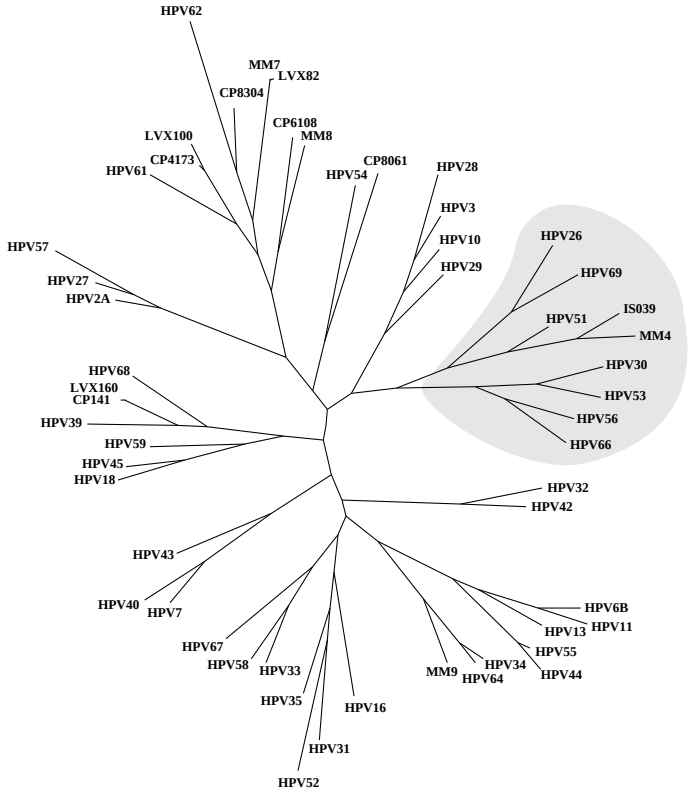
Group D Sequences

HPV26	**HPV30**
HPV51	**HPV53**
HPV56	**HPV66**
HPV69	**HPVIS039**
HPVAE2	**HPVMM4**



INTRODUCTION

Group D consists of human papillomavirus types 26, 30, 51, 53, 56, 66, 69, and the novel viruses IS039, AE2, and MM4, a group primarily associated with anogenital lesions, some with considerable oncogenic potential. Lorincz et al. classified HPV-56 as a "high-risk" virus and HPV-51 as an "intermediate-risk" virus [1]. However, Bergeron et al. placed HPV-56 in the intermediate-risk category; in the same study, Bergeron classified HPV-30 as an "intermediate risk" virus [2].

The group D viruses exclusively infect the anogenital area with the exception of HPV-30 and HPV-26 which have also been detected in respiratory and cutaneous lesions, respectively. HPV-53 and HPV-30 are rarely detected in the genital tract [3]. HPV-30 was initially isolated from a laryngeal carcinoma which suggested it might be the etiological agent of this form of cancer. An extensive effort was undertaken to probe laryngeal lesions for HPV-30 DNA. Unexpectedly, in 116 laryngeal squamous cell carcinomas, HPV-30 DNA was not detected [4]. Infrequent identification in genital lesions has been reported [3,5]. HPV-26 was first identified in numerous cutaneous lesions of an immunosuppressed patient [3]. Although the data is unpublished, at least one research lab has also identified HPV-26 DNA in numerous genital lesions [3]. HPV-53, originally isolated from a normal cervix, is also rarely found in the genital tract. When 189 asymptomatic women were examined for infection with HPV-53, only six positive samples were identified [6]. HPV-66 is frequently present in cases of multiple infection. For example, it was accompanied by HPV-16 and HPV-45 when originally isolated from an invasive carcinoma [7]. Additionally, it was detected in three out of four cervical intraepithelial cases accompanied by another HPV type [7]. HPV-69 was isolated and characterized from a cervical intraepithelial neoplasia [3].

The novel virus IS039 has been identified through international screening efforts. HPVIS039 was isolated from an invasive cervical carcinoma biopsy and is closely related to HPVMM4 (90.8%) [8]. MM4 was found in conjunction with HPV-53 in a normal cervix [9]. Initial prevalence data for MM4 is similar to that obtained for characterized "intermediate-risk" viruses [10]. Finally, the virus AE2 was detected in a cervical lavage sample obtained in a Bronx, N.Y. clinical trial [10]. Incidence rates and risk assessments have not been published for the recently identified IS039 and AE2.

Of the members of Group D, complete genomic sequences are available for HPV-26, HPV-30, HPV-51, HPV-53 and HPV-56. HPV-66, HPV-69 and the novel sequences HPVIS039, HPVMM4, and AE2 have been sequenced only over the My09-My11 region of L1. HPV-66 has been independently sequenced by two different groups and is represented by the two virtually identical sequences HPV66MY911 and HPV66L1AE3. The sequences of IS039 and AE2 are virtually identical. The members of Group D at times seem to break into two separate clusters, one including only HPV-26 and HPV-51 and the other including HPV-30, HPV-53 and HPV-56. Often HPV-26 and HPV-51 seem to show as much affinity to the members of Group C as they do to the remaining members of

Group D. For this reason, in some of the analyses done “by group” (Part III), we elected to treat Group D as composed of these two subgroups.

---

[1] Lorincz,A.T., Reid,R., Jenson,A.B., Greenberg,M.D., Lancaster,WD, and Kurman,R.J. Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types. *Obestet Gynecol* **79**: 328–337

[2] Bergeron,C., Barrasso,R., Beaudenon,S., Flamant,P., Croissant,O., and Orth,G. Human papillomaviruses associated with cervical intraepithelial neoplasia. Great diversity and distinct ditribution in low- and high-grade lesions. *Am J Surg Pathol* **16**: 641–649

[3] de Villiers,E.M. Human pathogenic papillomavirus types: an update. in *Human pathogenic papillomaviruses*, edited by Harald zur Hausen, Springer-Verlag, Heidelberg, pp 1–12 (1994)

[4] Syrjanen,S., Syrjanen,K., Mantyjarvi,R., Collan,Y., and Karja,J. Human papillomavirus DNA in squamous cell carcinomas of the larynx demonstrated by in situ DNA hybridization. *ORL J Otorhinolaryngol Relat Spec* **49**: 175–86

[5] Kahn,T., Schwarz,E. and zur Hausen,H. Molecular cloning and characterization of the DNA of a new human papillomavirus (HPV 30) from a laryngeal carcinoma. *Int J Cancer* **37**: 61–5 (1986)

[6] Gallahan,D., Muller,M., Schneider,A., Delius,H., Kahn,T., de Villiers, E.M. and Gissmann,L. Human papillomavirus type 53. *J Virol* **63**: 4911–2 (1989)

[7] Tawheed,A.R., Beaudenon,S., Favre,M., and Orth,G. Characterization of human papillomavirus type 66 from an invasive carcinoma of the uterine cervix. *J Clin Microbiol* **29**: 2656–60 (1991)

[8] Peyton,C.L., Jansen,A.M., Wheeler,C.M., Stewart,A.-C., Peto,J., Bosch,F.X., Munoz,N., Teyssie, A.R., Torroella, M., Wabinga, H.R., Sarjadi, Ngelangel,C., and Manos,M.M. A novel human papillomavirus sequence from an international cervical cancer study. *J Infect Dis* (1994) In press

[9] Manos,M.M., Waldman,J., Zhang,T. Greer,C., Eichinger,G., Schiffmann,M., and Wheeler, C. Epidemiology and partial nucleotide sequence of four novel genital human papillomaviruses. *J Infect Dis* (1994) In press

[10] Tachezy,R., Van Ranst,M.A., Cruz,Y. and Burk,R.D. Consensus primer mediated PCR allows identification of novel human papillomavirus PCR-types in cervicovaginal lavages. Unpublished

```
LOCUS       HPV26        7855 bp ds-DNA             VRL       04-OCT-1993
DEFINITION  Human papillomavirus type 26 (HPV-26), complete genome.
ACCESSION   X74472
SOURCE      Human papillomavirus type 26 DNA.
REFERENCE   1  (bases 1 to 7855)
  AUTHORS   Delius,H. and Hofmann,B.
  TITLE     Primer-directed sequencing of human papillomavirus types
  JOURNAL   Curr. Top. Microbiol. Immunol. 186, 13-31 (1994)
REFERENCE   2  (bases 1 to 7855)
  AUTHORS   Delius,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1993) to the EMBL/GenBank/DDBJ databases. H.
            Delius, Deutsches Krebsforschungszentrum, Abteilung ATV, Im
            Neuenheimer Feld 506, W 6900 Heidelberg, FRG
COMMENT     HPV-26 was first identified in numerous cutaneous lesions of an
            immunosuppressed patient.  Although the data is unpublished, at
            least one research lab has also identified HPV-26 DNA in numerous
            genital lesions.

BASE COUNT     2477 a   1449 c   1583 g   2346 t
ORIGIN      96 bp upstream from beginning of E6 cds
        1 taacaattat atgttactaa aagggtgTAA CCGAAAACGG TtgcaACCGA AACCGGTaca
                              E6 orf start ->
                                        -> E2 bind        -> E2 bind
       61 tataaaagta aaaggctagc tacgtgcaaa acagctATGt tcgaggatcc tagagaacga
                                         E6 cds ->
      121 cccagaacgc tacatgagct atgtgaaagc ttgaatacta ctttgcaaaa tttgcaggta
      181 cagtgtgtat attgcaagga aaccttacaa tgggctgatg tatataattt tgcaatttgt
      241 gacctaagag tagtatatag agataggagt ccgtatgctg catgcaaaag atgtgtaata
      301 ttttattcaa aaataacaga gtatagacgc tatacatgtt ctgtgtatgg tgcaacatta
      361 gaagccttaa ctaaaaaaag tttatgtaat ttgttaataa ggtgtcatag atgtcaaatg
      421 ccattggggc cagaagaaaa acaaagaatt gtggatgaaa agcgacgatt tcacgaaata
      481 gcagggcagt ggaaagggtt gtgtacaaat tgttggagac caaggcgcca aacagaaaca
      541 caagtgTAAa gaacaATGca tggaaacata attaatattg aagatgtaat actagatctg
                   E7 cds ->
      E7 orf start ->
                 <- E6 end
      601 gtgccgcaac ccgaaattga cctacgctgc tacgaacaat tggactatga acaatttgac
      661 agctcagatg aggatgaaac agataatatg cgtgaccagc aggccagaca agctggacaa
      721 gaagtgtgtt acagaattga agcacaatgt tgtatgtgta atagtatagt gcagctagct
      781 gtgcagagca gtcgacagaa cgttcgagtg ctggagcaga tgtTAAtgga agacgtgtcc
                                             E1 orf start ->
      841 ttggtgtgcc atcagtgtgc tgcacagTAA acctgcaATG gactgtgaag gtacaaatga
                                          E1 cds ->
                                       <- E7 end
      901 ggaggggcgg gggtgtacag ggtggttttc agtagaagct atagtggaaa aacatacagg
      961 ggacacaata tcagatgatg aaacagacaa tagtagtgat acagggtcgg acctaatagg
     1021 atttatagat gatagtagta taagtgatta tgcagaacag gaggtaaccc aggcattgtt
     1081 tcaggcacaa caaaaacagg caaatacaaa ggcagtgcgc aatttaaaac gaaagttact
     1141 aggtagtcag aacagcccgt tgcaagacat aacaaatcaa cacagacagc aaagcgacag
     1201 tcagcagaat acacaccaag taaataattc acaggccaaa aggagagccg tggacagtgt
     1261 accggacagc gggtatggct atactgaagt ggaaactctt acgcccgtac aggtagataa
     1321 acaatatgaa gaaaatggcg ggttgcctag tgtgtgtagt caggggggt caaatgcctc
     1381 agtggaagat atcgatgtag acacacatgt aaacagtgtt acacaaatat gtgaattatt
     1441 aaaatgtagt aatgtaaaag cagcattgtt aagtaaattt aaaacagtat atggtgtaag
     1501 ttttgcagaa ctagtacggg tgtttaaaag tgacaaaaca tgctgttcag attgggtgtg
     1561 tgcagcattc ggtgtggcag gctctgtagc agaaagtatt aaatcattaa tacaacaata
     1621 ttgtttatat tatcatatac aatgtttaac atgtaattgg ggagtaatag tactaatgct
     1681 agtgcgcttt acatgtgcaa aaaacagaac aacaattaaa aactgcctat gtatgttatt
     1741 aaatgtgcca gaaacgcaat tactaattga accaccaaaa ttgcgaagta cagcagtagc
     1801 attatatttt tataaaacag ggttgtccaa tataagtgag acatatggag atacaccaga
     1861 atggatagta cgacaaacac aattagaaca tagttttgat gatgctacat ttgatttatc
     1921 aaaaatggtg caatgggcgt tcgatcatga cataacagat gatagtgaaa ttgcatttaa
     1981 atatgcacag ttagctgaca tagatagtaa tgcagctgcc tttttaaaaa gcaattgtca
     2041 ggcaaaatat gtaaaagact gtgcaaccat gactagacat tataaaagag cacagaaacg
```

```
2101 atctatgtgt atgtcacaat ggctacaata tagatgttct aaaatagaag agggcgggtc
2161 gtggaaggaa attgccaaat ttttaaggtt tcaacatgta aactttattt atttttaca
2221 agtgttaaaa cagtttttaa agggtacccc aaagcacaat tgtatagtaa tatatggacc
2281 gccaaatact ggtaagtcac agtttgcaat gagttttata aaatttatgc aagggtcagt
2341 catttcatat gtaaattcaa atagccattt ttggctgcag cctttagaag atgcaaaagt
2401 tgcagtatta gatgatgcta catatagctg ctggttatat attgataaat atttacgtaa
2461 ctttttagat ggaaatccct gttgtataga cagaaaacat agaagcctac tgcaagttac
2521 atgcccccca ttaataatta cctcaaatat aaatcctcaa gaagataact cacttttgta
2581 tttacatagt agagtaacag tgataccatt tccaaataca tttccatttg acagcaatgg
2641 gaatcctgta tatgcattga ctgatgtaaa ttggaaaagc tttttttcca ccacctggtc
2701 cagatTAGat ttggaggagg acgcggacaa agaaaATGga gaacctttgc cagcgtttaa
E2 orf start ->                        E2 cds ->
2761 atgcgtgcca ggagaaaata ctagactatt aTGAactgga cagtaataaa ttaactgatc
                                       <- E1 end
2821 aaattgatta ttggaaactg gtacgatatg aatgtgcaat attttataaa gctcgtgaag
2881 gaaacatgca atgtataaac caccaggtgg tgccctctac tgttgtgtgt aaacaaaagg
2941 catggcaggc aattgaaata catatagcat tgcagtcgtt aataaacacg gactataata
3001 cagaagcttg gacaatgcga gacacaagct atgaaatgta tatgacagaa cctaaacatt
3061 gttttaaaaa agaaggaaca acggtaacag tggtatttga ttgtaataag gaaaatacaa
3121 tggattatat taggtggaaa tatgtgtatt ataaaactga tatagggtgg tgtaaaggta
3181 ctggagatgt tgatgcaaaa gggatatatt atacacaagg ggcatataag cagtattacg
3241 tggactttaa acaagaggcg gaaaaatatg ggacaggtgt gcaatgggct gtacatgtgt
3301 gtggtcaggT AAtctgttgt cctgaatttg tatctagtac ctgcagcagc aaccaaatat
     E4 orf start ->
     NH2 terminus unknown
3361 ccactgctaa aactgctgag ccagtatcaa acgccaccac ccagaccacc gaagcctacg
3421 tgcccgtggg caccaaggaa accgaggcgc catacccagg aaagcgacga cgactcagtg
3481 gacctgacac caccgtcacc acagtcacca ctgtcaccac agctgccaca cagcccggac
3541 agtcagtgga ctatacaaac aacaacctac acagtacaag tggaggccat cacccgggaa
3601 gggacacgag tagtgaccaa actgtgttta TAGtacacct aaaaggtgat acaaatagtt
                                       <- E4 end
3661 taaaatgttt aagatataga tttaaaaagc ataaaggatt gtattgcaat gtatcatcta
3721 cctggcattg gaccagtaat gataccaatc aacaaggcat tgtaacaatt acctttaaca
3781 gtataacaca acgtaataat tttttaacaa ctgttaaaat accacaaagt ataacttcaa
3841 cattgggaat aatgtcattg TAAtatagtg tatattttac caacacacaa gccaatatgt
                           <- E2 end
3901 gctgcTAAca cacctatata cctgtaatca ttattccttt tatagtttat gtgtttgtgc
E5 orf start ->
NH2 terminus unknown
3961 tttgcGTGTG TGTGTGTGTG Ttgctgtgtt gtttgttgcc acttttgctt tccatttatg
          ->   repeats
4021 tgtttgcagc ctcgctatta ttagtgtttt gtttttggtt tgtggtatct acatcatata
4081 taactaccta tattgtgtat atttgcttat tttttatacc tgcttgtttt ttacattttt
4141 atactgtaat ggTAAttgct actTAGtccc tttaataaac gtgttacaAT Ggtagctgtc
        L2 orf start ->        <- E5 end              L2 cds ->
4201 cgtgcccctc gacgcaaacg agcatcagct acagacttat acaaaacatg taaggccgct
4261 ggtacgtgcc ctcctgatgt tattcctaaa attgaaggtt ctacccttgc tgataagata
4321 ttacaatgga gtggtttggg aatattttta ggtggtcttg gtataggtac aggaactggg
4381 tctggtgggc gtactggata cattccccta ggagggggtg gtagaccctc tgttgtggat
4441 atcggcccta cccgtccgcc cattattatt gaacctgtgg gtcctacaga accttctata
4501 gttactttgg tggaggaatc tagtattata caatctggag cccctatacc tacatttagt
4561 ggtggcaatg gctttgaact taccacatcc tctgcaacaa cACCTGCTGT GTTggacatc
                                                 -> E2 bind
4621 acccccctctg ctggtactgt acatgtaaca agtaccaata tacaaaatcc attatatatt
4681 gaaccccccta tagatatacc acaggccggg gaagcatcag gtcatatatt tactacaacg
4741 tccacagctg gcacacatag ttatgaagaa attccaatgg aagtatttgc ttctactaat
4801 ggaacaggat tagaacctat tagtagtaca cctattcctg gtatacaacg agtgtcagct
4861 cctcgtttgt atagtaaggc ctatcaacag gtaaaggtta cagatcccaa ttttattggt
4921 aatccctcca catttgttac ctttgataat cctgcatatg aacctataga tgaaacactt
4981 acatatgctt ccagtagtac tgtagcacct gaccccgatt ttttggacat tattgcattg
5041 catcgtccgg cccttacatc tcgcaaaggt actgtacgct atagtaggtt gggtcaaaag
5101 gccactatga aaacacgtag tggaaaacaa attggagcta cagtacatta ttatcatgat
5161 attagtccta tacagtcttt tgctgaacac gaagaaattg aattgcagcc tttacataca
5221 tctacccatt catctgcacc tttgtttgat atatatgcag accctgatac agttcctagc
```

```
5281 atacatacgc cgcgcatgtc atattcccct acaacattac cagttccaag atatgcctcc
5341 aatgtgtttt cctctattaa tacttccact accaatgtta ctgtgccttt atccacctca
5401 tttgaactac ctgtatatag tgggtcagac atttacacgc ccacatcttc cccgacatgg
5461 ccatcattgc ccccccccacc caccactaac ttacctgcaa tagttgtgca tggggataat
5521 tattatttat ggccctatat ttattTAAtc cataaacgcc gtaaacgtat gccttatttt
                    L1 orf start ->
5581 ttttcagATG gctttgtggc gtacTAGtga cagcaaggta tatcttcctc ccacccctgt
        L1 cds ->             <- L2 end
5641 gtctcgggtt gtcaacacgg atgaatatgt aactcgcacc ggcatatatt attatgcggg
5701 cagctctcgt ttattaacat taggacatcc atatttttcc atacctaaaa ctggccaaaa
5761 ggccgaaatt cctaaggtat ctgcctatca gtacagggta tttagagtgc acctacctga
5821 tcctaataaa tttggattgc ctgatccaca gttatataat cctgacacag aacgcctggt
5881 gtgggcctgt gttggtgttg aagttggtag aggacagcca ttaggcattg gccttagtgg
5941 acatcctttg tttaataagt tggatgatac cgaaaactct catttggcta ctgtaaatgc
6001 agacactgac aacagggaca atgtttcagt tgataataaa caaacacagt tatgtattat
6061 aggttgtaca ccgcccttgg gagagcactg gggtattggc actatatgta aaaatacaca
6121 gacacaacgt ggggattgcc ccccttaga attaatttcc agcattattg aggatggcga
6181 tatgattgat acaggctttg gagctatgga ttttactgcc ttacaggcta ccaaatcaga
6241 tgtgcccatt gatattagtc aatccacatg taaatatcct gattatctta aaatgtctgc
6301 agatacatat ggaaacagca tgttttttt tcttcgccgg gaacaattat ttgccagaca
6361 tttttataat aaggcggggg ctgttgggga tgctataccc accactttgt atattaaagg
6421 tgctgaatca ggcagggagc cccctacatc ttctatttat tctgctacac ctagtggctc
6481 tatggttact tcggatgcac aactatttaa taagccatac tggttacaac gtgcacaggg
6541 tcataataat ggtatctgtt ggggcaatca attgtttgtt acctgtgttg ataccacccg
6601 cagtactaac cttaccatta gtacattatc tgcagcatct gcatccactc catttaaacc
6661 atctgattat aaacaattta taagacatgg cgaagaatat gaattacaat ttatatttca
6721 gttgtgtaaa ataacactta caacagatgt tatggcttac atacatttaa tgaatgcctc
6781 catattggag gattggaatt ttggactaac cttacctccc actgctagtt tggaagatgc
6841 ctataggttt attaaaaact ctgctactac ctgtcagcgt aacgcccctc ctgtgccaaa
6901 ggaagatcct tttcaaaaat ttaaattttg ggatgtagat ttaaaagaaa aattttctat
6961 tgatttggat caatttccac tagggcgtaa gtttatgtta caggccggca tacaacggcg
7021 gccgaaacta ggcaccaaac gtcccttatc ttctacctct tcctctacca aacgcaaaaa
```

HPV26

```
7081 acgtaaactt actaaaTAAt tccgcatgtt tgtgtgtgtg tatatgtttg tggtatgtta
                      <- L1 end
7141 tgtaaggtgt gtttgtatgt gtgtacaact gtatgtaatg taaaggcgtg ttttgtatgt
7201 atgtattata tgtatgcatg gttatgtgtt ttcctgtttg tatgcatact tgttatttaa
7261 taaagtatga atgtgtcttc atgcatggtt acatgtcttt actacactat ttgtcatttg
7321 ttttacccct gaggtaatgg gaggaACCTT AGGTGGTgtc ccttataatt attatattac
                             -> E2 bind
7381 acaagtttac cttttatatg tatttcacta aacttttgta gtgttatatt tttacttttt
7441 atattttcta ttactatcta ttgtcccaAC CGTTTTCGGT cgttccttat tttagttttt
                                  -> E2 bind
7501 atccaacttt catgctgtat cctgcaggaa cagttaatcc tttggcagac aacacatcct
7561 gcctcctacg cttggcttgc cattttggca ctataagtgg cgcgcctgta ttagtcacat
7621 atatttaaac aatacttaca taagcacttt ttcttacatt ataataaaac tgctgttagg
7681 cacatatttt tattttattt tataggtctt ttaagtgcat agttggctaa catatacact
7741 tttgtttgcc aactatgtgt ctgacacata ctgttgtaac ccatagttaa acacaggtgt
7801 gtatgtaacc gaaatgtgtt ttgttaatgc atgcaagttt ctttataata acttt
```

```
LOCUS       HPV30        7852 bp ds-DNA             VRL       04-OCT-1993
DEFINITION  Human papillomavirus type 30 (HPV-30), complete genome.
ACCESSION   X74474
SOURCE      Human papillomavirus type 30 DNA.
REFERENCE   1  (bases 1 to 7852)
  AUTHORS   Delius,H. and Hofmann,B.
  TITLE     Primer-directed sequencing of human papillomavirus types
  JOURNAL   Curr. Top. Microbiol. Immunol. 186, 13-31 (1994)
REFERENCE   2  (bases 1 to 7852)
  AUTHORS   Delius,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1993) to the EMBL/GenBank/DDBJ databases. H.
            Delius, Deutsches Krebsforschungszentrum, Abteilung ATV, Im
            Neuenheimer Feld 506, W 6900 Heidelberg, FRG
COMMENT     HPV30 was first isolated from a laryngeal carcinoma (Kahn et al.
            Int. J. Cancer 37, 61-65 ).  Once isolated, HPV30 was used as a
            probe to screen a large number of tissues of both mucosal and
            cutaneous origin.  In this study, HPV30 DNA was detected in one
            vulva papilloma and one case of condyloma acuminatum.  HPV30 was
            not present in any of the other tissues tested including 41
            laryngeal carcinomas, 31 other head and neck tumors, 10 bronchial
            carcinomas, 16 cervical, vulva and penile carcinomas, 23 head and
            neck papillomas, 3 rectal and bladder papillomas, 7 cervical
            dysplasias, 26 condylomata acuminata, 3 epidermodysplasia
            verruciformis and 2 Bowenoid papuloses. In a separate study,
            Bergeron et al. (Am J Surg Pathol 16: 641-649) classified HPV-30
            as an "intermediate-risk" virus.

BASE COUNT     2396 a   1504 c   1667 g   2285 t
ORIGIN      101 bp upstream from beginning of E6 cds
        1 tgaaagttac aagcatagtt tatagaaagg gagtgACCGA AATAGGTttt caACCGAAAA
                                               -> E2 bind         -> E2 bind
       61 CGGTacatat aaaagcactg taccaaacgg acagtgtacc cATGgctttc aaatttgaaa
                                               E6 cds ->
      121 atacaggcga gcgcccacgt actgtgcacc atctttgtga ggtacaagaa acatcgttgc
      181 tggagctaca gctacagtgt gtatattgca agaaggaatt atccagctca gaggtatata
      241 attttgcatg taaagattta agactggtat atagggagga cagcccatat gcagtgtgca
      301 atttctgttt attattttat agtaaagtaa gaaagattag acattacaac tattcattgt
      361 atggggcaag cctagtggca ttaactaaaa aagagttatt tgatttatta ataaggtgct
      421 acagatgtca acagccgttg acaccagagg aaaaacagtt acactgtgaa tataagaaac
      481 ggtttcacag aatatcacgt acgtggaccg ggttatgtct gcaatgctgg agacacacaa
      541 cgtccacTGA gacagcagta TAAtcATGca cggtaaagta acaactattc cagaatatat
        E7 orf start ->       E7 cds ->
                                 <- E6 end
      601 tttggacctt gtaccgcaaa ctgaaattga cctgcattgc tatgagcaat tgaacagctc
      661 agaggaagag gatgaggatg aagtagacaa tttacagaag cagccacagc aagctagaca
      721 agaagaacaa catccttgtt acctaattaa cacacagtgt tgtaggtgtg cgtctgcggt
      781 gcagttggct gttcagagtc ccacaaagga gctgcgtgcc ctacaacaga tgcttatggg
      841 cgcactggag ctagtgtgtc ccctgtgtgc aacaaggcgg TAAacggcaA TGgcgtcacc
                                                      E1 cds ->
                                          E1 orf start ->
                                                     <- E7 end
      901 tgaaggtaca gatgatgagg gaggggggtg cacgggatgg tttcatgttg aggctgtagt
      961 caagaaacgt acgggagata taatatcaga ggacgaaaca gaagaagatg agggcacagc
     1021 atcagattta gatgggtttc tagacaatag taatgtaata acaacacagg cagacaggga
     1081 gacagctcag cagttattgc atgcccaaaa cacatatgca gatacacaga cgctgcacaa
     1141 tttaaaacga aagtatttag gcagtccgtt gggagacatt agtaatcagc aatttgtgtg
     1201 ccgggaagga gtaaaacgaa ggataattga tacagacgtg gccgacagcg ggtatggcaa
     1261 tactttggaa gtggaagcaa cgcaacaggt acaggataat acgtatggta gtggtaaaca
     1321 gcaagatgga ggctcacaga ctagtgtgtg tagcagggaa aacagcatag aagcagacag
     1381 tgatatggat ataggcgcca cgccaccaca gcaaatacaa gaattattaa aatctagtaa
     1441 tgtacaggca aagctgtgtt ataaatttaa agagttattt ggaattccat tttcagaatt
     1501 ggtacgtaca tttaaaagtg acagcacatg ttgtcatgac tggatatgtg ccatgtttgg
     1561 tgttaatgaa accttagcag aggcattaaa aacaattatt aaatcacagt gcatgtatta
     1621 ccatatacag tgtttaacgt gtacatgggg cgttgtaata ttaatgttaa ttagatatac
```

```
1681 atgtggaaaa aatagaaaaa caattataaa atcactaagt tcaattgtaa atgtgcccag
1741 cgaacaaatg cttgtgcagc cacctaaaat acgtagccct gccgttgcat tatactttta
1801 caaaactgca atgtctaata ttagtgacat atatggtgag acaccagaat ggatacagcg
1861 acaaacacaa atacagcaca gttttcagga ctgccaattt gaactgtcga aaatggtgca
1921 atgggcattt gataatgacg taacagatga cagtgacatt gcattttact atgcacagtt
1981 agcagatgta gatagtaatg cacaggcttt tttaaaaagc aatatgcagg ccaaatatgt
2041 aaaggattgt ggaataatgt gcagacacta taaaagagca caacaacaac aaatgaacat
2101 gaaacaatgg attacacaca tatgtagtaa agtagatgag ggcggggact ggaggccaat
2161 tgtacagttt ttaagatatc aaggggtcga cttcatttca tttttaagtt attttaaatt
2221 atttttacga ggcacaccaa aacacaactg cctagtatta tatggaccac caaacacagg
2281 aaagtcatgt tttgctatga gccttataca atttttttcag gggtctgtta tttcatatgt
2341 aaattcacac agtcactttt ggttacaacc gttagacaat gctaaattgg gaatgctaga
2401 tgatgccaca gatgcatgct ggagatatat agatgaatat atgcgaaatt tattagatgg
2461 gaatcctgta agtttagata gaaaacataa acaactagta caaataaaat gtccaccagt
2521 aataataaca acaaatatta atccactgca cgatgcaaaa ttgcagtact tgcacagtag
2581 aatacatgtg gtgccatttc taaatccatt tccaattgat acaaatggta atcctgtata
2641 tcaattaaat aatgtaaact ggaaatgttt ttttgaaagg acatggtcca gattagattT
2701 GAacaacgac gaagacaagg aaaaccATGg agactctatg ccaacgttta gatgcgtgcc
E2 orf ->                E2 cds ->
 start
2761 aggagaaaat tctagactgt ttTGAaaacg atagtaaaaa aattgaagac catatagtgt
                             <- E1 end
2821 actggaaagc tgtacgacat gaaaatgttg tattatataa agcaagacaa aataatatta
2881 ctaaactacg ccaccaggtg gtgccatgtt tacaagtgtg taaagcaaag gcatgtgttg
2941 ctatagaaat acaaatggca ttggaatcat tatataaaac agagtataaa gtggaggagt
3001 ggacattaaa agatgtatgt gaaaatatgt ggcatacagc accaaaacag tgttttaaaa
3061 aaagtggaaa acgtatagaa gtgtggtttg atgggaaaaa agacaatcga actgaatatg
3121 ttgtgtggca atgggtgtat tactgtgggg acaatgggtg gactaaagtg ccttctgtag
3181 tagattacaa aggtatatat tatgtacatg acggtaacaa agtatattat acagacttta
3241 atgacgaggc agTAAagtAT Gggtataaag gcacatggga agtgcatatg ggaaatgaaa
               E4 cds ->
       E4 orf start ->
3301 gtatttattg tcccgactct gtgtctagta ccctcagatc caacgtatcc cctgttgaaa
3361 ctgttgtcga atacaacacc tacaacacct atcaaacccc caccacctcc acgcccgtgg
3421 gcgccaacga agccgcgtcc tccgcacggc cgggaaaacg tcctagaacc acagagcccg
3481 acagtacaga caccaccaga cagtccgctg ccagagagtc ccacgcaaac cgtgtcaaca
3541 caaacaacac aaacaacaga cagtgccttg gtggagctac atgttacaac acagaagtcg
3601 acggtggtta taaaactaca cctgTAGtgc atttaaaagg tgaaccaaac agattaaagt
                               <- E4 end
3661 gtttaagata taggtgtcaa aaacataagc acctatttgt aaatatatcg tctacatatc
3721 attggaccaa tacacataca gagtacagct acattactgt tgtatataaa gatgagaccc
3781 agcgtgccaa ctttttaaat gttgtaaaaa taccacccag tataaaaatt gtaatgggac
```

```
3841 atatgacagg tgttgatatg TAAcaacaca tatgttgtaa ttggtacaac aaacatgtaa
                          <- E2 end
3901 tattgtattc cacatgtaaa tattattgtg tactaTAGtt atacaacaac aaacgccata
                               E5 orf start ->
                               NH2 terminus unknown
3961 ttttgctgct acgcctgtta tatattgcaa ccattgtttt atttgtTAGt tttgtgtttt
                                        L2 orf start ->
4021 tgctttgctg ttttgtgtgc ctgtgtgtct gcctatatgt gccgcttttg ctgtctgcct
4081 cgttgttttc tacctgttta ttacttataa tattgttttg gtttgttgtt gcatcgtcct
4141 acattactgc atttactata tttttgctgt ttttttatat acctttatta cttgtatatg
4201 cccatgctgt gtggttaata aacacacaaT AAattgtttt gtaacatatt acacttttgc
                                     <- E5 end
4261 tttgtgtagg gcgtgtactA TGgtcgccca ccgggcacgc agacgaaaac gggcctctgc
               L2 cds ->
4321 tactcagctg tatcaaacgt gcaagcaggc tggcacatgt ccatctgatg ttattaacaa
4381 aattgaacac acaactttgg cagataagat attacaatgg ggtagtttgt ttacattttt
4441 tggtaattta ggtataggca ctggtgcagg ctctgggggg cgtgctggtt atgttccgtt
4501 aggtacacgg cctacaacag ttgtggacgc atcccctgct aggccaccta ttgttgttga
4561 atctgtgggt cctacggacc cttccattgt tacactggtt gaggagtcta gtgtggttaa
4621 tgctggagca tcattcccta attttacagg cactgcaggc tttgaggtta cgtcttcctc
4681 taccacaaca cctgctgtat tagacattac acctaccact ggctctgtcc atgttagtag
4741 tacccatttt accaatccct cctttgttga acccccagtt attgaggttc cacaaactgg
4801 ggaggtatcc ggccatatat tggttagcac tcccacatct ggtgtacata gttatgaaga
4861 aatacctatg caaacatttg ctgtgcatgg tacaggcact gaacctatta gtagtacccc
4921 tattcctggg ttacgacgta tagctgcccc tagattgtat caaagggcct ttcagcaggt
4981 aaaggtcact gaccccacat tccttaccaa acctgaaaca ttaattactg tggataatcc
5041 agtctttgag gatgctgaca caactttaac cttttcacca tcgggtgtgg cacctgatcc
5101 tgacttttta gatattgtgg ctttacacag acctgccttt acaacacgta ggggtggtgt
5161 gcgttttagt aggcttggta caaaggccac aatgcgcaca agaagtggca aacaaatagg
5221 tgctcgtgta cattattatt atgatgtgag tcctattgca cacactgagg aaattgaaat
5281 gcagccatta ttgtctgcaa ataattcatt tgatggccta tatgatattt atgcaaacct
5341 tgatgatgag gcgccagtgt catctcacct atccattgct acACCGTCCC GGTTgcctAC
                                                  -> E2 bind       ->
5401 CAACACTGTT cctttgtcct ttagtagtca aactaccaat gttactatac ctttgggtaa
     E2 bind
5461 atattgggat gttcctattt attcgggccc cgatatagta ttgcctactg gtcctaccac
5521 atggccctat gcccctcagg ccccatttga cactacacat gatgtggtta tacatggatc
5581 tacatttgct ttatggcctg tatacttttT AAggcgtagg cgtcgtaaac ATGttcccta
                           L1 orf start ->          L1 cds ->
5641 ttttcttgca gatggcggtg tggcggccTA Gtgaaaccaa ggtttacctg cctcctacac
                                    <- L2 end
5701 ctgtatcaaa ggtggtacca acagatgcat atgtaaagcg caccaatata ttttatcatg
5761 caggcagctc acgtttgctt gctgttggac atccatatta ttctatttct aaggctggta
5821 attccaaaac agatgttccc aaggtgtctg catttcagta tagggtcttt agggtccgct
5881 tgcccgatcc caataagttt gggttacctg ataccaatgt atttaatcct gagcaggagc
5941 ggctggtatg ggcctgtgtg ggcctggaaa taggccgtgg ccaaccttta ggtgttggtg
6001 ttagtggcaa tcctttattt aataaattgg atgacactga aagttccact atagctaatc
6061 aggatacagc agaggatagt agggacaaca tttctgttga tccaaagcaa acccaattgt
6121 gtattattgg gtgcactcct gctataggag aacattgggc taaaggcact gcctgtcgtt
6181 ctgcgccccc tgcacaaggg gattgtcctc ccttggaact tgttaattct cctatacagg
6241 atggcgacat ggttgatatt gggtttggtg ccatggattt taaaacatta caggaatcta
6301 aatcagatgt gcctttagat atttcacaat ccacctgtaa atatcctgat tatttaaaaa
6361 tgagtgcaga cgcctatggc gattctatgt ggttctattt gcgtagggaa cagttatttg
6421 ctaggcacta ctttaatagg gcaggtgcta ttggtgaaca attacctagc acattatata
6481 taaaaggtac aaataacagg gatcccccgc caagctcagt atatgttgct actcctagtg
6541 ggtctatggt aacctctgag gctcagttgt ttaataaACC TTACTGGTTg caacgcgcac
                                          -> E2 bind
6601 agggacacaa taatggcatt tgttggggca accaggtatt tgttactgtt gtggacacca
6661 ctaggaacac aaacatgact atatctgcaa ccacacaaac gttatccaca tataattcaa
6721 gccaaattaa acagtatgta agacatgtag aggaatatga attacagttt gtgtttcaac
6781 tgtgtaaaat ttcactgtct gcagaaacta tggcctattt acatactatg aactccactt
6841 tacttgaggg ctggaatatt ggattgtcac ccccagctgc cacaagctta gaggacaaat
6901 acagatatgt taaaagcctt gctataacct gtcaaaagga tcagcctcct gctgaaaaag
6961 aggacccact agctaaatat aagttttggg atgttaactt acaggacagt ttttctgctg
```

HPV30

```
7021 accttgacca attcccactg ggcagaaagt ttttaatgca acttggggtt cgtactaaac
7081 cttctactac tactaaaaaa cgctcggccc ccagttcctc tacctctaca ccatcagcca
7141 aacgcaagcg gcggTGAttg tttgtgtctg tgtatgtttc cttactattt attgtgcatg
                   <- L1 end
7201 aatgtatgtt tttgtatgga tgtatgtttt gtttatatgt ttctgtatga ctgtatgtat
7261 gtgtaatgac tgtatgtatg taataaatat gaatgagtct tacttttacg tgtggttaca
7321 taaactaagg tgcggttgtg tccctaggca tttaggtagc aatttaggtg gcgtccctat
7381 gtcctccACC CTTTTTGGTt gttgcacacc actgtgcagt tacttttttat atttatatta
           -> E2 bind
7441 taccaccaca gtagtgtcca ttttatgcat ttgtgccctc cattttacct ttcaACCGAT
                                                              -> E2 bind
7501 TTCGGTttcc tggcatgtat gagtgttttt tatatgtaca tgccaaaagt acattcagca
7561 aaacacttaa tccactggca tggtgccgtt tcctgcactt tatgtttttt tttgcaacgt
7621 aagacgcctc gccttattag tcatatatgt atgctgccaa ctatgctttt atcagcatac
7681 tttactgctg ttgggcataa gtttttattg caaacatctg caacacaatg atttggcttg
7741 cagcatatat tttggtagcc aactatgtgt cttgtaaagc aggtgtgtaa aaccttactc
7801 atatcccaca ACCGGTTACG GTtttgcagc aacaagttta ttttaTAAtt at
               -> E2 bind                  E6 orf start ->
```

```
LOCUS       HPV51        7808 bp ds-DNA             VRL       30-JUL-1991
DEFINITION  Human papillomavirus type 51 (HPV-51), complete genome.
ACCESSION   M62877
SOURCE      Human papillomavirus type 51 DNA recovered from a low-grade
            precancerous cervical lesion.
REFERENCE   1  (bases 1 to 7808)
  AUTHORS   Lungu,O., Crum,C.P. and Silverstein,S.J.
  TITLE     Biologic properties and nucleotide sequence analysis of human
            papillomavirus type 51
  JOURNAL   J. Virol. 65, 4216-4225 (1991)
COMMENT     Lorincz et al. ( Obestet Gynecol 79: 328-337) classified HPV-51 as
            an "intermediate-risk" virus.  Prevalence studies indicate that HPV-51
            is present in approximately 5% of biopsies from patients with abnormal
            Papanicolaou smears.

            The 7808 bp complete genome was cloned and sequenced from a
            low-grade precancerous lesion of the cervix.  By alignment with
            HPV-16, the first nucleotide of the genome was assigned.  As is
            common to many other papillomaviruses, the E5 ORF lacks a
            methionine start codon.  This genome also includes 4 copies of a
            decameric sequence of alternating GT residues.

            The E6 ORF of HPV-51 possesses four repeats of a cysteine doublet
            motif.  These repeats are separated by 29, 36 and 29 amino acids
            respectively.  It is associated with the coordination of Zn2+ and
            is conserved among all papillomaviruses.  Unlike the E6 regions of
            those HPVs associated with benign condylomata, the E6 gene of
            HPV-51 contains a putative splice acceptor/donor pair.

            The E7 gene of HPV-51 also possesses cysteine doublets.  Two
            repeats are present with a spacer of 29 intervening amino acids.
            In HPV-18, a string of seven amino acids has been shown to be
            necessary for binding pRB-105.  These seven nucleotides are also
            present in HPV-51.  Transformation experiments have shown HPV-51 to
            have transformation potential and that the transformed cells
            express RNAs homologous to E6 and E7.

            The URR of HPV-51 possesses three copies of the E2 binding site.
            There are seven copies of NF-1, and single sites for AP1, Oct-1,
            Sp1, and TFIID.  There is one copy of a glucocorticoid receptor
            within the URR and four homologs outside the URR; three outside of
            E5, and one within L2.

BASE COUNT     2404 a   1430 c   1621 g   2353 t
ORIGIN      96 bp upstream from beginning of E6 cds
        1 aacaattatc ttgtaaaaac tagggtgtaA CCGAAAAGGG TtatgACCGA AAACGGTgca
                                  E2 bind ->                -> E2 bind
       61 TATAAAAgtg cagtggtaaa agtaTAGaag aacaccATGt tcgaagacaa gagggaaaga
   signal ->              E6 orf start -> E6 cds ->
      121 ccacgaacgc tgcatgaatt atgtgaagct ttgaacgttt ctatgcacaa tatacaggta
      181 GTGTGTGTGT attgtaaaaa ggaattatgt agagcagatg tatataatgt agcatttact
          -> repeat
      241 gaaattaaga ttgtatatag ggataataat ccatatgcag tatgcaaaca atgtttactg
      301 ttttattcaa aaattagaga gtatagacgt tatagcaggt ctgtgtatgg tactacatta
      361 gaggcaatta ctaaaaaaag cttatatgat ttatcgataa ggtgtcatag atgtcaaaga
      421 ccacttgggc ctgaagaaaa gcaaaaattg gtggacgaaa aaaaaaggtt ccatgaaaTA
      481 Gcgggacgtt ggacggggca atgcgctaat tgctggcaac gtacacgaca acgtaacgaa
    E7 orf ->
     start
      541 acccaagtgT AAtaaagccA TGcgtggtaa tgtaccacaa ttaaaagatg tagtattgca
                      E7 cds ->
                    <- E6 end
      601 tttaacacca cagactgaaa ttgacttgca atgctacgag caatttgaca gctcagagga
      661 ggaggatgaa gtagataata tgcgtgacca gctaccagaa agacgggctg gacaggctac
      721 gtgttacaga atTGAagctc cgtgttgcag gtgttcaagt gtagtacaac tggcagtgga
             E1 orf start ->
```

```
 781 aagcagtgga gacacccttc gcgttgtaca gcagatgtta atgggcgaac taagcctggt
 841 ttgcccgtgt tgtgcgaaca acTAGcaacg gcgATGgact gtgaaggtac agaggatgag
                               E1 cds ->
                            <- E7 end
 901 ggggcggggt gtaatgggtg gtttttgtt gaagcaatag tagaaaaaaa aacaggagat
 961 aatgtttcgg atgatgagga tgaaaatgca gatgatacag gatctgattt aataaacttt
1021 atagatagtg aaactagtat ttgcagtcag gcggaacagg agacagcacg ggcgttgttt
1081 caggcccaag aattacaggc aaacaaagag gctgtgcatc agttaaaacg aaagtttcta
1141 gtcagcccgc gaagcagccc attaggagac attacaaatc aaaacaacac acacagccat
1201 agtcaggcaa acgagtcaca agttaaaagg agattactgg acagttatcc ggacagcgga
1261 tatggcaata cacaagtgga aactgtggaa gcaacgttgc aggtagatgg gcaacatggc
1321 ggttcacaga acagtgtgtg tagtagcggg gggggcagtg ttatggatgt ggaaacaaca
1381 gaaagctgtg caaatgtaga actaaacagt atatgtgaag tattaaaaag cagtaatgca
1441 aaagcaacgt taatggcaaa atttaaagag ttgtatggta ttagttataa tgagttggta
1501 cgggtgttta aaagtgataa aacatgttgt atagattggg tttgtgcatt gtttggcgtt
1561 tccccaatgg tagcagaaaa tttaaaaaca ctaattaagc cattttgcat gtactaccat
1621 atacaatgtt tatcatgtga ttggggcacc attgtattaa tgctaattag gttttcatgt
1681 gcaaaaaaca gaacaacaat tgctaagtgt ttaagtacat tagtaaatat cccacaatca
1741 caaatgttta tagaaccacc aaaattacgt agtacacctg tggcattata tttttataga
1801 acaggcatat caaacattag caatacatat ggagagacac ctgaatggat tacacgacaa
1861 acgcaactac aacatagttt tgaggatagt acctttgaat tatcacaaat ggtgcaatgg
1921 gcatttgacc atgaagtatt agatgatagt gaaatagcat ttcattatgc acaattagca
1981 gatatagata gtaatgctgc agcgttttta aagagtaatt gccaagcaaa atatgtaaaa
2041 gattgtggga ccatggcacg gcattacaaa cgagcacaaa gaaaatcatt atctatgtca
2101 gcctggataa ggtatagatg tgatagagca aaggatggag gcaactggag agaaattgct
2161 aaatttttaa gatatcaagg tgtaaacttt atgtccttta ttcaaatgtt taaacagttt
2221 ttaaaaggaa caccaaaaca caattgcata gtcatatatg gcccaccaaa cacaggcaag
2281 tcattatttg caatgagcct aatgaagttt atgcaagggt ccattatttc atatgtaaac
2341 tctggtagtc atttttggtt acagccacta gaggatgcta aaatagcatt gttagatgat
2401 gctacgtatg ggtgttggac atatattgat cagtatttaa gaaacttttt agatggtaat
2461 ccatgtagta tagatagaaa acataggagt ttaatacaat tagtatgtcc accattacta
2521 ataacgtcaa acataaatcc acaagaggat gcaaacctaa tgtatttaca tacaagggta
2581 acagtattaa agtttttaaa tacatttcca tttgataaca atgggaatgc tgtgtataca
2641 ttgaatgatg aaaattggaa aaatttttt tccaccacat ggtccagatT AGatttggag
                                              E2 orf start ->
2701 gaggaagagg acaaagaaaA TGgagaccct atgccaccgt ttaaatgtgt gccaggagaa
                E2 cds ->
2761 aatactagac tgttaTGAac tggacagtga taaattagta gatcaaatta actattggac
                     <- E1 end
2821 attgttacga tatgaagctg ctatgtttta tgcagcacgg gaaagaaact tacgaacaat
2881 caatcaccag gtagtaccag caacaacagt atcaaaacaa aaggcctgtc aagcaattga
2941 aatgcacatg gccttacaat cgcttaacaa atcagactat aacatggaac catggacaat
3001 gcgggagaca tgttatgaac tatggtgtgt ggctcccaag caatgtttca aaaaggggg
3061 cataactgta acagttatat ttgatggaaa taaggacaat gcaatggact atacaagctg
3121 gaaatttata tatatatatg ataatgataa gtgggtaaag acaaatggaa atgtggacta
3181 tacgggtata tattacactg taaattcaaa aaaagaatat tatgtacagt ttaaagatga
3241 agccaaaata tatggggcac aacagtggga ggtctatatg tatggtactg taaTAAcatg
                                                  E4 orf start ->
3301 tcctgaatAT Gtatctagta cctgcagcga cgcgttatcc actactacaa ctgttgaaca
      E4 cds ->
3361 actatcaaac accccaacga ccaatcccct taccacctgc gtgggcgcca aagaagccca
3421 gacacaacag cgaaaacgac agcgacttac tgagcccgac tcctccacaa tctccccact
3481 gtccgtggac aatacaaaca accaaataca ctgtggaagt ggaagcacta acactggagg
3541 gcaccaaagt gcaactcaga ctgcgtttaT AGtgcattta aaaggtgata caaattgttt
                                     <- E4 end
3601 aaaatgtttt agatacagat ttacaaaaca caaagggtta tataaaaacg tatcctcaac
3661 ctggcattgg accagtaata ctaaaacagg cattgttacc attgtgtttg acagtgcaca
3721 tcaacgggaa acatttataa aaaccattaa agtaccccca agtgtaacac tgtcattggg
3781 aattatgaca ctgTAActag tgtaatatat gtattgtACA TATATACTGT CACaagccaa
                  <- E2 end
             glucocorticoid response element ->
3841 tatgtgctgc TAAttgtata gacatattgt aaccattgca gtgtttatta ttttgctatt
      E5 orf start ->
      NH2 terminus unknown
```

```
  3901 tgtgctttgc ttGTGTGTGT GTcttgtgtt gtgttgtttg ttgccgctac tgctgtccca
         GT repeat ->
  3961 atacgtgttt gcagctgcct tattattaat tttatgtttt tggtttgttg ttgcaacatc
  4021 ccaattaACT ACATTTTTTG Tatatttgat tttttttac ttaccttgtt tacttttaca
             -> glucocorticoid response element
  4081 tctatataca tttttacttt tgcAATAAAc ttgttatatt tttgTGAtta aatATGgtgg
                         signal ->           L2 orf start ->
                                    <- E5 end              L2 cds ->
  4141 ctacacgtgc acggcgtcgg aagcgagcat ctgtaacaca attatattct acatgcaaag
  4201 ctgctggtac atgtcctcct gatgttgtga ataaggttga aggtactaca ttggccgata
  4261 aaatattaca gtggagtggg ttgggtatat ttttgggtgg cctaggtatt ggtactgggt
  4321 ctggatctgg ggggcgtact ggatatatcc ctttaggtgg tgggggtcgc ccaggcgtgg
  4381 tggatattgc tcctgcaagg ccacctatta taattgacct atggcaccat actgaacctt
  4441 ctatagtaaa tttggttgag gactctagta ttattcagtc tgggtctcct atacctACCT
                                                            E2 bind ->
  4501 TTACTGGTac cgatggcttt gaaattactt catcttccac aacaacccct gctgtgttgg
  4561 acatcacccc atctgctggt actgtacatg tttctagtac taacattgaa aatcctttat
  4621 atattgaacc tccatccatt gaggctccac aatctggaga agtgtcagat atatatttac
  4681 tagtacacta ctctggtact catgggtatg aagaaatacc tatggaagtg tttgcatcca
  4741 atgtcagtac tggtactgaa cctattagca gcacacctac tccaggggtt agtcgcatag
  4801 ctgctccccg cttgtatagt aagtcctaca cacaggttaa agttacaaat cctgatttta
  4861 ttagtaagcc atccacattt gttacattta ataatcctgc ttttgagcct attgacacat
  4921 ccataacttt tgaggaacct gatgctgttg cacctgatcc tgattttctg gatattatta
  4981 cactgcaccg ccctgccctt acatctcgta gaggcacagt acgctttagt aggttaggtc
  5041 aaaaggccac catgcgcact cgtagtggca aacaaattgg tgctcgTGTA CATTATTATC
                         glucocorticoid response element ->
  5101 ATgatattag tagaattgca ccagctgatg aacttgaaat gcagccttta ctttcacctt
  5161 ctaataatta tagttatgac atttatgctg atttagatga agctgaaaca ggttttatac
  5221 agcccacaca caccacacct atgtcacact cctctttgtc taggcagttg ccctccttat
  5281 cttcatctat gtcttcatct tatgcaaatg ttactattcc attttcaact acatattctg
  5341 ttcctattca tacagggcct gatgtggtat tgcccacatc tcctacagta tggccttatg
  5401 ttccccacac ttccattgac accaagcatt ctattgttat acTAGgtggg gattactatt
                                          L1 orf start ->
  5461 tgtggcccta tacacattta ctacgcaaac gccgtaaacg tataccctat ttttttacag
  5521 ATGgcattgt ggcgcacTAA tgacagcaag gtgtatttgc cacctgcacc tgtgtctcga
L1 cds ->                <- L2 end
  5581 attgtgaata cagaagaata tatcacacgc accggcatat attactatgc aggcagttcc
  5641 agactaataa cattaggaca tccctatttt ccaataccta aaacctcaac gcgtgctgct
  5701 attcctaaag tatctgcatt tcaatacagg gtatttaggg tacagttacc agatcctaac
  5761 aagtttggac tcccggatcc aaatttatat aatccagaca cagataggtt ggtgtggggt
  5821 tgtgtgggcg ttgaggtggg cagaggacag ccccttggtg ttggccttag tggtcatccc
  5881 ttatttaata aatatgatga cacagaaaat tcacgcatag caaatggcaa tgcacaacaa
  5941 gatgttagag ataacacatc tgttgacaac aaacagactc agttatgtat aataggctgt
  6001 gctccaccta ttggggaaca ctggggtatt ggcactacat gcaaaaacac acctgtacct
  6061 ccaggagact gccccccccc ggaacttgta tcctctgtca ttcaggatgg cgatatgatt
  6121 gatacagggt ttggagctat ggatttcgct gccctacagg ccaccaaatc agacgtccct
  6181 ttggatattt cacagtctgt ttgtaaatat cctgattatt taaaaatgtc tgcagacaca
  6241 tatggtaatt ccatgttttt tcatttacgc agggagcaaa tctttgctag gcactattat
  6301 aataaacttg taggtgttgg ggaagacatt cctaacgatt attatattaa gggtagtggt
  6361 aatggccgtg accctataga aagttatata tactctgcta ctcccagtgg gtctatgata
  6421 acatctgatt ctcaaatttt taataagcct tattggctcc accgtgcgca gggtcacaat
  6481 aatggcattt gctggaacaa tcagcttttt attacctgtg ttgatactac cagaagtaca
  6541 aatttaacta ttagcactgc cactgctgcg gtttccccaa catttactcc aagtaacttt
  6601 aagcaatata ttaggcatgg ggaagagtat gaattgcaat ttatttttca attatgtaaa
  6661 attactttaa ctacagaggt aatggcttat ttacacacaa tggatcctac cattcttgaa
  6721 cagtggaatt ttggattaac attacctccg tctgctagtt tggaggatgc atataggttt
  6781 gttagaaatg cagctactag ctgtcaaaag gacacccctc cacaggctaa gccagatcct
  6841 ttggccaaat ataaattttg ggatgttgat ttaaaggaac gattttcttt agatttagac
  6901 caatttgcat tgggtcgcaa gtttttgttg caggttggcg tacaacgcaa gcccagacca
  6961 ggccttaaac gcccggcctc atcggcatcc tcttcctctt cctcttcagc caaacgtaaa
  7021 cgtgttaaaa agTAAtgtat gttagttttt gtatgcttgt gcacactgtt gtatgcctgt
                    <- L1 end
  7081 atgtatatgt ttgtgtatgt actgtatgtg tttttGTGTG TGTGTGTGTt gttgttcctg
                                    GT repeat ->
```

```
   7141 tatgtatgag ttatgtatgt ttattattAA TAAActatgt ggtgtGTGTG TGTGTGTttt
                                signal ->       GT repeat ->
   7201 tgcatgactg catttgtatg acatgtacgg gtgtatgtgg gtattacatt atccccgtag
   7261 gtcaagggtg gtgtttcggt ggcgtcccta ttgccctacc catttttgc agcacaacag
   7321 tttatatttg tgctatttag ttatactttg tagcttccat tttgttacag ctgcagccat
   7381 tttgagtgca ACCGATTTCG GTTcgtgtac ttttagtata tttgccaagt tttaaaccac
           E2 bind ->         <-
 keratinocyte-dependent ->     <-
                enhancer
   7441 aactgccagt tgttTTTGGC ATaaaccatc atttttttat gacatagtgc atacatccgc
keratinocyte-dependent ->     <-
               enhancer
               NF-1 bind ->   <-
   7501 ccgcccacgc cttgtacttg gcgcgcctta ccggcgctag tcatacaacc tattagtcat
   7561 ttgtacttta acaattgTTG GCAcactgtt ttccgcccta taataattta actgcttata
                 NF-1 bind ->
   7621 ggcatgtatt ttTTGGCAta ttttatctta ctAATTGCAT agttggcagg tcaaatacta
           NF-1 bind ->        Oct-1 bind  ->
   7681 tgtttttagt gccaagtttc tatcctactt ataaaccatc ttactcatat gcaggtgtgc
   7741 tacacaaatg tgttacctaa ccgaTTTGTG TTctgcctat gcttgcaaca ttttttctta
  keratinocyte-dependent enhancer ->
   7801 taacattt
```

```
LOCUS       HPV53        7856 bp ds-DNA             VRL       04-OCT-1993
DEFINITION  Human papillomavirus type 53 (HPV-53), complete genome.
ACCESSION   X74482
SOURCE      Human papillomavirus type 53 DNA.
REFERENCE   1  (bases 1 to 7856)
  AUTHORS   Delius,H. and Hofmann,B.
  TITLE     Primer-directed sequencing of human papillomavirus types
  JOURNAL   Curr. Top. Microbiol. Immunol. 186, 13-31 (1994)
REFERENCE   2  (bases 1 to 7856)
  AUTHORS   Delius,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1993) to the EMBL/GenBank/DDBJ databases. H.
            Delius, Deutsches Krebsforschungszentrum, Abteilung ATV, Im
            Neuenheimer Feld 506, W 6900 Heidelberg, FRG
COMMENT     HPV-53, originally isolated from a normal cervix, is rarely found
            in the genital tract.  189 asymptomatic women were examined for
            infection with HPV-53.  Only six positive samples were identified
            (Gallahan et al. J Virol 63: 4911-2).

            The HPV53 E1 open reading frame is fragmented. Three probable points
            of deletion within E1 have been identified by aligning the E1 gene of
            HPV53 with E1 genes of related HPV types.
BASE COUNT     2393 a   1487 c   1666 g   2310 t
ORIGIN      101 bp upstream from beginning of E6 cds
        1 gaaagTAAca atcctacttt tatagacagg gagtaACCGA AATAGGTtta ggACCGAAAA
     E6 orf start ->                                -> E2 bind         -> E2 bind
       61 CGGTacatat aaaagcactg tgtacaacac ccaggacatc cATGgatcgt cagttatttg
                                                  E6 cds ->
      121 aaaatacaga agagcgacca cgtacattgc accagctatg tgaagttgtg aataaaccat
      181 tgctggagct gcaacttggc tgtgtgttct gcaagaaggc attgacagcg tcagaggtat
      241 ataattttgc atatacagat ctaagagtag tgtatagaga cgggtatccg tatggagtgt
      301 gcaaattctg tttgctattt tatagtaagg tccgaaaatt aagatattac aattgttcag
      361 tgtacggggc tagcctggaa gcactaacta aaaaaaagtt atctgattta tcaataaggt
      421 gctacagatg tcaacatccg ttgacaccag aggaaaaaca gttacactgt gactataaga
      481 aacggtttca caaaatttca catatgtgga ccgggtcgtg cctgacatgc tggagacaca
      541 caacagcaac TGAatcagca gtaTAAtcAT Gcacggtaac gtaccaacac ttccacaata
           E7 orf start ->       E7 cds ->
                                    <- E6 end
      601 tattatagaa cttataccac aaactgagat tgacctgcaa tgccatgagc aattgaacag
      661 ctcagaggat gaggatgagg atgaagtaga ccatctgcag gagcagccac agcaagctag
      721 acgggacgaa caacatcctt gttacctaat tgaaacacag tgttgtaggt gTGAgtcgtt
                                                    E1 orf start ->
      781 ggtgcagttg gctgttcaga gttcaacaaa agagctgcgt attttacaac aaatgcttat
      841 gggcacagtg gagcttgtgt gccccctctg cgcaacaagg cgaTAActgc aATGgcgtca
                                                        E1 cds ->
                                                      <- E7 end
      901 cctgaaggta cagatgatga ggggggatgc cggggatggt ttcacgtgga ggcaatagta
      961 aaaaaacgta caggggatgt aatatctgaa gatgaaacag acgaggaaag caccgaatct
     1021 gatttggatg ggtttataga caatagtaat ataatatcta cacaggcaga aagggagaca
     1081 gctcagcagt tgttacatgc caaaacacac atgcagatac acagacgctg cagaaatTAA
                                  /\ deletion                            <-
                                     causing premature                premature
                                     termination of E1 cds          termination
                                                                      of E1 cds
     1141 aacgaaagta tttaggcagt ccattaggtg atattagtaa tcagcaaaca gtgtgccggg
     1201 aagctgtaaa acgaaggcta attgatacag aagtgccgga cagcgggtat ggcaatactt
     1261 tggaaacgtg gaagcaacgc aacaggtaca ggaacagtat gTAAgggagg caagtggcgg
               /\ deletion                              <- premature
                  causing premature                        termination
                  termination of E1 cds                    of E1 cds
     1321 gaaagagcaa aatggaggct cacaacatag tgtgtgtagc agagatggga gcatagggtc
     1381 aggaagtgat atggatgtgg atagacagga tataatgcca ctgcaacaaa tacaagatat
     1441 attaaaatgt agtaatgtgc aggccaagct atattgtaaa tttaaagata tatttggcat
     1501 accattttca gagttggtac gaacatttaa aagtgatagt acatgttgtc atgattggat
     1561 atgtgctata tttggagtaa atgaacatTA Gcagaggcat taaaaacaat tattaaaaca
```

```
                    deletion /\    <- premature
            causing premature          termination
            termination of E1 cds      of E1 cds
1621 cagtgcatat attaccatat gcaatgttta acatgtacat ggggcgtggt tatattattg
1681 ttaattagat atacgtgtgg caaaaacaga aaaacaattg taaagtcttt agcatccata
1741 ttaaatgtac caactgaaca aatgcttgtg cagcctccaa aaatacgtag ccctgcggtt
1801 gcattatact tttataaaac atctatatcc aacattagtg acgtgtatgg gagtacacca
1861 gaatggatag aaagacaaac acagttacaa catagctttg aggactgtca atttgaacta
1921 tctaaaatgg tgcagtgggc atttgacaat gaagtaaccg atgatagtca aatagcattt
1981 cattatgcac agttagcaga tgtagacagt aatgcacaag cattttaaa aagtaatatg
2041 caagcaaaat atgtaaaaga ctgtggaata atgtgtagac actataaacg ggcacaacaa
2101 caacagatga atatgaaaca atggataaag catgtatgta gcaaggtgga tgatggtggg
2161 gactggaagc caatagtaca gttcctaagg tatcaagggg tcgagtttat ttcatttta
2221 agttatttta aattattttt acaaggtaca ccaaaacata attgtttagt tatttatggt
2281 cctccaaaca cgggtaaatc atgttttgct atgagtctta taaacttttt tcatgggtca
2341 gtcatttcat atgtaaattc acacagtcac ttttggttac aaccgttaga caatactaaa
2401 ttaggtatgc tagatgatgc cacagaggca tgttggaaat atattgatga atacctaaga
2461 aatttattag atggcaatcc tgtaagctta gatagaaaac ataagcaact agtacaaata
2521 aagtgtccac ctgtactaat aacaactaat ataaatccaa tgcaagatgc aaagctacgg
2581 tatttgcaca gtagaataca tgtgttacag tttttaaatc catttccaat tgatgtgaat
2641 ggaaatcctg tataccaatt aaataatgca aactggaaat gttttttga aaggacatgg
2701 tccagatTAG atttggataa cgacgaagac aaggaaaacg ATGgagacgc tatgccaacg
   E2 orf start ->                       E2 cds ->
2761 tttagatgcg tgccaggaga aaattctaga ctgtttTGAa agagatagca aaaatattac
                                            <- probable E1 end
2821 agaccatata gactactgga aagctgtgcg acaagaaaat gtaatatatt ataaggcaag
2881 agaaaataat atgactaaac taggccacca ggtggtgccc tgtttacaag tgtgtaaagc
2941 aaaggcatgc gttgctatag aactacaaat agcattggaa tcactttgta aaacagaata
3001 taatatggaa gagtggacat taagggatgt atgtgaaagt atgtggtata cagaacccaa
3061 acagtgtttt aaaaaacaag gacaacatat agaagtgtgg tttgatggca gcaaggacaa
3121 tcgggctgaa tatgttgtgt ggaaatgggt atattattgt ggggaggatg ggtggtgtaa
3181 agtgtcttct gcagTAAgct ATGagggcat atattatata catgacggcc ataaaacgta
                  E4 cds ->
         E4 orf start ->
3241 ttatacaaac tttaaagacg aggccaccaa atatgggtgt aaaggcacat gggaagtgca
3301 tatgggaaaa caaagtattt attgtcccga ctctgtgtcc agtaccttta gatccaacgt
3361 atcctctgtt gaaactgtta acgaatacta ctcccataag acccccacca cctccacgcc
3421 cgtgggcacc tacgaagcct catcatccct gcggccggga aaacgtccca gaaccacaga
3481 gcccgacagt actgactcca ccacacagtc cactaccaca gccagagagt cctacgcaga
3541 gtgtgtcgca aggaacacag acaacacaaa caacaacacc agaaaacacc tccctggtgg
3601 agcttcgtgt aacaacaccg aaatcgacag tggttataag actgcacctg TAGtacatat
                                                          <- E4 end
3661 aaaaggtgaa gcaaacagac taaagtgttt aaggtatcgg tttcaaaaac ataaacaact
3721 gtttgttact gtatcatcca cctatcattg gacaaatgta aactgtgctg taaataatag
3781 ttatattact gttgtatata aagatgaaac ccaacgccaa aaatttttgg atattgtaaa
3841 aatacccccc agtgtatcac tggtactggg acatatgaca tgtgttgata tgTAAtgtta
                                                             <- E2 end
3901 atgtaattgt tacattatgt aatattgtat tacaataata cctggtatat acaatactac
3961 acacgccata tatgctgcta ggcctgtaaa cttgcaacca ttgtgttgtg tgttagtttt
4021 gtgtttttgc tttgctgctt tgtgtgcctg tttgtctgct ttaatgtgcc gctattgcag
4081 tctgcctctg tgtttgcctc gtgtatatta cttatacttt tattttggtt tgttgtggca
4141 tcatcctaca taaccacatt tacactatat atacttttt tttacttccc tttattatgt
4201 ttctatagcc atgctgtgtg gtTAAttaac acacaataaa tcacagtttt tttgtattgt
                  L2 orf start ->
4261 atgtatttgt gtgttatata taATGgttgc ccaccgggca cgcagacgta agcgggcatc
                    L2 cds ->
4321 agcaacgcag ctatatcaaa catgtaaaca atctggcaca tgtcctgagg atgttattaa
4381 caaaattgaa cacaaaacct gggctgataa aatattgcaa tggggtagtt tatttacatt
4441 ttttggaggc cttggtattg gcactggctc tggcactggg gggcgtacag gatatattcc
4501 tttgggtact aggccttcca ctgttgttga tgtaacccct gctcggcccc ctattgttgt
4561 agaatcggta ggccccactg acccttcaat tgttacattg gttgaagaat ccagtgttat
4621 tgagtctgga gcttcgttcc ctaattttac aggtactgca ggatttgaag taacatcctc
4681 ctcaactacc acacctgcag tattagatat aacacccacc tctacctctg tacatgttag
4741 ttctactaca tatactaatc ccacatttgt agaccctcct gttattgagg tcccccaaac
```

```
4801 tggtgaggta tctgggaata tattgataag taccccaaca tctggggttc atagttatga
4861 ggaaatacct atgcaaacat ttgctgttca gggtacaggc aatgaaccta tcagtagtac
4921 tcctattcct ggtcttaggc gtatagctgc cccacgttta tataaaaaag cctttcaaca
4981 ggttaaggtc acagatcctg catttcttca caaacctgaa actttaatta atgttgataa
5041 tcccatattt caaaatgctg acacaacgtt aacctttca ccatcaggtg tggcacctga
5101 tcctgatttt ttagacattg tggctttaca cagacctgcc tttacaacac gtagaggtgg
5161 tgtgcgtttt agtaggcttg gcacaaaggc cactatgcgc accagaagtg gcaaacaaat
5221 aggtgctcgt gtacattatt attatgatgt gagtcctatt acacagactg aagaaattga
5281 aatgcagcca ttattgtcta cagataatac atttgatggc ctatatgata tttatgcaaa
5341 catagatgat gaggcacctg tgtcatctcg tttttctatt gctacACCTT CTAGGTTgcc
                                                  -> E2 bind
5401 tACCAACACT GTTccttttgt cttttagtgg tagtACCTCT AATGTTacaa tacccttttgg
      -> E2 bind                          -> E2 bind
5461 tacatcttgg gatgttccca tctattcagg tcccgatgta gtgttgccca caggaccccc
5521 tacatggcct tatgcacctc aatctccttt tgataccacc catgatgtag ttatacaggg
5581 ctccacattt gcgttatggc ctgtttattt ttTGAaacgt aggcgtcgta aacgtattcc
                            L1 orf start ->
5641 ctattttctt gcagATGgcg gtgtggcggc cTAGtgacag caaggtttat ctgcctccta
            L1 cds ->                  <- L2 end
5701 cccctgtatc aaaggtcatt acaacggatg cctatgtaaa acgcaccact atattttatc
5761 atgctggaag ctctcgcttg cttaccgtgg gacatcctta ttaccccatt tctaaatctg
5821 gtaaagcaga catccctaag gtgtctgcat ttcagtatag ggtgtttaga gtacgcctcc
5881 cagatcctaa taagtttggc cttcctgata caaatatatt taatcctgac caggagcggc
5941 ttgtatgggc ctgtgtaggg ctagagattg gccgtggaca gcctttaggt gttggtgtaa
6001 gtggccatcc actgttcaat aggttggatg acactgaaag ctctagtata gctattcagg
6061 atactgcccc ggacagtagg gataatgttt cagtggatcc taaacaaaca cagttatgca
6121 ttattggctg tgcacccgct ataggtgaac actggacaaa gggtacggcc tgtcgttcca
6181 cacctactac agcgggcgac tgccccccat tggaacttat caattcacct attgaggatg
6241 gagacatggt ggacacaggt tttggtgcat taaactttaa agctttgcag gaatctaaat
6301 ctgatgtgcc attggatatt gtacaatcca catgtaaata tcccgattat ttgaaaatga
6361 gtgcagatgc ctatggggat tctatgtggt tttatttacg tagggaacag ttgtttacca
6421 gacatttttt taatagggca ggcgttattg gtgaggaaat acctaatgac ttatatatta
6481 agggtagtaa tggcagggac ccgcccccta gctctgtata tgttgctaca cctagtgggt
6541 ctatgataac ttcagaggct caattgttta ataagccata ttggctgcaa cgtgcccagg
6601 gacataataa tggcatctgt tggaacaatc agttatttgt aactgttgtg gataccacca
6661 ggaatacaaa catgactctt tccgcaacca cacagtctat gtctacatat aattcaaagc
6721 aaattaaaca gtatgttaga catgcagagg aatatgaatt acaatttgtg tttcaactat
6781 gtaaaatatc cctgtctgct gaggttatgg cctatttaca tactatgaat tctaccttac
6841 tggaagactg gaatataggt ttgtcgcctc ctgttgccac tagcttagag gacaaataca
6901 gatatgtgaa aagtgcagct ataacctgtc aaaaggatca gccccctcct gaaaagcagg
6961 acccactatc taaatataaa ttttgggagg tcaatttgca aaacagtttt tctgctgatt
7021 tggatcagtt tcctcttggc aggaagtttt taatgcaggt tggggtccgt actaaaccgc
7081 ctgtatcctc taaaaaacgc tctgcttcta ctacatctac ctcagccccc tcctccaagc
7141 gcaaacgcaa aTAGtatgtt gtgtgtctgt attgtgtatg tatggttgta ttatgtacta
                 <- L1 end
7201 tttatttatt gtgtatactg tatgtgtgta tttgttgtct gtgttgtatg actgtatgta
7261 tgtgtaatgt ttgcatgttg ttaataaata tgaatgagtg ttacttttac gcgtggttgc
7321 ataaactaag gcgcggtggt gtccctaggc agttgggtgg catgttaggt ggcgtccctg
7381 tttatatacc aatctccacc ctgtattatt aatatatgta ctgttatcat gcagttacct
7441 ataccacata tagtatagct tataggtatc cattttagtt tggcgccttt ttggtaccct
7501 ccattttgcc ttgcaACCGG TTTCGGTtgt gcatatcttg taaataccac aagcatattc
                    -> E2 bind
7561 agcagaactg ttaatccttt ggcatagtgc cgtttcctgt gtttaatgtt tttttgtttt
7621 atacactatg ccttacctat tagtcactta ctgtaggctg ccaactatgc ttttacctgc
7681 ataccttagt gcttttgggc acacattttt attgcaaaca tctgcacttt agtgttttgg
7741 cttgcagcat acgttctgct agccaagtat ctgtctagta aaccaggtgt gcacctatta
7801 ctcatgcACC GCAACCGTTT ACGGTtttgc agcaacaggc ctttttttata attatt
                 -> E2 bind
           -> E2 bind
```

```
LOCUS       HPV56        7844 bp ds-DNA             VRL       04-OCT-1993
DEFINITION  Human papillomavirus type 56 (HPV-56), complete genome.
ACCESSION   X74483
SOURCE      Human papillomavirus type 56 DNA.
REFERENCE   1  (bases 1 to 7844)
  AUTHORS   Delius,H. and Hofmann,B.
  TITLE     Primer-directed sequencing of human papillomavirus types
  JOURNAL   Curr. Top. Microbiol. Immunol. 186, 13-31 (1994)
REFERENCE   2  (bases 1 to 7844)
  AUTHORS   Delius,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1993) to the EMBL/GenBank/DDBJ databases. H.
            Delius, Deutsches Krebsforschungszentrum, Abteilung ATV, Im
            Neuenheimer Feld 506, W 6900 Heidelberg, FRG
COMMENT     HPV-56 primarily causes anogenital lesions of differing severity.
            Lorincz et al. (Obestet Gynecol 79: 328-337) classified HPV-56 as a
            "high-risk" virus.  However, conflicting data reported by Bergeron
            et al.  ( Am J Surg Pathol 16: 641-649) placed HPV-56 in the
            "intermediate-risk" category .

            The HPV56 E1 open reading frame is fragmented.

BASE COUNT     2529 a   1335 c   1639 g   2341 t
ORIGIN      101 bp upstream from beginning of E6 cds
        1 GAaagtttca atcatacttt tatatattgg gagtgACCGA AAAGGGTtta agACCGAAAA
         -> E6 orf start (bp 7844)              -> E2 bind         -> E2 bind
       61 CGGTacatat aaaaggcagc ttattctgtg tggacatatc cATGgagcca caattcaaca
                                              E6 cds ->
      121 atccacagga acgtccacga agcctgcacc acttgagtga ggtattagaa atacctttaa
      181 ttgatcttag attatcatgt gtatattgca aaaaagaact aacacgtgct gaggtatata
      241 attttgcatg cactgaatta aaattagtgt atagggatga ttttccttat gcagtgtgca
      301 gagtatgttt attgttttat agtaaagtta gaaaatatag gtattatgac tattcagtgt
      361 atggagctac actagaaagt ataactaaaa aacagttatg tgatttatta ataaggtgct
      421 acagatgtca aagtccgtta actccggagg aaaagcaatt gcattgtgac agaaaaagac
      481 gatttcatct aatagcacat ggttggaccg ggtcatgttt ggggtgctgg agacaaacat
      541 ctagagaacc TAGagaatct acagtaTAAt cATGcatggt aaagtaccaa cgctgcaaga
           E7 orf start ->           E7 cds ->
                                       <- E6 end
      601 cgttgtatta gaactaacac ctcaaacaga aattgaccta cagtgcaatg agcaattgga
      661 cagctcagag gatgaggatg aggatgaagt agaccatttg caggagcggc cacagcaagc
      721 tagacaagct aaacaacata cgtgttacct aatacacgta ccttgttgtg agtgtaagtt
      781 tgtggtgcag ttggacattc agagtaccaa agaggacctg cgtgttgtac aacagctgct
      841 tatgggtgcg ttaacagtaa cgtgcccact ctgcgcatca agTAAcTAAc tgcaATGgcg
                                             E1 orf start -> E1 cds ->
                                                           <- E7 end
      901 tcacctgaag gtacagatgg ggaggggaag ggatgttgtg gatggtttga agtagaggca
      961 attgtagaaa aaaaaacagg agataaaata tcagatgatg aaagtgacga ggaggatgaa
     1021 atagatacag atttagatgg atttatagac gattcatata tacaaaatat acaggcagac
     1081 gcagaaacag tcaacaattg ttgcaagtac aaacagcaca tgcagataaa cagacgttgc
                                           /\ deletion causing
                                              premature termination of E1 cds
     1141 aaaaacTAAa acgaaagtat atagctagtc cattaaggga tattagtaat cagcaaactg
                <- premature termination of E1 cds
     1201 tgtgccggga aggagtaaaa cggaggctta ttttatcaga cctacaagac agcgggtatg
```

```
1261 gcaatacatt ggaaactctg gaaacACCAG AACAGGTaga tgaagaggta cagggacgtg
                                -> E2 bind <-
1321 ggtgcgggaa tacacaaaat ggaggctcac aaaacagtac ctatagtaac aatagtgagg
1381 actctgtaat acatatggat attgatagaa acaatgaaac gccaacacaa caattgcagg
1441 acttgtttaa aagtagcaat ttacaaggta aattatatta taaatttaaa gaagtgtatg
1501 gtattccatt ttcagaattg gtgcgtacgt ttaaaagtga tagtacatgt tgcaatgatt
1561 ggatatgtgc tatatttggt gttaatgaaa cattagccga ggcactaaaa actataataa
1621 aaccacactg tatgtattat catatgcaat gtttaacatg tacatggggg gttatagtaa
1681 tgatgctaat tagatataca tgtggcaaaa acagaaaaac aattgcaaaa gcattaagct
1741 caatattaaa tgtaccacag gagcaaatgt taattcaacc accaaaaata cgaagtcctg
1801 ctgtagcttt atatttttat aaaacagcaa tgtcaaatat tagtgatgtg tatggagaca
1861 caccagaatg gatacaaaga caaacacaat tgcaacacag tttacaggat agtcaatttg
1921 aattatctaa aatggtgcag tgggcatttg ataatgaagt aacagatgat agccaaattg
1981 cgtttcaata tgcacaatta gcagatgtag acagcaatgc acaagccttt ttaaaaagca
2041 atatgcaggc aaaatatgta aaggattgtg gaataatgtg tagacattat aaaagggcac
2101 aacagcaaca aatgaatatg tgccagtgga taaagcacat atgtagtaaa acagatgaag
2161 ggggtgattg gaaacccatt gtacaatttt taagatatca aggggtcgat ttcatttcat
2221 ttctaagtta ctttaaatta tttctacaag gaacacctaa acataactgt ttggtacttt
2281 gtggaccgcc aaatacaggt aaatcatgct ttgctatgag tcttataaag ttttttcaag
2341 ggtctgtcat ttcatttgtg aattcacaaa gccacttttg gttgcagcca ttagacaatg
2401 ctaaacttgg gttgttggat gatgcaacag aaatatgttg gaaatatata gacgattatt
2461 taaggaattt ggtagatgga aatcctataa gtttagatag aaaacataaa caattagtac
2521 aaataaaatg tccaccatta ctaattacaa ccaatataaa tcctatgcta gatgctaaat
2581 tacgatattt acacagtaga atgttagtgt ttcagtttca aaatccattt ccattagata
2641 ataatggtaa tcctgtatat gaattaagta atgtaaactg gaaatgtttc tttacaagga
2701 cgtggtccag atTAAatttg gataacgacg aggacaaaga aaacaATGga gacgctttcc
       E2 orf start ->                       E2 cds ->
2761 caacgtttaa atgcgtgcca gaacaaaata ctagactgtt tTGAaaaaag aTAGtagatg
                                probable E1 end <-   |     <- premature
                                                     |        termination
                                           insertion        of E2 cds
                                      causing premature
                                   termination of E2 cds
2821 tattgcagat catatagaat attggaaagc tgtgcgacat gaaaatgtgc tatactataa
2881 agcaagagaa aatgacatta ctgtactaaa ccaccagatg gtgccttgtt tacaagtatg
2941 taaagcaaaa gcatgtagtg caatagaagt gcaaatagca ctggaatcat taagtacaac
3001 aatatataac aatgaagagt ggacattaag agacacatgc gaggaactat ggcttactga
3061 acctaaaaaa tgctttaaaa aagaaggaca acatatagaa gtatggtttg atggtagtaa
3121 aaacaattgt atgcaatatg tagcctggaa atatatatat tacaatggag attgtgggtg
3181 gcaaaaagtg tgttctgggg TAGactatag aggtatatat tATGtacatg atggccacaa
               E4 orf start ->           E4 cds ->
3241 aacatactac acagactttg aacaagaggc caaaaaattt gggtgtaaaa acatatggga
3301 agtacatatg gaaaatgaga gtatttattg tcctgactct gtgtctagta cctgtagata
3361 caacgtatcc cctgttgaaa ctgttaacga atacaacacc cacaagacca ccaccaccac
3421 ctccacgtcc gtgggcaacc aagacgccgc agtatcccac agaccaggaa aacgacccag
3481 actacgggaa tcagaatttg actcctccag agagtcccac gcaaagtgtg tcacaacaca
3541 cacacacatc agcgacacag acaataccga cagtagaagt agaagtatca acaacaacaa
3601 ccaccctggt gataagacta cgcctgTAGt acatttaaaa ggtgaaccta acagattaaa
                                <- E4 end
3661 atgttgtaga tatcgatttc aaaaatataa aacattgttt gtggatgtaa catcaacata
3721 tcattggaca agtacagaca ataaaaatta tagcataatt acaattatat ataaggatga
3781 aacacaacga aacagctttt taagtcatgt aaaaattcca gtagtgtaca ggttagtttg
```

```
3841 ggacaaaTGA gttttccata aagtgctgta tatattgtat atacatttgt gttattgtaa
            <- probable E2 end
3901 cacacaaata cgtgaagtgt acctgccata cattgctgct acgcatatat attgcaacca
3961 ttgatttttg tgttattggt gtgtttgcgc tttgcttttg tgtttgtttg cttgtgtgtc
4021 atgttgtccc gcttttgcta tctgcctctg tgttttccag ttgtatatta ttaataatat
4081 tgttttggtt tgttatagcc acatcctttt ttaatacatt tataatattt ttgatatttt
4141 tttactgtcc tgtgctgtgt atatatttac atgctttgtg gataaTAAat aatatgtaaa
                                          L2 orf start ->
4201 tgtagtagta ctgttactac tATGgttgcc caccgtgcca cacgacgcaa acgcgcatct
                     L2 cds ->
4261 gcaacacaac tatataaaac atgtaagttg tctggtacat gtccagagga tgttgttaat
4321 aaaatagagc aaaaaacatg ggctgataaa atattgcaat ggggaagttt atttacatat
4381 tttggaggcc ttggcattgg tacaggaact gggtctgggg gtcgtgcagg ctatgttcca
4441 ttggggtcta ggccttccac aatagttgat gtaactccgg cgcgaccacc tattgttgtg
4501 gaatccgtag ggcctacaga cccttccatt gttacattag ttgaggagtc cagtgttata
4561 gaatctggtg cagggattcc taattttact gggtctgggg gatttgaaat tacatcctca
4621 tcaacaacta cACCTGCCGT GTTggatatt acaccaacct ctagtactgt acatgtcagt
                -> E2 bind
4681 agtacccata taaccaatcc gttatttatt gatccccctg ttattgaggc cccacaaaca
4741 ggcgaggtgt ctggcaatat tttaattagc acacccacat ctggtataca tagctatgaa
4801 gaaataccta tgcaaacatt tgctgttcac ggttctggta cagaacctat tagtagtact
4861 cctattccag gctttaggcg tattgcagct cctagattat atagaaaagc atttcagcag
4921 gttaaggtaa ctgaccctgc atttcttgat agacctgcaa cattagtatc tgctgataat
4981 ccacttttg aaggtactga cacatcttta gctttttctc cgtcgggtgt ggctcctgac
5041 cctgatttta tgaatatagt agcattacat aggcctgcat ttactacacg taggggtggt
5101 gtacgtttta gtaggcttgg cagaaaggct actatacaaa cacgtagagg cacacaaata
5161 ggtgcccgtg tgcattatta ttatgatata agtcctattg cacaggctga ggaaattgaa
5221 atgcagccat tattgtctgc aaataattca tttgatggcc tatatgatat ttatgcaaat
5281 atagatgatg aagcacctgg tttgtctagc cagtcagttg ctacaccttc tgcacactta
5341 cctataaagc cttccacatt gtcttttgct agtaacacca ctaatgtaac tgccccttta
5401 ggtaatgtgt gggaaacacc attttattca ggtcctgaca TAGtgttgcc tacaggcccc
                                     L1 orf start ->
5461 agtacgtggc cctttgttcc tcagtctcct tATGatgtta cccatgatgt atatatacag
                               L1 cds ->
5521 ggatcctcct ttgcattatg gcctgtgtat ttttttagac gtaggcgccg taaacgtatt
5581 ccctattttt ttgcagatgg cgacgtggcg gccTAGtgaa aataaggtgt atctacctcc
                                        <- L2 end
5641 aacacctgtt tcaaaggttg tggcaacgga ttcctatgta aaacgcacta gtatatttta
5701 tcatgcaggc agttcacgat tgcttgccgt aggacatccc tattactctg tgactaagga
5761 caataccaaa acaaacattc ccaaagttag tgcatatcaa tatagggtat ttagggtacg
5821 gttgcccgac cctaataagt ttgggcttcc agatactaat atttataatc cggACCAGGA
                                                              -> E2 bind
5881 ACGGTtagtg tgggcatgtg taggtttgga ggtaggccgc ggacagcctt taggtgctgg
5941 gctaagtggc catccattgt ttaataggct ggatgatact gaaagttcca atttagcaaa
6001 taataatgtt atagaagata gtagggacaa tatatcagtt gatggcaagc aaacacagtt
6061 gtgtattgtt ggatgtactc ccgctatggg tgaacattgg actaaaggtg ctgtgtgtaa
6121 gtccacacaa gttaccacag gggactgccc gcctcttgca ttaattaata cacctataga
6181 ggatggggac atgatagaca caggatttgg cgctatggac tttaaggtgt tgcaggaatc
6241 taaggctgag gtacctttag acattgtaca atccacctgt aaatatcctg actatttaaa
6301 aatgtctgca gatgcctatg gtgattctat gtggttttac ttacgcaggg aacaattatt
6361 tgccagacat tattttaata gggctggtaa agttggggaa acaatACCTG CAGAGTTata
                                                    -> E2 bind
6421 tttaaagggt agcaatggta gagaaccccc tccgagttct gtatatgttg ctacgcctag
```

```
6481 tgggtctatg attacgtctg aggcacagtt atttaataaA CCTTATTGGT Tgcaacgtgc
                                                -> E2 bind
6541 ccaaggccat aataatggca tttgctgggg taatcaatta tttgttactg tagtagatac
6601 tactagaagt actaacatga ctattagtac tgctacagaa cagttaagta aatatgatgc
6661 acgaaaaatt aatcagtacc ttagacatgt ggaggaatat gaattacaat ttgtttttca
6721 attatgcaaa attactttgt ctgcagaggt tatggcatat ttacataata tgaatgctaa
6781 cctactggag gactggaata ttgggttatc cccgccagtg gccaccagcc tagaagataa
6841 atatagatat gttagaagca cagctataac atgtcaacgg gaacagccac caacagaaaa
6901 acaggaccca ttagctaaat ataaattttg ggatgttaac ttacaggaca gtttttctac
6961 agacctggat caatttccac tgggtagaaa atttttaatg caactgggca ctaggtcaaa
7021 gcctgctgta gctacctcta aaaagcgatc tgctcctacc tccacctcta caccagcaaa
7081 acgtaaaagg cggTAGtgtg ttgttgtgtg tttgtgtaac tgtgtttgtg tgttgtatat
                    <- L1 end
7141 atggtatgtt tgtgtatgtg ctttatttta tactttgtat gtgtatgttg tgtttgtgta
7201 aatgtttgtg tgaaatgttt gtgtgtgtat tcattgtatg tatgactgta tatatgtgta
7261 atgtttgtgt gtctgtaata aacatgaatg agtgctttta cgcgtggttg cataaactaa
7321 ggtgtgtcat tattgtggct tttgttttgt aagttattgt gtacagtgta ctatgtgtat
7381 tgtgcataca tatatatacc ataacatact ccattttgtt gtttttccgc cattttgtac
7441 atgcaACCGA ATTCGGTtgc atggcctagt gccattattt aaactaaaag gaattcggtt
          -> E2 bind
7501 gcatggccta gtgccattat ttaaaccaaa aggccctttt cagcagaaca gttaatcctt
7561 tggcatattg ccgtttcctg tgttttatac ttgaattatg tacagtaccg caccctgtat
7621 tactcacagg tactatgact gccaactatg cttttatctg catactttag tgctgttggg
7681 cacacatttt tatacatgtg tctgcaactt tggtgttttg gcttgcagaa tacactatgt
7741 aggccaagta tctgtcagta tctgttttgc aaacatgtaa catacaatta ctcatttttt
7801 aaaACCGTTT ACGGTcgtgc aaaaacaggt ttcttttaat tgtT
        -> E2 bind                     E6 orf start ->
```

HPV66L1AE3

```
LOCUS       HPV66L1AE3    409 bp    DNA             VRL       25-MAY-1994
DEFINITION  Human papillomavirus type 66 (HPV-66), partial L1 cds,
            My09/My11 region.
ACCESSION   U01533
SOURCE      Human papillomavirus type 66 DNA, PCR amplified clone AE3
REFERENCE   1  (bases 1 to 409)
  AUTHORS   Tachezy,R., Van Ranst,M.A., Cruz,Y. and Burk,R.D.
  TITLE     Consensus primer mediated PCR allows identification of novel human
            papillomavirus PCR-types in cervicovaginal lavages
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 409)
  AUTHORS   Van Ranst,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1993) Marc A. Van Ranst, Albert Einstein College
            of Medicine, Dept. of Microbiology & Immunology, 1300 Morris Park
            Avenue, Bronx, NY 10461, USA
COMMENT     Isolate AE3 is a variant of HPV type 66 identified in a study
            conducted to screen cervical lavages from over 500 women for novel
            HPV types.  All women involved in the study were seen by physicians
            from clinics or private practices in the Bronx, N.Y. area.  Derived
            sequences are PCR products amplified over the My09/My11 primer
            region of L1.

BASE COUNT      138 a     74 c     80 g    117 t
ORIGIN
        1 catatgctgg ggtaatcagg tatttgttac tgttgtggat actaccagaa gcaccaacat
       61 gactattaat gcagctaaaa gcacattaac aaaatatgat gcccgtgaaa tcaatcaata
      121 ccttcgccat gtggaggaat atgaactaca gtttgtgttt caactttgta aaataacctt
      181 aactgcagaa gttatggcat atttgcataa tatgaataat actttattag acgattggaa
      241 tattggctta tccccaccag ttgcaactag cttagaggat aaatataggt atattaaaag
      301 cacagctatt acatgtcaga gggaacagcc ccctgcagaa aagcaggatc ccctggctaa
      361 atataagttt tgggaagtta atttacagga cagcttttct gcagacctg
```

```
LOCUS       HPV66MY911    449 bp ds-DNA             VRL       16-OCT-1994
DEFINITION  Human papillomavirus type 66 (HPV-66), partial L1 cds, My09/My11
            region.
ACCESSION   U12498
SOURCE      Human papillomavirus type 66 DNA.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Bernard,H.-U., Chan,S.-Y., Manos,M.M., Ong,C.-K., Villa,L.L.,
            Delius,H., Peyton,C.L., Bauer,H.M., and Wheeler,C.M.
  TITLE     Identification and assessment of known and novel human
            papillomaviruses by PCR amplification, restriction fragment
            length polymorphisms, nucleotide sequence, and phylogenetic
            algorithms
  JOURNAL   J. Infect. Dis. (1994) In press
COMMENT     HPV-66 was isolated on two separate occasions by independent
            researchers.  In 1991, Dr. Tawheed isolated the virus from a
            cervical carcinoma.  The novel virus was subsequently assigned
            the HPV type 66.  Dr. M. Manos isolated the same strain from
            a patient with normal cytology and designated it PAP88.  It
            was later discovered that the PAP88 sequence was identical to
            the sequence of HPV-66 (M. Van Ranst personal communication).
            Primer regions are annotated in the sequence; information in
            this region is not accurate due to primer degeneracy.  This
            sequence is similar to HPV66L1AE3 (U01533).

BASE COUNT      149 a     80 c     90 g    130 t
ORIGIN
        1 gcacagggtc ataataatgg catatgctgg ggtaatcagg tatttgttac tgttgtggat
  L1 cds ->
          -> MY11 PCR primer <-
       61 actaccagaa gcaccaacat gactattaat gcagctaaaa gcacattaac taaatatgat
      121 gcccgtgaaa tcaatcaata ccttcgccat gtggaggaat atgaactaca gtttgtgttt
      181 caactttgta aaataacctt aactgcagaa gttatggcat atttgcataa tatgaataat
      241 actttattag acgattggaa tattggatta tccccaccag ttgcaactag cttagaggat
      301 aaatataggt atattaaaag cacagctatt acatgtcaga gggaacagcc ccctgcagaa
      361 aagcaggatc ccctggctaa atataagttt tgggaagtta atttacagaa cagcttttct
      421 gcagacctgg atcagtttcc ttttggacg
                                 L1 cds ->
                  -> MY09 PCR primer  <-
```

```
LOCUS       HPV69MY911    455 bp ds-DNA             VRL       16-OCT-1994
DEFINITION  Human papillomavirus type 69 (HPV-69), partial L1 cds, My09/My11
            region.
ACCESSION   U12497
SOURCE      Human papillomavirus type 69 DNA.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Bernard,H.-U., Chan,S.-Y., Manos,M.M., Ong,C.-K., Villa,L.L.,
            Delius,H., Peyton,C.L., Bauer,H.M., and Wheeler,C.M.
  TITLE     Identification and assessment of known and novel human
            papillomaviruses by PCR amplification, restriction fragment
            length polymorphisms, nucleotide sequence, and phylogenetic
            algorithms
  JOURNAL   J. Infect. Dis. (1994) In press
COMMENT     This HPV-69 partial genomic fragment was isolated by J. Brandsma,
            C. Greer and M. Manos from a dysplastic lesion of the tongue.  It
            was designated ‘‘JB10" and was later discovered that the sequence
            was virtually identical to the sequence of HPV-69.  Primer regions
            are annotated in the sequence; information in this region is not
            accurate due to primer degeneracy.

BASE COUNT      136 a     87 c     83 g    149 t
ORIGIN
        1 gcacagggac ataacaatgg catttgttgg ggcaaccaat tgtttgttac ttgtgtagat
  L1 cds ->
          -> MY11 PCR primer <-
       61 actacccgca gtaccaacct cactattagt actgtatctg cacaatctgc atctgccact
      121 tttaaaccat cagattataa gcagtttata aggcatggtg aggaatatga attacagttt
      181 atatttcaat tgtgtaaaat tactcttacc actgatgtaa tggcctatat ccatacaatg
      241 aattctgcta ttttggaaaa ttggaatttt ggccttacct tgcctcctac tgctagtttg
      301 gaagatgcat ataggtttat taaaaattca gctactacat gtcaacgcga tgcccctgca
      361 cagcccaagg aggatccatt tagtaaatta aaattttggg acgttgatct taaagaaaag
      421 ttttctattg atttagatca gtatcccctt ggacg
                                          L1 cds ->
                          -> MY09 PCR primer  <-
```

```
LOCUS       HPVIS039      455 bp ds-DNA             VRL       16-OCT-1994
DEFINITION  Human papillomavirus, partial L1 cds, My09/My11 region, isolate
            IS039.
ACCESSION   U12481
SOURCE      Human papillomavirus DNA recovered from a cervical carcinoma
            biopsy from an Argentinian woman, isolate IS039.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Peyton,C.L., Jansen,A.M., Wheeler,C.M., Stewart,A.-C., Peto,J.,
            Bosch,F.X., Munoz,N., Teyssie,A.R., Torroella, M., Wabinga, H.R.,
            Sarjadi, Ngelangel,C., and Manos,M.M.
  TITLE     A novel human papillomavirus sequence from an international
            cervical cancer study
  JOURNAL   J. Infect. Dis. (1994) In press
COMMENT     HPVIS039 was isolated from an invasive cervical carcinoma biopsy as
            part of the International Biological Study of Cervical Cancer. Over
            one thousand specimens from cervical carcinomas were amplified
            using PCR with the My09/My11 primers, identifying 837 specimens
            positive for HPV DNA. Of these, twelve yielded HPV DNA fragments
            that failed to hybridize with probes specific to known HPV types.
            Among these was HPVIS039, from Argentina. This sequence was
            virtually identical to that of another specimen from Cuba, and
            otherwise most closely related to HPVMM4 (90.8%), three of its
            variants, and HPV51, as well as a previously characterized sequence
            HPVL1AE2 (U01532).  Primer regions are annotated in the sequence;
            information in this region is not accurate due to primer degeneracy.
            All similarity calculations exclude data from this region.

BASE COUNT      135 a     95 c     88 g    137 t
ORIGIN
        1 gcacagggac ataataatgg catttgctgg aataatcagc tttttattac ttgtgttgac
  L1 cds ->
          -> MY11 PCR primer <-
       61 actaccagaa gtactaattt aaccattagc actgctgcta ctccatcagt tgcacagaca
      121 ttcactccaa caaactttaa gcagtatatt aggcacgggg aagaatatga attacaattt
      181 atttttcaat tgtgtaaaat tactctaaca actgaggtaa tggcttacct gcacaccatg
      241 gattctacaa tactagagca gtggaatttt ggattaacat tgcctccttc agctagtttg
      301 gaggatgcct atcgctttgt gaaaaatgca gcaacctctt gtcaacggga cagccctcca
      361 caggctaaac aggacccttt ggcaaaatat aagttttgga ccgtggacct taaggaacgc
      421 ttttctttgg atttagatca gtttccttct ggacg
                                        L1 cds ->
                          -> MY09 PCR primer  <-
```

HPVL1AE2

```
LOCUS       HPVL1AE2      415 bp    DNA             VRL       25-MAY-1994
DEFINITION  Human papillomavirus, partial L1 cds, MY09/MY11 region.
ACCESSION   U01532
SOURCE      Human papillomavirus DNA, PCR amplified clone AE2
REFERENCE   1  (bases 1 to 415)
  AUTHORS   Tachezy,R., Van Ranst,M.A., Cruz,Y. and Burk,R.D.
  TITLE     Consensus primer mediated PCR allows identification of novel human
            papillomavirus PCR-types in cervicovaginal lavages
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 415)
  AUTHORS   Van Ranst,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1993) Marc A. Van Ranst, Albert Einstein College
            of Medicine, Dept. of Microbiology & Immunology, 1300 Morris Park
            Avenue, Bronx, NY 10461, USA
COMMENT     Isolate AE2 is a novel HPV type identified in a study conducted to
            screen cervical lavages from over 500 women for novel HPV types.
            All women involved in the study were seen by physicians from clinics
            or private practices in the Bronx, N.Y. area.  Derived sequences are
            PCR products amplified over the My09/My11 primer region of L1. The
            sequence for AE2 is very close to that of HPVIS039 (U12481).

BASE COUNT      125 a     87 c     76 g    127 t
ORIGIN
        1 catttgctgg aataatcagc tttttattac ttgtgttgac actaccagaa gtactaattt
       61 aaccattagc actgctgcta ctccatcagt tgcacagaca ttcactccaa caaactttaa
      121 gcagtatatt aggcacgggg aagaatatga attacaattt atttttcaat tgtgtaaaat
      181 tactctaaca actgaggtaa tggcttacct gcacaccatg gattctacaa tactagagca
      241 gtggaatttt ggattaacat tgcctccttc agctagtttg gaggatgcct atcgctttgt
      301 gaaaaatgca gcaacctctt gtcaacggga cagccctcca caggctaaac aggacccttt
      361 ggcaaaatat aagttttgga ccgtggacct taaggaacgc ttttctttag attta
```

```
LOCUS       HPVMM4        455 bp ds-DNA             VRL       16-OCT-1994
DEFINITION  Human papillomavirus, isolate MM4, partial L1 cds, My09/My11
            region.
ACCESSION   U12488
SOURCE      Human papillomavirus DNA recovered from a cervical swab specimen
            from a woman with normal cytology, isolate MM4.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Manos,M.M., Waldman,J., Zhang,T. Greer,C., Eichinger,G.,
            Schiffmann,M., and Wheeler, C.
  TITLE     Epidemiology and partial nucleotide sequence of four novel genital
            human papillomaviruses
  JOURNAL   J. Infect. Dis. (1994) In press
COMMENT     MM4, also known as W13B, was isolated from a genital swab sample.
            Samples were obtained from over 500 patients examined at either the
            Shasta/Diablo planned parenthood clinic or at a private practice
            in the state of California over the course of seventeen months.
            Each of the samples were cervical or vulvar/intraoital in origin.
            DNA was PCR amplified over the MY09/MY11 region and subsequently
            sequenced if the HPV digested products yielded unique RFLP patterns.
            This procedure resulted in the identification of four novel HPV
            types: W13B, PAP291, PAP155, and PAP238a, which have subsequently
            been renamed MM4, MM7, MM8, and MM9. Oligonucleotide probes over
            the MY9/MY11 region from these viruses have been reported by
            Hildesheim et al. (J Infect Dis 169: 235-40).   These probes were
            used to determine prevalence in different populations.  Prevalence
            for each of these viruses was similar to that seen in other
            characterized "intermediate risk" viruses probed for in these
            studies.  It should be noted that MM4 is extremely similar (90.8%)
            to novel HPVIS39 (U12481) and MM7 is virtually identical to LVX82
            (U12487).  Primer regions are annotated in the sequence;
            information in this region is not accurate due to primer degeneracy.
            All similarity calculations exclude data from this region.

BASE COUNT      142 a     89 c     81 g    143 t
ORIGIN
        1 gcacagggac ataataatgg catttgctgg aataatcagc tttttattac ttgtgttgac
  L1 cds ->
          -> MY11 PCR primer <-
       61 actactagaa gtaccaattt aaccattagc actgctgtta ctcaatctgt tgcacaaaca
      121 tttactccag caaactttaa gcaatacatt aggcatgggg aagaatatga attgcaattt
      181 atatttcaat tgtgtaaaat cactttaact actgaaatta tggcttacct gcacaccatg
      241 gattctacaa ttttagaaca gtggaatttt ggattaacct tgccccctc agctagtttg
      301 gaggatgcct atcgatttgt aaaaaatgca gcaacatcct gtcacaagga cagtcctcca
      361 caggctaaac aagacccttt ggcaaaatat aaattttgga atgtagacct taaggaacgc
      421 ttttctttgg atttggatca gtttccttt ggacg
                                        L1 cds ->
                          -> MY09 PCR primer  <-
```