accession number SRP014009, SRP017790, and SRP012412).

RNA sequencing analysis

The hg19 genome version was used for transcript annotation. Differential expression analysis was computed through EdgeR with the raw counts of all transcripts as input. We set the cutoff value of $P < 0.001$ and |fold change| > 2. In the correlation analysis, we define the statistical significance with the |Spearman Rho| > 0.3 and P-value < 0.05.

Functional enrichment annotation

Gene ontology enrichment and pathway enrichment annotation were computed through Molecular Signature Database (http://software.broadinstitute.org/gsea/msigdb/index.jsp) and CLUEGO.

Survival analysis and principle component analysis

Patients were split according to the viral status and the longest survival time was set as two years. The P-value for survival analysis was computed through the log-rank test. Principle component analysis was analyzed based on ggfortify package in R. APA events with null values were discarded for principal component analysis urged by ggfortify package.

CONFLICTS OF INTEREST

The authors have no competing interests.

FUNDING

REFERENCES

1. Jayasekara H, MacInnis RJ, Room R, English DR. Long-Term Alcohol Consumption and Breast, Upper Aero-Digestive Tract and Colorectal Cancer Risk: A Systematic Review and Meta-Analysis. Alcohol Alcohol. 2016; 51:315–30.

https://doi.org/10.1093/alcalc/agv110

2. Berger NA. Obesity and cancer pathogenesis. Ann N Y Acad Sci. 2014; 1311:57–76. https://doi.org/10.1111/nyas.12416

3. Krump NA, You J. Molecular mechanisms of viral oncogenesis in humans. Nat Rev Microbiol. 2018; 16:684–98. https://doi.org/10.1038/s41579-018-0064-6

4. de Groot PM, Wu CC, Carter BW, Munden RF. The epidemiology of lung cancer. Transl Lung Cancer Res. 2018; 7:220–33. https://doi.org/10.21037/tlcr.2018.05.06

5. zur Hausen H, de Villiers EM. Cancer "causation" by infections--individual contributions and synergistic networks. Semin Oncol. 2014; 41:860–75. https://doi.org/10.1053/j.seminoncol.2014.10.003

6. Moore PS, Chang Y. Why do viruses cause cancer? Highlights of the first century of human tumour virology. Nat Rev Cancer. 2010; 10:878–89. https://doi.org/10.1038/nrc2961

7. Doorbar J. Molecular biology of human papillomavirus infection and cervical cancer. Clin Sci (Lond). 2006; 110:525–41. https://doi.org/10.1042/CS20050369

8. Sano D, Oridate N. The molecular mechanism of human papillomavirus-induced carcinogenesis in head and neck squamous cell carcinoma. Int J Clin Oncol. 2016; 21:819–26. https://doi.org/10.1007/s10147-016-1005-x

9. Memon MM, Amin E. Cervical cancer caused by human papillomavirus: highly preventable yet highly under-diagnosed. J Pak Med Assoc. 2018; 68:831.

10. McLaughlin-Drubin ME, Munger K. Viruses associated with human cancer. Biochim Biophys Acta. 2008; 1782:127–50. https://doi.org/10.1016/j.bbadis.2007.12.005

11. Kim EK, Choi EJ. Compromised MAPK signaling in human diseases: an update. Arch Toxicol. 2015; 89:867–82. https://doi.org/10.1007/s00204-015-1472-2

12. Zhang Q, Lenardo MJ, Baltimore D. 30 Years of NF-κB: A Blossoming of Relevance to Human Pathobiology. Cell. 2017; 168:37–57. https://doi.org/10.1016/j.cell.2016.12.012

13. Wong JP, Damania B. Modulation of oncogenic signaling networks by Kaposi's sarcoma-associated herpesvirus. Biol Chem. 2017; 398:911–18. https://doi.org/10.1515/hsz-2017-0101

14. Cheng H, Yang X, Si H, Saleh AD, Xiao W, Coupar J, Gollin SM, Ferris RL, Issaeva N, Yarbrough WG, Prince ME, Carey TE, Van Waes C, Chen Z. Genomic and Transcriptomic Characterization Links Cell Lines with Aggressive Head and Neck Cancers. Cell Reports. 2018; 25:1332–1345.e5. https://doi.org/10.1016/j.celrep.2018.10.007

15. Erson-Bensan AE, Can T. Alternative Polyadenylation: Another Foe in Cancer. Mol Cancer Res. 2016; 14:507–17. https://doi.org/10.1158/1541-7786.MCR-15-0489

16. Xia Z, Donehower LA, Cooper TA, Neilson JR, Wheeler DA, Wagner EJ, Li W. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. Nat Commun. 2014; 5:5274. https://doi.org/10.1038/ncomms6274

17. Thivierge C, Tseng HW, Mayya VK, Lussier C, Gravel SP, Duchaine TF. Alternative polyadenylation confers Pten mRNAs stability and resistance to microRNAs. Nucleic Acids Res. 2018; 46:10340–52. https://doi.org/10.1093/nar/gky666

18. Park HJ, Ji P, Kim S, Xia Z, Rodriguez B, Li L, Su J, Chen K, Masamha CP, Baillat D, Fontes-Garfias CR, Shyu AB, Neilson JR, et al. 3' UTR shortening represses tumor-suppressor genes in trans by disrupting ceRNA crosstalk. Nat Genet. 2018; 50:783–89. https://doi.org/10.1038/s41588-018-0118-8

19. Lajer CB, Garnæs E, Friis-Hansen L, Norrild B, Therkildsen MH, Glud M, Rossing M, Lajer H, Svane D, Skotte L, Specht L, Buchwald C, Nielsen FC. The role of miRNAs in human papilloma virus (HPV)-associated cancers: bridging between HPV-related head and neck cancer and cervical cancer. Br J Cancer. 2012; 106:1526–34. https://doi.org/10.1038/bjc.2012.109

20. Zhang Y, Sturgis EM, Sun Y, Sun C, Wei Q, Huang Z, Li G. A functional variant at miRNA-122 binding site in IL-1α 3' UTR predicts risk and HPV-positive tumours of oropharyngeal cancer. Eur J Cancer. 2015; 51:1415–23. https://doi.org/10.1016/j.ejca.2015.04.016

21. Xiang Y, Ye Y, Lou Y, Yang Y, Cai C, Zhang Z, Mills T, Chen NY, Kim Y, Muge Ozguc F, Diao L, Karmouty-Quintana H, Xia Y, et al. Comprehensive Characterization of Alternative Polyadenylation in Human Cancer. J Natl Cancer Inst. 2018; 110:379–89. https://doi.org/10.1093/jnci/djx223

22. Jenal M, Elkon R, Loayza-Puch F, van Haaften G, Kühn U, Menzies FM, Oude Vrielink JA, Bos AJ, Drost J, Rooijers K, Rubinsztein DC, Agami R. The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell. 2012; 149:538–53. https://doi.org/10.1016/j.cell.2012.03.022

23. Masamha CP, Wagner EJ. The contribution of alternative polyadenylation to the cancer phenotype.

Carcinogenesis. 2018; 39:2–10. https://doi.org/10.1093/carcin/bgx096

24. Yi Y, Zhao Y, Li C, Zhang L, Huang H, Li Y, Liu L, Hou P, Cui T, Tan P, Hu Y, Zhang T, Huang Y, et al. RAID v2.0: an updated resource of RNA-associated interactions across organisms. Nucleic Acids Res. 2017; 45:D115–18. https://doi.org/10.1093/nar/gkw1052

25. Sanchez-Mejias A, Tay Y. Competing endogenous RNA networks: tying the essential knots for cancer biology and therapeutics. J Hematol Oncol. 2015; 8:30. https://doi.org/10.1186/s13045-015-0129-1

26. Brenna SM, Syrjänen KJ. Regulation of cell cycles is of key importance in human papillomavirus (HPV)-associated cervical carcinogenesis. Sao Paulo Med J. 2003; 121:128–32. https://doi.org/10.1590/S1516-31802003000300009

27. Pyeon D, Pearce SM, Lank SM, Ahlquist P, Lambert PF. Establishment of human papillomavirus infection requires cell cycle progression. PLoS Pathog. 2009; 5:e1000318. https://doi.org/10.1371/journal.ppat.1000318

28. Krishna S, Ulrich P, Wilson E, Parikh F, Narang P, Yang S, Read AK, Kim-Schulze S, Park JG, Posner M, Wilson Sayres MA, Sikora A, Anderson KS. Human papilloma virus specific immunogenicity and dysfunction of CD8+ T cells in head and neck cancer. Cancer Res. 2018; 78:6159–70. https://doi.org/10.1158/0008-5472.CAN-18-0163

29. Conesa-Zamora P, Doménech-Peris A, Orantes-Casado FJ, Ortiz-Reina S, Sahuquillo-Frías L, Acosta-Ortega J, García-Solano J, Pérez-Guillermo M. Effect of human papillomavirus on cell cycle-related proteins p16, Ki-67, Cyclin D1, p53, and ProEx C in precursor lesions of cervical carcinoma: a tissue microarray study. Am J Clin Pathol. 2009; 132:378–90. https://doi.org/10.1309/AJCPO0WY1VIFCYDC

30. Satapathy S, Batra J, Jeet V, Thompson EW, Punyadeera C. MicroRNAs in HPV associated cancers: small players with big consequences. Expert Rev Mol Diagn. 2017; 17:711–22. https://doi.org/10.1080/14737159.2017.1339603

31. Wang X, Wang HK, Li Y, Hafner M, Banerjee NS, Tang S, Briskin D, Meyers C, Chow LT, Xie X, Tuschl T, Zheng ZM. microRNAs are biomarkers of oncogenic human papillomavirus infections. Proc Natl Acad Sci USA. 2014; 111:4262–67. https://doi.org/10.1073/pnas.1401430111

32. Yang L, Yi K, Wang H, Zhao Y, Xi M. Comprehensive analysis of lncRNAs microarray profile and mRNA-lncRNA co-expression in oncogenic HPV-positive cervical cancer cell lines. Oncotarget. 2016; 7:49917–29. https://doi.org/10.18632/oncotarget.10232

33. Pulikkan JA, Dengler V, Peramangalam PS, Peer Zada AA, Müller-Tidow C, Bohlander SK, Tenen DG, Behre G. Cell-cycle regulator E2F1 and microRNA-223 comprise an autoregulatory negative feedback loop in acute myeloid leukemia. Blood. 2010; 115:1768–78. https://doi.org/10.1182/blood-2009-08-240101

34. Fogal V, Hsieh JK, Royer C, Zhong S, Lu X. Cell cycle-dependent nuclear retention of p53 by E2F1 requires phosphorylation of p53 at Ser315. EMBO J. 2005; 24:2768–82. https://doi.org/10.1038/sj.emboj.7600735

35. Schuldt A. Cell cycle: E2F1 ensures the endocycle. Nat Rev Mol Cell Biol. 2011; 12:768–69. https://doi.org/10.1038/nrm3232

36. Zauli G, Voltan R, di Iasio MG, Bosco R, Melloni E, Sana ME, Secchiero P. miR-34a induces the downregulation of both E2F1 and B-Myb oncogenes in leukemic cells. Clin Cancer Res. 2011; 17:2712–24. https://doi.org/10.1158/1078-0432.CCR-10-3244

37. Zhong R, Bechill J, Spiotto MT. Loss of E2F1 Extends Survival and Accelerates Oral Tumor Growth in HPV-Positive Mice. Cancers (Basel). 2015; 7:2372–85. https://doi.org/10.3390/cancers7040895

38. Slebos RJ, Jehmlich N, Brown B, Yin Z, Chung CH, Yarbrough WG, Liebler DC. Proteomic analysis of oropharyngeal carcinomas reveals novel HPV-associated biological pathways. Int J Cancer. 2013; 132:568–79. https://doi.org/10.1002/ijc.27699

39. Colgan DF, Manley JL. Mechanism and regulation of mRNA polyadenylation. Genes Dev. 1997; 11:2755–66. https://doi.org/10.1101/gad.11.21.2755

40. Millevoi S, Vagner S. Molecular mechanisms of eukaryotic pre-mRNA 3′ end processing regulation. Nucleic Acids Res. 2010; 38:2757–74. https://doi.org/10.1093/nar/gkp1176

41. Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3′-end processing. Cell Mol Life Sci. 2008; 65:1099–122. https://doi.org/10.1007/s00018-007-7474-3

42. Ji X, Kong J, Liebhaber SA. An RNA-protein complex links enhanced nuclear 3′ processing with cytoplasmic mRNA stabilization. EMBO J. 2011; 30:2622–33. https://doi.org/10.1038/emboj.2011.171

43. Richard P, Manley JL. Transcription termination by nuclear RNA polymerases. Genes Dev. 2009; 23:1247–69. https://doi.org/10.1101/gad.1792809

44. Vinciguerra P, Stutz F. mRNA export: an assembly line from genes to nuclear pores. Curr Opin Cell Biol. 2004; 16:285–92. https://doi.org/10.1016/j.ceb.2004.03.013

45. Barreau C, Paillard L, Osborne HB. AU-rich elements

and associated factors: are there unifying principles? Nucleic Acids Res. 2006; 33:7138–50. https://doi.org/10.1093/nar/gki1012

46. Fabian MR, Sonenberg N, Filipowicz W. Regulation of mRNA translation and stability by microRNAs. Annu Rev Biochem. 2010; 79:351–79. https://doi.org/10.1146/annurev-biochem-060308-103103

47. Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell. 2009; 138:673–84. https://doi.org/10.1016/j.cell.2009.06.016

48. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science. 2008; 320:1643–47. https://doi.org/10.1126/science.1155390

49. Tan S, Li H, Zhang W, Shao Y, Liu Y, Guan H, Wu J, Kang Y, Zhao J, Yu Q, Gu Y, Ding K, Zhang M, et al. NUDT21 negatively regulates PSMB2 and CXXC5 by alternative polyadenylation and contributes to hepatocellular carcinoma suppression. Oncogene. 2018; 37:4887–900. https://doi.org/10.1038/s41388-018-0280-6

50. Passacantilli I, Panzeri V, Bielli P, Farini D, Pilozzi E, Fave GD, Capurso G, Sette C. Alternative polyadenylation of ZEB1 promotes its translation during genotoxic stress in pancreatic cancer cells. Cell Death Dis. 2017; 8:e3168. https://doi.org/10.1038/cddis.2017.562

51. Yan H, Tian R, Wang W, Zhang M, Wu J, He J. Aberrant *Ki-67* expression through 3'UTR alternative polyadenylation in breast cancers. FEBS Open Bio. 2018; 8:332–38. https://doi.org/10.1002/2211-5463.12364

52. Feng X, Li L, Wagner EJ, Li W. TC3A: The Cancer 3' UTR Atlas. Nucleic Acids Res. 2018; 46:D1027–30. https://doi.org/10.1093/nar/gkx892

53. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res. 2014; 42:D92–97. https://doi.org/10.1093/nar/gkt1248
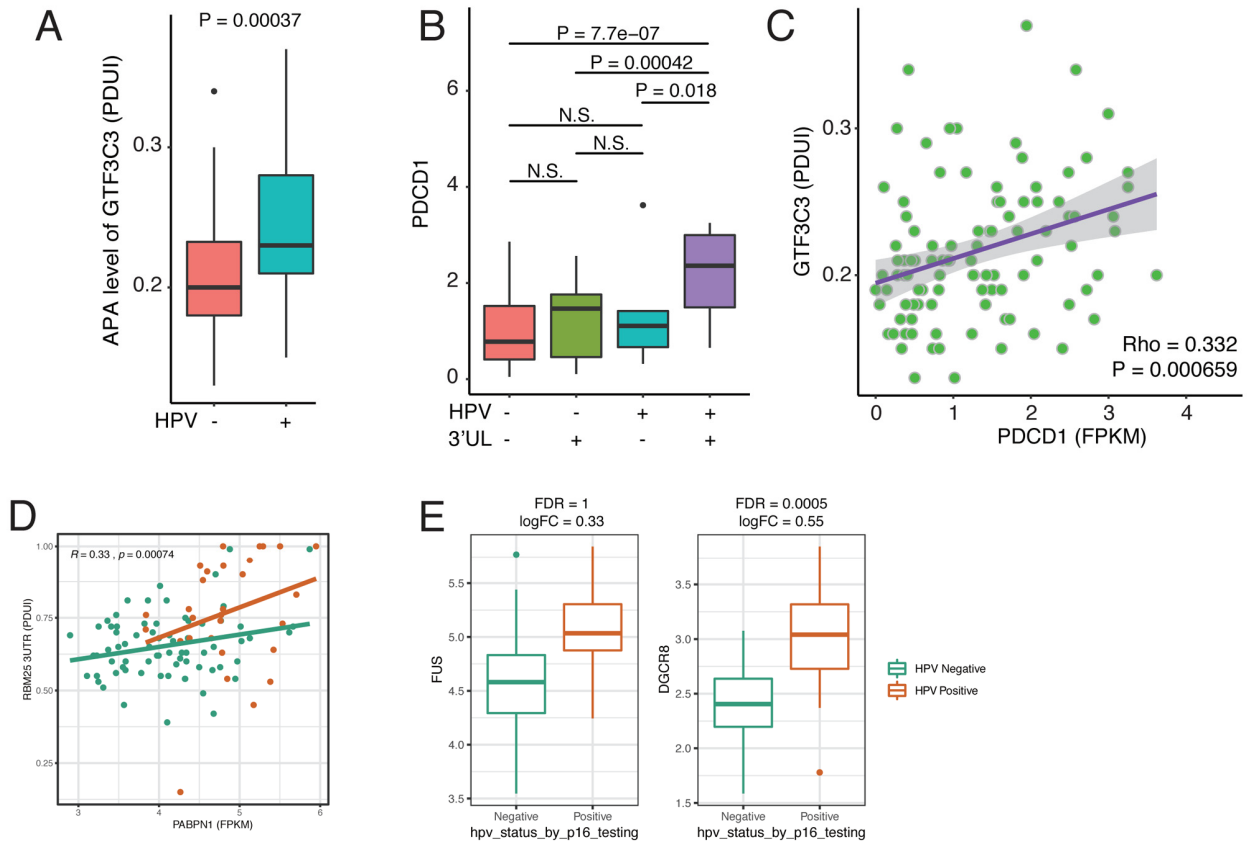
Supplementary Figures
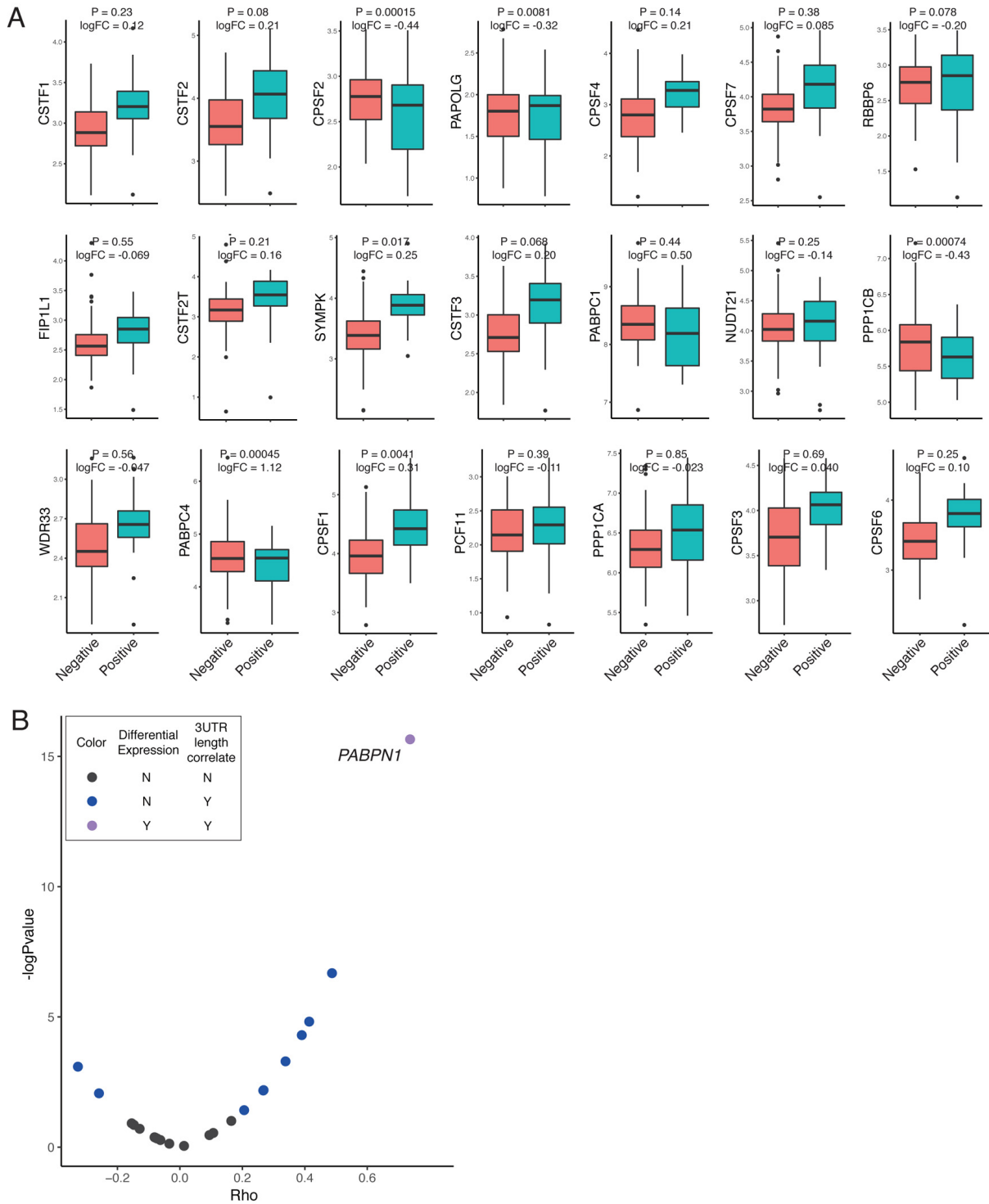
Supplementary Figure 1. The expression profile of key APA regulation components. (**A**) The list of regulators was fetched from the literature (Ref. 20). The Y-axis refers to the FPKM value of each mRNA. (**B**) Scatter plot of all components shows PABPN1 plays key roles in HPV-mediated APA. Each dot represents each known APA regulators, colored according to whether this regulator is differentially expressed in HPV+ samples or whether the expression is correlated with the median 3'UTR length of all APA events (|Spearman Rho| > 0.3 and P-value < 0.05).

A
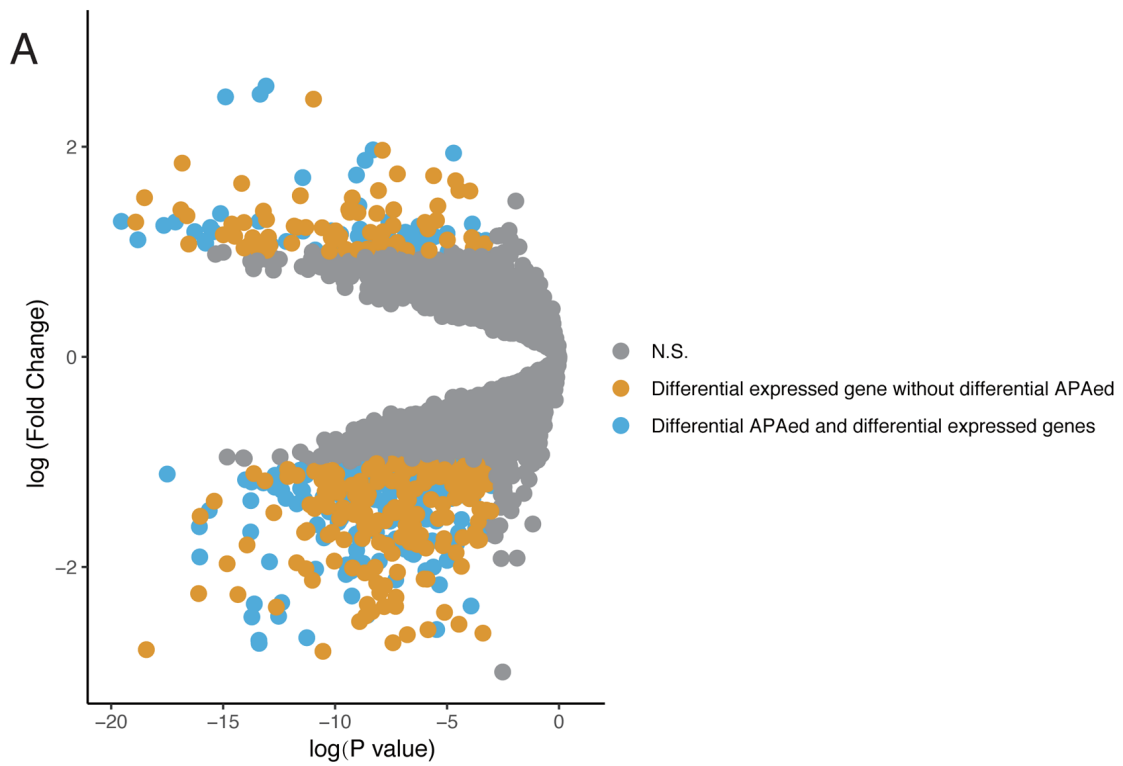
B

| Gene sets	P value (hypergeometric test)	Significance
Tumor supressor (Uniprot)	0.4992319	N.S.
Oncogene (Uniprot)	0.7815282	N.S.
Tumor supressor (TSGD)	0.9227407	N.S.
Oncogene (NCG)	0.998491	N.S.
Immune genes (Bindea 2013)	0.7493049	N.S.
HPV changed genes (MSigDB)	0.7229499	N.S.

Supplementary Figure 2. 3'UTR-longer genes cannot explain the differential expression and are not enriched in cancer-associated pathways. (**A**) Differentially APA genes are not enriched in differentially expressed genes. (**B**) 3'UL genes with direct down-regulated gene expression were not associated with cancer and immunity. All 6 signature shows no significant results (hypergeometric test).

Supplementary Figure 3. HPV upregulates PD-1 via impacting RBPs. (**A**) 3'UTR level of GTF3C3 is significantly longer in HPV+ samples (Wilcoxon test). (**B**) 3'UL+ and HPV+ group shows the highest PDCD1 expression (FPKM). (**C**) 3'UTR length of GTF3C3 is positively correlated with PDCD1 expression. (**D**) Correlation between 3'UTR length of RBM25 and PABPN1 RNA expression. (**E**) The expression difference of FUS and DGCR8 between HPV positive and negative samples.
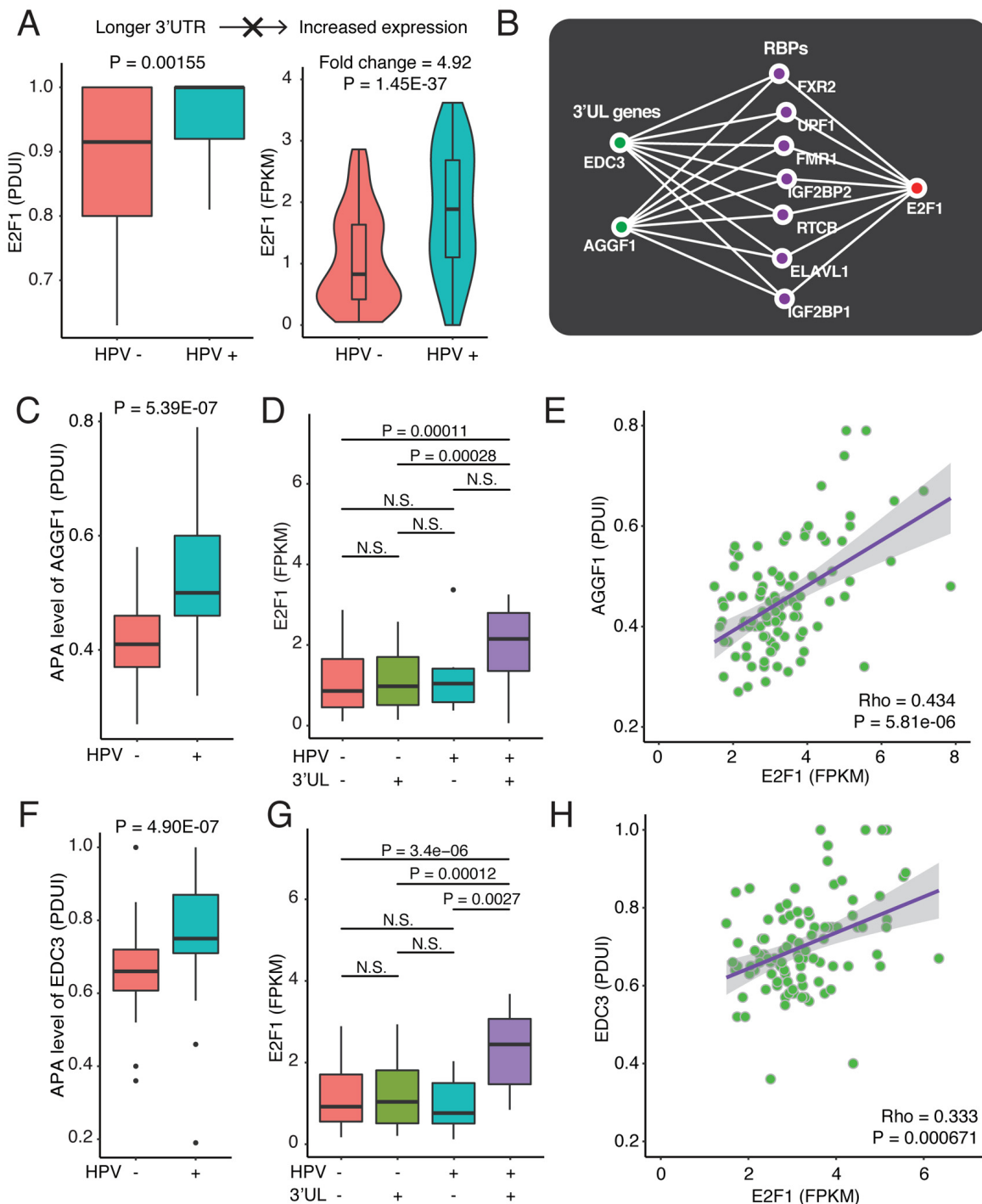
Supplementary Figure 4. HPV-mediated longer 3'UTR of AGGF1, EDC3 upregulates E2F1 via RBPs. (**A**) HPV+ samples show a longer 3'UTR length of E2F1. However, E2F1 is upregulated in HPV+ samples. The down-regulated expression of E2F1 may not be caused by 3'UL in cis. The right plot means PDCD1 is upregulated in the HPV+ samples. Fold change and P-value were computed through EdgeR. The right plot means. (**B**) Based on CLIP-seq, the 3'UL—RBP—E2F1 network was computed. (**C**) 3'UTR level of AGGF1 is significantly longer in HPV+ samples (Wilcoxon test). (**D**) 3'UL+ and HPV+ group shows the highest E2F1 expression (FPKM). (**E**) 3'UTR length of AGGF1 is positively correlated with E2F1 expression. (**F**) 3'UTR level of EDC3 is significantly longer in HPV+ samples (Wilcoxon test). (**G**) 3'UL+ and HPV+ group shows the highest E2F1 expression (FPKM). (H) 3'UTR length of EDC3 is positively correlated with E2F1 expression.

Supplementary Table 1. The clinical information of patients included in the analysis.

			HPV positive (n = 38)		HPV negative (n = 73)	
Age						
		Median (range)	56 (40-71)		60 (19-83)	
Stage						
	I	2	5.3%	4		5.5%
	II	6	15.8%	9		12.3%
	III	5	13.2%	13		17.8%
	IV	25	65.8%	46		63.0%
	Unknown	0	0.0%	1		1.4%
Gender						
	Male	35	92.1%	56		76.7%
	Female	3	7.9%	17		23.3%
Smoking history						
	Never	11	28.9%	11		15.1%
	Former	11	28.9%	31		42.5%
	Current	15	39.5%	30		41.1%
	Unknown	1	2.6%	1		1.4%
Alcohol history						
	YES	29	76.3%	54		74.0%
	NO	9	23.7%	17		23.3%
	unknown	0	0.0%	2		2.7%