HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation

Cecilia Johansson^{1,6}, Monika Somberg^{2,6},
Xiaoze Li¹, Ellenor Backström Winquist¹,
Joanna Fay³, Fergus Ryan⁴, David Pim⁵,
Lawrence Banks⁵ and Stefan Schwartz^{1,*}

¹Section of Medical Microbiology, Department of Laboratory Medicine, Lund University, Lund, Sweden, ²Department of BioNut, Karolinska Institute, Stockholm, Sweden, ³National Children's Research Centre, Our Lady's Children's Hospital, Crumlin, Ireland, ⁴Dublin Institute of Technology, Dublin, Ireland and ⁵International Centre for Genetic Engineering and Biotechnology, Trieste, Italy

We provide evidence that the human papillomavirus (HPV) E2 protein regulates HPV late gene expression. High levels of E2 caused a read-through at the early polyadenylation signal pAE into the late region of the HPV genome, thereby inducing expression of L1 and L2 mRNAs. This is a conserved property of E2 of both mucosal and cutaneous HPV types. Induction could be reversed by high levels of HPV-16 E1 protein, or by the polyadenylation factor CPSF30. HPV-16 E2 inhibited polyadenylation *in vitro* by preventing the assembly of the CPSF complex. Both the N-terminal and hinge domains of E2 were required for induction of HPV late gene expression in transfected cells as well as for inhibition of polyadenylation *in vitro*. Finally, overexpression of HPV-16 E2 induced late gene expression from a full-length genomic clone of HPV-16. We speculate that the accumulation of high levels of E2 during the viral life cycle, not only turns off the expression of the pro-mitotic viral E6 and E7 genes, but also induces the expression of the late HPV genes L1 and L2.

The EMBO Journal (2012) 31, 3212–3227. doi:10.1038/emboj.2012.147; Published online 22 May 2012

Subject Categories: chromatin & transcription; microbiology & pathogens

Keywords: E2; HPV; papillomavirus; polyadenylation; splicing

Introduction

Persistent infection with high-risk human papillomaviruses (HPVs), such as HPV-16, is a risk factor for the development of cervical cancer (zur Hausen, 2002; Howley and Lowy, 2006). In common with the majority of DNA viruses that establish persistence, the HPV-16 life cycle is divided into

early and late phases (Doorbar, 2005; Chow *et al.*, 2010; Bodily and Laimins, 2011). The delay in late gene expression is likely a mechanism of immune evasion that contributes to establishment of HPV persistence. None of the early HPV proteins has yet been shown to regulate HPV late gene expression. However, multiple cellular factors that regulate papillomavirus late gene expression have been identified (Zheng and Baker, 2006; Schwartz, 2008; Graham, 2010; Johansson *et al.*, 2010). We have previously shown that the adenovirus E4orf4 protein can induce HPV-16 late gene expression (Somberg *et al.*, 2009), by interfering with cellular SR proteins (Kanopka *et al.*, 1996, 1998). HPV-5 and HPV-16 E2 interact with SR proteins (Lai *et al.*, 1999; Bodaghi *et al.*, 2009), HPV-16 E2 inhibits splicing *in vitro* (Bodaghi *et al.*, 2009) and HPV-16 E2 activates expression of the *SFRS1* gene (Mole *et al.*, 2009). We speculated that E2 might play the role of a viral activator of HPV late gene expression.

The E2 protein is modular and consists of an N-terminal transactivation domain (TAD), a hinge region, and a C-terminal DNA-binding domain (DBD; Androphy *et al.*, 1987; Giri and Yaniv, 1988; Li *et al.*, 1989; Bedrosian and Bastia, 1990). Binding of E2 to the viral DNA is required for replication of the viral DNA, regulation of transcription from the early promoter and for segregation of the viral episomal DNA genome at cell division (Stenlund, 2003; You *et al.*, 2004; Baxter *et al.*, 2005; Howley and Lowy, 2006; McBride *et al.*, 2006; Parish *et al.*, 2006; McBride, 2008; Kadaja *et al.*, 2009; Thierry, 2009). Inhibition of HPV transcription from the early HPV promoter by E2 is caused by steric hindrance of cellular transcription factors (Dong *et al.*, 1994; Thierry, 2009), but it too depends on the interaction of E2 with cellular factors, including Brd4 (Ilves *et al.*, 2006; McPhillips *et al.*, 2006; Wu *et al.*, 2006; Schweiger *et al.*, 2007; Smith *et al.*, 2010; Yan *et al.*, 2010), SMCX (Smith *et al.*, 2010) and EP400 (Smith *et al.*, 2010). In the viral life cycle, high levels of E2 protein accumulate immediately prior to late gene expression, during a stage of efficient replication of the viral DNA genome (Xue *et al.*, 2010), and one may speculate that E2 activates HPV late gene expression.

We show that HPV E2 induces HPV late gene expression by inhibiting polyadenylation at the early polyA signal on the viral genome. E2 interacts with the cellular polyadenylation factor CPSF30 *in vitro* and perturbs the polyadenylation complex that forms at the polyA signal. We conclude that HPV-16 E2 contributes to induction of HPV-16 late gene expression by inhibiting early polyadenylation

Results

HPV-16 E2 induces HPV-16 late gene expression

To investigate if any of the HPV-16 early proteins can induce HPV-16 late gene expression, we overexpressed HPV-16 early proteins with the subgenomic HPV-16 plasmid pBELM (Figure 1A). This plasmid produces high levels of early

*Corresponding author. Section of Medical Microbiology, Department of Laboratory Medicine, Lund University, BMC-B13, Lund 221 84, Sweden. Tel.: +46 46 222 0628; Fax: +46 46 130064;

E-mail: stefan.schwartz@med.lu.se

⁶These authors contributed equally to this work

Received: 25 October 2011; accepted: 24 April 2012; published online: 22 May 2012

Figure 1 HPV E2 induces HPV-16 late gene expression. **(A)** Schematic representation of the HPV-16 genome and the subgenomic HPV-16 expression plasmids pBEL (Zhao *et al*, 2004) and pBELM (Zhao *et al*, 2004). Nucleotide positions refer to the HPV-16R sequence (Baker and Calef, 1997). Mut indicates a number of nucleotide substitutions that inactivated splicing silencers downstream of SA5639 in plasmid pBELM (Collier *et al*, 2002; Zhao *et al*, 2004). A subset of mRNAs that can be produced by pBEL and pBELM are indicated (Zhao *et al*, 2005). LCR, long control region. **(B)** Northern blot on cytoplasmic RNA extracted from HeLa cells transfected with pBELM in the absence or presence of plasmids expressing HPV-16 E1, E2, E4, E6 and E7, and probed with the L1 probe. **(C)** Northern blot on cytoplasmic RNA extracted from HeLa cells transfected with pBELM in the absence or presence of pCMV16E2, and probed with the E4 probe. **(D)** RT-PCR using primers 3515s and L1a or 757s and L1a on cDNA produced from cytoplasmic RNA extracted from HeLa cells transfected with pBELM in the absence or presence of pCMV16E2. MW, molecular weight marker. **(E)** Western blot on cell extracts from HeLa cells transfected with the expression plasmids for HPV-16 E1, E2 or E4. **(F)** Northern blot on cytoplasmic RNA extracted from HeLa cells transfected with pBEL (Zhao *et al*, 2004) and a two-fold serial dilution of pCMV16E2 expression plasmid, starting at 4 μ g, and probed with the L1 probe. **(G)** Northern blot on cytoplasmic RNA extracted from HeLa cells transfected with pBELM (Zhao *et al*, 2004) and pCMV16E2 in the absence or presence of plasmids expressing HPV-16 E1 (Kadaja *et al*, 2007), E6, E7 (Pim *et al*, 2005) or L2 protein (Oberg *et al*, 2003), and probed with the L1 probe. **(H)** Northern blot on cytoplasmic RNA extracted from HeLa cells transfected with pBEL or pBELM (Zhao *et al*, 2004) in the absence or presence of plasmid pCMV11E2, and probed with the L1 probe. **(I)** Northern blot on cytoplasmic RNA extracted from HeLa cells transfected with pBELM (Zhao *et al*, 2004) in the absence or presence of plasmid pCMV16E2 or pCMV5E2, and probed with the L1 probe. Duplicate transfections are shown. Numbers indicate fold induction of L2/L1 or L1 mRNAs by E2. Figure source data can be found with the Supplementary data.

mRNAs, primarily E4 mRNAs upon transfection of HeLa cells. Co-transfection of plasmid pBELM with plasmids expressing each of the HPV-16 E1, E2, E4, E6 or E7 proteins

revealed that the HPV-16 E2 protein induced high levels of L1 mRNAs from pBELM, whereas none of the other early HPV-16 proteins did (Figure 1B). In contrast, E4 mRNA levels were

relatively unaffected by E2 (Figure 1C). Cloning and sequencing of Reverse Transcription (RT)-PCR products revealed that the L1 mRNAs induced by E2 were spliced from SD880 to SA3358 and further from SD3632 to SA5639 (Figure 1D). pBELM contains mutations that destroy splicing silencers in the L1 coding region, which allows usage of late 3'-splice site SA56329 in these plasmids. The E2 induction of L1 mRNAs from pBELM was not a result of the mutations in L1, as E2 also induced late gene expression from pBEL-OPSA (Supplementary Figure S1), in which sequences downstream of SA5639 are intact, while the 3'-splice site itself has been optimised by extension of the polypyrimidine tract upstream of SA5639. Western blot analysis of the transfected HeLa cells confirmed that E1, E2 and E4 protein production was dependent on transfection with the various expression plasmids (Figure 1E). We were unable to detect E6 and E7 expression with available sera. One cannot exclude that the lack of effect of E6 and E7 on late gene expression may be due to the relatively low expression levels of E6 and E7. Next, we tested the effect of E2 on pBEL (Figure 1A), in which splicing silencers at SA5639 are intact and late mRNA splicing is efficiently suppressed (Zhao *et al*, 2004). In contrast to the induction of spliced L1 mRNA from pBELM, E2 primarily induced production of the L2/L1 mRNA from pBEL (Figure 1F). This suggested that E2 did not activate HPV-16 late mRNA splicing, but rather caused a transcriptional read-through at pAE, into the late region of the HPV-16 genome (Figure 1A). Since E2 functioned on pBEL plasmids that lack the LCR, E2 acted independently of E2 binding sites located in the LCR.

HPV-16 E1 inhibits induction of HPV-16 late gene expression by HPV-16 E2

The ability of E2 to induce HPV-16 late gene expression could possibly be regulated by other HPV proteins that interact with E2. To investigate this, we transfected pBELM and the E2 plasmid with a two-fold excess of expression plasmids that produce proteins that have been shown previously to interact with E2. These plasmids expressed HPV-16 E1 (Hibma *et al*, 1995), E6 (Grm *et al*, 2005), E7 (Gammoh *et al*, 2006) or L2 (Day *et al*, 1998; Heino *et al*, 2000; Okoye *et al*, 2005). As can be seen in Figure 1G, only E1 reduced the ability of E2 to induce HPV-16 late gene expression. These results demonstrated that HPV-16 E1 can inhibit E2-mediated induction of HPV-16 late gene expression, and suggested that high levels of E1 during the viral infection will favour E1- and E2-mediated genome replication over induction of HPV late gene expression by E2. As E1 and E2 are co-expressed during the HPV-16 infection, one may speculate that a high E1/E2 ratio may prevent induction of late gene expression by E2 and favour replication of the HPV genome, whereas a low E1/E2 ratio may favour induction of HPV late gene expression by E2.

Induction of HPV late gene expression is a conserved property of HPV E2

Next, we investigated if E2 from mucosal low-risk HPV-11 as well as cutaneous HPV-5 could activate HPV-16 late gene expression. We found that E2 from HPV types 5 and 11 could induce HPV-16 late gene expression to the same extent as HPV-16 (Figure 1H and I), demonstrating that the ability of E2

to induce late gene expression is a conserved property among both cutaneous and low-risk and high-risk mucosal HPVs.

To further investigate if the ability of E2 to induce late gene expression is a general property of HPV, we wished to analyse the effect of HPV-16 E2 on a distantly related, cutaneous HPV type, not associated with cancer. To this end, we generated an HPV type 1 subgenomic expression plasmid named pBEL-1, in which a CMV promoter was inserted at nucleotide position 684 of the HPV-1 genome (Figure 2A). Similarly to the HPV-16 plasmid pBEL (Figure 2B), pBEL-1 produced primarily the E4 mRNA, detected by the CMV probe that hybridised to the leader sequence in both HPV-16 pBEL and HPV-1 pBEL-1 (Figure 2B and C). Sequencing of the RT-PCR product shown in Figure 2C confirmed that it was the result of splicing from SD827 to SA3200 in HPV-1. The HPV-1 polyadenylation site was mapped to nucleotide positions 4023–25 by 3'-rapid amplification of cDNA ends (3'-RACE; Figure 2D and E). Overexpression of the HPV-16 E2 protein induced HPV-1 late gene expression (Figure 2F), while the effect of HPV-16 E2 on the HPV-1 E4 mRNA was minor (Figure 2F). We have shown previously that both adenovirus E4orf4 and PTB induce HPV-16 late gene expression (Somberg *et al*, 2008, 2009). Adenovirus E4orf4 induced HPV-1 late gene expression, while PTB did not (Figure 2F). High levels of PTB were produced by the pCMVPTB plasmid (Figure 2G). Therefore, HPV-1 and HPV-16 late gene expression may respond differentially to various factors, but the ability of E2 to induce HPV late gene expression is conserved.

HPV-16 E2 induces late gene expression from a full-length genomic clone of HPV-16

To determine if E2 could act on a full-length genomic clone of HPV-16, we transfected the E2 plasmid together with pHPV16ALuc (Figure 3A). This plasmid encodes the entire HPV-16 genome with the exception of the 3'-end of the L1 gene that has been replaced with the poliovirus 2A internal ribosome entry site (IRES) followed by the luciferase reporter gene (Figure 3A). The luciferase gene acts as a surrogate marker for HPV-16 late gene expression. As expected, this plasmid produced only low levels of luciferase upon transfection of C33A cells (Figure 3B). Co-transfection of a serial dilution of the E2 plasmid induced luciferase expression from this construct in a plasmid dose-dependent manner (Figure 3C). Analysis of L1 mRNA levels by real-time RT-PCR confirmed that E2 induced production of late mRNAs from pHPV16ALuc (Figure 3D). To determine if E2 could induce HPV-16 late gene expression from episomal copies of the genome, the HPV-16 genome flanked by loxP sites in pHPV16ANsL (Supplementary Figure S2A) was released by the cre recombinase, as recently described for HPV-18 (Wang *et al*, 2009). Overexpression of HPV-16 E2 induced late gene expression (Supplementary Figure S2C). Only episomal copies of the HPV-16 genome could be detected (Supplementary Figure S2B). We concluded that HPV-16 E2 could induce late gene expression from a full-length HPV-16 genome.

Both the transactivation and hinge domains of HPV-16 E2 are required for induction of HPV-16 late gene expression

Next, we investigated the ability of various E2 mutants to induce HPV-16 late gene expression (Figure 4A). These mutants were efficiently expressed in transfected cells, with

Figure 2 HPV E2 induces HPV-1 late gene expression. **(A)** Schematic representation of the HPV-1 genome and the subgenomic HPV-1 expression plasmid pBEL-1. **(B)** Northern blot on cytoplasmic RNA extracted from HeLa cells transfected with HPV-16 plasmid pBEL (Zhao *et al*, 2004) or HPV-1 plasmid pBEL-1. The blot was probed with the CMV probe that detects both HPV-16 mRNAs produced from pBEL (Zhao *et al*, 2004) and HPV-1 mRNAs produced from pBEL-1 (Collier *et al*, 2002; Zhao *et al*, 2004). **(C)** RT-PCR using HPV-1 primers 827s and E4a on cDNA generated from cytoplasmic RNA extracted from HeLa cells transfected with pBEL-1. **(D)** 3'-RACE products derived from cytoplasmic RNA extracted from HeLa cells transfected with pBEL-1. Arrow points to the band that represents cDNA synthesised from mRNAs that are polyadenylated at HPV-1 pAE. **(E)** Sequencing of 3'-RACE products shown in (D) demonstrated that HPV-1 early mRNAs were cut and polyadenylated at genomic nucleotide positions 4023–25. **(F)** Northern blot on cytoplasmic RNA extracted from HeLa cells transfected with pBEL-1 in the absence or presence of pCMV16E2 (encoding HPV-16 E2), pE4orf4 (encoding adenovirus E4orf4) (Shtrichman *et al*, 1999) or pCMVPTB (encoding polypyrimidine tract binding protein 1). Blots were probed with HPV-1 L1 or E4 probe. **(G)** Western blot on HeLa cells transfected with empty CMV vector (–) or a CMV-PTB expression plasmid (+). Figure source data can be found with the Supplementary data.

the exception of the hinge domain (Figure 4B). The mutational analyses of E2 showed that both the N-terminal TAD and hinge domains of HPV-16 E2 were required for induction of late gene expression from pBELM (Figure 4C). We also tested two previously described and characterised E2 mutants. M11 has an amino-acid substitution that makes E2 unable to support replication of viral DNA, but still able to

activate transcription. M12 has a four amino-acid deletion and was shown to be defective in transcription regulation, but could replicate viral DNA (M12) (Piccini *et al*, 1997; Figure 4A). The M11 mutant activated HPV-16 late gene expression, whereas M12 did not (Figure 4D). These results demonstrated that the ability of HPV-16 E2 to induce HPV-16 late gene expression was distinct from its effects on

Figure 3 HPV-16 E2 induces late gene expression from the HPV genome. (A) Schematic representation of the HPV-16 genome and genomic HPV-16 expression plasmid pHPV16ALuc. Poliovirus IRES followed by the luciferase coding sequence was inserted between *Bam*HI in HPV-16 L1 and the L1 stop codon. The most common E6/E7 mRNA is indicated. In addition, the read-through mRNAs induced from this plasmid by E2 overexpression are indicated (late/Luc mRNA). (B) Luciferase expression in C33A cells transfected with pHPV16ALuc or a CMV promoter-driven luciferase expression plasmid (pLuc). (C) Relative luciferase levels monitored in C33A cells transfected with pHPV16ALuc and a 10-fold serial dilution of the pCMV16E2 expression plasmid, starting at 1 μ g. Bars represent fold increase in luciferase production from pHPV16ALuc in the presence of increasing concentrations of transfected pCMV16E2 expression plasmid. Error bars show standard deviation from triplicate transfections. (D) Real-time PCR on total RNA extracted from the C33A cells shown in (C). Primers used for PCR of spliced L1 mRNA are shown in (A). Mean values and standard deviation from two parallel transfection experiments, each performed in triplicates, are shown. The qPCR was normalised to qPCR on the GAPDH mRNAs in the same samples.

viral DNA replication, but might be coupled to the effects on transcription.

The effects of the E2 mutants were also analysed using a reporter plasmid named pBELMCAT, in which the poliovirus 2A IRES element followed by the CAT reporter gene had been inserted in the 3'-end of the L1 gene (Figure 4E). CAT protein levels serve as a surrogate marker for HPV-16 late gene expression. Co-transfection of pBELMCAT with the various E2 mutants confirmed the northern blot results, in that E2, E2NH and M11 induced CAT production, while E2HC and M12 did not (Figure 4F and G). To confirm these results in human primary keratinocytes, the natural target cell of HPV-16, we replaced the CAT reporter gene with the gene for luciferase, for which there is a more sensitive assay, resulting in pBELLuc and pBELMLuc (Figure 4E). Co-transfection of these plasmids with pCMV16E2, or plasmids expressing E2 mutants, into human primary keratinocytes, confirmed the results obtained in HeLa cells (Figure 4H, left and right panel). However, in keratinocytes, the levels of induction were generally higher and the E2NH mutant induced higher levels of late gene expression than the full-length E2 protein

(Figure 4H). We concluded that the hinge domain and the N-terminal TAD were required for induction of HPV-16 late gene expression, both in human primary keratinocytes and in cervical cancer cells, while the C-terminal DBD was not.

HPV-16 E2 acts on a 300-nucleotide fragment of the HPV-16 genome that encodes the early polyA signal pAE

Data shown above indicated that HPV-16 E2 caused a read-through into the late region of the HPV-16 genome, indicating that E2 inhibited the HPV-16 early polyA signal pAE. To narrow-down the target sequence of HPV-16 E2 on the HPV-16 genome, a 520- or 300-nucleotide sequence encompassing the HPV-16 pAE was inserted between the CMV promoter and the CAT gene in the pM2-CAT plasmid, generating pM2A-CAT and pM2ADE-CAT (Figure 5A), respectively. These insertions both inhibited expression of CAT mRNA (Figure 5B). Overexpression of HPV-16 E2 induced production of the CAT mRNA from both pM2A-CAT and pM2ADE-CAT (Figure 5B), whereas CAT expression from pM2-CAT, which lacks HPV sequences, was unaffected by HPV-16 E2 (Figure 5B). These results demonstrated that a

Figure 4 The E2 N-terminus and hinge are required to induce HPV-16 late gene expression. **(A)** Schematic representation of the HPV-16 E2 protein. The transactivation domain (TAD), hinge domain and DNA-binding domain (DBD) are indicated. Numbers represent amino-acid positions at the borders of the E2 protein domains. Various E2 mutants are shown. **(B)** Western blot on cell extracts from HeLa cells transfected with the indicated plasmids. The E2HC protein also migrates as a dimer (*). **(C, D)** Northern blots on cytoplasmic RNA extracted from HeLa cells transfected with pBELM (Zhao *et al*, 2004) in the absence or presence of plasmids expressing HPV-16 E2 or E2 mutants shown in **(A)**. The blot was probed with L1 probe (Figure 1A). Fold induction of L1 mRNA by E2 is indicated below the gels. **(E)** Schematic representation of the HPV-16 subgenomic expression plasmids pBELMCAT, pBELLuc and pBELMLuc. The IRES is followed by the CAT gene or the luciferase gene (Luc). A subset of mRNAs produced by these plasmids is indicated. **(F, G)** Levels of CAT protein produced from pBELMCAT transfected into HeLa cells in the absence or presence of plasmids expressing HPV-16 E2, or E2 mutants. Arbitrary CAT units monitored by CAT ELISA are shown. **(H)** Relative levels of luciferase activity produced from pBELLuc or pBELMLuc transfected into primary human keratinocytes in the absence or presence of E2 expression plasmids. Error bars represent standard deviations from triplicate transfections.

300-nucleotide sequence encompassing HPV-16 pAE was sufficient to subject the CAT gene to E2 regulation. We subjected RNA from the pM2A-CAT or pM2ADE-CAT transfected cells to 3'-RACE, which confirmed that pAE was efficiently utilised (Figure 5C). Cloning and sequencing of

the 3'-RACE product mapped the cleavage and polyadenylation site to position 4232/4233/4234, as previously described (Baker and Calef, 1997; Zhao *et al*, 2005). In addition to the CAT reporter plasmid, we generated the luciferase reporter plasmid pE5Luc (Figure 5D), in which an HPV-16 sequence

Figure 5 HPV-16 E2 targets the early polyA signal. (A) Schematic representation of the HPV-16 genome and the pM2A-CAT, pM2ADE-CAT and pM2-CAT expression plasmids. The HPV-16 sequences present in the CAT plasmids are indicated. IRES, poliovirus 2A internal ribosome entry site. (B) Northern blots on cytoplasmic RNA extracted from HeLa cells transfected with pM2A-CAT and pM2ADE-CAT, or pM2-CAT in the absence or presence of pCMV16E2, probed with CAT probe. RNAs from duplicate transfections are shown. (C) 3'-RACE products derived from cytoplasmic RNA extracted from HeLa cells transfected with the indicated plasmids. Arrow points to the band that represents cDNA synthesised from mRNAs that are polyadenylated at HPV-16 pAE. (D) Schematic representation of the HPV-16 genome and the pE5Luc expression plasmid. Luc, luciferase gene. The borders of the HPV-16 sequences present in the pE5Luc plasmid are indicated. The E5 and Luc mRNAs that can be produced by pE5Luc are indicated. (E) 3'-RACE products derived from cytoplasmic RNA extracted from A549 cells transfected with pE5Luc. Arrow points to the band that represents cDNA synthesised from mRNAs that are polyadenylated at HPV-16 pAE. (F) Relative luciferase activity in A549 cells (left panel) or human primary keratinocytes (HEK293) transfected with plasmid pE5Luc in the absence or presence of plasmids encoding HPV-16 E2, E2NH or E2HC. Error bars show standard deviation from at least three independent transfection experiments. Figure source data can be found with the Supplementary data.

encoding the entire HPV-16 E5 open reading frame, the HPV-16 early untranslated region (UTR) and the pAE was inserted upstream of the poliovirus 2A IRES and the luciferase

reporter gene (Figure 5D; Backström Winquist *et al*, 2012). Cloning and sequencing of 3'-RACE products revealed that pAE was used in pE5Luc (Figure 5E). Overexpression of

HPV-16 E2 or E2NH significantly enhanced production of luciferase activity from pE5Luc while E2HC had no detectable effect (Figure 5F). Similar results were obtained by transfection of primary human keratinocytes (HEKn; Figure 5F), with the exception that E2NH induced higher levels of luciferase than E2 in primary keratinocytes (Figure 5F). These results supported the conclusion that HPV-16 E2 induced late gene expression by inhibiting the early polyA site pAE.

Inactivation of HPV-16 pAE by nucleotide substitutions or overexpression of known cellular, polyadenylation inhibitory factors mimics the effect of HPV-16 E2 on HPV-16 late gene expression

If E2 inhibits pAE, then mutational inactivation of pAE should induce L2/L1 mRNAs from pBEL, and L1 mRNAs from pBELM, as E2 did (Figure 1B, C and F). Inactivation of pAE with six nucleotide substitutions in either pBEL or pBELM, resulting in plasmids pBELDP and pBELDPM (Zhao *et al*, 2005), respectively, caused an induction of L2/L1 or L1 mRNA (Supplementary Figure S3A), similarly to the effects of E2 on pBEL and pBELM (Supplementary Figure S3B). In addition, the HuR and the influenza virus NS1 proteins with documented inhibitory effect on polyadenylation (Nemeroff *et al*, 1998; Zhu *et al*, 2007; Dai *et al*, 2012) exhibited the same effect on HPV-16 late gene expression as HPV-16 E2 (Supplementary Figure S3C and D), while the SR

protein 9G8 had no detectable effect on HPV-16 late gene expression (Supplementary Figure S3C). In addition, NS1 induced luciferase production from the reporter plasmid pE5Luc (Supplementary Figure S3E) in a manner similar to HPV-16 E2 (Figure 5F). Collectively, these results strengthened the conclusion that HPV-16 E2 inhibits HPV-16 early polyadenylation.

HPV-16 E2 inhibits HPV-16 pAE more efficiently than pAL

Since our results indicated that E2 induced late gene expression by inhibiting the HPV-16 early polyA signal, we speculated that E2 exerted a greater inhibitory effect on pAE than on pAL, as that would cause a higher induction of late gene expression, than if E2 inhibited both polyA signals to the same extent. To this end, we used a reporter plasmid named p2xpAL, in which the pAE in pBSpliceM had been replaced with pAL (Figure 6A; Somberg *et al*, 2008). pBSpliceM is a previously described plasmid that contains the early polyA signal and only two splice sites named SD3632 and SA5639 that are used exclusively to produce L1 mRNAs. This plasmid does not produce L1 mRNAs unless the pAE is deleted, or the splice sites are optimised. As pAL is efficiently used in HeLa cells, plasmid p2xpAL did not produce L1 mRNA either. Overexpression of HPV-16 E2 efficiently induced L1 mRNA from pBSpliceM (Figure 6B and C), but acted less efficiently

Figure 6 Stronger inhibition of HPV-16 pAE than pAL by E2. **(A)** Schematic representation of the HPV-16 subgenomic expression plasmids pBSpliceM (Somberg *et al*, 2008) and p2xpAL (Somberg *et al*, 2008). In p2xpAL, the sequence between nucleotide positions 4211 and 4534, which encodes the early polyA signal pAE, has been deleted and replaced by a sequence spanning positions 7924–7454 that contains the late polyA signal pAL. A subset of mRNAs that may be produced by these plasmids are indicated (Somberg *et al*, 2008). **(B)** Northern blot on cytoplasmic RNA extracted from HeLa cells transfected with pBSpliceM (Somberg *et al*, 2008) or p2xpAL (Somberg *et al*, 2008) in the absence or presence of pCMV16E2 (triplicate transfections with p2xpAL) or pCMVPTB (duplicate transfections p2xpAL). The blot was probed with the L1 probe. **(C)** Fold induction of L1 mRNA levels by E2 or PTB obtained in two independent experiments.

on p2xpAL (Figure 6B and C), indicating that E2 inhibited polyadenylation at pAE more efficiently than polyadenylation at pAL. In contrast, PTB induced high levels of the L1 mRNA in the presence of pAL (Figure 6B and C). Note that the L1 mRNAs produced from pBSpliceM and p2xpAL are identical as a result of splicing (Figure 6A). Therefore, differences in RNA stability cannot explain the different effect of E2 on the two plasmids. The preferred inhibitory effect of E2 on pAE over pAL offers an explanation for the induction of late mRNA production since late mRNAs use pAL.

HPV-16 E2 inhibits polyadenylation *in vitro*

Transfection experiments indicated that HPV-16 E2 inhibited the HPV early polyadenylation site to induce HPV-16 late gene expression. We therefore wished to determine whether E2 could inhibit polyadenylation *in vitro*. We used the well-characterised adenovirus L3 polyadenylation signal (Skolnik-David *et al*, 1987). *In-vitro* polyadenylation with HPV-16 pAE was unsuccessful, probably due to the requirement for downstream sequences (Terhune *et al*, 1999, 2001; Oberg *et al*, 2003, 2005). Radiolabelled RNA was synthesised from pT7L3short (Figure 7A) by run-off transcription, and *in-vitro* polyadenylation assays using HeLa cell nuclear extract and various recombinant GSTE2 proteins (Figure 7B) were performed for 0.5–2 h at 30°C. The results revealed that E2 inhibited *in-vitro* polyadenylation of the L3short RNA substrate (Figure 7C). Longer incubation times yielded a higher degree of polyadenylation, but the effect of E2 was seen at all time points. Mean values and standard deviations of three independent *in-vitro* polyadenylation assays are shown (Figure 7C, right). The full E2 protein appeared to inhibit *in-vitro* polyadenylation to a greater extent than E2NH and E2HC, but the difference was not significant (Figure 7D). Mean values and standard deviations of three independent *in-vitro* polyadenylation assays are shown (Figure 7D, right). These results are in line with the previously reported transfection experiments shown in Figures 4 and 5.

HPV-16 E2 interferes with polyadenylation complexes *in vitro*

The inhibitory effect of E2 on polyadenylation *in vitro* suggested that E2 targets the core polyadenylation machinery. Formation of polyadenylation complexes were analysed in RNA gel shift experiments using HeLa cell nuclear extracts and the L3 RNAs (Figure 7A). Upon addition of nuclear extract to the RNA, two slower migrating complexes were formed (U and PA) (Figure 8A, left panel). The slowest migrating complex named PA proved to be the specific polyadenylation complex, as it did not form on a mutant polyadenylation substrate named L3G (Figure 8A, right panel), in which the polyA signal had been mutationally inactivated (Figure 7A). The unspecific complex was named U (Figure 8A and B). A competition experiment with unlabelled L3 or L3G competitor RNA revealed that only L3 RNA competed with the PA complex on the wild-type L3 substrate (Figure 8A).

To further demonstrate that the PA complex contained polyadenylation factors, antibodies against CPSF160, CPSF73, CPSF30 and CstF64 were added to the L3 RNA together with nuclear extract. Addition of antibodies against CPSF30 and CstF64 resulted in a supershift of the specific PA

Figure 7 HPV-16 E2 inhibits polyadenylation *in vitro*. (A) Schematic representation of adenovirus L3 polyadenylation constructs that were used to produce the L3 (Skolnik-David *et al*, 1987), L3G (Skolnik-David *et al*, 1987) and L3short RNAs. (B) Coomassie Brilliant Blue stained SDS-PAGE gel with GST and various GSTE2 fusion proteins purified from *E. coli*. MW, molecular weight marker. (C, D) *In-vitro* polyadenylation assays on the L3short transcript performed in HeLa nuclear extracts in the presence of ATP. The assay was performed in the absence or presence of the indicated GST or GSTE2 fusion proteins. The positions of the unprocessed RNA and the polyadenylated product are indicated. The levels of mRNA and polyadenylated mRNAs were quantified in a phosphorimager and the ratio of polyadenylated RNA versus mRNAs was calculated for each sample. The ratios obtained in the presence of various GSTE2 fusion proteins were divided by the ratio obtained in the presence of GST alone, resulting in the relative amount of poly(A) product (shown below the gels). The mean values and standard deviation of the relative amount of poly(A) product from three independent experiments are shown to the right of each *in-vitro* polyadenylation gel.

complex (Figure 8B; Supplementary Figure S4A). The antibody against CPSF73 had a small effect, if any, on the PA complex (Figure 8B; Supplementary Figure S4A). In addition, an antibody against the RNA binding polyadenylation factor CPSF160 prevented complex formation (Figure 8B; Supplementary Figure S4A). None of these antibodies affected the unspecific complex U (Figure 8B). These results supported the conclusion that the slowest migrating PA complex was the specific polyadenylation complex and contained the core polyadenylation machinery.

Next, we wished to determine the effect of the HPV-16 E2 protein on the polyadenylation complex. Addition of serially diluted GSTE2 protein to the nuclear extract before incubation

Figure 8 HPV-16 E2 targets the cellular polyadenylation complex *in vitro*. (A) RNA gel shift experiments using radiolabelled L3 or L3G RNAs (see Figure 7A) incubated in HeLa nuclear extracts, in the absence or presence of unlabelled L3 or L3G RNA competitors. Numbers represent fold excess of unlabelled L3 or L3G RNA versus radiolabelled L3 or L3G RNA probe. (B) Radiolabelled L3 RNA was incubated in HeLa nuclear extracts in the presence of antibodies against polyadenylation factors CPSF160, CPSF73, CPSF30 or CstF64. See also Supplementary Figure S3. (C) Radiolabelled L3 RNA was incubated in HeLa nuclear extracts supplemented with purified GST or a two-fold serial dilution of GSTE2 protein. (D) Radiolabelled L3short RNA was incubated in HeLa nuclear extracts supplemented with purified GST, GSTE2, GSTE2NH or GSTE2HC protein. PA, specific polyadenylation complex; U, unspecific complex; P, radiolabelled RNA probe.

with L3 or L3short RNA (Figure 7A) significantly affected the migration of the PA complex in the gel (Figure 8C and D; Supplementary Figure S4B and C). Addition of GSTE2 caused an altered mobility of the specific PA complex in a GSTE2 dose-dependent manner, while the unspecific U complex was unaffected (Figure 8C; Supplementary Figure S4B and C). We concluded that addition of GSTE2 altered the specific PA complex.

Since E2 and E2NH, but not E2HC, inhibited polyadenylation in transfected cells, we speculated that GSTE2 and GSTE2NH should have the same effect on the PA complex and that GSTE2HC should not affect the PA complex. Analysis of various E2 deletion mutants on polyadenylation complex formation demonstrated that full-length GSTE2 and GSTE2NH altered the mobility of the specific PA complex in a similar manner, while GSTE2HC did not (Figure 8D; Supplementary Figure S4C). Together, these results demonstrated that E2 targets the core polyadenylation complex.

E2 binds to the CPSF complex and overexpression of CPSF30, but not CPSF73 or Fip1, prevents induction of HPV-16 late gene expression

In order to investigate if HPV-16 E2 interacts with polyadenylation factors directly, we performed *in-vitro* pull-down

experiments with GSTE2 and *in-vitro* translated and radiolabelled polyadenylation factors CPSF160, CPSF100, CPSF73 and CPSF30. GST served as a negative control for the bait. GSTE2 interacted with all members of the CPSF complex, but to various extent (Figure 9A). CPSF30 and CPSF100 were pulled down by GSTE2 (Figure 9A). Weaker bands were obtained with CPSF160 and CPSF73 (Figure 9A). GSTE2 did not interact with the cellular transcription factor CtBP (Figure 9A). GSTE2 interacted less well with the polyadenylation factors than with HPV-16 E1 (Figure 9A). We also performed a pull-down experiment with GSTE2 and CPSF30 in the absence or presence of RNaseA/T1 to show that the interaction between the two was not bridged by RNA. Similar results were obtained in the absence or presence of RNaseA/T1 (Figure 9B). To investigate if the same interaction could be detected in living cells, co-immunoprecipitations of HPV-16 E2 and HA-tagged CPSF30 or CPSF73 expressed from plasmids in transfected HeLa cells were performed. The results revealed that immunoprecipitation of E2 brought down CPSF30, but not CPSF73, although both polyadenylation factors were expressed to a similar extent in the transfected cells (Figure 9C; Supplementary Figure S5A). We concluded that HPV-16 interacted with cellular polyadenylation factors

Figure 9 HPV-16 E2 interacts with cellular polyadenylation factors. (A) Pull-down assay using GST or GSTE2 fusion protein coupled to sepharose beads incubated with *in-vitro* translated and ^{35}S -Met/Cys-radiolabelled polyadenylation factors, HPV-16 E1 or CtBP. Proteins in purified complexes were separated on SDS-PAGE gels and visualised by autoradiography. Input represents 5% of labelled protein used in the pull-down assay. (B) GSTE2 pull down of CPSF30 in the absence or presence of RNaseA/T1 treatment. (C) Co-immunoprecipitation of E2 and CPSF30 or CPSF73 from HeLa cells transfected with pCMV16E2 and HA-tagged CPSF30 or CPSF73 expression plasmids. Immunoprecipitation was performed with E2 antiserum TVG261 and western blotting was performed with anti-HA antibody. (D–F) Northern blot on cytoplasmic RNA extracted from HeLa cells transfected with pBELM (Zhao *et al*, 2004) in the absence or presence of pCMV16E2 and/or CPSF30, CPSF73 or Fip1 expression plasmid. The blots were probed with the L1 probe. (G) Northern blot on cytoplasmic RNA extracted from HeLa cells transfected with pBELM (Zhao *et al*, 2004) in the absence or presence of 1 μg of pCMV16E2 or pCMV16E2 and a two-fold serial dilution of the CPSF30 expression plasmid, starting at 4 μg , and probed with the L1 probe. Figure source data can be found with the Supplementary data.

in vitro and in transfected cells, albeit with relatively low efficiency.

If HPV-16 E2 acted on the polyadenylation complex, then overexpression of one or more of the polyadenylation factors with E2 in transfection experiments might restore early polyadenylation and prevent activation of late gene expression by E2. We therefore co-transfected the HPV-16 E2 expression plasmid and the HPV-16 reporter plasmid pBELM with expression plasmids encoding a subset of

cellular polyadenylation factors, including CPSF30, CPSF73 and Fip1. Western blot analysis showed that CPSF30, CPSF73 and Fip1 were expressed in the transfected cells (Supplementary Figure S5A). HPV-16 L1 mRNA production from pBELM was monitored by northern blotting. The results revealed that overexpression of CPSF30 inhibited induction of HPV-16 late gene expression by HPV-16 E2, whereas Fip1 and CPSF73 did not (Figure 9D–F). This effect was dependent on the ratio of CPSF30 to HPV-16 E2 expression plasmid

co-transfected with pBELM (Figure 9G). As the results with CPSF73 and Fip1 were negative, we cannot exclude that higher levels would inhibit E2.

It is well established that the influenza virus NS1 protein interacts with CPSF30 to inhibit mRNA polyadenylation (Nemeroff *et al*, 1998; Noah *et al*, 2003). To validate the results from the transfection competition assay, we co-transfected the influenza virus NS1 expression plasmid with serially diluted CPSF30 or CstF64 expression plasmid and monitored the effect on the induction of luciferase expression from the reporter plasmid pE5Luc (Figure 5D) by NS1. As expected, overexpression of CPSF30, but not CstF64, prevented induction of luciferase from pE5Luc by NS1 (Supplementary Figure S5B). This supported our conclusion that induction of luciferase from pE5Luc is dependent on pAE inhibition, and that pAE inhibition by E2 or NS1 can be revealed by overexpression of CPSF30. Collectively, our results indicated that E2 targets the CPSF-containing polyadenylation complex by interacting directly with CPSF30, or with a factor binding to CPSF30. This interaction causes inhibition of HPV-16 early polyadenylation signal pAE, and results in induction of HPV-16 late gene expression.

Discussion

Since E2 targets the general polyadenylation factor CPSF30 that is required for polyadenylation of the majority of the mRNAs in the cell, the question arises as how inhibition of polyadenylation through CPSF30 can specifically inhibit HPV pAE to cause a read-through into HPV the late region and produce late mRNAs that are polyadenylated at pAL, without inhibiting pAL too. The simplest explanation is that a general inhibition of polyadenylation will favour HPV late gene expression. For example, if 95% of the mRNAs are polyadenylated at pAE, and E2 causes a two-fold reduction of early polyadenylation, early gene expression would drop from 95 to 47.5% whereas late gene expression would increase from 5 to 47.5%, a near 10-fold activation in late gene expression. However, if E2 has a general inhibitory effect on polyadenylation, it would reduce polyadenylation at pAL by a factor of 2 as well, down to 23.75%. Late gene expression would still increase from 5 to 23.75%, a near five-fold increase. This requires that E2 is relatively inefficient in inhibiting polyadenylation, or it would be cytotoxic. In conclusion, a partial inhibitory effect of E2 on polyadenylation could cause a significant increase in HPV late gene expression, while simultaneously allowing production of mRNAs that are polyadenylated at pAE, which is what is seen at the late stage in the HPV life cycle. Although HPV-16 pAE appears to be more sensitive to E2 than pAL, the inhibition of the adenovirus L3 polyA signal *in vitro* and the interactions of E2 with general polyadenylation factors, argue for a general, but relatively low inhibitory effect on polyadenylation by E2. In contrast, influenza virus NS1 is a relatively potent inhibitor of polyadenylation that shuts down host cell polyadenylation in order to favour influenza virus mRNA utilisation and to inhibit interferon production during the short influenza virus replication cycle. Therefore, to activate HPV-16 late gene expression by NS1 plasmid transfection, 10- to 20-fold lower concentration of NS1 plasmid than E2 plasmid was used in transfections. We speculate that E2 is a relatively weak inhibitor of polyadenylation that allows production of

early mRNAs, but at a reduced level, while at the same time causing a significant increase in late gene expression. Alternatively, E2 targets another cellular factor that specifically enhances polyadenylation at pAE, such as hnRNP H (see Discussion). Further analysis of the PA complex in the gel shift experiments with the help of RNA affinity chromatography may identify the factor targeted by E2.

The C-terminal DBD of E2 was not necessary for the induction of late gene expression, separating the effect of E2 on late gene expression from those on transcription and replication, which require the combined activity of TAD and DBD. It has been shown previously that E2 can bind to RNA (Bodaghi *et al*, 2009), and it is conceivable that E2 might interfere with polyadenylation by binding to RNA. However, this can be excluded as the C-terminal DBD that binds RNA is not required for polyadenylation inhibition. Experiments shown here suggest that E2 binds polyadenylation factor CPSF30, or a binding partner of CPSF30. Interactions with CPSF are strategically important since CPSF binds to and recognises the polyA signal. A pool of CPSF is associated with RNA polymerase and causes pausing of the polymerase if a functional polyA signal is encountered, and this interaction is mediated by CPSF30. Influenza virus NS1 inhibits pausing through interactions with CPSF30. It is reasonable to speculate that E2 would inhibit pausing at a polyA signal too, as that would allow the polymerase to continue to the downstream late polyA signal pAL. On the other hand, another pool of CPSF associates with polyadenylation factor CstF. As both CPSF and CstF are present in the PA complexes in the gel shift assays in Figure 8, one may speculate that HPV E2 interferes with the conformation of the CPSF-CstF complex that forms on the polyA signal, thereby reducing the efficiency of polyadenylation. However, other scenarios are possible. Splice sites located immediately upstream of a polyA signal can stimulate polyadenylation, sometimes through SR proteins (Lou *et al*, 1998; Valente *et al*, 2009). An upstream splice site stimulates polyadenylation at HPV-16 pAE (Rush *et al*, 2005; Somberg and Schwartz, 2010). HPV-5 and HPV-16 E2 proteins have been shown to interact with SR proteins *in vitro* (Lai *et al*, 1999; Bodaghi *et al*, 2009), which could potentially interfere with polyadenylation. In addition, previous experiments have shown that polyadenylation at HPV-16 or HPV-31 pAE is stimulated by a downstream polyadenylation enhancer element (Terhune *et al*, 1999, 2001; Oberg *et al*, 2003, 2005). In HPV-31, CstF64 contributed to the activity of the element (Terhune *et al*, 1999, 2001), whereas in HPV-16 the polyadenylation element was shown to interact with hnRNP H (Oberg *et al*, 2003, 2005). E2 could potentially interact with hnRNP H, thereby preventing the interaction of hnRNP H with the CstF-polyadenylation complex, or by binding to CstF directly.

Experiments with full-length genomes of HPV-31 carrying either a replication defective or a transactivation defective E2 mutant revealed that the transactivation mutant replicated DNA well and expressed similar levels of E4 protein as the wild-type HPV-31 genome (Stubenrauch *et al*, 1998). However, HPV-31 late gene expression was down by 80% and the E2 transactivation mutant expressed only a fifth of the quantities of L1 and L2 mRNA expressed by the wild-type HPV-31 genome (Stubenrauch *et al*, 1998). In our HPV-16-based experimental system, E2 induced late gene expression three- to five-fold, which fits well with the five-fold reduction

in late gene expression reported for HPV-31 (Stubenrauch *et al*, 1998).

Premature activation of HPV late gene expression could be detrimental as it would uncover the virus-infected cells for the immune system of the host that would eliminate the infection prior to efficient virus production. To avoid premature induction of late gene expression by the E2 protein, which at earlier stages are involved in replication of the viral DNA genome, it is probably necessary to regulate this function of E2. We show that the ability of E2 to induce late gene expression is inhibited by E1, its partner in DNA replication. This would prevent E2 proteins involved in replication of the viral DNA genome, to prematurely induce late gene expression. It has previously been shown that the relative levels of E1 versus E2 affect the efficiency of viral DNA replication (Hubert and Laimins, 2002; Deng *et al*, 2003). Alternatively, E2 levels may be regulated post-translationally by changes in the stability of the E2 protein (Davy *et al*, 2009; Johansson *et al*, 2009; Zheng *et al*, 2009). An effect of the HPV E4 protein on HPV late gene expression has been suggested (Nakahara *et al*, 2005; Wilson *et al*, 2005, 2007). Formation of heterodimers between E2 and E8 may also affect the ability of E2 to induce late gene expression (Lace *et al*, 2008; Kurg *et al*, 2010). Finally, one may speculate that alterations in subcellular distribution of the shuttling E2 protein may affect the E2 concentration in the cell nucleus (Blachon *et al*, 2005).

We propose that HPV-16 E2 induces HPV-16 late gene expression by inhibiting polyadenylation at pAE, and that this occurs through a mechanism whereby E2 perturbs the polyadenylation complex. We speculate that polyadenylation fails to occur because E2 blocks cleavage and that this effect is mediated through interactions with CPSF30, and/or factors interacting with CPSF30.

Materials and methods

A more detailed Materials and methods is provided in Supplementary Data.

Transfection and cell culture

HeLa, C33A and A549 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) containing 10% fetal bovine calf serum and Penicillin-Streptomycin. Neonatal human epidermal keratinocytes (HEKn) were purchased from Gibco Invitrogen Cell Culture, and were propagated in EpiLife Medium supplemented with Human keratinocyte Growth Supplement (Gibco Invitrogen Cell Culture) as recommended by the manufacturer. Transfections were carried out by using Eugene 6 (Roche) or Turbofect (Fermentas) according to manufacturers' instructions. Each plasmid was analysed in a minimum of three independent transfection experiments.

RNA extraction, northern blotting and radiolabelled DNA probe synthesis

Cytoplasmic RNA was extracted with NP-40 buffer using standard methods. Total RNA was extracted using TRI reagent (Life Technologies) according to manufacturer's instructions. The RNA was treated with DNaseI prior to northern blotting. Northern blot analysis was carried out by size separation of 10 µg cytoplasmic RNA on 1.2% agarose gels containing 2.2 M formaldehyde. The RNA was transferred overnight onto a nitrocellulose filter followed by hybridisation with L1 (Zhao *et al*, 2004), E4 (Zhao *et al*, 2004), CMV (Collier *et al*, 2002), CAT (Zhao *et al*, 2007) or 28S rRNA (Collier *et al*, 2002) probe. GAPDH probe was synthesised from a PCR fragment generated by primers GAPDHs (5'-AGGTCCGAGTCAACGGATTTGG-3') and GAPDHs (5'-ACAGTCTTCTGGGTGGCA GTGATG-3'). DNA probes were radiolabelled with [α -³²P]dCTP using the Decaprime kit (Ambion).

RT-PCR and 3'-RACE

In all, 200 ng of cytoplasmic RNA, or 2 µg of total RNA, was reverse transcribed at 42°C using Superscript II (Invitrogen) and random hexamers according to manufacturer's protocol. cDNA was PCR amplified using the oligonucleotides indicated in the figures. GAPDH cDNA was amplified as control using previously described primers (Johansson *et al*, 2005). Real-time PCR was performed in a MiniOpticon (Bio-Rad) using the Sso Advanced SYBR Green Supermix (Bio-Rad) according to manufacturer's instructions. The results were normalised to GAPDH mRNA levels.

For 3'-RACE, total RNA was first reverse transcribed with oligodTGC (5'-GCGAGCTCCGCGCCGCGTTTCTTTTCTTTT-3'), as previously described (Oberg *et al*, 2003). An aliquot of the cDNA product was PCR amplified with oligonucleotides oligodTGC and 827S to identify the HPV-1 early polyadenylation site in pBEL-1, with primer E5S (Backström Winquist *et al*, 2012) to detect polyadenylation at HPV-16 pAE in pE5-Luc and with primer CATS for pM2A-CAT and pM2ADE-CAT. The amplified products were cloned into a pCRII-topo vector (Invitrogen) and individual clones were subjected to sequencing.

CAT ELISA and luciferase assay

CAT levels were determined using CAT-ELISA (Roche) according to manufacturer's instructions. For Luciferase assay, cells were lysed in 25 mM Tris, pH 7.8, 2 mM DTT, 1% Triton X-100 and 10% glycerol. Luciferase activity was measured in a TriStar luminometer using Beetle luciferin (Promega) as substrate.

GST pull-down assay

Polyadenylation factors, CtBP and HPV-16 E1 were *in-vitro* translated and ³⁵S-Met/Cys labelled using TNT T7 Coupled Reticulocyte Lysate System (Promega) according to manufacturer's instructions. Plasmids used were pcDNA3CPSF160(bovine)-myc, pcDNA3CPSF100(bovine)-myc, pcDNA3CPSF73(bovine)-myc, pcDNA3CPSF30(bovine)-myc, pcDNA3Fip1(mouse)-myc, pcDNA3CtBP and pGADT7-16E1. For GST pull-down assays, GST proteins on sepharose beads were incubated with *in-vitro* translated proteins in binding buffer (25 mM Hepes, pH 7.9, 50 mM NaCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 0.025% Triton X-100) for 90 min at 4°C. After washing five times in washing buffer (20 mM Hepes, pH 7.9, 50 mM NaCl, 2.5 mM MgCl₂, 0.1 mM EDTA, 0.05% Triton X-100), proteins were separated on SDS-PAGE and visualised by autoradiography. When indicated, lysates were pretreated with 2500 U RNaseT1 and 25 µg RNaseA for 3 h at 30°C.

In-vitro transcription

Plasmids pT3L3 and pT3L3G were linearised with *DraI* to synthesise the L3 and L3G RNAs, plasmid pT7L3short with *Bam*HI to make the L3short RNA using T3 or T7 RNA Polymerase (Fermentas). To produce radiolabelled RNAs, [α -³²P]UTP was included in the synthesis reaction. The RNA was purified by Proteinase K treatment followed by ethanol precipitation.

Preparation of nuclear extract, RNA gel shifts and supershift experiments

Nuclear extracts was prepared from HeLa spinner cells (Muhlemann and Akusjarvi, 2007). Gel shift reactions contained 50–200 fmol ³²P-labelled transcript, 3–10 µg HeLa NE, 1–2 µg GST proteins, 1% polyvinylalcohol and 4 U Ribolock RNase inhibitor (Fermentas) in a total volume of 15 µl of Buffer D (20 mM Hepes, pH 7.9, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT). Supershift reactions also contained 1–2 µg of specific antibody and competition reactions contained cold RNA up to 100-fold molar excess. All components were mixed on ice and then incubated for 20 min at room temperature. Heparin was added to a final concentration of 5 mg/ml and reactions incubated for an additional 10 min. Reactions containing GST proteins and/or antibodies were preincubated for 10 min at room temperature after which the radioactive RNA was added. Complexes were separated on non-denaturing 4% polyacrylamide/bisacrylamide gels. Gels were dried and exposed to phosphorimager. For supershift experiments, antibodies against CPSF160 (H-300) and CPSF30 (E-25) were purchased from Santa Cruz Biotechnology and antibody against CPSF73 (D01P) was purchased from Abnova. Antibody 3A7 against CstF64 (Takagaki *et al*, 1990) was generously provided by Clinton MacDonald.

In-vitro polyadenylation assays

Polyadenylation assays (25 µl) contained 50 fmol ³²P-labelled L3short transcript, 30 µg HeLa NE, 15 pmoles of GST protein, 8.8 mM Hepes, pH 7.9, 8.8% glycerol, 0.1 mM EDTA, 0.2 mM DTT, 100 mM KCl, 0.44 mM MgCl₂, 20 mM creatine phosphate, 4% polyvinylalcohol, 2 U Ribolock RNase inhibitor (Fermentas), 250 ng tRNA and 0.1 mM ATP. Reactions were incubated for 0.5–2 h at 30°C. Reaction products were purified using by Proteinase K treatment followed by ethanol precipitation and resolved on 8% denaturing urea-polyacrylamide gels.

Statistical analyses

All experiments have been performed a minimum of three times. Representative experiments are shown in Figures 1, 2, 3B, 4B–D, F, 5B, C, E, 6B, 7B, 8A–D, 9A–G. Values are mean ± s.d. (*n* ≥ 3) in Figures 3C, D, 4G, H, 5F, 6C, 7C and D.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

References

- Androphy EJ, Lowy DR, Schiller JT (1987) Bovine papillomavirus E2 trans-activating gene product binds to specific sites in papillomavirus DNA. *Nature* **325**: 70–73
- Backström Winquist E, Abdurahman S, Tranell A, Lindström S, Tingsborg S, Schwartz S (2012) Inefficient splicing of segment 7 and 8 mRNAs is an inherent property of influenza virus A/Brevig Mission/1918/1 (H1N1) that causes elevated expression of NS1 protein. *Virology* **422**: 46–58
- Baker C, Calef C (1997) Maps of papillomavirus mRNA transcripts. In *Human Papillomaviruses: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*, Billakanti SR, Calef CE, Farmer AD, Halpern AL, Myers GL (eds). Los Alamos: Los Alamos National Library
- Baxter MK, McPhillips MG, Ozato K, McBride AA (2005) The mitotic chromosome binding activity of the papillomavirus E2 protein correlates with interaction with the cellular chromosomal protein, Brd4. *J Virol* **79**: 4806–4818
- Bedrosian CL, Bastia D (1990) The DNA-binding domain of HPV-16 E2 protein interaction with the viral enhancer: protein-induced DNA bending and role of the nonconserved core sequence in binding site affinity. *Virology* **174**: 557–575
- Blachon S, Bellanger S, Demeret C, Thierry F (2005) Nucleocytoplasmic shuttling of high risk human Papillomavirus E2 proteins induces apoptosis. *J Biol Chem* **280**: 36088–36098
- Bodaghi S, Jia R, Zheng ZM (2009) Human papillomavirus type 16 E2 and E6 are RNA-binding proteins and inhibit *in vitro* splicing of pre-mRNAs with suboptimal splice sites. *Virology* **386**: 32–43
- Bodily J, Laimins LA (2011) Persistence of human papillomavirus infection: keys to malignant progression. *Trends Microbiol* **19**: 33–39
- Chow LT, Broker TR, Steinberg BM (2010) The natural history of human papillomavirus infections of the mucosal epithelia. *APMIS* **118**: 422–449
- Collier B, Oberg D, Zhao X, Schwartz S (2002) Specific inactivation of inhibitory sequences in the 5' end of the human papillomavirus type 16 L1 open reading frame results in production of high levels of L1 protein in human epithelial cells. *J Virol* **76**: 2739–2752
- Dai W, Zhang G, Makeyev EV (2012) RNA-binding protein HuR autoregulates its expression by promoting alternative polyadenylation site usage. *Nucleic Acids Res* **40**: 787–800
- Davy C, McIntosh P, Jackson DJ, Sorathia R, Miell M, Wang Q, Khan J, Soneji Y, Doorbar J (2009) A novel interaction between the human papillomavirus type 16 E2 and E1-E4 proteins leads to stabilization of E2. *Virology* **394**: 266–275
- Day PM, Roden RB, Lowy DR, Schiller JT (1998) The papillomavirus minor capsid protein, L2, induces localization of the major capsid protein, L1, and the viral transcription/replication protein, E2, to PML oncogenic domains. *J Virol* **72**: 142–150

Acknowledgements

We wish to thank Anki Mossberg for expert technical assistance and other members of Stefan Schwartz's group for reagents and discussion. We also wish to thank the members of Göran Magnusson's, Göran Akusjärvi's and Catharina Svensson's research groups at Uppsala University for providing reagents and for discussion. We are grateful to Francoise Thierry, Mart Ustav, Ethel-Michele de Villiers, Sally Roberts, Clare Gooding, Clinton MacDonald, Iain Morgan, Akihiro Mizutani, Zbigniew Dominski, Anders Virtanen and Andras Nagy for generously providing us with various reagents. Research sponsored by Science Foundation Ireland, by the Swedish Cancer Society and by the Swedish Research Council/Medicine.

Author contributions: SS designed and supervised experiments together with CJ and MS; CJ, MS, DP, XL, EBW and JF performed experiments and analysed data; SS, LB, DP and FR analysed data; CJ and SS wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

- Deng W, Jin G, Lin BY, Van Tine BA, Broker TR, Chow LT (2003) mRNA splicing regulates human papillomavirus type 11 E1 protein production and DNA replication. *J Virol* **77**: 10213–10226
- Dong G, Broker TR, Chow LT (1994) Human papillomavirus type 11 E2 proteins repress the homologous E6 promoter by interfering with the binding of host transcription factors to adjacent elements. *J Virol* **68**: 1115–1127
- Doorbar J (2005) The papillomavirus life cycle. *J Clin Virol* **32**(Suppl 1): S7–S15
- Gammoh N, Grm HS, Massimi P, Banks L (2006) Regulation of human papillomavirus type 16 E7 activity through direct protein interaction with the E2 transcriptional activator. *J Virol* **80**: 1787–1797
- Giri I, Yaniv M (1988) Structural and mutational analysis of E2 trans-activating proteins of papillomaviruses reveals three distinct functional domains. *EMBO J* **7**: 2823–2829
- Graham SV (2010) Human papillomavirus: gene expression, regulation and prospects for novel diagnostic methods and antiviral therapies. *Future Microbiol* **5**: 1493–1506
- Grm HS, Massimi P, Gammoh N, Banks L (2005) Crosstalk between the human papillomavirus E2 transcriptional activator and the E6 oncoprotein. *Oncogene* **24**: 5149–5164
- Heino P, Zhou J, Lambert PF (2000) Interaction of the papillomavirus transcription/replication factor, E2, and the viral capsid protein, L2. *Virology* **276**: 304–314
- Hibma MH, Raj K, Ely SJ, Stanley M, Crawford L (1995) The interaction between human papillomavirus type 16 E1 and E2 proteins is blocked by an antibody to the N-terminal region of E2. *Eur J Biochem* **229**: 517–525
- Howley PM, Lowy DR (2006) Papillomaviridae. In *Virology*, Knipe DM, Howley PM (eds) Vol. 25th edn 62, pp 2299–2354. Philadelphia, PA: Lippincott/The Williams & Wilkins Co
- Hubert WG, Laimins LA (2002) Human papillomavirus type 31 replication modes during the early phases of the viral life cycle depend on transcriptional and posttranscriptional regulation of E1 and E2 expression. *J Virol* **76**: 2263–2273
- Ilves I, Maemets K, Silla T, Janikson K, Ustav M (2006) Brd4 is involved in multiple processes of the bovine papillomavirus type 1 life cycle. *J Virol* **80**: 3660–3665
- Johansson C, Graham SV, Dornan ES, Morgan IM (2009) The human papillomavirus 16 E2 protein is stabilised in S phase. *Virology* **394**: 194–199
- Johansson C, Somberg M, Schwartz S (2010) Proteins involved in HPV-16 pre-mRNA processing. *Curr Top Virol* **8**: 17–27
- Johansson C, Zhao H, Bajak E, Granberg F, Pettersson U, Svensson C (2005) Impact of the interaction between adenovirus E1A and CtBP on host cell gene expression. *Virus Res* **113**: 51–63

- Kadaja M, Silla T, Ustav E, Ustav M (2009) Papillomavirus DNA replication - from initiation to genomic instability. *Virology* **384**: 360–368
- Kadaja M, Sumerina A, Verst T, Ojarand M, Ustav E, Ustav M (2007) Genomic instability of the host cell induced by the human papillomavirus replication machinery. *EMBO J* **26**: 2180–2191
- Kanopka A, Muhlemann O, Akusjarvi G (1996) Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* **381**: 535–538
- Kanopka A, Muhlemann O, Petersen-Mahrt S, Estmer C, Ohrmalm C, Akusjarvi G (1998) Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. *Nature* **393**: 185–187
- Kurg R, Uusen P, Vosa L, Ustav M (2010) Human papillomavirus E2 protein with single activation domain initiates HPV18 genome replication, but is not sufficient for long-term maintenance of virus genome. *Virology* **408**: 159–166
- Lace MJ, Anson JR, Thomas GS, Turek LP, Haugen TH (2008) The E8–E2 gene product of human papillomavirus type 16 represses early transcription and replication but is dispensable for viral plasmid persistence in keratinocytes. *J Virol* **82**: 10841–10853
- Lai MC, Teh BH, Tarn WY (1999) A human papillomavirus E2 transcriptional activator. The interactions with cellular splicing factors and potential function in pre-mRNA processing. *J Biol Chem* **274**: 11832–11841
- Li R, Knight J, Bream G, Stenlund A, Botchan M (1989) Specific recognition nucleotides and their DNA context determine the affinity of E2 protein for 17 binding sites in the BPV-1 genome. *Genes Dev* **3**: 510–526
- Lou H, Neugebauer KM, Gagel RF, Berget SM (1998) Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Mol Cell Biol* **18**: 4977–4985
- McBride AA (2008) Replication and partitioning of papillomavirus genomes. *Adv Virus Res* **72**: 155–205
- McBride AA, Oliveira JG, McPhillips MG (2006) Partitioning viral genomes in mitosis: same idea, different targets. *Cell Cycle* **5**: 1499–1502
- McPhillips MG, Oliveira JG, Spindler JE, Mitra R, McBride AA (2006) Brd4 is required for e2-mediated transcriptional activation but not genome partitioning of all papillomaviruses. *J Virol* **80**: 9530–9543
- Mole S, Milligan SG, Graham SV (2009) Human papillomavirus type 16 E2 protein transcriptionally activates the promoter of a key cellular splicing factor, SF2/ASF. *J Virol* **83**: 357–367
- Muhlemann O, Akusjarvi G (2007) Preparation of soluble extracts from adenovirus-infected cells for studies of RNA splicing. *Methods Mol Med* **131**: 33–46
- Nakahara T, Peh WL, Doorbar J, Lee D, Lambert PF (2005) Human papillomavirus type 16 E1circumflexE4 contributes to multiple facets of the papillomavirus life cycle. *J Virol* **79**: 13150–13165
- Nemeroff ME, Barabino SM, Li Y, Keller W, Krug RM (1998) Influenza virus NS1 protein interacts with the cellular 30kDa subunit of CPSF and inhibits 3' end formation of cellular pre-mRNAs. *Mol Cell* **1**: 991–1000
- Noah DL, Twu KY, Krug RM (2003) Cellular antiviral responses against influenza A virus are countered at the posttranscriptional level by the viral NS1A protein via its binding to a cellular protein required for the 3' end processing of cellular pre-mRNAs. *Virology* **307**: 386–395
- Oberg D, Collier B, Zhao X, Schwartz S (2003) Mutational inactivation of two distinct negative RNA elements in the human papillomavirus type 16 L2 coding region induces production of high levels of L2 in human cells. *J Virol* **77**: 11674–11684
- Oberg D, Fay J, Lambkin H, Schwartz S (2005) A downstream polyadenylation element in human papillomavirus type 16 L2 encodes multiple GGG motifs and interacts with hnRNP H. *J Virol* **79**: 9254–9269
- Okoye A, Cordano P, Taylor ER, Morgan IM, Everett R, Campo MS (2005) Human papillomavirus 16 L2 inhibits the transcriptional activation function, but not the DNA replication function, of HPV-16 E2. *Virus Res* **108**: 1–14
- Parish JL, Bean AM, Park RB, Androphy EJ (2006) ChlR1 is required for loading papillomavirus E2 onto mitotic chromosomes and viral genome maintenance. *Mol Cell* **24**: 867–876
- Piccini A, Storey A, Romanos M, Banks L (1997) Regulation of human papillomavirus type 16 DNA replication by E2, glucocorticoid hormone and epidermal growth factor. *J Gen Virol* **78**(Part 8): 1963–1970
- Pim D, Massimi P, Dilworth SM, Banks L (2005) Activation of the protein kinase B pathway by the HPV-16 E7 oncoprotein occurs through a mechanism involving interaction with PP2A. *Oncogene* **24**: 7830–7838
- Rush M, Zhao X, Schwartz S (2005) A splicing enhancer in the E4 coding region of human papillomavirus type 16 is required for early mRNA splicing and polyadenylation as well as inhibition of premature late gene expression. *J Virol* **79**: 12002–12015
- Schwartz S (2008) HPV-16 RNA processing. *Front Biosci* **13**: 5880–5891
- Schweiger MR, Ottinger M, You J, Howley PM (2007) Brd4-independent transcriptional repression function of the papillomavirus e2 proteins. *J Virol* **81**: 9612–9622
- Shtrichman R, Sharf R, Barr H, Dobner T, Kleinberger T (1999) Induction of apoptosis by adenovirus E4orf4 protein is specific to transformed cells and requires an interaction with protein phosphatase 2A. *Proc Natl Acad Sci USA* **96**: 10080–10085
- Skolnik-David H, Moore CL, Sharp PA (1987) Electrophoretic separation of polyadenylation-specific complexes. *Genes Dev* **1**: 672–682
- Smith JA, White EA, Sowa ME, Powell ML, Ottinger M, Harper JW, Howley PM (2010) Genome-wide siRNA screen identifies SMCX, EP400, and Brd4 as E2-dependent regulators of human papillomavirus oncogene expression. *Proc Natl Acad Sci USA* **107**: 3752–3757
- Somberg M, Rush M, Fay J, Ryan F, Lambkin H, Akusjarvi G, Schwartz S (2009) Adenovirus E4orf4 induces HPV-16 late L1 mRNA production. *Virology* **383**: 279–290
- Somberg M, Schwartz S (2010) Multiple ASF/SF2 sites in the human papillomavirus type 16 (HPV-16) E4-coding region promote splicing to the most commonly used 3'-splice site on the HPV-16 genome. *J Virol* **84**: 8219–8230
- Somberg M, Zhao X, Frohlich M, Evander M, Schwartz S (2008) Polypyrimidine tract binding protein induces human papillomavirus type 16 late gene expression by interfering with splicing inhibitory elements at the major late 5' splice site, SD3632. *J Virol* **82**: 3665–3678
- Stenlund A (2003) Initiation of DNA replication: lessons from viral initiator proteins. *Nat Rev Mol Cell Biol* **4**: 777–785
- Stubenrauch F, Lim HB, Laimins LA (1998) Differential requirements for conserved E2 binding sites in the life cycle of oncogenic human papillomavirus type 31. *J Virol* **72**: 1071–1077
- Takagaki Y, Manley JL, MacDonald CC, Wilusz J, Shenk T (1990) A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. *Genes Dev* **4**: 2112–2120
- Terhune SS, Hubert WG, Thomas JT, Laimins LA (2001) Early polyadenylation signals of human papillomavirus type 31 negatively regulate capsid gene expression. *J Virol* **75**: 8147–8157
- Terhune SS, Milcarek C, Laimins LA (1999) Regulation of human papillomavirus type 31 polyadenylation during the differentiation-dependent life cycle. *J Virol* **73**: 7185–7192
- Thierry F (2009) Transcriptional regulation of the papillomavirus oncogenes by cellular and viral transcription factors in cervical carcinoma. *Virology* **384**: 375–379
- Valente ST, Gilmartin GM, Venkatarama K, Arriagada G, Goff SP (2009) HIV-1 mRNA 3' end processing is distinctively regulated by eIF3f, CDK11, and splice factor 9G8. *Mol Cell* **36**: 279–289
- Wang HK, Duffy AA, Broker TR, Chow LT (2009) Robust production and passaging of infectious HPV in squamous epithelium of primary human keratinocytes. *Genes Dev* **23**: 181–194
- Wilson R, Fehrmann F, Laimins LA (2005) Role of the E1–E4 protein in the differentiation-dependent life cycle of human papillomavirus type 31. *J Virol* **79**: 6732–6740
- Wilson R, Ryan GB, Knight GL, Laimins LA, Roberts S (2007) The full-length E1E4 protein of human papillomavirus type 18 modulates differentiation-dependent viral DNA amplification and late gene expression. *Virology* **362**: 453–460
- Wu SY, Lee AY, Hou SY, Kemper JK, Erdjument-Bromage H, Tempst P, Chiang CM (2006) Brd4 links chromatin targeting to HPV transcriptional silencing. *Genes Dev* **20**: 2383–2396
- Xue Y, Bellanger S, Zhang W, Lim D, Low J, Lunny D, Thierry F (2010) HPV16 E2 is an immediate early marker of viral infection, preceding E7 expression in precursor structures of cervical carcinoma. *Cancer Res* **70**: 5316–5325
- Yan J, Li Q, Lievens S, Tavernier J, You J (2010) Abrogation of the Brd4-positive transcription elongation factor B complex by

- papillomavirus E2 protein contributes to viral oncogene repression. *J Virol* **84**: 76–87
- You J, Croyle JL, Nishimura A, Ozato K, Howley PM (2004) Interaction of the bovine papillomavirus E2 protein with Brd4 tethers the viral DNA to host mitotic chromosomes. *Cell* **117**: 349–360
- Zhao X, Fay J, Lambkin H, Schwartz S (2007) Identification of a 17-nucleotide splicing enhancer in HPV-16 L1 that counteracts the effect of multiple hnRNP A1-binding splicing silencers. *Virology* **369**: 351–363
- Zhao X, Oberg D, Rush M, Fay J, Lambkin H, Schwartz S (2005) A 57-nucleotide upstream early polyadenylation element in human papillomavirus type 16 interacts with hFip1, CstF-64, hnRNP C1/C2, and polypyrimidine tract binding protein. *J Virol* **79**: 4270–4288
- Zhao X, Rush M, Schwartz S (2004) Identification of an hnRNP A1-dependent splicing silencer in the human papillomavirus type 16 L1 coding region that prevents premature expression of the late L1 gene. *J Virol* **78**: 10888–10905
- Zheng G, Schweiger MR, Martinez-Noel G, Zheng L, Smith JA, Harper JW, Howley PM (2009) Brd4 regulation of papillomavirus protein E2 stability. *J Virol* **83**: 8683–8692
- Zheng ZM, Baker CC (2006) Papillomavirus genome structure, expression, and post-transcriptional regulation. *Front Biosci* **11**: 2286–2302
- Zhu H, Zhou HL, Hasman RA, Lou H (2007) Hu proteins regulate polyadenylation by blocking sites containing U-rich sequences. *J Biol Chem* **282**: 2203–2210
- zur Hausen H (2002) Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer* **2**: 342–350