

Construction of a Full Transcription Map of Human Papillomavirus Type 18 during Productive Viral Infection[∇]§

Xiaohong Wang,¹ Craig Meyers,² Hsu-Kun Wang,³ Louise T. Chow,³ and Zhi-Ming Zheng^{1*}

Tumor Virus RNA Biology Section, HIV and AIDS Malignancy Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland¹; Department of Microbiology and Immunology, Penn State University College of Medicine, Hershey, Pennsylvania²; and Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, Alabama³

Received 4 April 2011/Accepted 27 May 2011

Human papillomavirus type 18 (HPV18) is the second most common oncogenic HPV genotype, responsible for ~15% of cervical cancers worldwide. In this study, we constructed a full HPV18 transcription map using HPV18-infected raft tissues derived from primary human vaginal or foreskin keratinocytes. By using 5' rapid amplification of cDNA ends (RACE), we mapped two HPV18 transcription start sites (TSS) for early transcripts at nucleotide (nt) 55 and nt 102 and the HPV18 late TSS frequently at nt 811, 765, or 829 within the E7 open reading frame (ORF) of the virus genome. HPV18 polyadenylation cleavage sites for early and late transcripts were mapped to nt 4270 and mainly to nt 7299 or 7307, respectively, by using 3' RACE. Although all early transcripts were cleaved exclusively at a single cleavage site, HPV18 late transcripts displayed the heterogeneity of 3' ends, with multiple minor cleavage sites for late RNA polyadenylation. HPV18 splice sites/splice junctions for both early and late transcripts were identified by 5' RACE and primer walking techniques. Five 5' splice sites (donor sites) and six 3' splice sites (acceptor sites) that are highly conserved in other papillomaviruses were identified in the HPV18 genome. HPV18 L1 mRNA translates a L1 protein of 507 amino acids (aa), smaller than the 568 aa residues previously predicted. Collectively, a full HPV18 transcription map constructed from this report will lead us to further understand HPV18 gene expression and virus oncogenesis.

Cervical cancer is a leading cause of death for women in the developing world, with about 493,000 new cases and nearly 273,000 deaths each year (www.who.int/hpvcancer/en/). Oncogenic human papillomavirus (HPV) infection is widely recognized as a principal cause of cervical, penile, and anal cancers (62). Among over 120 genotypes with human origin (3), infection with HPV16 and HPV18, the two most common oncogenic HPV types, leads to the development of ~70% of all cervical and other anogenital cancers (55, 62). HPV18 alone accounts for more than 15% of all cervical cancer cases worldwide (14). Persistent HPV16 infection is responsible for development of both cervical squamous cell carcinoma and adenocarcinoma. In contrast, persistent HPV18 infection is a preferential risk factor for the development of cervical adenocarcinoma (7, 10).

HPVs are a group of small DNA viruses which have a high degree of conservation with respect to genome structure and organization as well as gene expression (44, 91). The circular, double-stranded viral genome is approximately 8 kb in size and contains eight open reading frames (ORFs) which are all transcribed from the same strand. In general, the viral genome can be divided into three major regions: an early region, a late region, and a long control region (LCR) or a noncoding region (NCR). The early region is positioned in the 5' half of the virus

genome and encodes six common ORFs (E1, E2, E4, E5, E6, and E7) with regulatory functions in viral replication, gene expression, and pathogenesis. The late region lies downstream of the early region and encodes the viral major and minor capsid proteins L1 and L2. The ~850-bp LCR spanning the segment between the end of the late region and the beginning of the early region has no coding function but contains the origin of viral DNA replication and transcription factor binding sites (2).

HPVs infect the basal layer of a stratified squamous epithelium, and their life cycle is closely linked to keratinocyte differentiation. Thus, an organotypic raft culture system has been employed to recapitulate and investigate the productive HPV life cycle (22, 46, 50, 51, 83). Expression of five (E1, E2, E5, E6, and E7) of the six viral regulatory proteins from the early region of the virus genome takes place in undifferentiated or intermediately differentiated keratinocytes in the basal/parabasal and spinous layers of the epithelium. In contrast, viral DNA replication, expression of E1, E4, L1, and L2, and assembly of virions occur exclusively in the upper spinous and more differentiated granular layers of the epithelium (23, 44). This differential expression of viral early and late genes in oncogenic HPV infections is attributable to activation of a viral early promoter in undifferentiated keratinocytes and a viral late promoter in highly differentiated keratinocytes (30, 32, 36, 52, 83).

Extensive RNA mapping and analysis of oncogenic HPV16- and HPV31-infected cells and raft tissues and nononcogenic HPV6 and HPV11 recovered from patient lesions have led to

* Corresponding author. Mailing address: HIV and AIDS Malignancy Branch, National Cancer Institute, National Institutes of Health, 10 Center Dr., Rm. 6N106, Bethesda, MD 20892-1868. Phone: (301) 594-1382. Fax: (301) 480-8250. E-mail: zhengt@exchange.nih.gov.

§ Supplemental material for this article may be found at <http://jvi.asm.org/>.

[∇] Published ahead of print on 15 June 2011.

the establishment of a transcription map from each virus genome (1, 13), which provides guidance for various HPV studies today. It is worth noting from the described transcription maps that each virus genome contains two major promoters. The early promoter P97 of HPV16 (77, 78) and P99 of HPV31 (33, 59) lie upstream of the E6 ORF and are responsible for the transcription of almost all viral early ORFs except E4. Although the viral late region is in the 3' half of the virus genome, the late promoter P670 of HPV16 (28) and P742 of HPV31 (33, 59) reside in the E7 ORF of the early region and are responsible for viral E4, L1, and L2 expression. Consequently, these transcription mappings have led to the conclusion that each transcript from the virus genome is bicistronic or polycistronic and undergoes extensive alternative RNA splicing and polyadenylation (13, 33, 58, 69, 91). In contrast, little progress has been made in HPV18 transcription and gene expression in the context of virus infection after complete sequencing of the entire HPV18 genome (15) and successful mapping of an HPV18 early promoter P105 (68, 72, 81). Although HPV18 ranks second of the most important oncogenic HPV types causing anogenital cancers, with preferential association with cervical adenocarcinoma, so far, there is no HPV18 transcription map to assist our understanding of HPV18 molecular biology, pathogenesis, diagnosis, and clinical management. In the present study, we analyzed the expression of HPV18 early and late genes by using HPV18-infected raft tissues derived from human keratinocytes. We identified a new promoter, P55, for the expression of HPV18 early ORFs and a late promoter in the E7 ORF region for the late ORF expression in addition to the mapping of viral RNA splice junctions and polyadenylation cleavage sites.

MATERIALS AND METHODS

Cells and raft tissues. Human foreskin keratinocytes (HFKs) or vaginal keratinocytes (HVKs) maintaining episomal HPV18 genomic DNA were cultured with E media in the presence of mitomycin C (4 µg/ml)-treated J2 3T3 feeder cells as previously described (51). Raft cultures derived from HPV18-immortalized HVKs or HFKs or HPV18-infected HFKs without immortalization by Cre-mediated recombination were prepared as described previously (47, 83). Differentiation of HPV18-infected HFKs in monolayer cultures reaching 90% confluence was induced in the presence of 1.5 mM calcium chloride (54). HPV18⁺ HeLa cells from the American Type Culture Collection (ATCC, Manassas, VA) were grown in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (FBS) at 37°C and 5% CO₂.

PCR and reverse transcription (RT)-PCR. DNA was extracted from HPV18-infected HFKs or HVKs in monolayer cultures and raft tissues using a QIAamp DNA blood minikit (Qiagen, Gaithersburg, MD). Total RNA was extracted from the cells described above using TRIzol reagent (Invitrogen), followed by treatment with RQ1 DNase I (Promega, Madison, WI) for 5 min at 37°C, heat inactivation, and precipitation. The gene-specific primers (Fig. 1A) were designed according to the HPV18 ORFs previously assigned (15), are listed in Table S1 in the supplemental material, and were used for detection of the following products: oZMZ252 (Pr121) and oZMZ229 (Pr967) for E1, oZMZ252 and oZMZ253 (Pr850) for E6 DNA and mRNA, oXHW45 (Pr3599) and oXHW44 (Pr5793) for L1 mRNA, and oXHW47 (Pr5487) and oXHW48 (Pr5935) for L2 mRNA. Primers oZMZ269 (5'-GTCATCAATGGAAATCC CATCACC-3') and oZMZ270 (5'-TGAGTCCTCCACGATACCAAAA-3') were used for detection of GAPDH RNA.

Northern blot analysis. Each sample containing 1 µg of the poly(A)⁺ RNA from HFK-derived, 16-day-old rafts with or without HPV18 infection was mixed with NorthernMax formaldehyde load dye (Ambion, Austin, TX) and denatured at 75°C for 15 min. The samples were then separated in 1% (wt/vol) formaldehyde-agarose gels in 1× morpholinepropanesulfonic acid running buffer and then transferred onto a GeneScreen Plus hybridization transfer membrane (Perkin Elmer, Fremont, CA) and immortalized by UV light. The membrane was

prehybridized with PerfectHyb Plus hybridization buffer (Sigma, St. Louis, MO) for 2 h, and then hybridization was carried out for 24 h at 42°C. The [α -³²P]ATP-labeled L1/L2-specific probe transcribed from a DNA template prepared by PCR using the primer pair oXHW98 and oXHW99 (see Table S1 in the supplemental material) was used to detect HPV18 L1 and L2 transcripts. The [α -³²P]dCTP-labeled, E4-specific probe prepared using a DECA Primer II random primed DNA labeling kit (Ambion) from a PCR template amplified by a primer pair oXHW91 and oXHW42 (Table S1) was used to detect HPV18 E4 transcript. Cyclophilin RNA was used as an internal loading control, and its probe was labeled as described above using a cyclophilin DNA template provided by the DECA Primer II random primed DNA labeling kit.

Western blot analysis. Protein samples were prepared by direct lysis of 10-day-old raft tissues in 2× RIPA buffer by homogenization and followed by centrifugation. The supernatant was mixed with 2× SDS protein sample buffer containing 5% 2-mercaptoethanol, denatured by heating at 90°C for 5 min, and separated in a NuPAGE 4 to 12% Bis-Tris gel (Invitrogen) in 1× NuPAGE MES SDS running buffer. After transfer, the nitrocellulose membrane was blocked with 5% nonfat milk in Tris-buffered saline (TBS) for 1 h at room temperature. After rinsing with TBS, the membrane was incubated overnight at 4°C with HPV Ab-3 (clone K1H8) mouse monoclonal antibody (Thermo Scientific, Fremont, CA) to detect HPV18 L1. The membrane was washed 3 times with TTBS (TBS with the addition of Tween 20 at a final concentration of 0.1% [vol/vol]). Horseradish peroxidase-labeled goat-anti mouse secondary antibody (Sigma) diluted in 1:5,000 in TTBS was incubated for 1 h at room temperature. After three washes with TTBS, the immunoreactive proteins were detected with enhanced chemiluminescence substrate (Pierce, Rockford, IL). The signal was captured on X-ray film. Before being reprobed with β -tubulin antibody, the membrane was stripped with Restore Western blot stripping buffer (Pierce) according to the manufacturer's instructions and blocked with 5% nonfat milk in TBS.

5' and 3' RACE. 5' and 3' rapid amplification of cDNA ends (RACE) assays were carried out to amplify the ends of HPV18 transcripts using a Smart Race cDNA amplification kit (Clontech, Mountain View, CA) according to the manufacturer's instructions. Total RNA or poly(A)-selected RNA (1 µg/reaction) from HeLa cells or HPV18-infected rafts at day 8, 10, or 16 were used as templates. The following primers (see Table S1 in the supplemental material) were used for RACE: oXHW86 (Pr233) for 5' RACE to map HPV18 start sites for early transcripts, oZMZ253 (Pr850) and oZMZ433 (Pr904) for 5' RACE to map HPV18 start sites for late transcripts, and oXHW90 (Pr3976) and oXHW97 (Pr7038) for 3' RACE to map HPV18 polyadenylation sites for respective early and late transcripts. Primer oXHW38 (Pr3517) and Primer oXHW44 (Pr5793) were also used for 5' RACE to identify HPV18 splice junctions and other potential promoters in the early and late regions of the virus genome. Each RACE product was gel purified, cloned, and sequenced.

RESULTS

Productive HPV18 infection of human keratinocyte-derived rafts. To initiate HPV18 transcription mapping, two sets of HPV18-infected HFK or HVK rafts were used. In one approach, HFKs and HVKs were transfected by electroporation of EcoRI-linearized HPV18 genome and the transfected keratinocytes were grown in monolayer and raft cultures as described previously (47). The linearized HPV18 genomes in the transfected cells self-ligated and became circular episomal forms, presumably by a cellular DNA repair mechanism. In another approach, HFKs were cotransfected with HPV18 genome containing a loxP-Neo-loxP insert between nucleotide (nt) 7473 and nt 7474 and a Cre expression vector in the presence of FuGene 6. The extrachromosomal HPV18 genomic plasmid generated by Cre-mediated recombination in the transfected HFKs has a 34-bp loxP insert between nt 7473 and nt 7474 (83). Total cell DNA and RNA were extracted from HPV18 transfected keratinocytes grown in monolayers and raft tissues to determine viral productive infection and to map HPV18 transcription. We found that the Cre-mediated recombinant HPV18 genome is identical to the wild-type (wt) HPV18 genome in viral gene expression and infectious virus production, as reported (83). Both raft systems gave similar

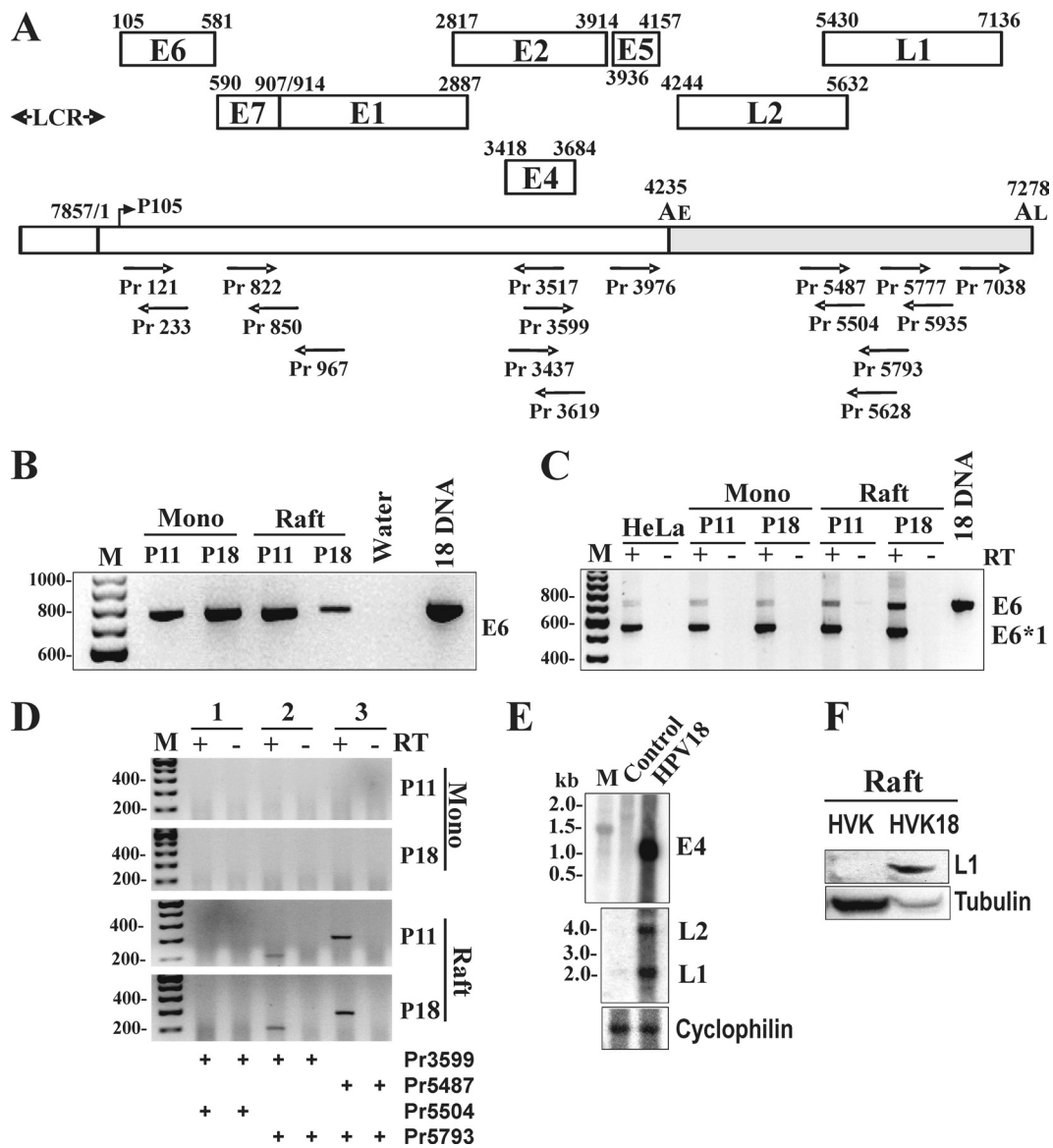


FIG. 1. Virus infection and gene expression in raft tissues. (A) Schematic diagrams of HPV18 ORFs initially deduced from the HPV18 genome (15). The numbers above each ORF are the first nt of the start codon and the last nt of the termination codon in the HPV18 genome illustrated by a bracket line below the ORFs as a linear form, with the head-to-tail junction, the promoter P105, and early (AE) and late (AL) polyadenylation sites. LCR, long control region. Primers (arrows) used for PCR, RT-PCR, 5' RACE, and 3' RACE are shown below the bracket line and are named by the positions of their 5' ends in the virus genome. (B to E) HPV18 infection and viral gene expression in keratinocytes. Total cell DNA or RNA extracted from HPV18-infected human primary vaginal keratinocytes (HVK) in monolayer cultures (mono) at passage 11 (P11) or 18 (P18) and their derived, 10-day-old raft tissues (raft) from the corresponding passages was used for HPV18 DNA (B) or RNA (C) detection at the E6E7 ORF region by PCR (B) or RT-PCR (C) with a primer pair of Pr121 and Pr850. The total cell RNA was also used for L1/L2 RNA detection by RT-PCR with indicated primer pairs below the panel (D). A plasmid containing HPV18 genomic DNA (18DNA) served as a positive control in panel B or a size marker for E6 RNA in panel C. HeLa cell total RNA served as a positive control for E6E7 mRNA expression in panel C. M, molecular weight marker. + and - RT, with (+) or without (-) reverse transcription. (E) Expression of HPV18 late mRNAs in HPV18-infected HFV raft tissues. Total RNA from HFV-derived, 16-day-old raft tissues with or without HPV18 infection was examined for E4, L2, and L1 mRNAs by Northern blotting with a ³²P-labeled E4 probe (nt 3437 to 3619) or L1/L2 probe (nt 5777 to 5939). Cyclophilin mRNA served as a loading control. (F) L1 protein expression in HPV18-infected, 10-day-old HVK (HVK18) raft tissues. Cell lysates in RIPA buffer prepared from the HVK-derived (at passage 11) raft tissues with or without HPV18 infection were analyzed by Western blotting. β -tubulin served as a loading control.

and reproducible results in HPV18-infected HFVs and HVKs. The transfected HVKs in both monolayer and raft cultures at the tested passages contained HPV18 DNA (Fig. 1B) and expressed viral early transcripts as exemplified by RT-PCR detection of both spliced and unspliced E6 RNAs (Fig. 1C). As

expected, viral late transcripts, L1 and L2, were detectable only from raft tissues derived from HPV18-infected HVKs and HFVs by RT-PCR (Fig. 1D) and by Northern blotting (Fig. 1E). We further confirmed HPV18 L1 expression in the infected HVK raft tissues by Western blotting (Fig. 1F). The raft

tissues at day 10 or older were used in this study for detection of HPV18 late gene expression because viral L1 is not detectable in day 8 or younger rafts by immunohistochemical staining (83, 84). Together, these data indicate that both HPV18-infected HFK and HVK rafts supported productive HPV infection and were equally suitable for the construction of an HPV18 transcription map.

Mapping of HPV18 transcription start sites (TSS) for early and late transcripts. Total cell RNA extracted from HPV18-infected, 8-day-old rafts containing an episomal viral genome or HPV18⁺ HeLa cells containing an integrated HPV18 genome was used for mapping of early transcript TSS by 5' RACE in the presence of a virus-specific antisense primer, Pr233 (Fig. 2A). We included HeLa cell RNA for comparison because the initial TSS mapping of HPV18 E6E7 using primer extension and S1 digestion was carried out in HeLa cells (72). As shown in Fig. 2A, two 5' RACE products were found from both RNA preparations, suggesting that viral early transcripts are initiated from two different TSS. Following gel purification, cloning, and sequencing, we found that viral early transcription was started at a purine A or G as reported in eukaryotes (8). The majority of the product 1 from HPV18-infected rafts initiated at nt 102 (13/28 clones). Two alternative TSS at nt 100 (8/28 clones) or nt 105 (6/28 clones) were also present. HeLa cell product 1 initiated at nt 105 (7 out of 12 clones) as described previously (68, 72, 81) and at nt 100 (4 out of 12 clones). A scattered initiation site from the product 1 was also detected at nt 92 in HPV18 rafts and at nt 90 in HeLa cells. Cloning and sequencing of the product 2 showed a TSS primarily at nt 55 both in HPV18-infected rafts (7/8 colonies) and in HeLa cells (7/8 colonies), with only one scattered TSS at nt 100 (raft) or at nt 70 (HeLa). Analyses of the region 5' to each TSS showed a TATA box (a eukaryotic core promoter motif) 27 bp upstream of the nt 55 TSS and 25 bp upstream of the nt 102 TSS, suggesting that the mapped TSS are authentic. In a separate 5' RACE experiment by using a primer in the E4 ORF on an RNA preparation from HPV18-infected, 10-day-old rafts, we also determined the early TSS in 10 bacterial clones, with 7 at nt 102, 2 at nt 105, and 1 at nt 100 but none at nt 55 in the virus genome.

We next wished to map the TSS for viral late transcripts of HPV18. Given that the TSS of both HPV16 and HPV31 late transcripts were mapped to nt 670 (28) and nt 742 (58), respectively, in the corresponding viral E7 ORFs, we performed a 5' RACE with an HPV18-specific primer at nt 850 or 904 using poly(A)⁺ total RNA extracted from 16-day-old, HPV18-infected HFK rafts or from HPV18-infected HFKs in monolayer cultures in the presence of calcium. The detected 5' RACE product in these reactions was a broad band between ~100 and ~150 bp in size (Fig. 2B). Following gel purification and cloning, we randomly sequenced 51 bacterial colonies, of which 22 were from the Pr850 5' RACE product and 29 from the Pr904 5' RACE product from HPV18-infected HFK rafts. We found that the late TSS of HPV18 in HFK rafts mapped to multiple positions in the E7 ORF. As shown in Fig. 2B, mapped viral TSS included 9 colonies with a start site at nt 765, 6 at nt 806, 14 at nt 811, 6 at nt 814, and 8 at nt 829; other scattered TSS not in Fig. 2B were 3 at nt 800, 1 at nt 808, 2 at nt 821, 1 at nt 826, and 1 at nt 840. Consistent with this, mapped viral TSS in HPV18-infected HFKs in monolayer cul-

tures in the presence of 1.5 mM calcium chloride after sequencing of 11 bacterial colonies were 1 at nt 674, 2 at nt 765, 1 at nt 772, 1 at nt 800, 4 at nt 806, and 2 at nt 829. All of the mapped TSS started with an A or G (Fig. 2B). Therefore, we concluded that the majority of HPV18 late transcripts have a TSS around nt 811. Analysis of the 5' region to the nt 765, 811 or 829 start site revealed no consensus TATA or TATA-like box, which perhaps accounted for the heterogeneity of the late TSS. The heterogeneity of the late TSS is common in HPV16, HPV31, HPV6, and HPV11 late transcription (13, 28, 34, 56, 60, 85).

Mapping of HPV18 polyadenylation cleavage sites for early and late transcripts. RNA polyadenylation of adding a poly(A) tail of ~150 to 200 adenosine residues at the RNA 3' end is an important posttranscriptional step in gene expression. RNA transcripts without a poly(A) tail are degraded or not efficiently exported to the cytoplasm. Efficient RNA polyadenylation requires at least two *cis* elements, a poly(A) signal (PAS) AAUAAA and a U/GU-rich motif downstream, and five polyadenylation machinery components to interact with these two *cis* elements (53, 88). Additional auxiliary elements flanking PAS and U/GU motifs are regulatory elements for efficient RNA 3' processing (27, 64, 87). As HPV18 has a putative early PAS downstream of the viral E5 ORF at nt 4235 in the virus genome (Fig. 1A), we wished to determine whether this PAS is authentic in addition to mapping the cleavage site of the viral early transcripts for poly(A) addition. Poly(A)⁺ total mRNA isolated from HPV18-infected rafts was analyzed by 3' RACE using an HPV18 E5-specific primer, Pr3976. Following gel purification, cloning, and sequencing of a 3' RACE product of ~400 bp (Fig. 3A), we found that all 7 sequenced bacterial colonies exhibited a product with a 3' end at nt 4270, 30 nt downstream of the putative nt 4235 PAS, indicating that HPV18 early transcripts were cleaved at the nt 4270 for RNA polyadenylation using the nt 4235 PAS AAUAAA. Analysis of the region 3' to this cleavage site shows a GU-rich element with four overlapping U/GU motifs (from nt 4334 to 4339), highly conserved recognition sites of CSF (cleavage stimulation factor) in RNA polyadenylation (26, 53).

From the analysis of the HPV18 genome sequence, we predicted a putative PAS AAUAAA at nt 7278 for polyadenylation of viral L2 and L1 transcripts (Fig. 1A). This putative PAS is positioned 141 nt downstream of the L1 ORF. Using 3' RACE, we analyzed poly(A)⁺ total RNA prepared from HPV18-infected rafts with L1-specific primer Pr7038 for mapping the polyadenylation cleavage site of viral late transcripts. Following gel purification and cloning of 3' RACE products of ~320 bp, we sequenced 34 bacterial colonies and found the usage of multiple cleavage sites for polyadenylation of HPV18 late transcripts. Among the mapped cleavage sites, 26 were mapped to a pyrimidine U or C, with nt 7299 and nt 7307 being the two most common cleavage sites (38% in total) followed in frequency by nt 7294, 7297, and 7306 (38% in total). The 5'-most cleavage site of the late transcripts was mapped to a U at nt 7294, 11 nt downstream of the putative late PAS. nt 7307 is the 3'-most cleavage site, 24 nt downstream of the late PAS. Thus, heterogeneity of the late transcript cleavage sites spanned a short region from nt 7294 to nt 7307. Analysis of the sequences 3' to the nt 7307 cleavage site showed multiple U/GU motifs, featuring five tandem U/GU motifs from nt 7341

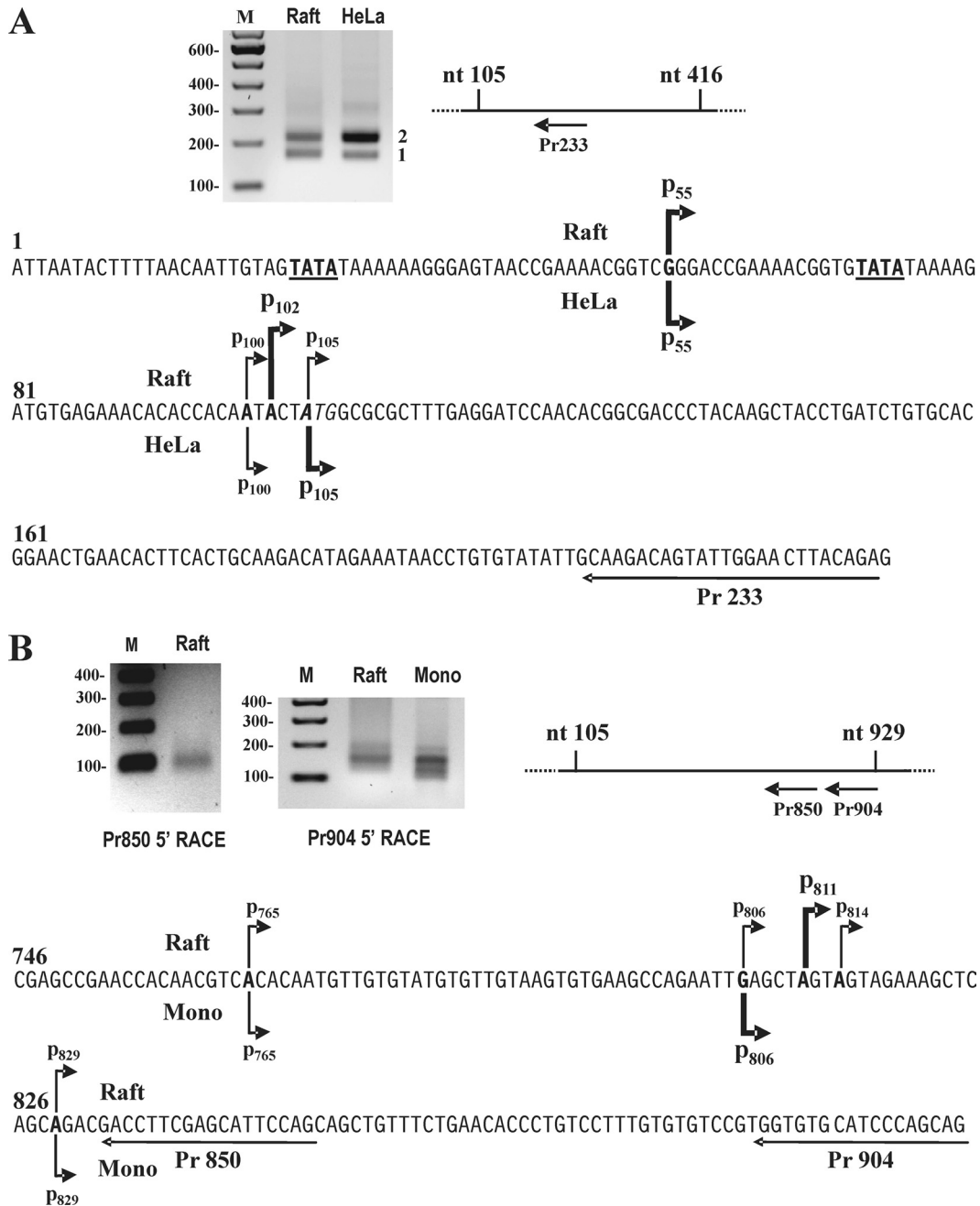


FIG. 2. Mapping of HPV18 TSS for viral early and late transcripts. (A) Mapping of HPV18 TSS for early transcripts. 5' RACE was conducted with an HPV18-specific primer, Pr233, on poly(A)⁺ total RNA isolated from 8-day-old, HPV18-infected rafts (raft). 5' RACE on total RNA from HeLa cells was used as a control. The RACE product 1 and product 2 were gel purified, cloned, and sequenced. The sequence below the gel with arrows shows identified TSS (boldface) from product 1 and product 2, with the start site named by its nucleotide position in the virus genome in HPV18 rafts (above the sequence) or HeLa cells (below the sequence). Heavier arrows indicate a TSS predominantly found in a sequenced RACE product. The E6 initiation codon AUG is italic. Boldface and underlined sequences are TATA boxes upstream of TSS. (B) Mapping of HPV18 TSS for late transcripts. 5' RACE was performed with an HPV18-specific primer, Pr850 or Pr904, on total poly(A)⁺ mRNA isolated from 16-day-old, HPV18-infected HFK rafts (raft) or from HPV18-infected HFKs in monolayer cultures (mono) in the presence of 1.5 mM calcium chloride for 48 h. Other details are as described for panel A.

to nt 7350. These data suggest that the nt 7278 PAS and its downstream U/GU repeats are responsible for polyadenylation of HPV18 L2 and L1 transcripts.

Identification of HPV18 splice junctions in viral early and late transcripts. Because HPV early and late transcripts con-

tain multiple splice sites (11, 13, 58, 69, 91), we hypothesized that HPV18 RNA transcripts would be structurally similar. Both 5' RACE and primer walking RT-PCR using poly(A)⁺ total mRNA extracted from HPV18-infected rafts were used to identify possible splice sites being used in production of

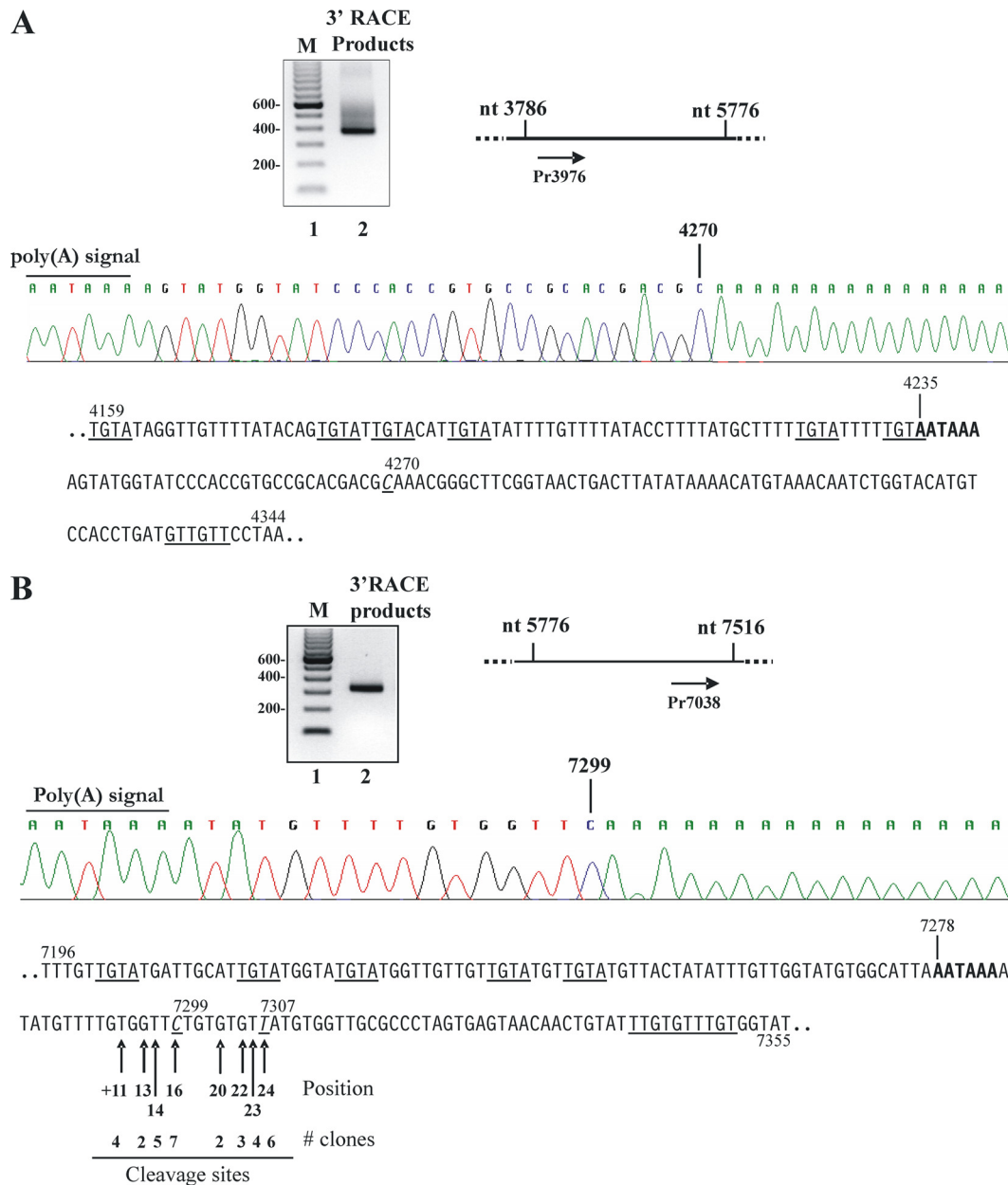


FIG. 3. Mapping of HPV18 polyadenylation cleavage sites for early and late transcripts. (A) Mapping of HPV18 early polyadenylation cleavage site. 3' RACE was conducted with an HPV18-specific primer, Pr3976, on poly(A)⁺ total mRNA isolated from 16-day-old, HPV18-infected rafts. RACE products were gel purified, cloned, and sequenced. A sequence reading below the gel shows the early PAS and a mapped cleavage site of early transcripts. Below the sequencing reading is the HPV18 sequence surrounding the 4235 PAS (bolded), with the sequences underlined for putative UGUA motifs, a mapped cleavage site (italic) and putative U/GU motifs. (B) Mapping of HPV18 late polyadenylation cleavage sites. 3' RACE was performed with an HPV18-specific primer, Pr7038, on the same poly(A)⁺ total mRNA used for mapping of the cleavage site of early transcripts in panel A. The RACE products were gel purified, cloned, and sequenced. A sequence reading below the gel shows the late PAS and a representative of the mapped late cleavage site. Below the sequencing reading is the HPV18 sequence surrounding the 7278 PAS (boldface), with the sequences underlined for putative UGUA motifs, the mapped major cleavage sites (italic), and putative U/GU motifs. Arrows show a mapped cleavage site relative to the last nucleotide of the late PAS and its frequency (number of clones) of usage in late polyadenylation. Not shown is a mapped cleavage site at nt 7295 cloned only once in this experiment.

HPV18 early and late transcripts. As shown in Fig. 4A, 5' RACE with antisense primer Pr3517 in the E4 ORF produced seven visible RACE products, dominating with product 1, product 2, and product 6. Gel purification, cloning, and sequencing of each RACE product enabled us to identify the splice junctions from viral early or late transcripts. Product 1

did not contain specific HPV18 sequence and disappeared in another RNA preparation. Product 2, of ~250 bp, comprised mixed products derived primarily from early transcripts using P102 with a splice junction of nt 233/3434 and late transcripts from P811/P829 with a splice junction of nt 929/3434. A minor transcript initiated at nt 1202 was also found to have a splice

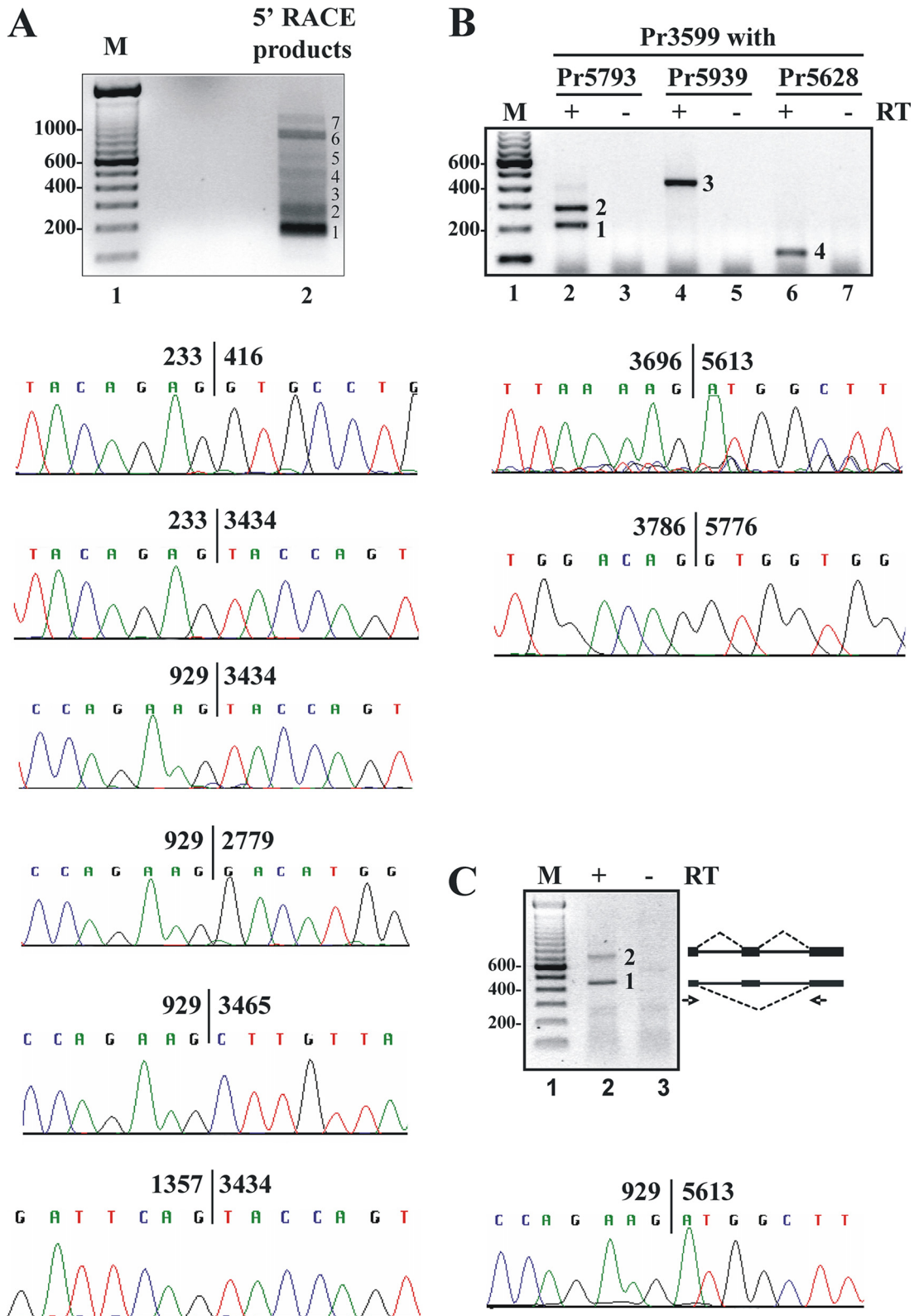


FIG. 4. Identification of HPV18 splice junctions by 5' RACE and by primer walking RT-PCR. (A) Identification of HPV18 splice junctions by 5' RACE. 5' RACE was performed on total poly(A)⁺ mRNA isolated from HPV18-infected, 10-day-old rafts using an HPV18-specific primer, Pr3517. Multiple RACE products were gel purified, cloned, and sequenced. Shown below 5' RACE products 1 to 7 are the splicing junctions identified by sequencing. (B and C) Identification of HPV18 splice junctions by primer walking RT-PCR. The primer walking RT-PCRs were performed on total RNA isolated from HPV18-infected rafts using three different pairs of HPV18-specific primers (Pr3599 plus Pr5793, Pr5939, or Pr5628) (B) and one additional pair of HPV18-specific primers Pr822 and Pr5939 (C). Products 1 to 4 (B) and product 1 (C) were gel purified, cloned, and sequenced. Product 2 in panel C is a double-spliced L1 mRNA as diagrammed on the gel's right, with heavy lines for exons, thin lines for introns, and dashed lines for splicing directions. Arrows below the line diagram are primer positions. Shown below the gels in panels B and C are the corresponding splicing junctions identified by sequencing.

junction of nt 1357/3434. Product 3, of ~300 bp, and product 4, of ~430 bp, were late transcription products with a splice junction of nt 929/3434 from either P765 or a minor TSS at nt 586. Product 5, of ~600 bp, contained mixed late transcription products initiated occasionally at nt 459 or 498 with a splice junction of nt 929/3434. Product 6, of ~760 bp, was double-spliced early transcripts from P102 with two splice junctions of nt 233/416 and nt 929/3434. Product 7, of ~930 bp, comprised mainly early transcripts from P102 with a single splice junction of nt 929/3434 or nt 929/3465 and had no E6 intron splicing. A minor transcript spliced from nt 929 to nt 2779 was also detected from product 7. In total, six splice junctions were identified by this approach. Although the majority of the 5' RACE products were spliced from nt 233 to 416 and/or nt 929 to 3434, others were spliced from nt 233 to 3434 or from nt 929 to 3465. The 5' RACE products with a splice junction of nt 929/2779 or nt 1357/3434 were detectable but rare. This conclusion was confirmed by multiple 5' RACE experiments using a different gene-specific primer or by primer walking RT-PCR in the early region of the HPV18 genome.

The splice junctions for the HPV18 late transcripts were also mapped by primer walking RT-PCR using a forward primer, Pr3599, within the E4 ORF, in combination with backward primers, Pr5628, Pr5793, or Pr5939, from different positions within the L1 ORF. Using poly(A)⁺ total RNA isolated from 16-day-old, HPV18-infected rafts, we found that the majority of late transcripts were spliced from nt 3696 to 5613 (Fig. 4B, products 2 and 3 in lanes 2 and 4). This splice junction could be verified by using a backward splice junction primer, Pr5628, of which the 3' end has 2 nt identical to nt 3696 to 3695 (Fig. 4B, product 4 in lane 6). In addition, a smaller RT-PCR product (Fig. 4B, product 1 in lane 2) detected with a primer pair of Pr3599 and Pr5793 was identified by sequencing as a product spliced from nt 3786 to 5776. This minor transcript was detectable only with a primer pair of Pr3599 and Pr5793 because the 3'-end 3 nt of Pr5793 are identical to the sequences from nt 3786 to 3784 and thereby the Pr5793 can function as a splice junction 3786/5776 primer.

To provide further evidence that the splice junction of nt 3696/5613 is used for transcriptional products from the viral late promoters, forward primer Pr822 within the E7 ORF, downstream of the major late promoter P811, was used in primer walking RT-PCR in combination with Pr5939. Two major RT-PCR products with sizes of ~433 bp (product 1) and ~695 bp (product 2) were detected from poly(A)⁺ total RNA isolated from 16-day-old, HPV18-infected rafts (Fig. 4C, lane 2). The size of product 1 was equivalent to a single spliced product of nt 929/5613, and product 2 was a double-spliced product of nt 929/3434 and nt 3696/5613. The product spliced from nt 929 to nt 5613 was confirmed by gel purification, cloning, and sequencing (Fig. 4C). Together with the data in Fig. 4B, the primer walking RT-PCR strategies were able to identify three additional splice junctions within viral late transcripts.

HPV18 sequences within each splice junction are classical eukaryotic introns with splice sites highly conserved in other members of papillomaviruses. We analyzed the sequences within each splice junction and found that these sequences contain a 5' GU dinucleotide (5' splice site or donor site) and a 3' AG dinucleotide (3' splice site or acceptor site) characteristic of an eukaryotic intron (90). Thereby, three major

introns spanning over the entire HPV18 genome were identified, with each having an alternative 5' or 3' splice site, as seen in the HPV16 or HPV31 genome (34, 91). Accordingly, the HPV18 genome contains at least five major 5' splice sites and six major 3' splice sites to facilitate RNA splicing of viral early and late gene transcripts.

Each intron has at least three *cis* elements: a 5' splice site, a branch site, and a 3' splice site with a run of 15 to 40 pyrimidines upstream. However, other sequences upstream of the 5' splice site and downstream of the 3' splice site are also important in the U1 recognition of the 5' splice site and the U2 and U2AF recognition of the 3' splice site (90, 92). To further validate the identified HPV18 5' splice sites and 3' splice sites, we compared the exon-intron boundary sequences of each 5' splice site and 3' splice site identified in the HPV18 genome with the corresponding sequences of all known 5' splice sites and 3' splice sites in BPV-1, HPV16, HPV31, and HPV11 by Pictogram (<http://genes.mit.edu/pictogram.html>). As shown in Fig. 5, the HPV18 5' splice sites and 3' splice sites we identified are highly conserved in all mapped papillomaviral splice sites. In this regard, the five HPV18 5' splice sites are most closely homologous to the four HPV16 5' splice sites. All six HPV18 3' splice sites are similar to other 3' splice sites identified from BPV-1, HPV16, HPV31, and HPV11 and have an upstream suboptimal polypyrimidine track interspersed by purines. However, their exon boundaries are not identical and contain either purine A or purine G or even occasionally a pyrimidine C or T. These data indicate that these 3' splice sites are suboptimal and would not facilitate a strong recognition by RNA splicing machinery, consequently leading to selection of an alternative 3' splice site during RNA splicing. This conclusion is consistent with the findings that the nt 233 5' splice site could be alternatively spliced to a nt 416 or nt 3434 3' splice site and the nt 929 5' splice site to a nt 3434, nt 3465, or even nt 5613 3' splice site.

Construction of a full HPV18 transcription map. Based on the mapping results of TSS, polyadenylation cleavage sites, and RNA splice sites of HPV18 early and late transcripts, we constructed a full transcription map of the HPV18 genome. As shown in Fig. 6, there are three major introns spanning over the entire HPV18 genome. Viral early transcripts are started either at nt 55 or at nt 102 and are polyadenylated at nt 4270 using a PAS at nt 4235. Viral late transcripts start commonly at nt 811 but also can use several other start sites, similar to what has been observed in HPV16 and HPV31. Viral late transcripts are commonly polyadenylated at nt 7299 or nt 7307 using a PAS at nt 7278. The majority of abundant viral early and late transcripts have three exons and two introns, with the 5' portion of the viral late transcripts overlapping with the 3' portion of the viral early transcripts. The transcripts were alternatively spliced, and the coding capacities of each RNA species may be inferred from the ORF included in the mRNA (Fig. 6).

Viral early transcripts have an intron from nt 234 to 415 in the E6 ORF and a second intron from nt 930 to 3433 spanning over almost the entire E1 ORF and part of the E2 ORF, depending on which alternative 3' splice site is chosen. Splicing of the first intron (182 nt) and the second intron (2,504 nt) from viral early transcripts would remove the E6, E1, and E2 ORFs but not affect the expression of E7 and E5 positioned separately in the second and third exons. However, both the

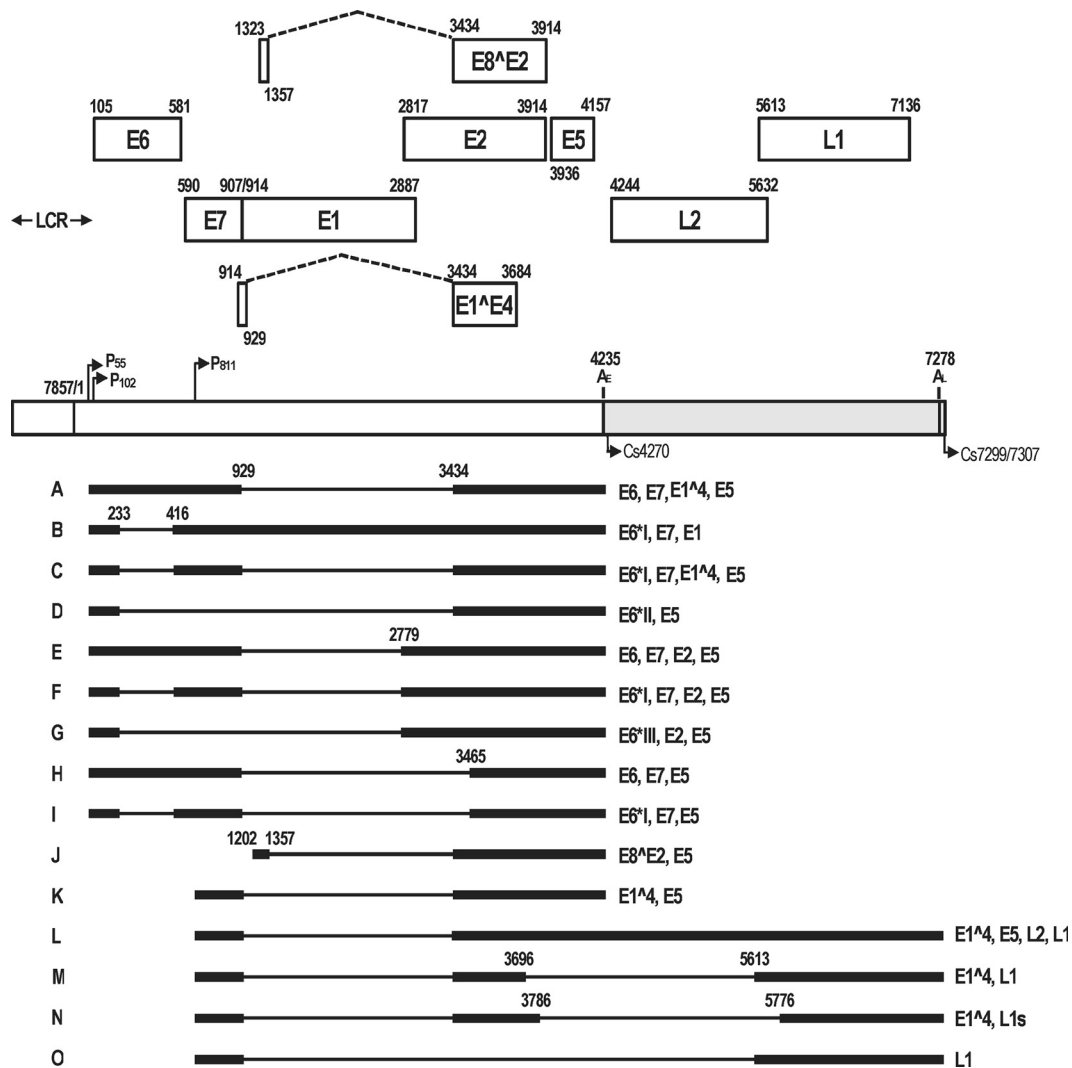


FIG. 6. Full transcription map of HPV18 in keratinocytes with productive viral infection. The bracket line in the middle of the panel represents a linear form of the virus genome for better presentation of head-to-tail junction, promoters (arrows), early (AE) and late (AL) PAS, and mapped cleavage sites (CS). The ORFs (open boxes) are diagrammed above the bracket, and the numbers above each ORF (E6, E7, E1, E2, E5, L2, and L1) are the positions of the first nucleotide of the start codon and the last nucleotide of the stop codon assigned to the HPV18 genome according to a previous report (15) and our results from this study. The E1^{E4} and E8^{E2} ORF span over two exons with the nucleotide positions indicated. Because the first AUG of E1^{E4} and E8^{E2} is positioned in the first exon, formation of an intact E1^{E4} or E8^{E2} ORF requires RNA splicing (dashed lines). LCR indicates a long control region. Below the bracket line are the RNA species derived from alternative promoter usage and alternative RNA splicing. Exons (heavy lines) and introns (thin lines) are illustrated for each species of the RNA, with the mapped splice site positions being numbered by nucleotide positions in the virus genome. Coding potentials for each RNA species are shown on the right. RNA G was not detected in this study but was reported in another study (51).

693, 706, 713, and 766 were also mapped by primer extension (28).

We are intrigued by the identification of promoter P55 as it lies in the viral replication origin (18, 19, 43, 66). In HPV16, a P14 promoter in the origin of viral replication has been recently identified to initiate viral E1 transcripts (41). Conceivably, viral DNA replication creating a replication fork at the origin would separate two strands of the P55 promoter, thereby inactivating the P55 promoter and leaving the P102 promoter as the only one to function. This hypothesis is an area for further investigation.

In eukaryotes and viruses, 3'-end cleavage of transcripts and polyadenylation are tightly coupled nuclear events and are

triggered by recognition of two primary RNA elements by the cellular polyadenylation machinery involving ~85 proteins (76). These RNA elements include a highly conserved PAS hexamer (AAUAAA or its variants [82]), 10 and 30 nt upstream of the actual cleavage for binding of cleavage and polyadenylation specificity factor (CPSF), and a U/GU-rich element located ~30 nt downstream of the cleavage site for binding of CstF (cleavage stimulation factor). Other auxiliary elements upstream of AAUAAA or downstream of U/GU, such as an element of UGUA for binding of cleavage factors CFI_m and CFI_{II}_m, are regulatory elements for stable CPSF binding and cleavage efficiency of CPSF-73 (6, 20, 37, 45, 71, 87). These RNA-protein interactions with partition of poly(A)

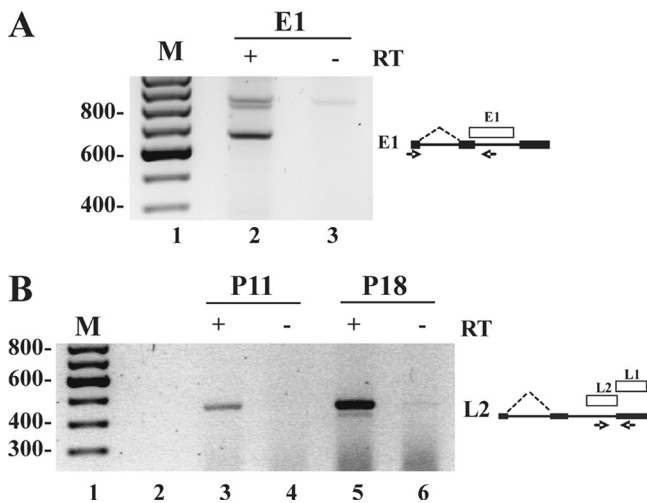


FIG. 7. Identification of HPV18 E1 and L2 transcripts. (A) Identification of HPV18 E1 transcript by RT-PCR. RT-PCR was performed on total RNA isolated from 10-day-old rafts derived from HPV18-infected keratinocytes at passage 3 using a primer pair of Pr121 and Pr967. The diagram on the right shows E1 RNA structure (heavy line, exons; thin lines, introns), RNA splicing direction and the locations of the E1 ORF and the primers used in the assay. (B) Identification of HPV18 L2 transcript by RT-PCR. RT-PCR using primers Pr5487 and Pr5935 was performed on total RNA isolated from 10-day-old, HPV18-infected rafts derived from HVK cultures at passages 11 and 18. The diagram on the right shows L2 RNA structure, RNA splicing direction, and the locations of the L2 ORF and the primers used in the assay.

polymerase in the protein-RNA complex result in the addition of adenosine residues immediately after CPSF-73-mediated RNA cleavage (45, 70). In this report, we mapped the cleavage sites of both viral early and late transcripts for RNA polyadenylation and demonstrated that viral early transcripts are polyadenylated exclusively at nt 4270 by using a PAS at nt 4235, but viral late transcripts are polyadenylated at multiple sites, with nt 7297 and 7307 sites being preferred, by using a PAS at nt 7278.

The 3'-end heterogeneity of HPV18 late transcripts appears to be unrelated to a series of UGUA motifs upstream of the nt 7278 late PAS. These motifs are in general responsible for CFIm binding and polyadenylation efficiency (6, 20, 37, 71, 87). Analyses of the 80-nt sequences 5' to the nt 7278 late PAS showed that this region contains five UGUA motifs (Fig. 3B). Similarly, there are six UGUA motifs in the same region upstream of the nt 4235 early PAS (Fig. 3A). However, the sequence 5' to the late PAS or 3' to late poly(A) cleavage sites has a potential U1 binding site (a cryptic 5' splice site) from nt 7217 to 7225 upstream of the PAS and four putative overlapped U1 binding sites (nt 7292 to 7302, nt 7300 to 7310, nt 7309 to 7319, nt 7321 to 7331) downstream of the late cleavage sites. Previous studies indicated that the presence of a U1 binding site in the 3' untranslated region of BPV-1, HPV16, and HPV31 suppresses RNA polyadenylation of viral late transcripts (16, 17, 21, 25) via its interaction with U1, U2AF⁶⁵, CstF64, HuR, and SF2/ASF (29, 39, 48). In contrast, the presence of a U-rich element upstream and six G triplets downstream of the HPV16 early PAS promotes polyadenylation of viral early transcripts (57, 89). Whether the similar structures

upstream and downstream of the HPV18 early PAS play a role as described in HPV16 remains to be understood.

Mapping of RNA splicing junctions in each species of HPV18 transcripts led us to identify all major 5' and 3' splice sites in both viral early and late transcripts. These 5' splice sites and 3' splice sites are highly conserved splice sites among all mapped HPVs (1, 12, 13, 24, 33, 34, 58, 61, 69, 74, 75, 91). These data, together with the mapping data of TSS and polyadenylation sites, provide a basis for us to construct an HPV18 transcription map. In this transcription map, we provide the direct RNA evidence that HPV18 E1 mRNA is a partially spliced mRNA without an intact E6 ORF upstream, HPV18 E4 mRNA is a spliced E1⁺E4 mRNA transcribed from viral late promoter P811, and HPV18 L1 mRNA initiates its translation from the nt 5613 AUG during virus infection, and thus, the encoded L1 consists of 507 amino acid (aa) residues, 61 aa residues shorter than the one originally predicted (15) (see Fig. S1 in the supplemental material). In addition, we identified a spliced transcript starting from nt 1202, and this spliced transcript might encode a 171-aa E8⁺E2 (E2C) by analogy to other papillomaviruses (11, 35, 40, 42, 69, 79). Altogether, we believe this carefully mapped landscape of HPV18 transcripts will provide a solid foundation for future understanding of HPV18 molecular biology, pathogenesis, and prevention.

ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the NCI, Center for Cancer Research, National Institutes of Health, and NIH grant R01 AI057988-01 to C. Meyers and NIH grant CA83679 to L. T. Chow.

We thank Jennice Gullett for preparation of raft culture tissues and other members of the Zheng laboratory for their assistance and critical comments in the course of the study.

REFERENCES

- Baker, C. C., and C. Calef. 1996. Maps of papillomavirus mRNA transcripts, p. III-3-III-13. In *Los Alamos National Laboratory* (ed.), *The human papillomaviruses compendium*. Los Alamos National Laboratory, Los Alamos, NM.
- Bernard, H. U. 2002. Gene expression of genital human papillomaviruses and considerations on potential antiviral approaches. *Antivir. Ther.* 7:219-237.
- Bernard, H. U., et al. 2010. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* 401:70-79.
- Bishop, B., et al. 2007. Crystal structures of four types of human papillomavirus L1 capsid proteins: understanding the specificity of neutralizing monoclonal antibodies. *J. Biol. Chem.* 282:31803-31811.
- Brady, J., et al. 1982. Site-specific base substitution and deletion mutations that enhance or suppress transcription of the SV40 major late RNA. *Cell* 31:625-633.
- Brown, K. M., and G. M. Gilmartin. 2003. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol. Cell* 12:1467-1476.
- Bulk, S., et al. 2006. Preferential risk of HPV16 for squamous cell carcinoma and of HPV18 for adenocarcinoma of the cervix compared to women with normal cytology in The Netherlands. *Br. J. Cancer* 94:171-175.
- Butler, J. E., and J. T. Kadonaga. 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev.* 16:2583-2592.
- Carninci, P., et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* 38:626-635.
- Chan, P. K., et al. 2009. Distribution of human papillomavirus types in cervical cancers in Hong Kong: current situation and changes over the last decades. *Int. J. Cancer* 125:1671-1677.
- Chiang, C. M., T. R. Broker, and L. T. Chow. 1991. An E1M-E2C fusion protein encoded by human papillomavirus type 11 is a sequence-specific transcription repressor. *J. Virol.* 65:3317-3329.
- Chow, L. T., and T. R. Broker. 2007. Human papillomavirus RNA transcription, p. 109-144. In R. L. Garcea and D. DiMaio (ed.), *The papillomaviruses*. Springer, New York, NY.

13. **Chow, L. T., M. Nasser, S. M. Wolinsky, and T. R. Broker.** 1987. Human papillomavirus types 6 and 11 mRNAs from genital condylomata acuminata. *J. Virol.* **61**:2581–2588.
14. **Clifford, G., S. Franceschi, M. Diaz, N. Munoz, and L. L. Villa.** 2006. Chapter 3: HPV type-distribution in women with and without cervical neoplastic diseases. *Vaccine* **24**(Suppl. 3):S3/26–S3/34.
15. **Cole, S. T., and O. Danos.** 1987. Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. *J. Mol. Biol.* **193**:599–608.
16. **Cumming, S. A., M. G. McPhillips, T. Veerapraditsin, S. G. Milligan, and S. V. Graham.** 2003. Activity of the human papillomavirus type 16 late negative regulatory element is partly due to four weak consensus 5' splice sites that bind a U1 snRNP-like complex. *J. Virol.* **77**:5167–5177.
17. **Cumming, S. A., et al.** 2002. The human papillomavirus type 31 late 3' untranslated region contains a complex bipartite negative regulatory element. *J. Virol.* **76**:5993–6003.
18. **Del Vecchio, A. M., H. Romanczuk, P. M. Howley, and C. C. Baker.** 1992. Transient replication of human papillomavirus DNAs. *J. Virol.* **66**:5949–5958.
19. **Demeret, C., M. Le Moal, M. Yaniv, and F. Thierry.** 1995. Control of HPV 18 DNA replication by cellular and viral transcription factors. *Nucleic Acids Res.* **23**:4777–4784.
20. **de Vries, H., et al.** 2000. Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J.* **19**:5895–5904.
21. **Dietrich-Goetz, W., I. M. Kennedy, B. Levins, M. A. Stanley, and J. B. Clements.** 1997. A cellular 65-kDa protein recognizes the negative regulatory element of human papillomavirus late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **94**:163–168.
22. **Dollard, S. C., et al.** 1992. Production of human papillomavirus and modulation of the infectious program in epithelial raft cultures. *Genes Dev.* **6**:1131–1142.
23. **Doorbar, J.** 2005. The papillomavirus life cycle. *J. Clin. Virol.* **32**(Suppl. 1):S7–S15.
24. **Doorbar, J., et al.** 1990. Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line. *Virology* **178**:254–262.
25. **Furth, P. A., W. T. Choe, J. H. Rex, J. C. Byrne, and C. C. Baker.** 1994. Sequences homologous to 5' splice sites are required for the inhibitory activity of papillomavirus late 3' untranslated regions. *Mol. Cell. Biol.* **14**:5278–5289.
26. **Gilmartin, G. M.** 2005. Eukaryotic mRNA 3' processing: a common means to different ends. *Genes Dev.* **19**:2517–2521.
27. **Graham, S. V.** 2008. Papillomavirus 3' UTR regulatory elements. *Front. Biosci.* **13**:5646–5663.
28. **Grassmann, K., B. Rapp, H. Maschek, K. U. Petry, and T. Iftner.** 1996. Identification of a differentiation-inducible promoter in the E7 open reading frame of human papillomavirus type 16 (HPV-16) in raft cultures of a new cell line containing high copy numbers of episomal HPV-16 DNA. *J. Virol.* **70**:2339–2349.
29. **Gunderson, S. I., M. Polycarpou-Schwarz, and I. W. Mattaj.** 1998. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol. Cell* **1**:255–264.
30. **Guo, M., et al.** 2008. Evaluation of a commercialized in situ hybridization assay for detecting human papillomavirus DNA in tissue specimens from patients with cervical intraepithelial neoplasia and cervical carcinoma. *J. Clin. Microbiol.* **46**:274–280.
31. **Homa, F. L., J. C. Glorioso, and M. Levine.** 1988. A specific 15-bp TATA box promoter element is required for expression of a herpes simplex virus type 1 late gene. *Genes Dev.* **2**:40–53.
32. **Hopman, A. H., et al.** 2005. HPV in situ hybridization: impact of different protocols on the detection of integrated HPV. *Int. J. Cancer* **115**:419–428.
33. **Hummel, M., J. B. Hudson, and L. A. Laimins.** 1992. Differentiation-induced and constitutive transcription of human papillomavirus type 31b in cell lines containing viral episomes. *J. Virol.* **66**:6070–6080.
34. **Hummel, M., H. B. Lim, and L. A. Laimins.** 1995. Human papillomavirus type 31b late gene expression is regulated through protein kinase C-mediated changes in RNA processing. *J. Virol.* **69**:3381–3388.
35. **Jeckel, S., E. Loetzsch, E. Huber, F. Stubenrauch, and T. Iftner.** 2003. Identification of the E9/E2C cDNA and functional characterization of the gene product reveal a new repressor of transcription and replication in cottontail rabbit papillomavirus. *J. Virol.* **77**:8736–8744.
36. **Jia, R., et al.** 2009. Control of the papillomavirus early-to-late switch by differentially expressed Rsp20. *J. Virol.* **83**:167–180.
37. **Kaufmann, I., G. Martin, A. Friedlein, H. Langen, and W. Keller.** 2004. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J.* **23**:616–626.
38. **Kehn, K., et al.** 2005. The HTLV-I Tax oncoprotein targets the retinoblastoma protein for proteasomal degradation. *Oncogene* **24**:525–540.
39. **Koffa, M. D., S. V. Graham, Y. Takagaki, J. L. Manley, and J. B. Clements.** 2000. The human papillomavirus type 16 negative regulatory RNA element interacts with three proteins that act at different posttranscriptional levels. *Proc. Natl. Acad. Sci. U. S. A.* **97**:4677–4682.
40. **Lace, M. J., J. R. Anson, G. S. Thomas, L. P. Turek, and T. H. Haugen.** 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.
41. **Lace, M. J., J. R. Anson, L. P. Turek, and T. H. Haugen.** 2008. Functional mapping of the human papillomavirus type 16 E1 cistron. *J. Virol.* **82**:10724–10734.
42. **Lambert, P. F., B. C. Monk, and P. M. Howley.** 1990. Phenotypic analysis of bovine papillomavirus type 1 E2 repressor mutants. *J. Virol.* **64**:950–956.
43. **Lee, K. Y., T. R. Broker, and L. T. Chow.** 1998. Transcription factor YY1 represses cell-free replication from human papillomavirus origins. *J. Virol.* **72**:4911–4917.
44. **Longworth, M. S., and L. A. Laimins.** 2004. Pathogenesis of human papillomaviruses in differentiating epithelia. *Microbiol. Mol. Biol. Rev.* **68**:362–372.
45. **Mandel, C. R., et al.** 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**:953–956.
46. **McLaughlin-Drubin, M. E., N. D. Christensen, and C. Meyers.** 2004. Propagation, infection, and neutralization of authentic HPV16 virus. *Virology* **322**:213–219.
47. **McLaughlin-Drubin, M. E., and C. Meyers.** 2005. Propagation of infectious, high-risk HPV in organotypic "raft" culture. *Methods Mol. Med.* **119**:171–186.
48. **McPhillips, M. G., et al.** 2004. SF2/ASF binds the human papillomavirus type 16 late RNA control element and is regulated during differentiation of virus-infected epithelial cells. *J. Virol.* **78**:10598–10605.
49. **McWaters, B. J., R. M. Stenberg, and J. A. Kerry.** 2002. Characterization of the human cytomegalovirus UL75 (glycoprotein H) late gene promoter. *Virology* **303**:309–316.
50. **Meyers, C., M. G. Frattini, J. B. Hudson, and L. A. Laimins.** 1992. Biosynthesis of human papillomavirus from a continuous cell line upon epithelial differentiation. *Science* **257**:971–973.
51. **Meyers, C., T. J. Mayer, and M. A. Ozbun.** 1997. Synthesis of infectious human papillomavirus type 18 in differentiating epithelium transfected with viral DNA. *J. Virol.* **71**:7381–7386.
52. **Middleton, K., et al.** 2003. Organization of human papillomavirus productive cycle during neoplastic progression provides a basis for selection of diagnostic markers. *J. Virol.* **77**:10186–10201.
53. **Millevoi, S., and S. Vagner.** 2010. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res.* **38**:2757–2774.
54. **Moody, C. A., and L. A. Laimins.** 2009. Human papillomaviruses activate the ATM DNA damage pathway for viral genome amplification upon differentiation. *PLoS Pathog.* **5**:e1000605.
55. **Munoz, N., X. Castellsague, A. B. de Gonzalez, and L. Gissmann.** 2006. Chapter 1: HPV in the etiology of human cancer. *Vaccine* **24**(Suppl. 3):S3/1–S3/10.
56. **Nasser, M., R. Hirochika, T. R. Broker, and L. T. Chow.** 1987. A human papilloma virus type 11 transcript encoding an E1–E4 protein. *Virology* **159**:433–439.
57. **Oberg, D., J. Fay, H. Lambkin, and S. Schwartz.** 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.
58. **Ozbun, M. A., and C. Meyers.** 1997. Characterization of late gene transcripts expressed during vegetative replication of human papillomavirus type 31b. *J. Virol.* **71**:5161–5172.
59. **Ozbun, M. A., and C. Meyers.** 1998. Temporal usage of multiple promoters during the life cycle of human papillomavirus type 31b. *J. Virol.* **72**:2715–2722.
60. **Ozbun, M. A., and C. Meyers.** 1999. Two novel promoters in the upstream regulatory region of human papillomavirus type 31b are negatively regulated by epithelial differentiation. *J. Virol.* **73**:3505–3510.
61. **Palermo-Dilts, D. A., T. R. Broker, and L. T. Chow.** 1990. Human papillomavirus type 1 produces redundant as well as polycistronic mRNAs in planar warts. *J. Virol.* **64**:3144–3149.
62. **Parkin, D. M., and F. Bray.** 2006. Chapter 2: the burden of HPV-related cancers. *Vaccine* **24**(Suppl. 3):S3/11–S3/25.
63. **Pastrana, D. V., et al.** 2004. Reactivity of human sera in a sensitive, high-throughput pseudovirus-based papillomavirus neutralization assay for HPV16 and HPV18. *Virology* **321**:205–216.
64. **Phillips, C., N. Pachikara, and S. I. Gunderson.** 2004. U1A inhibits cleavage at the immunoglobulin M heavy-chain secretory poly(A) site by binding between the two downstream GU-rich regions. *Mol. Cell. Biol.* **24**:6162–6171.
65. **Preker, P., et al.** 2008. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**:1851–1854.
66. **Remm, M., R. Brain, and J. R. Jenkins.** 1992. The E2 binding sites determine the efficiency of replication for the origin of human papillomavirus type 18. *Nucleic Acids Res.* **20**:6015–6021.
67. **Remm, M., A. Remm, and M. Ustav.** 1999. Human papillomavirus type 18 E1

- protein is translated from polycistronic mRNA by a discontinuous scanning mechanism. *J. Virol.* **73**:3062–3070.
68. Romanczuk, H., F. Thierry, and P. M. Howley. 1990. Mutational analysis of cis elements involved in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters. *J. Virol.* **64**:2849–2859.
 69. Rotenberg, M. O., L. T. Chow, and T. R. Broker. 1989. Characterization of rare human papillomavirus type 11 mRNAs coding for regulatory and structural proteins, using the PCR. *Virology* **172**:489–497.
 70. Ryan, K., O. Calvo, and J. L. Manley. 2004. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA*. **10**:565–573.
 71. Sartini, B. L., H. Wang, W. Wang, C. F. Millette, and D. L. Kilpatrick. 2008. Pre-mRNA cleavage factor I (CFIm): potential role in alternative polyadenylation during spermatogenesis. *Biol. Reprod.* **78**:472–482.
 72. Schneider-Gadicke, A., and E. Schwarz. 1986. Different human cervical carcinoma cell lines show similar transcription patterns of human papillomavirus type 18 early genes. *EMBO J.* **5**:2285–2292.
 73. Serio, T. R., N. Cahill, M. E. Prout, and G. Miller. 1998. A functionally distinct TATA box required for late progression through the Epstein-Barr virus life cycle. *J. Virol.* **72**:8338–8343.
 74. Sherman, L., N. Alloul, I. Golan, M. Durst, and A. Baram. 1992. Expression and splicing patterns of human papillomavirus type-16 mRNAs in pre-cancerous lesions and carcinomas of the cervix, in human keratinocytes immortalized by HPV 16, and in cell lines established from cervical cancers. *Int. J. Cancer* **50**:356–364.
 75. Sherman, L., Y. Golan, S. Mitrani-Rosenbaum, and A. Baram. 1992. Differential expression of HPV types 6 and 11 in condylomas and cervical preneoplastic lesions. *Virus Res.* **25**:23–36.
 76. Shi, Y., et al. 2009. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell* **33**:365–376.
 77. Smotkin, D., H. Prokoph, and F. O. Wettstein. 1989. Oncogenic and non-oncogenic human genital papillomaviruses generate the E7 mRNA by different mechanisms. *J. Virol.* **63**:1441–1447.
 78. Smotkin, D., and F. O. Wettstein. 1986. Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the E7 protein. *Proc. Natl. Acad. Sci. U. S. A.* **83**:4680–4684.
 79. Stubenrauch, F., M. Hummel, T. Iftner, and L. A. Laimins. 2000. The E8E2C protein, a negative regulator of viral transcription and replication, is required for extrachromosomal maintenance of human papillomavirus type 31 in keratinocytes. *J. Virol.* **74**:1178–1186.
 80. Tang, S., K. Yamanegi, and Z. M. Zheng. 2004. Requirement of a 12-base-pair TATT-containing sequence and viral lytic DNA replication in activation of the Kaposi's sarcoma-associated herpesvirus K8.1 late promoter. *J. Virol.* **78**:2609–2614.
 81. Thierry, F., J. M. Heard, K. Dartmann, and M. Yaniv. 1987. Characterization of a transcriptional promoter of human papillomavirus 18 and modulation of its expression by simian virus 40 and adenovirus early antigens. *J. Virol.* **61**:134–142.
 82. Tian, B., J. Hu, H. Zhang, and C. S. Lutz. 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* **33**:201–212.
 83. Wang, H. K., A. A. Duffy, T. R. Broker, and L. T. Chow. 2009. Robust production and passaging of infectious HPV in squamous epithelium of primary human keratinocytes. *Genes Dev.* **23**:181–194.
 84. Wang, X., et al. 2009. Oncogenic HPV infection interrupts the expression of tumor-suppressive miR-34a through viral oncoprotein E6. *RNA*. **15**:637–647.
 85. Ward, P., and P. Mounts. 1989. Heterogeneity in mRNA of human papillomavirus type-6 subtypes in respiratory tract lesions. *Virology* **168**:1–12.
 86. Wing, B. A., R. A. Johnson, and E. S. Huang. 1998. Identification of positive and negative regulatory regions involved in regulating expression of the human cytomegalovirus UL94 late promoter: role of IE2-86 and cellular p53 in mediating negative regulatory function. *J. Virol.* **72**:1814–1825.
 87. Yang, Q., G. M. Gilmartin, and S. Double. 2010. Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. *Proc. Natl. Acad. Sci. U. S. A.* **107**:10062–10067.
 88. Zhang, X., A. Virtanen, and F. E. Kleiman. 2010. To polyadenylate or to deadenylate: that is the question. *Cell Cycle* **9**:4437–4449.
 89. Zhao, X., et al. 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.
 90. Zheng, Z. M. 2004. Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. *J. Biomed. Sci.* **11**:278–294.
 91. Zheng, Z. M., and C. C. Baker. 2006. Papillomavirus genome structure, expression, and post-transcriptional regulation. *Front. Biosci.* **11**:2286–2302.
 92. Zheng, Z. M., M. Huynen, and C. C. Baker. 1998. A pyrimidine-rich exonic splicing suppressor binds multiple RNA splicing factors and inhibits spliceosome assembly. *Proc. Natl. Acad. Sci. U. S. A.* **95**:14088–14093.