

Maps of Papillomavirus mRNA Transcripts

Carl Baker^a and Charles Calef^b

^a Basic Research Laboratory, DBS, National Cancer Institute, NIH, Bethesda, MD 20892-5055

^b MS K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Explanation of Maps

In the 1995 compendium, we presented maps showing the open reading frames (ORFs) and mRNA transcripts of seven papillomaviruses (BPV-1R, HPV-5, HPV-8, HPV-11, HPV-16R, HPV-31, and HPV-47. In 1996 we presented a new map for HPV-1a and updated maps for HPV-11, HPV-16R, and HPV-31. This year we have revised and reprinted the HPV-31 and HPV-1a maps. All these maps can be obtained via the WWW at http://hpv-web.lanl.gov/HTML_FILES/HPVcompintro3.html.

In each map the significant ORFs are shown in their proper reading frames as colored rectangles. At the upper left end of the rectangle are two numbers. The first corresponds to the nucleotide (nt) position of the ORF start, the first nucleotide following a stop codon. The second number records the nt position of the first ATG, which is also indicated by a dotted line within the rectangle. The position of the last nt in the stop codon of each ORF is printed at the lower right corner of the rectangles. The rectangle's colored fill begins at the ATG and extends to the end; thus it corresponds to the coding sequence, not the entire ORF. In the cases where no ATG exists in the ORF, only one number is present in the upper left corner and the rectangle is completely filled with color. Below the ORFs is a scale of the genome divided into thousands. On the scale are placed the positions of promoters (represented by arrows) and the poly(A) signals. The exact position of the poly(A) signal is indicated. Located below the genome scale are diagrams of mRNA species, most of which are spliced. The exons are illustrated by narrow black rectangles, while the introns are indicated by black hairlines between. The numbers printed below the lines indicate the 5' and 3' termini of the RNAs, and the 5' and 3' splice junction positions. The splice junction numbers give the position of the last nucleotide in the exon before the splice and the position of the first nucleotide of the exon following the splice. Splice junctions in parentheses were deduced from the genomic sequence and have not been confirmed by cDNA sequencing. Where 5' or 3' ends of the RNAs are uncertain, no nt position is given. Superimposed on the exons are colored rectangles representing the gene, or part thereof, coded by that portion of the exon. The coding potential of each transcript is also listed at the right and includes all proteins which could be translated from a given mRNA. In that list a ^ symbol between two gene names (e.g., E1^E4) indicates a fusion product. The * symbol indicates a truncated form of a protein. Alternative splicing can join the same upstream exon to different downstream exons. The resulting truncated proteins are identified by Roman numerals, such as E6*I, E6*II, and E6*III as illustrated by HPV16R species B, C, and D (published in 1996).

Occasionally an intron occurs wholly within an ORF. In such cases it is possible that the normal beginning of the ORF may be spliced to a downstream, in-frame portion of the orf. This results in a truncated version of the complete protein, bearing a normal N-terminal portion ("head") and normal C-terminal portion ("tail"), but missing the middle. The E6*I protein from species B in HPV16R is an example.

Truncation can also occur when an upstream exon is spliced to a downstream exon such that the downstream exon is translated in a reading frame that results in a stop codon soon being encountered. The result is a protein composed of a head, translated in one of the "normal" ORFs such as E6 or E1, followed by a short tail that is out of frame. We have chosen to draw this out-of-frame (OOF) tail as a gray box to emphasize it is not in the frame of one of the normal (colored) ORFs. (Note: in the 1996 maps we showed the OOF tail as a narrower bar in the color of the upstream intron to which it is fused--a convention we now feel was misleading.) An example of this situation is the E5a* protein from species q-t in HPV31. Here a short piece of normal E5a is spliced to position 5555, the region of overlap of the L2 and L1 genes. But, because of the splice position, the 5555 region is actually translated in neither the L2 or L1 frame, but in frame three which results in a stop codon being quickly encountered. The OOF bits of RNA need not always terminate the translation of a transcript. Species l of HPV31 could be translated such that 15 amino acids (aa) of E1^E4 are fused to 18 aa from the L2 region, but not in the L2 frame, which are in turn fused to a 21 aa tail in the L1 region but not the L1 frame.

The next level of complexity is exemplified by the E1M fusion protein coded by species B in HPV-1a. Here the first 139 aa are normal E1 followed by 20 aa translated from RNA in the E2/E4 overlapping region of the genome. However, these 20 amino acids are neither E2 or E4 because they happen to be translated in frame 2, and consequently are shown in gray. In the same mRNA species, (species B in HPV 1a), the "converse" situation is also illustrated; the E2C protein bears an OOF head (beginning at the AUG codon at position 1200 in the E1 region) joined to a normal E2 tail. The narrow gray bar in exon 1 goes with the thick (normal) blue bar in exon 2. Occasionally a very short normal head is joined to a very short OOF tail, such as in the E1* protein of species C in HPV1a. We illustrate these merely because they have the potential to be translated.

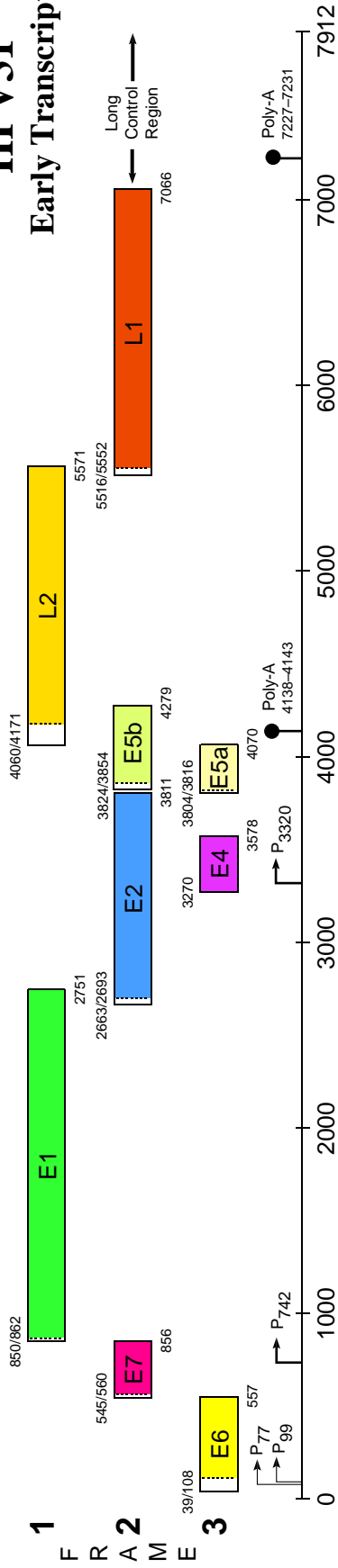
HPV-31

HPV-31 mRNAs were investigated in CIN612 cells grown in monolayer cultures and in organotypic raft cultures treated with TPA or C8 [1–3; Ozbun, M. A. and C. Meyers, manuscript in preparation]. CIN612 cells contain extrachromosomal HPV-31. Four promoters have been identified by primer extension, RNase protection, and nuclease S1 and ExoVII analyses. P_{99} is active in both monolayer and organotypic raft cultures; P_{742} is active only in differentiated raft cultures [1]. A promoter was also identified upstream of P_{99} with a start site at nt 77 [3]. Although this promoter transcribes several late mRNAs, it also transcribes early mRNAs and has therefore been named P_{77} rather than PL [Ozbun, M. A. and C. Meyers, manuscript in preparation]. Finally, a smaller number of transcripts appear to be transcribed from a promoter at nt 3320 (P_{3320}) [3]. No promoter could be identified in the E6 ORF [1]. The early region mRNA structures are shown in the facing map; late mRNA structures are shown in the following map.

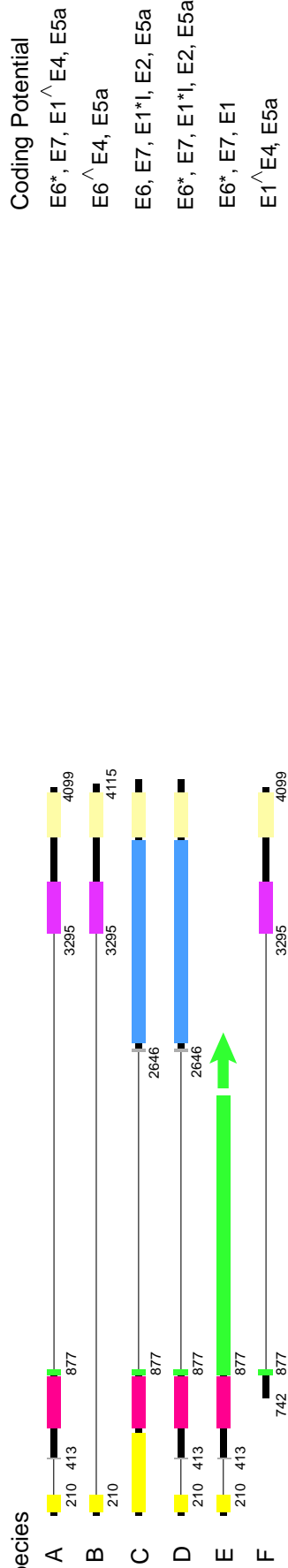
Early region mRNAs: Species A–E are most likely transcribed from both P_{77} and P_{99} , with P_{99} contributing more mRNA than P_{77} . Species F is transcribed from P_{742} . There is a polyadenylation signal at nt 4138–4143. However, early mRNAs have been shown to be polyadenylated upstream of this signal between nt 4099 and 4125 [1].

HPV31

Early Transcripts



Species

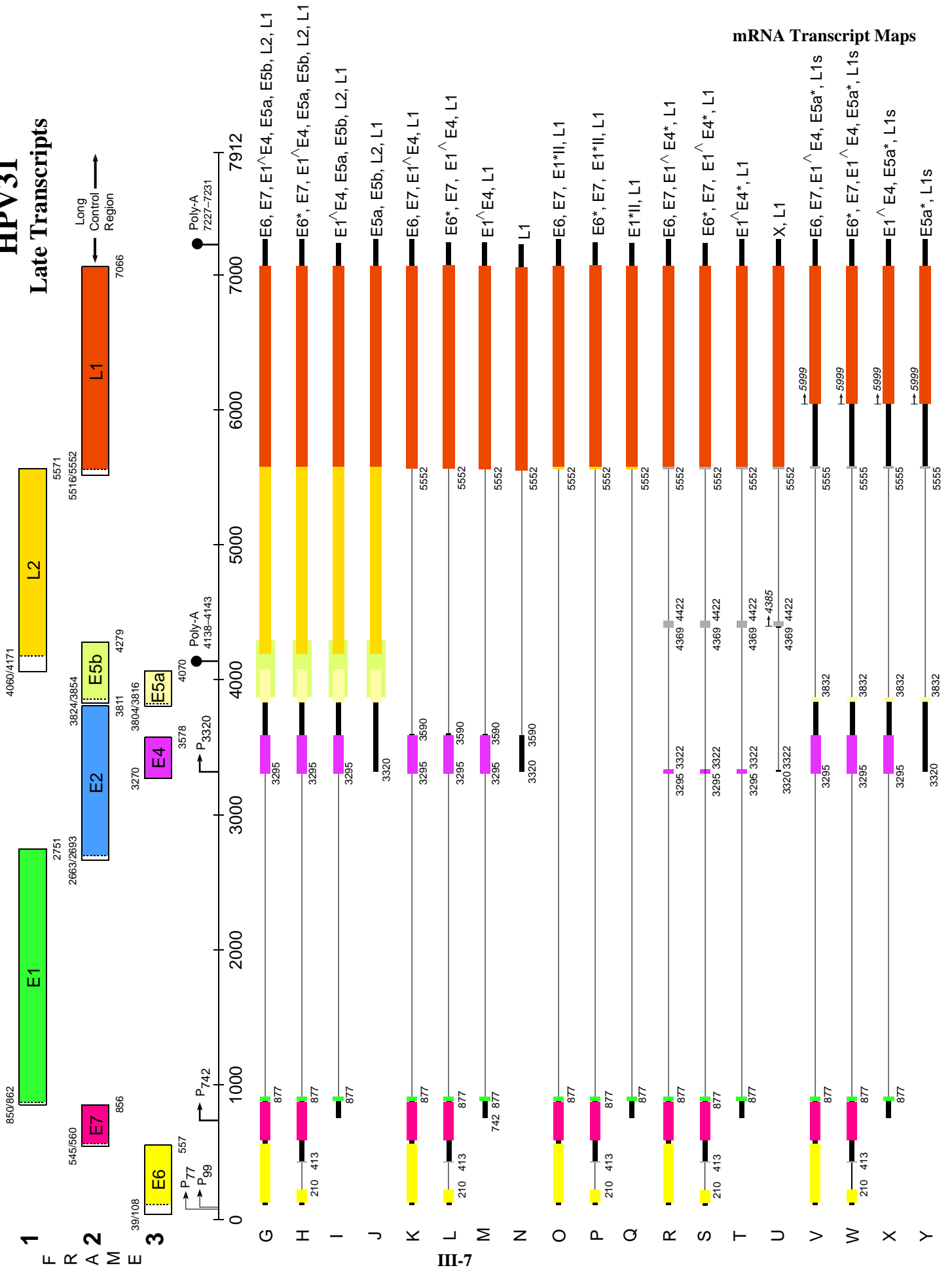


HPV-31 (cont.)

HPV-31 Late region mRNAs: The structures of the late region mRNAs on the facing page were determined by RT-PCR, cloning, and sequencing [2,3]. S1 and exoVII analyses demonstrated that mRNAs with the same basic structure were transcribed from at least three different promoters: P₇₇ (and possibly P₉₉), P₇₄₂, and P₃₃₂₀. The reason for the extreme diversity of late region mRNAs is not known. However, it is possible that some species represent nuclear splicing intermediates or the products of aberrant splicing and are not biologically relevant. The exact sites of polyadenylation at the late poly(A) site were not determined in these studies, but a putative polyadenylation signal exists at nt 7227.

HPV31

Late Transcripts

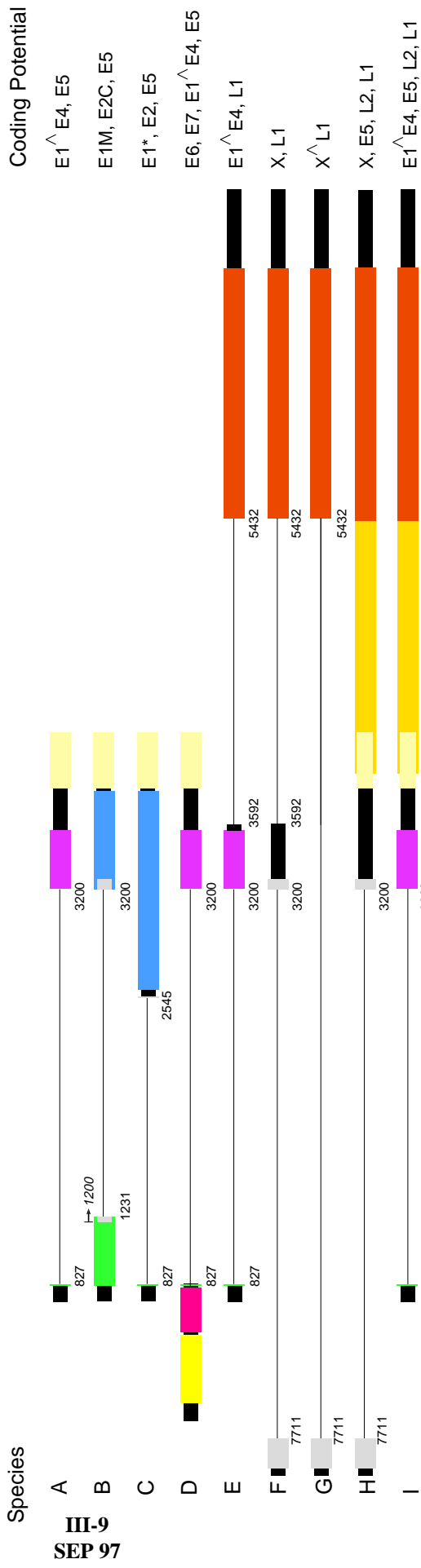
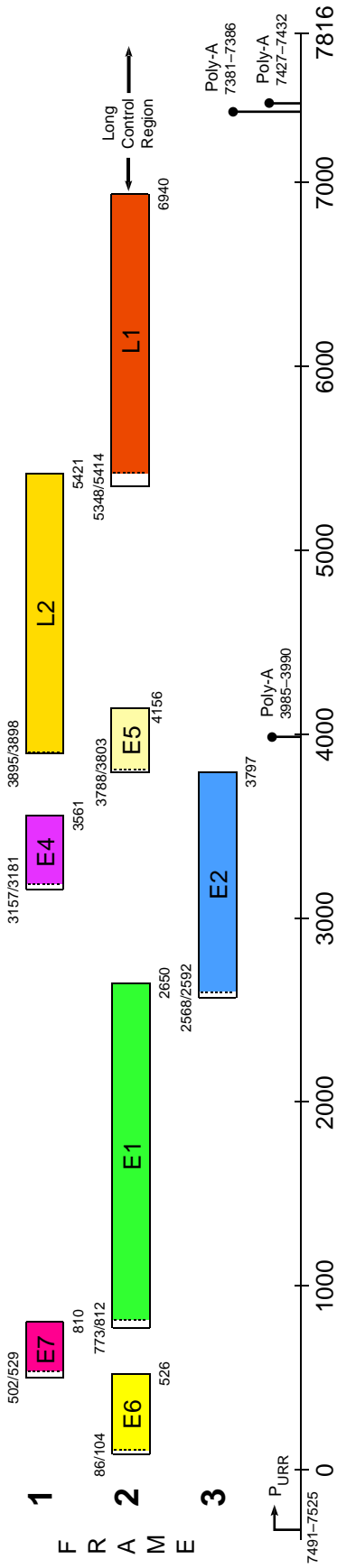


HPV-1a

The map of HPV-1a is redrawn this year to reflect corrections to the reference sequence, the most important of which, an insertion of an A at position 3884, reunites the putative E5a and E5b ORFs into a single E5 ORF. The insertion also causes the numbering of all downstream features to be increased by one. Otherwise the map remains unchanged from last year.

HPV-1a mRNAs from plantar warts and cultured keratinocytes infected with HPV-1 were analyzed by electron microscopy of R-loops and by RT-PCR [4,5]. The deduced mRNA structures are shown in the facing map. There are three putative promoters, only one of which has been accurately mapped. The major promoter is located in the E7 ORF and presumably transcribes species A-C, E, and I. The most abundant mRNA in the plantar wart is species A which encodes the E1^E4 mRNA [4]. A minor promoter precedes the E6 ORF and presumably transcribes at least species D which can encode the E6 and E7 proteins. A very minor wart specific promoter was mapped to the URR by primer extension analysis and has start sites from nt 7491-7525 and major start sites at nt 75010, 7511, and 7512 [5]. The late mRNAs F-H are transcribed from this promoter. Species F, G, and H contain two weak translation initiation codons in exon 1 which, in species G, are in frame with the L1 ORF and could therefore encode longer L1 proteins. However, the most abundant L1 mRNA (E) is transcribed from the promoter in the E7 ORF. The potential peptides starting in exon 1 of species F, G, and H, here called "X," are shown as gray boxes because they correspond to none of the recognized ORFs of HPV1a. The early and late poly(A) sites have not been determined experimentally. However, polyadenylation signals (AAUAAA) are present at nt 3985-3990, 7381-7386, and 7427-7432.

HPV1aR



mRNA Transcript Maps

mRNA Transcript Maps

Acknowledgments

We would like to thank Thomas Broker and Louise Chow for providing materials for the HPV-1a and HPV-11 maps and Michelle Ozbun and Craig Meyers for providing HPV-31 transcript data prior to publication.

References

- [1] 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.
- [2] 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.
- [3] 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.
- [4] Chow, L. T., S. S. Reilly, T. R. Broker, and L. B. Taichman. 1987. Identification and mapping of human papillomavirus type 1 RNA transcripts recovered from plantar warts and infected epithelial cell cultures. *J. Virol.* **61**:1913–1918.
- [5] Palermo-Dilts, D. A., T. R. Broker, and L. T. Chow. 1990. Human papillomavirus type 1 produces redundant as well as polycistronic mRNAs in plantar warts. *J. Virol.* **64**:3144–3149.