Published in final edited form as:

Curr Protoc Microbiol. ; 40: 14B.6.1–14B.6.29. doi:10.1002/9780471729259.mc14b06s40.

Construction of a Transcription Map for Papillomaviruses using RACE, RNase Protection and Primer Extension Assays

Xiaohong Wang and Zhi-Ming Zheng

Tumor Virus RNA Biology Section, Gene Regulation and Chromosome Biology Laboratory,
Center for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick,
Maryland

Abstract

Papillomaviruses are a family of small, non-enveloped DNA tumor viruses. Knowing a complete transcription map from each papillomavirus genome can provide guidance for various papillomavirus studies. This unit provides detailed protocols to construct a transcription map of human papillomavirus type 18. The same approach can be easily adapted to other transcription map studies of any other papillomavirus genotype due to the high degree of conservation in the genome structure, organization and gene expression among papillomaviruses. The focused methods are 5'- and 3'- rapid amplification of cDNA ends (RACE), which are the techniques commonly used in molecular biology to obtain the full length RNA transcript or to map a transcription start site (TSS) or an RNA polyadenylation (pA) cleavage site. Primer walking RT-PCR is a method for studying splicing junction of RACE products. In addition, RNase protection assay and primer extension are also introduced as alternative methods in the mapping analysis.

Keywords

papillomaviruses; transcription map; RACE; primer walking; RPA; primer extension

INTRODUCTION

Papillomaviruses are a family of small, non-enveloped DNA tumor viruses. Several hundreds of papillomavirus genotypes have been identified from almost all mammals (Bernard et al., 2010) and some amniotes (Herbst et al., 2009; Drury et al., 1998; Lange et al., 2011). Papillomaviruses infect only keratinocytes of the skin or mucous and their infections are highly species- and tissue-specific (Mistry et al., 2008). Human papillomaviruses (HPV) consist of ~200 genotypes and their infections are in general asymptomatic, but may cause benign papillomas or warts. Certain types of HPVs in persistent infection may lead to the development of cancers at the infected sites, including cervix, anus, vulva, and oropharynx. Thus, HPVs can be clinically grouped into low-risk HPVs, which mainly cause genital warts or papilloma, and high-risk HPVs, which are frequently associated with invasive cervical cancer. Among those high-risk HPVs, HPV16 and HPV18 are two most common genotypes of which their persistent infections are responsible for ~70% of cervical cancer cases worldwide (Munoz et al., 2006).

Although cervical cancer is preventable today, the underlined mechanism of how HPV infection leads to development of cervical cancer needs more careful studies. In this regard, elucidation of HPV gene expression and posttranscriptional regulation in cervical lesion progression and in virus life cycle would provide pivotal clue for us to understand molecular basis of HPV pathogenesis. To date, a complete transcription map has been constructed from several HPV genomes, including HPV 1 (Palermo-Dilts et al., 1990), HPV5 (Sankovski et al., 2014), HPV11 (Chiang et al., 1991; Renaud and Cowser, 1996), HPV16 (Zheng and Baker, 2006), HPV18 (Wang et al., 2011), and HPV31 (Hummel et al., 1992; Ozbun and Meyers, 1997) and partial transcription maps are also available from the early region of several other HPV genomes, including HPV47 (Kiyono et al., 1989) and HPV58 (Li et al., 2013). For most of HPVs, however, their genome transcription map remains largely unknown.

Because the productive HPV life cycle is tightly linked to squamous cell differentiation, organotypic raft cultures have been developed to grow these viruses in vitro (Banerjee et al., 2005; Wang et al., 2008; Wang et al., 2009b; Wang et al., 2014) and had been widely used to study both early and late stages of the viral life cycle (Wang et al., 2009a; Dollard et al., 1992; McLaughlin-Drubin et al., 2004; Meyers et al., 1992; Meyers et al., 1997). The protocols presented here are based on the expression analysis of HPV18 early and late genes using HPV18-infected raft tissues derived from human foreskin keratinocytes which had been used to construct a complete transcription map of HPV18 from productive viral infection (Wang, Meyers, Wang, Chow, and Zheng, 2011). A 5' rapid amplification of cDNA ends (5' RACE) is to map where the first nucleotide or transcription start site (TSS) of the early or late viral transcripts is started from the virus genome. A 3' RACE is used to map where the last nucleotide or a cleavage site for RNA polyadenylation (pA) of the early or late transcripts in the course of viral RNA transcription. A primer walking RT-PCR in combination with DNA sequencing is to define RNA splice junction. Both an RNase protection assay and a primer extension assay in combination with DNA sequencing are two other methods being used to map the RNA transcription start site.

BASIC PROTOCOL 1

Rapid Amplification of cDNA Ends (RACE)

Rapid amplification of cDNA ends (RACE) is a technique widely used in molecular biology to obtain the full length sequence of an RNA transcript (Chenchik et al., 1996; Zhu et al., 2001; Zhumabayeva et al., 2001). RACE results in the production of a cDNA copy of the RNA sequence of interest, which is produced through reverse transcription, followed by PCR amplification of the cDNA copy. The amplified cDNA copies are then sequenced and mapped to a genomic region. RACE is often applied to map a transcript 5' end (5' RACE) or 3' end (3' RACE) of a known gene when a gene-specific primer is used for amplification.

5' RACE is commonly used to map a transcript 5' end or transcription start site (TSS) and 3' RACE is often applied to map a pA cleavage site (3' end) of a particular RNA of interest. The basic RACE protocol comprises total RNA or poly (A)-selected RNA isolation, generating RACE-ready cDNA, rapid amplification of cDNA ends, and characterization of RACE products.

5' RACE mapping a TSS depends on a SMARTer IIA oligonucleotide and the SMARTScribe RT enzyme which bears terminal transferase activity. When reaching the end of an RNA template, the RT enzyme adds 3-5 nucleotides to the 3' end of the first strand cDNA. The SMARTer IIA oligonucleotide contains a 3' terminal stretch of modified bases that anneal to the extended cDNA tail added by the RT enzyme and allow the oligo to serve as a template for the RT enzyme. SMARTScribe RT switches template from the mRNA molecule to the SMARTer IIA oligo, generating a complete, double-stranded cDNA copy of the original RNA with the additional SMARTer IIA sequence at the end. Since the template switching activity of the RT occurs only when the enzyme reaches the end the RNA template, the SMARTer IIA is typically only incorporated into the first strand cDNA at the 5' end and the cDNA with this oligo sequence can be subsequently amplified in the following PCR reactions. This technology had been used to construct a complete transcription map of HPV18 from productive viral infection (Wang, et al, 2011).

Materials

HeLa cells

HPV infected Raft tissues from human keratinocytes; see UNIT 14B.3

TRIzol Reagent (Invitrogen)

Electric homogenizer (Omni International)

Chloroform

Isopropyl alcohol

75% ethanol (in DEPC-treated water)

RNase-free water

Microcentrifuge tubes

NanoDrop 1000 Spectrophotometer

QuickPrep Micro mRNA Purification Kit (GE Healthcare)

The kit includes the following materials:

Oligo(dT)-Cellulose

Extraction Buffer (Buffered aqueous solution containing guanidinium thiocyanate and N-lauroyl sarcosine)

High Salt Buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.5M NaCl)

Low Salt Buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.1M NaCl)

Elution Buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA)

Glycogen Solution (5-10 mg/ml glycogen in DEPC-treated water)

Potassium Acetate Solution (2.5 M potassium acetate (pH 5.0))

MicroSpin Columns

SMARTer RACE cDNA Amplification Kit (Clontech)

The kit includes the following materials:

5× First-Strand Buffer (250 mM Tris-HCl (pH 8.3), 375 mM KCl, 30 mM MgCl₂) (RNase-Free)

Dithiothreitol (DTT; 20 mM)

dNTP Mix (dATP, dCTP, dGTP, and dTTP, each at 10 mM)

5'-RACE CDS Primer A (5'-CDS; 12 μM)

5'-(T)₂₅VN-3' (N = A, C, G, or T; V = A, G, or C)

3'-RACE CDS Primer A (3'-CDS; 12 μM)

5'-AAGCAGTGGTATCAACGCAGAGTAC(T)₃₀ VN-3' (N = A, C, G, or T; V = A, G, or C)

Deionized H₂O

SMARTer IIA oligonucleotide (12 μM)

5'-AAGCAGTGGTATCAACGCAGAGTACXXXXX-3' (X, undisclosed base including I and G)

RNase Inhibitor (40 U/μl)

SMARTScribe Reverse Transcriptase (100 U/μl)

Tricine-EDTA Buffer (10 mM Tricine-KOH, pH 8.5, 1.0 mM EDTA)

Advantage 2 PCR kit (Clontech)

The kit includes the following materials:

PCR-Grade Water

10× Advantage 2 PCR Buffer (400 mM Tricine-KOH (pH 8.7 at 25°C), 150 mM KOAc, 35 mM Mg(OAc)₂, 37.5 μg/ml BSA, 0.05% Tween 20, 0.05% Nonidet-P40)

50× dNTP Mix

50× Advantage 2 Polymerase Mix

5'-RACE-Ready cDNA

3'-RACE-Ready cDNA

10× Universal Primer A Mix (UPM)

Long (0.4 μM): 5'-

CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3'

Short (2 μM): 5'-CTAATACGACTCACTATAGGGC-3')

Nested Universal Primer A (NUP; 10 μM)

5'-AAGCAGTGGTATCAACGCAGAGT-3'

Materials not included in the kit:

HPV18-specific Primer (10 μ M) see Table 1.

QIAquick Gel Extraction Kit (Qiagen)

The kit includes the following materials:

QIAquick Spin Columns

Buffer QG (5.5 M guanidine thiocyanate (GuSCN), 20 mM Tris HCl (pH 6.6))

Buffer PE (10 mM Tris-HCl (pH 7.5), 80% ethanol)

Buffer EB (10 mM Tris-Cl, pH 8.5)

Materials not included in the kit:

3M Sodium Acetate (pH 5.0)

Isopropanol

pCR2.1-TOPO TA Cloning Kit (Invitrogen)

The kit includes the following materials:

Salt Solution (1.2 M NaCl, 0.06 M $MgCl_2$)

Sterile Water

pCR2.1-TOPO vector

dNTPs

M13 Forward Primer (5'-GTAAAACGACGGCCAG-3')

M13 Reverse Primer (5'-CAGGAAACAGCTATGAC-3')

Materials not included in the kit:

Purified gel product

Transforming One Shot TOP10 Competent cells

S.O.C. medium

LB plates containing 50 μ g/ml ampicillin

DEPC water

10 $\times$ PCR buffer II (200 mM Tris-HCl (pH 8.4), 500 mM KCl)

$MgCl_2$ (25mM)

Taq DNA polymerase (5U/ μ l)

QIAprep Spin Miniprep Kit (Qiagen)

The kit includes the following materials:

Buffer P1 (50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 100 μ g/ml RNaseA)

Buffer P2 (200 mM NaOH, 1% SDS)

Buffer N3 (4.2 M Gu-HCl , 0.9 M potassium acetate, pH 4.8)

Buffer PE (10 mM Tris-HCl (pH 7.5), 80% ethanol)

Buffer EB (10 mM Tris-Cl, pH 8.5)

Plasmid DNA Sequencing

BigDye terminator v1.1 ready reaction mix (Applied Biosystems, Life Technologies)

2.5× sequencing buffer (200 mM Tris-HCl, 5 mM MgCl₂, pH 9.0)

Plasmid DNA

HPV18 specific primer (20 μM)

Water

Centri Spin Columns (Princeton Separations)

3130XL DNA Analyzer system (Applied Biosystems, Life Technologies)

Agarose gel electrophoresis

Agarose

Ethidium bromide

Owl EasyCast B2 Mini Gel Electrophoresis Systems

100bp DNA marker

1XTBE buffer (89 mM Tris base, 89 mM boric acid, 2 mM EDTA)

Total RNA Isolation: When working with RNA, wear gloves at all times. After putting on gloves, avoid touching contaminated surfaces and equipment with the gloved hands. Wipe benches with 100% ethanol each time prior to use, in order to rid the area of microorganisms. Treat surfaces of benches, pipettes, and glassware, etc, with RNase inactivating agents-RNase Zap Wipes. Place all samples on ice because RNA is very susceptible to degradation when left at room temperature.

1. Add 1 ml TRIzol Reagent directly to the PBS-rinsed HeLa cell monolayers in a 35-mm culture dish, lyse the cells by pipetting up and down several times; Alternatively, add 1 ml TRIzol Reagent to the HPV18-infected keratinocyte raft tissues in an Eppendorf tube and homogenize the tissues using an electric homogenizer with a separate disposable probe.
2. Incubate the homogenized sample for 5 minutes at room temperature to permit complete dissociation of the nucleoprotein complex.
3. Add 0.2 ml of chloroform per 1 ml of TRIzol Reagent to each sample.
4. Shake or vortex tube vigorously for 15 seconds, incubate for 2-3 minutes at room temperature.

5. Centrifuge the sample at $12,000 \times g$ for 15 minutes at 4°C .
6. Transfer the aqueous phase on the top to a fresh RNase-free 1.5-ml microfuge tube, add 0.5 ml isopropyl alcohol to the fresh tube, mix well by inverting ~ 40 times, and incubate the mixture at room temperature for 10 minutes.
7. Centrifuge the mixture at $12,000 \times g$ for 10 minutes at 4°C to form pellet.
8. Remove and discard supernatant, leaving only the RNA pellet; wash the RNA pellet by adding 1 ml 75% ethanol. Mix the pellet by vortexing, and centrifuge at $7500 \times g$ for 5 minutes at 4°C .
9. Air dry the RNA pellet, dissolve the RNA pellet in DEPC-treated water, and then check the concentration and purity of RNA using a NanoDrop 1000 Spectrophotometer. Store the RNA in -70°C freezer until use.

Note: The purity of the total RNA is the key factor for successful cDNA syntheses and RACE reaction. The presence of residual organics, metal ions, salt or nucleases in the RNA sample could have a large impact on downstream enzymatic applications by inhibiting enzymatic activity or degrading the RNA.

Poly(A⁺)-selected RNA isolation: using QuickPrep Micro mRNA Purification Kit (GE Healthcare)

Note: Although total RNA can be used for reverse transcription of cDNA, Poly(A⁺) RNA is preferentially used to decrease background.

10. Preparation of oligo(dT)-cellulose: approximately 20-30 minutes before starting the extraction, remove kit from storage at $2-8^{\circ}\text{C}$ and place it at room temperature. Place the Extraction Buffer at 37°C until all crystalline material is dissolved and then cool to room temperature. Gently swirl the oligo(dT)-cellulose slurry to obtain a uniform suspension. Immediately pipette 1 ml aliquots of oligo(dT)-cellulose into a microcentrifuge tube for each purification (sample).
11. Add 0.4 ml of Extraction Buffer to 100 μg total RNA, vortex until a homogeneous suspension is achieved.
12. Dilute the sample by adding 0.8 ml of Elution Buffer and mix by vortexing. Place 0.5 ml of Elution Buffer (per purification) at 65°C until needed.
13. Centrifuge the sample and the tube containing 1 ml oligo(dT)-cellulose for 1 minute at top speed, remove the supernatant from the oligo(dT)-cellulose pellet by pipetting.
14. Transfer the supernatant of the sample (cleared cellular homogenate) onto top of the pellet of oligo(dT)-cellulose. Close the tube and invert to resuspend the oligo(dT)-cellulose.
15. Gently mix the tube for 10 minutes by inverting manually or by placing it on a rocking table. Spin at top speed for 10 seconds.

16. Remove and discard the supernatant. Add 1 ml of High Salt Buffer to the oligo(dT)-cellulose pellet, resuspend the resin by inversion, spin at top speed for 10 seconds, remove and discard the supernatant. Repeat the washing with High Salt Buffer four more times, for a total of five times.
17. Add 1 ml of Low Salt Buffer to oligo(dT)-cellulose pellet, resuspend the resin by inversion, spin at top speed for 10 seconds, remove and discard the supernatant. Repeat the washing with Low Salt Buffer one more time, for a total of two times.
18. Resuspend the resin in 0.3 ml of Low Salt buffer and transfer the slurry to a MicroSpin Column. Spin at full speed for 5 seconds.
19. Discard the effluent in the collection tube, replace the column in the tube and add 0.5 ml Low Salt Buffer, centrifuge for 5 seconds at full speed. Repeat this step for additional two times, empty the collection tube between each step, for a total of three times.
20. Place the column in a new sterile microcentrifuge tube, add 0.2 ml of prewarmed Elution Buffer (from step 12) to the top of the resin bed, centrifuge at top speed for 5 seconds.
DO NOT DISCARD THE ELUTE-IT CONTAINS PURIFIED mRNA!
21. Add another 0.2 ml aliquot of warm Elution Buffer, centrifuge at top speed for another 5 seconds. Collect the elute to the same tube in step 20.
22. Add 10 μ l of Glycogen Solution and 40 μ l of Potassium Acetate Solution to the 400 μ l of sample, add 1 ml of 95% ethanol (chilled to -20°C) and place the sample at -20°C .
23. Collect the precipitated mRNA by centrifugation at 4°C , wash once with 75% ethanol. Dissolve the pellet in 10 μ l of DEPC-treated water, aliquot into 3 tubes (3 μ l/tube)
24. Store the tubes at -70°C freezer until use.

First-Strand cDNA Synthesis using SMARTer RACE cDNA Amplification Kit (Clontech)

25. For each 10 μ l cDNA synthesis reaction, mix the following reagents and spin briefly in a microcentrifuge, then set aside at room temperature.

5 $\times$ First-Strand Buffer	2 μ l
DTT (20 mM)	1 μ l
dNTP Mix (10 mM)	1 μ l
Total	4 μ l

26. Combine the following reagents in separate microcentrifuge tubes:

For preparation of 5'-RACE-Ready cDNA using total RNA (step 9):

Total RNA (1 µg/µl)	1 µl
5'-CDS Primer A	1 µl
Sterile water	1.75 µl
Total	3.75 µl

For preparation of 5'-RACE-Ready cDNA using Poly(A)-selected RNA (step 23):

Poly(A)-selected RNA	2.75 µl
5'-CDS Primer A	1 µl
Total	3.75 µl

For preparation of 3'-RACE-Ready cDNA using total RNA (step 9):

Total RNA (1µg/µl)	1 µl
3'-CDS Primer A	1 µl
Sterile water	2.75 µl
Total	4.75 µl

For preparation of 3'-RACE-Ready cDNA using Poly(A)-selected RNA (step 23):

Poly(A)-selected RNA	3.75 µl
3'-CDS Primer A	1 µl
Total	4.75 µl

27. Mix contents and spin the tubes briefly in a microcentrifuge. Incubate the tubes at 72°C for 3 minute, and then cool the tubes to 42°C for 2 minute. Spin the tubes briefly after cooling.
28. Add 1 µl of the SMARTer IIA oligo to 5'RACE cDNA synthesis reaction.
29. Prepare enough of the following Master Mix for all 5'-& 3'-RACE-Ready cDNA synthesis reactions:

Buffer Mix from step 25	4 µl
-------------------------	------

RNase Inhibitor (40 U/μl)	0.25 μl
SMARTScribe Reverse Transcriptase (100 U/μl)	1 μl
<hr/>	
Total	5.25 μl

30. Add 5.25 μl of the Master Mix from step 29 to the denatured RNA from step 27 (3'-RACE cDNA) and step 28 (5'-RACE cDNA), for a total volume of 10 μl.
31. Mix the contents of the tubes by gently pipetting, and spin the tubes briefly to collect the contents at the bottom.
32. Incubate the tubes at 42°C for 90 minutes, and then heat at 70°C for 10 minutes using PCR machine.
33. Dilute the cDNA product with Tricine-EDTA Buffer by adding 100 μl to RACE reaction products using total RNA and adding 250 μl to RACE reaction products using Poly(A⁺)-selected RNA.
34. Aliquot the RACE products into 5 tubes, store 2 of them at -20°C for current use, place other tubes at -70°C freezer for longer storage.

Rapid Amplification of cDNA Ends (5' & 3' RACE PCR) using Advantage 2 PCR kit (Clontech)

35. Prepare PCR Master Mix, for each 50-μl PCR reaction, mix the following reagents:

PCR-grade water	34.5 μl
10× Advantage 2 PCR buffer	5 μl
d NTP mix (10 mM)	1 μl
50× Advantage 2 Polymerase mix	1 μl
<hr/>	
Total	41.5 μl

36. Mix well by vortexing, and then briefly spin the tube in a microcentrifuge.
37. Prepare PCR reactions in 0.2-ml PCR tubes as following:

<hr/>	
For 5' RACE:	
5' RACE-Ready cDNA (from step 34)	2.5 μl
10× Universal Primer A Mix (UPM)	5 μl
HPV18 5' RACE primer (10 μM)	1 μl
Master mix (from step 35)	41.5 μl
<hr/>	
Total	50 μl

For 3' RACE:	
3' RACE-Ready cDNA (from step 34)	2.5 µl
10× Universal Primer A Mix (UPM)	5 µl
HPV18 3' RACE primer (10 µM)	1 µl
Master Mix (from step 35)	41.5 µl
<hr/>	
Total	50 µl

- 38.** Run thermal cycling using following program:

20 cycles (Poly(A)-selected RNA) or 25 cycles (Total RNA)	
94°C	30 sec
68°C	30 sec
72°C	3 min (for fragments <3kb, add 1 min for each additional 1 kb)

- 39.** Run the PCR products on a 1.2-2% agarose gel (depends on expected fragment size). If the primary PCR reaction fails to give the distinct bands, run additional 5 cycles or perform a secondary “nested” PCR.
- 40.** For secondary “nested” PCR, dilute the primary PCR product 1:50 with Tricine-EDTA Buffer, prepare PCR reaction:

Diluted primary PCR product	5 µl
Nested Universal Primer A (NUP; 10 µM)	1 µl
HPV18 RACE primer (10 µM)	1 µl
PCR Grade Water	1.5 µl
Master Mix (from step 35)	41.5 µl
<hr/>	
Total	50 µl

- 41.** Run thermal cycling using following program:

15 cycles (Poly(A)-selected RNA) or 20 cycles (Total RNA)	
94°C	30 sec
68°C	30 sec
72°C	3 min (for fragments <3kb, add 1 min for each additional 1 kb)

- 42.** Run the PCR products on a 1.2-2% low melting agarose gel (depends on expected fragment sizes).

To avoid low melting agarose gel crack, mix the small part of LE agarose with low melting agarose, for example, to prepare 1.5% low melting gel, mix 1g low melting agarose with 0.5g LE agarose in 100 ml water.

Extract and Purify DNA Fragments from Agarose Gel using QIAquick Gel Extraction Kit (Qiagen)

43. Excise the DNA fragment from the agarose gel using a clean blade, place into a colorless microcentrifuge tube.

44. Weigh the gel slice, add 3 volumes of Buffer QG to 1 volume of gel (100 mg = ~100 μ l).

For example, add 360 μ l Buffer QG to 120 mg of gel.

45. Incubate the tubes at 50°C for 10 min; mix by vortexing every 2-3 min during the incubation to help the gel slices completely dissolve.

It is important to completely solubilize agarose.

46. After the gel slice has dissolved completely, the color of the mixture should be yellow, similar to Buffer QG without dissolved agarose. If the color of the mixture is orange or violet, add 10 μ l of 3 M sodium acetate (pH 5.0) and mix. The color of the mixture will turn to yellow.

47. Add 1 gel volume of isopropanol to the samples and mix.

For example, if the agarose gel slice is 120 mg, add 120 μ l isopropanol. Do not centrifuge the sample at this stage.

48. Place a QIAquick spin column in a 2-ml collection tube, add the sample to the QIAquick column, and spin at 13,000 rpm for 1 min.

49. Discard the flow-through and place the QIAquick column back in the same collection tube, add 0.5 ml of Buffer QG to the QIAquick column and spin at 13,000 rpm for 1 min.

50. Wash the sample by adding 0.75 ml of Buffer PE to the QIAquick column, let the column stand for 5 min and then spin at 13,000 rpm for 1 min. Discard the flow-through and repeat the washing one more time by adding 0.5 ml of Buffer PE to the QIAquick column and spin at 13,000 rpm for 1 min.

51. Discard the flow-through and spin the QIAquick column for an additional 1 min at 13,000 rpm to completely remove the residual ethanol from Buffer PE.

52. Place the QIAquick column into a clean 1.5-ml microcentrifuge tube, add 30 μ l of water (pH 7.0-8.5) to the center of the QIAquick column membrane, let the column stand for 1-2 min, and then spin for 1 min. Collect the flow-through in the tube and keep the tube with the extracted DNA at 4°C.

TOPO TA Cloning using pCR2.1-TOPO Kit (Invitrogen)

53. Mix the following reagents gently and incubate for 25 min at room temperature:

Purified gel product (from step 52)	4 μ l
Salt solution	1 μ l
TOPO vector	1 μ l
Total	6 μ l

54. Place the reaction on ice and proceed to competent cells transformation.
55. Thaw one vial of One Shot TOP10 chemically competent cells on ice for each transformation.
56. Add 3 μ l of the TOPO TA Cloning reaction product (from step 53) into a vial of OneShot cells and mix gently. Incubate the vials on ice for 30 minutes.
57. Heat shock the cells for 30 seconds at 42°C water bath without shaking and then place on ice for 2-3 minutes.
58. Add 250 μ l of pre-warmed (room temperature) S.O.C Medium to each vial, transfer the entire content to a 14-ml round-bottom Falcon tube, and then shake in a shaking incubator at 37°C for 1 hour at 225 rpm.
59. Spread 20 μ l or 100 μ l two different volumes from each transformation on LB plates containing ampicillin at 50 μ g/ml and incubate at 37°C overnight. Store the remaining transformation mix at 4°C.
60. On next day, pick 10-20 colonies from each transformation; grow each colony in 5 ml LB Broth containing 50 μ g/ml ampicillin at 37°C for 6 hours at 225 rpm.
61. Perform PCR-based screening for correct insertion using HPV18 specific primers and a M13 Forward or Reverse Primer.

DEPC water	35.25 μ l
10 $\times$ PCR buffer II	5 μ l
MgCl ₂ (25 mM)	5 μ l
dNTP mix (10 mM each)	1 μ l
Taq DNA polymerase (5 U/ μ l)	0.25 μ l
HPV18 specific primer (20 μ M)	0.25 μ l
M13 Forward or Reverse Primer (20 μ M)	0.25 μ l
Bacterial suspension from step 60	3 μ l
Total	50 μ l

62. Run PCR products on a 1.5% agarose gel, select those ones with correct insertions for plasmid isolation

Minipreparation of Plasmid DNA using QIAprep Spin Miniprep Kit (Qiagen)

- 63.** Pellet 5 ml bacterial cultures (from step 60) by centrifuge at 9000 rpm for 3 min at room temperature. Resuspend the pelleted bacterial cells in 250 μ l Buffer P1 and transfer to a microcentrifuge tube.
- 64.** Add 250 μ l Buffer P2 and mix thoroughly by inverting the tubes 4-6 times until the solution become clear.
- Do not vortex. Do not allow the lysis reaction to proceed for more than 5 minutes.
- 65.** Add 350 μ l Buffer N3 and mix immediately and thoroughly by inverting the tubes 4-6 times.
- 66.** Centrifuge for 10 minutes at 13,000 rpm in a table-top microcentrifuge.
- 67.** Transfer the supernatant from step 66 to the QIAprep spin column, centrifuge for 30-60 seconds at 13,000 rpm, and discard the flow-through.
- 68.** Wash the QIAprep spin column by adding 0.75 ml Buffer PE. Centrifuge for 30-60 seconds at 13,000 rpm, and discard the flow-through.
- 69.** Wash the QIAprep spin column one more time by adding 0.5 ml Buffer PE. Centrifuge for 30-60 seconds at 13,000 rpm, and discard the flow-through.
- 70.** Spin the QIAquick column for an additional 1 min at 13,000 rpm to completely remove the residual ethanol from Buffer PE.
- 71.** Place the QIAprep column in a clean 1.5 ml microcentrifuge tube, add 50 μ l Buffer EB (10mM Tris-Cl, pH 8.5) to the center of the QIAprep column membrane, let stand for 1-2 minutes, and then centrifuge for 1 minute at 13,000 rpm. Collect the flow-through with the extracted plasmid DNA in the tube.
- 72.** Check the purity and concentration of plasmid DNA using NanoDrop 1000 Spectrophotometer. Store the DNA in -20°C freezer until use.

Plasmid DNA Sequencing

- 73.** All plasmid DNA purified above are then sequenced by conventional Sanger sequencing from both directions using either M13 Forward Primer or M13 Reverse Primer. Sequencing reactions are set up as following:

Plasmid DNA (0.5 $\mu\text{g}/\mu\text{l}$) (from step 72)	1 μl
BigDye v1.1 ready reaction mix	2 μl
2.5 $\times$ sequencing buffer	2 μl
M13 Forward Primer or Reverse Primer (20 μM)	1 μl
Water	4 μl
Total	10 μl

Run program as following:

74. Sequencing reactions are purified using Centri.Spin-20 Columns to remove unincorporated dye terminators and residual salts.
- After the sequencing reaction, it is important to remove unincorporated dye terminators and residual salts because 1) the unincorporated dye terminators tend to compromise the accuracy of the sequencing base calling and 2) excess salts and other ion-carrying molecules in the sequencing mix act to lower the signal intensity by competing with the dye-labeled DNA sequencing products.
75. Gently tap the Centri.Spin-20 Column to insure that the dry gel has settled in the bottom of the spin column.
76. Reconstitute the column by adding 0.65 ml of molecular grade water and vortex vigorously for 5 seconds. Remove air bubbles by tapping the bottom of the column.
77. Let the columns stand for at least 30 minutes at room temperature.
78. After the gel has hydrated and is free of bubbles, first remove the top column cap, and then remove the column end stopper from the bottom.
79. Spin the column at 750 g for 2 minutes; discard the flow-through and spin the column for one more minute.
80. Place the column into a 1.5-ml microcentrifuge tube; add the 10 μ l sequencing reaction directly onto the center of the gel bed without disturbing the gel surface, spin at 750 g for 2 minutes. Collect the flow-through in the tube.
81. Dry the sample in the tube using Speed Vacuum and then proceed to Sanger DNA sequencing by using an Applied Biosystems® 3130XL DNA Analyzer or submit to a sequencing service provider.

Sequence analysis

82. Align the obtained sequence against HPV18 reference genome using NCBI Blast tool to identify HPV18 TSS for early and late transcripts, pA cleavage sites for early and late transcripts, and HPV18 splice junctions from 5' or 3' RACE results.

BASIC PROTOCOL 2

PRIMER WALKING RT-PCR TO DETECT A SPLICE JUNCTION

Primer walking is a step-by-step sequencing approach for sequencing long DNA templates from end to end. The long fragments that cannot be sequenced in a single sequencing reaction are divided into several consecutive short ones and sequenced separately. After an initial round of sequencing from a known sequence at one end of the template, each subsequent round is initiated from a new primer, which is based on the end of the sequence obtained from the previous reaction. Primer walking method combined with reverse transcription polymerase chain reaction (RT-PCR) can be used to identify RNA splice junctions of HPV18 transcripts (Wang, et al, 2011) .

Materials—Total RNAs (from step 9, Basic Protocol 1)

RQ1 RNase-Free DNase 10× Reaction Buffer (400 mM Tris-HCl, pH 8.0, 100 mM MgSO₄, 10 mM CaCl₂ (Promega)

RQ1 RNase-Free DNase (Promega)

RQ1 DNase Stop Solution (20 mM EGTA, pH 8.0, Promega)

Nuclease-free water

DEPC-Treated Water

Following reagents from Life Technologies

10× PCR Buffer II (200 mM Tris-HCl, pH 8.4, 500 mM KCl)

MgCl₂ (25 mM)

dNTP Mix (10 mM each)

Taq DNA polymerase (5 U/μl)

RNase Inhibitor (40 U/μl)

Random Hexamer (50 μM)

MuLV Reverse Transcriptase (50 U/μl)

Primer design

1. Design primers for primer walking RT-PCR based on HPV reference genome. Figure 1 shows the primers designed for HPV18 splice junction identification.

Remove viral DNA from total RNA preps

2. Set up DNase digestion reaction as following:

Total RNA (30 μg)	x μl
RNase-Free DNase 10× Reaction Buffer	4 μl

RQ1 RNase-Free DNase	2 μ l
Nuclease-free water	y μ l
<hr/>	
Total	40 μ l

3. Incubate the reactions at 37°C for 30 minutes.
4. Add 1 μ l of RQ1 DNase Stop Solution to terminate the reaction, and then incubate at 65°C for 10 minutes to inactivate DNase.
5. Add 8 μ l of DEPC-treated water, 0.5 volume (25 μ l) of 5 M Ammonium Acetate, 1 μ l of glycogen and 3 volume (225 μ l) of cold 100% ethanol to each reaction, incubate at -20°C overnight.
6. Centrifuge the samples at 13,000 g for 15 minutes at 4°C, discard the supernatant, wash RNA pellet once with 1 ml of 75% ethanol.
7. Centrifuge the samples at 13,000 g for 5 minutes at 4°C, discard the supernatant, spin the tubes briefly, remove the residual ethanol, air dry, and resuspend the RNA pellet in 30 μ l cold DEPC-treated water.
8. Check the concentration and purity of RNA using NanoDrop 1000 Spectrophotometer. Store the RNA in -70°C freezer until use.

RT- PCR

9. Set up RT-PCR reaction as following:

DEPC-treated water	35.25 μ l
10 $\times$ PCR Buffer II	5 μ l
MgCl ₂ (25 mM)	5 μ l
dNTP Mix (10 mM each)	1 μ l
Taq DNA polymerase (5U/ μ l)	0.25 μ l
RNase Inhibitor (40 U/ μ l)	1 μ l
Random Hexamer (50 μ M)	0.5 μ l
MuLV Reverse Transcriptase (50 U/ μ l)	0.5 μ l
DNase-Treated Total RNA (1 μ g/ μ l)	2 μ l
<hr/>	
Total	50 μ l

Note: To replace the Random Hexamer and MuLV Reverse Transcriptase for minus RT control, add DEPC-treated water instead.

10. Mix and run at 42°C for 60 minutes, then add the following PCR primers to each corresponding tube.

HPV Specific Forward Primer (20 μ M)	0.25 μ l
HPV Specific Reverse Primer (20 μ M)	0.25 μ l

11. Run the PCR reaction using the following program:

12. Run the PCR products on a 1.2-2% low melting agarose gel (depends on expected fragment sizes) as described in the Basic Protocol 1 step 42.
13. Extract and purify DNA fragments from agarose gel, TOPO TA cloning, miniprep of plasmid DNA, plasmid DNA sequencing and sequencing data analysis follow procedures described for Basic Protocol 1 for RACE steps 43-82.

BASIC PROTOCOL 3

RNase PROTECTION ASSAY

RNase protection assay (RPA) is a technique used to identify individual RNA molecules in samples of total cellular RNA. The extracted RNA is first mixed with an antisense, radiolabeled RNA probe that is complementary to the sequence of interest to form a RNA:RNA hybrid. The mixture is then exposed to ribonucleases that specifically cleave only single-stranded RNA, but have no activity against double-stranded RNA hybrid. When the reaction runs to completion, susceptible RNA regions are degraded to very short oligomers or to individual nucleotides and the surviving RNA fragments in the presence of ribonucleases are those that were complementary to the added antisense strand probe labeled with ^{32}P and thus contained the sequence of interest. The size of the resistant fragment can be determined by electrophoresis on a high-resolution, denaturing polyacrylamide gel and can be aligned on the same gel against DNA sequencing read being generated by the same oligo primer used for the antisense probe preparation. This would allow determining the exact position being protected. RPA has been successfully used to map the transcription start site, the ends of RNA molecules or exon-intron boundaries by using the RNA probe-generating primer for DNA sequencing (Tang and Zheng, 2002; Majerciak et al., 2006).

Materials—Riboprobe in vitro Transcription Systems (Promega):

Materials included in the system:

5× Transcription Buffer (200 mM Tris-HCl (pH 7.9), 30 mM MgCl_2 , 10 mM Spermidine, 50 mM NaCl)

DTT (100 mM)
RNase Inhibitor (20 U/μl)
rATP (10 mM)
rUTP (10 mM)
rGTP (1 mM)
rCTP (10 mM)
m7G (5')ppp(5') (10 mM)
α-³²P rGTP (10 mCi/ml, 3000 Ci/mmol)
T7 RNA polymerase (20U/μl)
Nuclease-free water

Materials not included:

Linearized DNA template or a PCR product with a T7 promoter attached
DEPC-treated water
5 M Ammonium Acetate
100% Ethanol
tRNA or Yeast RNA (Ambion)
10× TURBO DNase Buffer (200 mM Tris-HCl pH 7.5, 100 mM MgCl₂, and 5 mM CaCl₂, pH 7.5, Ambion)
TURBO DNase (Ambion)
DNase Inactivation Reagent (0.5 mM EDTA (pH 8.0), 10 mM Tris-HCl (pH 7.5))
Formamide Loading dye (49 ml formamide, 1 ml 0.5 M EDTA, 0.013 g Bromophenol blue, 0.013 g xylene cyanol)
6% 7.5 M Urea-gel (see Support Protocol 1 for details)
PAGE Gel Elution Buffer (0.5 M Ammonium Acetate, 1 mM EDTA, 0.2% SDS)

RPA III System (Ambion)

The system includes the following materials:

Hybridization III Buffer (80% formamide, 400 mM NaCl, 40 mM PIPES, pH 6.4, 1 mM EDTA)
RNase A/RNase T1 mix (A mixture of 250 units/ml RNase A and 10,000 units/ml RNase T1)
RNase Digestion III Buffer (5 mM EDTA, 300 mM NaCl, 10 mM Tris-HCl (pH 7.5))

RNase Inactivation/Precipitation III Solution(4 M guanidine isothiocyanate, 0.5% n-lauroyl sarcosine, 25 mM sodium citrate (pH 7.0), 0.1 M β -mercaptoethanol, 4 μ g/ml yeast tRNA)

RNA Loading Dye (95% formamide, 0.025% xylene cyanol and bromophenol blue, 18 mM EDTA, 0.025% SDS)

Human pTRI-cyclophilin DNA template

Materials not included in the system:

0.5 $\times$ TBE (44.5 mM Tris base, 44.5 mM boric acid, 0.95 mM EDTA)

75% Ethanol

Beckman LS 6500 Scintillation counter

Prepare antisense RNA probes by in vitro transcription and gel purification: Note: A human cyclophilin probe transcribed from Ambion pTRI-cyclophilin DNA template (Ambion) should be used as an internal control for RPA and should be prepared using similar procedures.

1. Mix the following reagents:

5 $\times$ Transcription Buffer	10 μ l
DTT (100 mM)	5 μ l
RNase Inhibitor (20 U/ μ l)	1 μ l
rATP (10 mM)	2.5 μ l
rUTP (10 mM)	2.5 μ l
rGTP (1 mM)	2.5 μ l
rCTP (10 mM)	2.5 μ l
m7G (5')ppp(5') (10 mM)	2.5 μ l
α - ³² P rGTP (10 mCi/ml, 3000 Ci/mmol)	8 μ l
T7 RNA polymerase (20 U/ μ l)	4 μ l
*DNA template (~1 μ g/ μ l)	1 μ l
DEPC-treated water	8.5 μ l
Total	50 μ l

*Template DNA could be a linearized plasmid DNA with insertion of the interested DNA fragment under control by a T7 promoter. Alternatively, a PCR product generated by a chimeric T7-gene-specific primer may be used as a DNA template.

2. Incubate the reaction at 37°C for 2 hour.
3. Add 25 μ l of 5 M Ammonium Acetate and 200 μ l of 100% Ethanol to each reaction.
4. Keep the tube on dry ice for 30 minutes, and then centrifuge at 12000 rpm for 15 minutes at 4°C.

5. Wash the pellet by adding 1 ml of 75% ethanol, and then centrifuge at 12000 rpm for 5 minutes at 4°C.
6. Dissolve the pellet in 44 µl of DEPC-treated water.
7. Add 5 µl of 10× TURBO DNase Buffer and 1 µl of TURBO DNase; incubate at 37°C for 45 minutes.
8. Add 25 µl of 5 M Ammonium Acetate and 200 µl of 100% Ethanol to each reaction.
9. Keep the tube on dry ice for 30 minutes, and then centrifuge at 12000 rpm for 15 minutes at 4°C.
10. Wash the pellet by adding 1 ml of 75% ethanol, and then centrifuge at 12000 rpm for 5 minutes at 4°C. Discard the supernatant, dry the pellet.
11. Dissolve the pellet in 5 µl of Formamide Loading dye.
12. Prepare a 6%, 7.5 M Urea-gel as described in Support Protocol 1, and pre-run the gel at 150 V for 30 min.
13. Load all 5 µl of the sample from step 11 to each pre-washed gel-well of the pre-run Urea-gel
14. Run the sample probe in the 6%, 7.5 M Urea-gel at 200-400V for 90 min.
15. Lift one side of the gel-supporting glass and wrap the gel on the other-side supporting glass with saran plastic wrap. Expose the gel to a X-ray film in a cassette in a dark room and develop the film in a X-ray film developer to identify the radioactive transcript band in the expected size.
16. Align the exposed X-ray film with the gel carefully and cut the transcript band and put it into a fresh 1.5-ependorf tube. Crush the excised gel into fine fragments with a pipette-tip.
17. Suspend the gel slice in 50 µl of PAGE Gel Elution Buffer; incubate at 37°C overnight.
18. Centrifuge 12,000 rpm for 1 min, and transfer supernatant to a new tube.
19. Add 25 µl of elution buffer to the gel, centrifuge at 12,000 rpm for 1 min, and then transfer the supernatant to the above collection tube
20. Transfer total 75 µl of supernatant onto Millipore 0.44 µm filtration tube, centrifuge at 12,000 rpm for 2 min, and collect the flow-through into a new tube.
21. Add 2 volumes (150 µl) of pre-cold 100% Ethanol. Incubate for 15 min at 4°C, and then centrifuge at 12,000 rpm for 15 min.
22. Washing the RNA pellets by adding 500 µl of 75% ethanol and centrifuge at 12,000 rpm for 5 min.
23. Discard the supernatant and wash the pellet again with 500 µl of 75% ethanol.

24. Discard the supernatant, spin down, completely remove residual ethanol and air dry for 5 min.
25. Dissolve the pellet in 20 μ l of DEPC-treated water.
26. Count the radioactivity of the probe using 2 μ l of the sample using a scintillation counter.
27. Calculate the approximate RNA concentration using the following formula:
$$\text{Average CPM}/35000 \times \text{Dilution Factor} \approx \text{ng}/\mu\text{l}$$
28. Adjust the concentration of antisense RNA probe to 1 ng/ μ l.

Set up RNase Protection Assay reaction

29. Set up individual reaction as below:

Total RNA	20-30 μ g
Antisense RNA probe (1 ng/ μ l) (from step 28)	4 μ l
Cyclophilin RNA probe (1 ng/ μ l)	3 μ l
Adjust the total volume with DEPC-treated water to	20 μ l

Note: Set up 2 minus-template control reactions for each probe using 2 μ g of tRNA or Yeast RNA equivalent to the highest amount of sample RNA.

30. Add 2 μ l of 5 M Ammonium Acetate and 2.5 volumes (55 μ l) of 100% ethanol to each reaction, and mix thoroughly.
31. Keep the tubes on dry ice for at least 30 minutes, and then spin at 16 000 rpm for 10 min at 4°C.
32. Discard the supernatant, wash the pellet by adding 1 ml of 75% ethanol, centrifuge at 13,500 rpm for 5 minutes at 4°C.
33. Remove the supernatant completely, and air-dry the pellet.
34. Resuspend the pellet in 10 μ l Hybridization Buffer, vortex for 5-10 sec, and then spin briefly.
35. Incubate the tubes at 95°C for 4 minutes to denature the RNA and aid in its solubilization. Centrifuge briefly after incubation to collect the contents in the bottom of the tube.
36. Incubate the tubes at 42°C overnight in a water bath or in a water-filled heat block to hybridize probe to its complement in the RNA sample.

RNase digestion of hybridized probe and RNA sample

37. Dilute RNase A/RNase T1 1:100 in RNase Digestion III Buffer.
38. Add 150 μ l diluted RNase solution to each RNA sample tube and to one of the minus-template controls (step 36).

39. Add 150 μ l RNase Digestion III Buffer (without RNase) to the remaining minus-template control.

Note: minus-template control with only diluted RNase solution will be used as probe control; minus-template control with only RNase Digestion III Buffer will serve as a control for probe integrity.

40. Mix and incubate the reactions at 37°C for 30 minutes.

During this incubation, unprotected single-stranded RNA will be digested.

41. After incubation, add 225 μ l of RNase Inactivation/Precipitation III Solution to each reaction, vortex and spin briefly.
42. Keep the tubes at -20°C for 30 minutes or overnight.
43. Centrifuges the tubes at 13,500 rpm for 15 minutes at 4°C.
44. Carefully remove all supernatant, wash once with 75% ethanol and then air-dry the pellet.
45. Dissolve the pellet in 10 μ l of RNA Loading dye.
46. Prepare an 8% denaturing PAGE gel (see Support Protocol 1 below).
47. Incubate the reactions (step 45) at 75°C for 7 minutes.
48. Load 3 μ l of each reaction into each well of the gel and run the gel at 65 Watt in 0.5 $\times$ TBE buffer for 2.5 hours. Alternatively, an RPA-template primer generated sequencing reaction (see Basic Protocol 4, Sanger sequencing reaction) can be loaded in parallel for electrophoresis.
49. Lift one supporting glass from the gel, transfer the gel onto a 3M filter paper in the size of the gel, cover the gel with saran plastic wrap, and dry the gel onto the filter paper on a gel dryer (Savant stacked gel dryer SGD300) for 2 hours.
50. Gel image is captured by exposing the dried gel to an X-ray film or by a PhosphorImager and analyzed for RPA products and their positions along with sequencing ladders derived from the same primer being used for RPA-DNA template preparation.

BASIC PROTOCOL 4

PRIMER EXTENSION

Primer extension is an alternative to an RNase protection assay which can be used to determine the transcription start site of RNA molecules with known sequence. A radiolabeled antisense primer at the position of ~100-150 nts near to the RNA 5' end is annealed to the RNA and reverse transcriptase is used to synthesize cDNA from the RNA until it reaches the 5' end cap of the RNA. By denaturing the hybrid and using the same antisense, radiolabeled primer for DNA sequencing of the corresponding genome region as a marker on an electrophoretic gel, the TSS can be determined by comparing its location on the gel with the DNA sequence. Usually, the primer extension product is stopped at one nucleotide before reaching to the RNA 5' end because of the 5' cap.

Materials

T4 Polynucleotide Kinase (New England Biolab)

T4 Polynucleotide Kinase 10× Buffer (500 mM Tris-HCl (pH 7.5), 100 mM MgCl₂, 50 mM DTT, 1 mM spermidine)

γ -³²P ATP (10 mCi/ml, 3000 Ci/mmol, GE Healthcare Life Sciences)

Nuclease-Free Water

100-bp DNA ladder (Invitrogen)

MicroSpin G-25 Column

5 M Ammonium Acetate

100% ethanol (cold)

75% ethanol

AMV Primer Extension 2× Buffer (100 mM Tris-HCl (pH 8.3 at 42°C), 100 mM KCl, 20 mM MgCl₂, 20 mM DTT, 2 mM each d NTP, 1 mM spermidine)

Sodium Pyrophosphate (40 mM)

AMV Reverse Transcriptase (Promega)

DEPC-treated water

Loading Dye (98% formamide, 10 mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue)

Primer labeling with γ -³²P

1. Design a primer (usually 20-50 nucleotides in length) without a 5' phosphate, which is complementary to a region near the 5' end of the gene.
2. Set up primer labeling reaction as following:

Primer (20 μ M)	0.5 μ l
T4 Polynucleotide Kinase 10× Buffer	1 μ l
T4 Polynucleotide Kinase	1 μ l
γ - ³² P ATP (10 mCi/ml, 3000 Ci/mmol)	3 μ l
Nuclease-Free Water	4.5 μ l
Total	10 μ l

3. Incubate at 37°C for 10 minutes.
4. Heat to 90°C for 2 minutes to inactivate the T4 polynucleotide kinase, then centrifuge briefly. Bring the final concentration of labeled primer to 0.1 pmol/ μ l by adding 90 μ l of Nuclease-free water. Store at -20°C for later use.

DNA Marker labeling with γ - ^{32}P

5. Set up DNA marker labeling reaction as following:

100-bp DNA ladder (250 ng)	0.5 μl
T4 Polynucleotide Kinase 10 $\times$ Buffer	1 μl
T4 Polynucleotide Kinase	1 μl
γ - ^{32}P ATP (10 mCi/ml, 3000 Ci/mmol)	3 μl
<hr/>	
Nuclease-Free Water	4.5 μl Total
10 μl	

6. Incubate at 37°C for 10 minutes.
7. Heat to 90°C for 2 minutes to inactivate the T4 polynucleotide kinase, then centrifuge briefly. Add 190 μl of Nuclease-free water. Store at -20°C for later use.

RNA sample precipitation

8. Bring the volume of 10-100 μg total RNA to 50 μl with nuclease-free water.

Note: The amount of RNA to use varies with its copy number. 10 μg of total RNA per reaction is a good starting point, but it may be necessary to use as much as 100 μg per reaction for analysis of rare messages. Try a range, such as 10, 50 and 100 μg of total RNA.

9. Add 0.5 volumes of 5 M Ammonium Acetate (25 μl), add 3 volumes of cold 100% ethanol (225 μl).
10. Place the sample on dry ice for at least 30 minutes or -20°C overnight.
11. Centrifuge at 13,000 g for 15 minutes at 4°C.
12. Remove the supernatant, wash the pellet once by adding 1 ml of 75% ethanol and centrifuge at 13,000 g for 5 minutes at 4°C.
13. Air-dry pellet. Dissolve the pellet in 5 μl DEPC-treated water.

Primer Extension Reaction

14. Combine the following components for primer annealing:

Precipitated total RNA (step 13)	5 μl
AMV Primer Extension 2 $\times$ Buffer	5 μl
^{32}P labeled primer (step 4)	1 μl

15. Heat the sample reaction in step 14 at 58°C for 20 minutes, and then place the tubes at room temperature for 10 minutes.
16. Pre-warm Sodium Pyrophosphate to 37°C.

17. Prepare master mix of the following reagents:

AMV Primer Extension 2× Buffer	5 µl
Sodium Pyrophosphate (40 mM)	1.4 µl
AMV Reverse Transcriptase	1 µl
Nuclease-free water	1.6 µl
Total	9 µl

18. Immediately add 9 µl of the above mixture to each RNA/primer tube (step 15).
19. Incubate at 42°C for 30 minutes.
20. Add 30 µl of DEPC-treated water to each reaction.
21. Add 0.5 volumes of 5 M Ammonium Acetate (25 µl) and then add 3 volumes of cold 100% ethanol (225 µl) to each reaction.
22. Place the sample reactions on dry ice for at least 30 minutes or –20°C overnight.
23. Centrifuge at 13,000 g for 15 minutes at 4°C.
24. Remove the supernatant, wash the pellet once by adding 1 ml of 75% ethanol and centrifuge at 13,000 g for 5 minutes at 4°C.
25. Air dry pellet. Dissolve the pellet in 10 µl DEPC-treated water.
26. Add 10 µl of loading dye to each reaction.
27. Heat samples at 90°C for 10 minutes and the sample are now ready for gel loading.
28. To prepare loading-ready DNA marker, mix 1 µl labeled DNA marker with 9 µl TE buffer and 10 µl loading dye. Heat to 90°C for 10 minutes before gel loading.

Sanger sequencing reaction: This is achieved by using DNA chain terminators, 2', 3'-dideoxynucleotide triphosphates (ddNTPs). ddNTP lacks the 3'-OH group of dNTPs that is essential for polymerase-mediated strand elongation in sequencing reaction.

USB Sequenase Version 2.0 DNA Sequencing Kit contains the following reagents:

Sequenase Version 2.0 DNA Polymerase (13 unites/µl)

Inorganic Pyrophosphatase (4 units/µl)

Enzyme Dilution Buffer (10 mM Tris-HCl, pH7.5, 5 mM DTT, 0.1 mM EDTA)

Glycerol Enzyme Dilution Buffer (20 mM Tris-HCl, pH7.5, 2 mM DTT, 0.1 mM EDTA, 50% glycerol)

Sequenase Reaction Buffer (5×, 200 mM Tris-HCl, pH7.5, 100 mM MgCl, 250 mM NaCl)

Dithiothreitol Solution (0.1 M)

Mn Buffer (0.15 M sodium isocitrate, 0.1 M MnCl)

Control DNA M13 mp18 (0.2 µg/µl)

Primer (−40 M13) (0.5 pmol/µl, 5'-GTTTTCCCAGTCACGAC-3')

Labeling Mix (5×, 7.5 µM dGTP, 7.5 µM dCTP, 7.5 µM dTTP)

ddGTP Termination Mix (80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddGTP, 50 mM NaCl)

ddATP Termination Mix (80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddATP, 50 mM NaCl)

ddTTP Termination Mix (80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddTTP, 50 mM NaCl)

ddCTP Termination Mix (80 µM dGTP, 80 µM dATP, 80 µM dCTP, 80 µM dTTP, 8 µM ddCTP, 50 mM NaCl)

Sequencing Extending Mix (180 µM each dGTP, dATP, dCTP, dTTP, 50 mM NaCl)

Stop Solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF)

Reagents not included in the Kit:

γ -³²P labeled HPV primer (step 4)

29. Denature double-stranded templates and anneal with HPV primer

HPV DNA(approx.0.5 pmol)	up to 7 µl
H ₂ O	µl (to adjust total volume)
Sequenase Reaction Buffer (5×)	2 µl
γ - ³² P labeled HPV primer (step 4)	1 µl
Total	10 µl

Anneal by heating 2 minutes at 65°C, then cool slowly to <35°C over 15-30 minutes. Chill on ice.

30. While cooling, label, fill and cap tubes with 2.5 µl of each termination mix.

Keep covered at room temperature.

G	(2.5 µl)
A	(2.5 µl)
T	(2.5 µl)
C	(2.5 µl)

31. Dilute labeling mix 1:5 to working concentration with water.
32. Pre-warm 4 termination tubes 'G', 'A', 'T' and 'C' from step 30 in a 37°C bath.
33. Dilution of sequenase DNA polymerase

Inorganic pyrophosphatase	25 µl
Glycerol enzyme dilution buffer	150 µl
Sequenase DNA polymerase (13 U/µl)	25 µl
Total	200 µl

34. Set up sequencing reaction

To ice-cold annealed DNA mixture (10 µl) from step 29, add:

Dithiothreitol (DTT), 0.1 M	1 µl
Diluted labeling mix from step 31	2 µl
dATP (10 µM)*	0.5 µl
Diluted Sequenase Polymerase (step 33)	2 µl
Total	15.5 µl

Mix and incubate at room temperature 2-5 minutes.

*Can be replaced for other sequencing purposes by using [α - 32 P]dATP or [α - 35 S]dATP at 10 µCi/µl and 1000-1500 Ci/mmol if the sequencing primer at step 29 is not isotope-labeled.

35. Transfer 3.5 µl of the labeling reaction to each termination tube ('G', 'A', 'T' and 'C') at step 30, mix and continue incubation of the termination reactions at 37°C for 5 minutes.
36. Stop the termination reactions by adding 4 µl of Stop Solution.
37. Heat samples to 75°C for 2 minutes immediately before loading onto sequencing gel.

Electrophoresis

38. Prepare a denaturing polyacrylamide gel containing 8% acrylamide, 7 M Urea and 1× TBE buffer (89 mM Tris base, 89 mM boric acid, 1.9 mM EDTA (see Support Protocol 1 for details) and pre-run the gel at 65 watt for 30 min.
39. Load 2 µl of the DNA marker in step 28, all 10 µl of each primer-extended sample reaction in step 27, along with 3 µl of each ddATP, ddTTP, ddCTP, and ddGTP sequencing reaction in step 37, into each corresponding well of the gel, and run the gel at 65 Watt in 0.5× TBE buffer for 2.5 hours.

40. Lift one supporting glass from the gel, transfer the gel onto a 3M filter paper in the size of the gel, cover the gel with saran plastic wrap, and dry the gel onto the filter paper on a gel dryer (Savant stacked gel dryer SGD300) for 2 hours.
41. Gel image is captured by expose the dried gel to an X-ray film or by a PhosphorImager and analyzed for the extension products along with sequencing ladders derived from the same primer used for primer extension.

SUPPORT PROTOCOL 1: PREPARATION OF 8% DENATURING PAGE GEL AND GEL ANALYSIS OF PRIMER EXTENSION AND RPA PRODUCTS

A denaturing polyacrylamide gel electrophoresis (PAGE) is commonly performed to separate RPA products, primer extension products, and terminated sequencing products. The gel is transferred to a 3M filter paper and exposed to an X-ray film by autoradiography so that only the bands with the radioactive label will appear. In PAGE, the shortest fragments will migrate faster than the large ones. Therefore, the bottom-most band in each lane indicates the smallest products protected from RPA or extended or synthesized from a labeled primer. By reading the products in size order crossing over the G, A, T and C sequencing-label lane from the gel bottom up (reading the 5' to 3' sequence of the strand complementary to the sequenced strand) and comparing its size with the primer extension or RPA products, one can continue reading in this fashion over the position of a primer extension or RPA product to determine at which nucleotide the primer extension was stopped or the RPA product was protected.

Materials

Apparatus and accessories:

Sequencing glass plates (Gibco BRL) Sigma cote

Combs (0.4 mm thick, Gibco-BRL)

Spacers (0.4 mm thick, Gibco-BRL)

Model S2 Sequencing Gel Electrophoresis Apparatus (Gibco-BRL)

Savant stacked gel dryer SGD300

3M filter paper

Saran plastic wrap

Scissors

Kodak X-ray film

Kodak X-ray cassettes

SequaGel UreaGel System (National diagnostics):

UreaGel Concentrate (237.5 g/L of Acrylamide, 12.5 g/L of methylene bisacrylamide, 7.5 M urea in deionized water)

UreaGel Diluent (7.5 M urea in deionized water)

UreaGel Buffer [0.89 M Tris-Borate, 20 mM EDTA, pH 8.3 (10× TBE) and 7.5 M urea]

Materials not included in the UreaGel System:

10% ammonium persulfate

TEMED

1× TBE (89 mM Tris base, 89 mM boric acid, 1.9 mM EDTA)

1. Clean two glass plates (one longer and one shorter in length) thoroughly using 70% ethanol first, and then rinse thoroughly with tap water.
2. Apply Sigma cote to one glass plate to ensure the gel will release from one side of the glass plate after electrophoresis. Overnight Sigma cote treatment is preferred.
3. Wipe the plate with distilled water first, and then with 70% ethanol. Wait the glass plate to completely dry and then assemble as a glass cassette with a space on each side, but not the bottom. Lay the assembled glass cassette top at ~25° with the short glass on the top.
4. Prepare gel solution using SequaGel UreaGel System. For example, to make 60 ml of an 8% gel, add 19.2 ml UreaGel Concentrate, 6 ml UreaGel Buffer and 34.8 ml UreaGel Diluent to an Erlenmeyer flask. Swirl gently to mix.
5. Add 400 µl FRESHLY PREPARED 10% ammonium persulfate, swirl gently to mix.
6. Right before casting the gel, add 40 µl of TEMED to the gel solution, swirl gently to mix. Pour the gel solution into the assembled glass cassette from the top IMMEDIATELY. After the gel solution in the cassette flow to the bottom, lay down the glass cassette flat and position the comb in place at the top of the gel.
7. Allow the gel in the glass cassette to polymerize for at least 1-2 hours at room temperature.
8. Mount the glass cassette in the running apparatus. Fill the upper chamber with 0.5× TBE buffer and remove the gel comb. Fill the lower chamber with 0.5× TBE BUFFER as well.
9. Pre-run the gel for 30 minutes before loading each experimental sample from step 47 (Basic Protocol 3), steps 27 and/or 28 (Basic Protocol 4), along with samples from step 37.
10. Clean each well by pipetting the TBE buffer in each well of the gel and then load at least half of each reaction into each well in order as marked and run the gel at 65 Watt in 0.5× TBE buffer for 2.5 hours.
11. Lift one supporting glass from the gel, transfer the gel onto a 3M filter paper in the size of the gel, cover the gel with saran plastic wrap, and dry the gel onto the filter paper on a gel dryer (Savant stacked gel dryer SGD300) for 2 hours.
12. Gel image is captured by expose the dried gel to an X-ray film or by a PhosphorImager and analyzed for the extension or RPA product along with

sequencing ladders derived from the same primer used for primer extension or RPA DNA template preparation.

COMMENTARY

Background Information

Papillomaviruses have a high degree of conservation with respect to viral genome structure and organization as well as gene expression (Longworth and Laimins, 2004;Zheng and Baker, 2006). The circular, double-stranded viral genome in size of ~8 kb encodes in general eight open reading frames (ORFs) which are all transcribed from the same strand. The viral genome can be divided into three major regions: an early region, a late region, and an upstream regulatory region (URR) or a long control region (LCR). The early region in the 5' half of the virus genome encodes six common ORFs (E1, E2, E4, E5, E6 and E7) with regulatory functions in viral replication, gene expression and pathogenesis. Expression of five (E1, E2, E5, E6 and E7) of the six viral regulatory proteins takes place in undifferentiated or intermediately differentiated keratinocytes in the basal/parabasal layers of the epithelium. The late region lies downstream of the early region and encodes the viral major L1 and minor L2 capsid proteins. The LCR spanning the segment between the end of the late region and the beginning of the early region has no coding function, but contains the origin of viral DNA replication and transcription factor binding sites (Bernard, 2002). Viral DNA replication, expression of E4, L1 and L2, and assembly of virions occur exclusively in the upper spinous and more differentiated granular layers of the epithelium (Longworth and Laimins, 2004;Doorbar, 2005).

Activation of viral promoters is required for expression of the viral early region in undifferentiated keratinocytes and of the late region in highly differentiated keratinocytes (Guo et al., 2008;Hopman et al., 2005;Jia et al., 2009;Middleton et al., 2003;Wang, et al., 2009a). A TATA box or a TATA-like sequence is a promoter core sequence and can be found in a region 25 to 35 bases upstream of a TSS. This core sequence binds a general transcription factor, TFIID, to form a transcription initiation complex which includes RNA polymerase II. Recent studies indicate that the canonical TATA box or TATA-like elements can be found only in ~24% of human genes (Yang et al., 2007). Thus, to better define a viral early or late promoter in the virus genome is to map where the first nucleotide or TSS of the viral transcripts is started from the virus genome by 5' RACE. 3' RACE is to map the last nucleotide or a cleavage site for RNA polyadenylation of the transcripts in the course of viral RNA transcription.

RACE

RACE is a powerful technique not only being used to clone a full length cDNA transcript of a gene, but also being widely used to map a TSS (5' RACE) or a pA cleavage site (3' RACE) of a particular transcript. Both 5' RACE and 3' RACE have been successfully applied, respectively, to specifically amplify the 5' end or TSS and the 3' pA cleavage site of HPV18 transcripts (Wang, et al., 2011). Each RACE method may display one or more RACE products due to alternative usage of multiple TSS or pA cleavage sites during gene transcription and RNA processing. The multiple or alternative usage of TSS and/or pA cleavage sites is common to the polymerase II transcription of almost all mammalian

(Kawaji et al., 2006; Pauws et al., 2001) and many viral genes. For example, we identified two TSS for the expression of HPV18 early genes and multiple TSS and pA cleavage sites for the expression of HPV18 late genes (Wang, et al., 2011). However, it becomes clear that an alternative TSS usage relies on the conditions of cell differentiation and viral replication status. Thus, which TSS could be a major one in a particular cell/tissue condition could be different from one to another experiment. Similar results are also expected when mapping an alternative pA site.

Alternative approaches

Paired-end analysis of transcription start sites (PEAT)—PEAT strategy for deep sequencing of capped RNA was recently developed to explore the landscape of transcription initiation in *Drosophila melanogaster* embryo (Ni et al., 2010). The detailed protocol can be found at <http://www.nature.com/protocolexchange/protocols/662> and can also be applied to investigate the TSS of papillomaviruses. In brief, the RNA is first treated with alkaline phosphatase to remove 5'-phosphates on non-capped RNA and further treated with tobacco acid pyrophosphatase (TAP) to remove the 5' cap on capped RNA. A ribonucleotide-DNA hybrid oligo (r-oligo) with a MmeI site is then ligated to the decapped 5'-end of RNA by T4 RNA ligase. This oligocapping strategy (Maruyama and Sugano, 1994) specifically labels the decapped end of the mRNAs. After RT with a random hexamer tailed with a second MmeI site and five-cycle PCR, cDNA products are circularized by bridge ligation followed by exonuclease digestion to eliminate linear DNA fragments. The resulting DNA circles are amplified by rolling circle amplification, digested with MmeI, and ligated with sequencing adaptors. The final PEAT library is sequenced by an Illumina HiSeq2000 sequencer with paired-end capability. Compared to conventional single-end TSS mapping strategies, the PEAT approach improves the alignment yield and accuracy and provides additional information on local transcript structure (such as linking 5' TSS tag to known genes).

Using a polyadenylation-deep sequencing method (PA-Seq) to map RNA 3' end—Genome-wide analysis on RNA polyadenylation (pA) of viral transcripts from infected cells could be also performed today by PA-Seq technology (Ni et al., 2013; Majerciak et al., 2013). PA-Seq could be applied to map the pA cleavage sites of papillomaviruses or any other DNA tumor viruses. Briefly, DNA-free total RNA are sheared into 200-300 nt fragments by heating with magnesium and converted to cDNAs by RT using a modified oligo(dT) primer. The resulted dsDNAs pulled down by Dynabeads are dephosphorylated with alkaline phosphatase, released from the beads by USER enzyme digestion, end-repaired by an "A" base addition at the ends, and ligated to bar-coded Illumina paired-end Y linker. Ligation products between 250 bp and 450 bp are gel purified and amplified by 16-cycle PCR with high-fidelity DNA polymerase. The obtained libraries are then sequenced by an Illumina HiSeq2000 sequencer. The obtained reads along the virus genome can be visualized using IGV genome browser (www.broadinstitute.org/igv/). Individual viral pA site is designated by peak calling using F-Seq program (Boyle et al., 2008) for a nucleotide position with the highest number of reads within individual peaks. A relative expression level can be calculated by total reads-counts per million to overall reads mapped to virus and human (Majerciak, et al., 2013).

Both PEAT for TSS mapping and PA-Seq for RNA 3'-end mapping are two powerful high-throughput technologies. Both technologies create millions of the reads and have been applied to determine TSS and pA cleavage site, but require someone familiar with computer softwares to analyze the sequence reads for genome mapping. Moreover, the TSS mapped by PEAT and the pA sites mapped by PA-Seq should be validated by RACE, RPA, or primer extension.

Critical parameters

Each method introduced above has its advantages and some disadvantages. The RACE methods and primer walking RT-PCR are quite simple and efficient for mapping TSS, pA cleavage sites, and splice junctions, but they are also time consuming, mainly because each RACE product or primer walking RT-PCR product needs to be gel-purified, cloned and sequenced. Because RACE is a RT-PCR-based reaction, various cautions in a RT-PCR reaction would also apply to a RACE reaction, including RNA sample preparation, protecting RNA from RNase contamination and degradation, gene-specific primer designing and mis-annealing, RT enzyme selection, amplification of GC-rich template, etc. High quality, poly-A selected RNA would be essential for a successful result of each RACE protocol. Designing a gene-specific primer for a RACE product in size or ~ 500 bp or less would be useful for the rest of the downstream cloning and sequencing steps. Because of the heterogeneity of TSS and pA cleavage site usage of a given viral transcript, 10 or more bacterial colonies for each RACE product should be picked up for sequencing to determine a proportion of individual TSS or pA cleavage site usage. RPA and primer extension methods have higher sensitivity, but have to use radioisotope and their products need to be run against DNA sequencing reactions carried out with the same primer in RPA probe preparation and in the primer extension in order to know the mapped site position in a virus genome. Both methods require high quality of total RNA free from protein binding and a special room/area designated for handling radioactive materials. The RNA with bound proteins will prevent the antisense probe from hybridization to the targeted region and reduce the probe accessibility and thus the assay sensitivity. The primer extension could be also affected by the RNA secondary structures if the detecting RNA in the sample is not denatured well. This will result in production of a premature extension product stopping in the front of the secondary structure.

Anticipated results

These protocols are simple and practical for construction of a full transcription map of any given viral genome. Normally, one or two RACE products are anticipated, but sometime three or more products would appear due to the presence of multiple alternative TSS or pA sites when a promoter sequence or pA signal sequence is not strong. In this regard, a major product would represent the predominant usage of the TSS or pA sites for a given gene transcription and polyadenylation. These mapping results would be also seen in RPA or primer extension. Primer walking RT-PCR in combination of cDNA sequencing is useful for identification of RNA splice junction. It may end up with several products because of the presence of an alternative 5' splice site or 3' splice site in the detection region.

Time Considerations

A full-mapping of TSS and pA site for a given gene transcript will take at least 2-3 weeks to complete. Primer-walking RT-PCT in combination with cDNA sequencing may take a week. A complete RPA from probe preparation, hybridization, RNase digestion to gel electrophoresis would take 4-5 days. Primer extension will also take 3-4 days. However, time to capture radioactive signal for the RPA or primer extension product may vary from overnight to several days depending on the abundance of the detecting transcript.

Acknowledgements

This work was supported by the intramural research programs of the National Institutes of Health, the National Cancer Institute, and the Center for Cancer Research.

Literature Cited

- Banerjee NS, Chow LT, Broker TR. Retrovirus-mediated gene transfer to analyze HPV gene regulation and protein functions in organotypic “raft” cultures. *Methods Mol. Med.* 2005; 119:187–202. [PubMed: 16353336]
- Bernard HU. Gene expression of genital human papillomaviruses and considerations on potential antiviral approaches. *Antivir. Ther.* 2002; 7:219–237. [PubMed: 12553476]
- Bernard HU, Burk RD, Chen Z, van Doorslaer K, Hausen H, de Villiers EM. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology.* 2010; 401:70–79. [PubMed: 20206957]
- Boyle AP, Guinney J, Crawford GE, Furey TS. F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics.* 2008; 24:2537–2538. [PubMed: 18784119]
- Chenchik A, Diachenko L, Moqadam F, Tarabykin V, Lukyanov S, Siebert PD. Full-length cDNA cloning and determination of mRNA 5' and 3' ends by amplification of adaptor-ligated cDNA. *Biotechniques.* 1996; 21:526–534. [PubMed: 8879595]
- Chiang CM, Broker TR, Chow LT. An E1M--E2C fusion protein encoded by human papillomavirus type 11 is a sequence-specific transcription repressor. *J. Virol.* 1991; 65:3317–3329. [PubMed: 1851879]
- Dollard SC, Wilson JL, Demeter LM, Bonnez W, Reichman RC, Broker TR, Chow LT. Production of human papillomavirus and modulation of the infectious program in epithelial raft cultures. *Off. Genes Dev.* 1992; 6:1131–1142. [PubMed: 1321068]
- Doorbar J. The papillomavirus life cycle. *J. Clin. Virol.* 2005; 32(Suppl 1):S7–15. [PubMed: 15753007]
- Drury SE, Gough RE, McArthur S, Jessop M. Detection of herpesvirus-like and papillomavirus-like particles associated with diseases of tortoises. *Vet. Rec.* 1998; 143:639. [PubMed: 9881444]
- Guo M, Gong Y, Deavers M, Silva EG, Jan YJ, Cogdell DE, Luthra R, Lin E, Lai HC, Zhang W, Sneige N. Evaluation of a commercialized in situ hybridization assay for detecting human papillomavirus DNA in tissue specimens from patients with cervical intraepithelial neoplasia and cervical carcinoma. *J. Clin. Microbiol.* 2008; 46:274–280. [PubMed: 17977987]
- Herbst LH, Lenz J, Van DK, Chen Z, Stacy BA, Wellehan JF Jr, Manire CA, Burk RD. Genomic characterization of two novel reptilian papillomaviruses, *Chelonia mydas* papillomavirus 1 and *Caretta caretta* papillomavirus 1. *Virology.* 2009; 383:131–135. [PubMed: 18973915]
- Hopman AH, Kamps MA, Smedts F, Speel EJ, Herrington CS, Ramaekers FC. HPV in situ hybridization: impact of different protocols on the detection of integrated HPV. *Int. J. Cancer.* 2005; 115:419–428. [PubMed: 15688369]
- Hummel M, Hudson JB, Laimins LA. Differentiation-induced and constitutive transcription of human papillomavirus type 31b in cell lines containing viral episomes. *J. Virol.* 1992; 66:6070–6080. [PubMed: 1326657]

- Jia R, Liu X, Tao M, Kruhlak M, Guo M, Meyers C, Baker CC, Zheng ZM. Control of the papillomavirus early-to-late switch by differentially expressed SRp20. *J. Virol.* 2009; 83:167–180. [PubMed: 18945760]
- Kawaji H, Frith MC, Katayama S, Sandelin A, Kai C, Kawai J, Carninci P, Hayashizaki Y. Dynamic usage of transcription start sites within core promoters. *Genome Biol.* 2006; 7:R118. [PubMed: 17156492]
- Kiyono T, Nagashima K, Ishibashi M. The primary structure of major viral RNA in a rat cell line transfected with type 47 human papillomavirus DNA and the transforming activity of its cDNA and E6 gene. *Virology.* 1989; 173:551–565. [PubMed: 2556842]
- Lange CE, Favrot C, Ackermann M, Gull J, Vetsch E, Tobler K. Novel snake papillomavirus does not cluster with other non-mammalian papillomaviruses. *Virol. J.* 2011; 8:436. [PubMed: 21910860]
- Li Y, Wang X, Ni T, Wang F, Lu W, Zhu J, Xie X, Zheng ZM. Human papillomavirus type 58 genome variations and RNA expression in cervical lesions. *J. Virol.* 2013; 87:9313–9322. [PubMed: 23785208]
- Longworth MS, Laimins LA. Pathogenesis of human papillomaviruses in differentiating epithelia. *Microbiol. Mol. Biol. Rev.* 2004; 68:362–372. [PubMed: 15187189]
- Majerciak V, Ni T, Yang W, Meng B, Zhu J, Zheng ZM. A viral genome landscape of RNA polyadenylation from KSHV latent to lytic infection. *PLoS. Pathog.* 2013; 9:e1003749. [PubMed: 24244170]
- Majerciak V, Yamanegi K, Zheng ZM. Gene structure and expression of Kaposi's sarcoma-associated herpesvirus ORF56, ORF57, ORF58, and ORF59. *J. Virol.* 2006; 80:11968–11981. [PubMed: 17020939]
- Maruyama K, Sugano S. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene.* 1994; 138:171–174. [PubMed: 8125298]
- McLaughlin-Drubin ME, Christensen ND, Meyers C. Propagation, infection, and neutralization of authentic HPV16 virus. *Virology.* 2004; 322:213–219. [PubMed: 15110519]
- Meyers C, Frattini MG, Hudson JB, Laimins LA. Biosynthesis of human papillomavirus from a continuous cell line upon epithelial differentiation. *Science.* 1992; 257:971–973. [PubMed: 1323879]
- Meyers C, Mayer TJ, Ozbun MA. Synthesis of infectious human papillomavirus type 18 in differentiating epithelium transfected with viral DNA. *J. Virol.* 1997; 71:7381–7386. [PubMed: 9311816]
- Middleton K, Peh W, Southern S, Griffin H, Sotlar K, Nakahara T, El Sherif A, Morris L, Seth R, Hibma M, Jenkins D, Lambert P, Coleman N, Doorbar J. Organization of human papillomavirus productive cycle during neoplastic progression provides a basis for selection of diagnostic markers. *J. Virol.* 2003; 77:10186–10201. [PubMed: 12970404]
- Mistry N, Wibom C, Evander M. Cutaneous and mucosal human papillomaviruses differ in net surface charge, potential impact on tropism. *Virol. J.* 2008; 5:118. [PubMed: 18854037]
- Munoz N, Castellsague X, de Gonzalez AB, Gissmann L. Chapter 1: HPV in the etiology of human cancer. *Vaccine.* 2006; 24(Suppl 3):S3–1–S310.
- Ni T, Corcoran DL, Rach EA, Song S, Spana EP, Gao Y, Ohler U, Zhu J. A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat. Methods.* 2010; 7:521–527. [PubMed: 20495556]
- Ni T, Yang Y, Hafez D, Yang W, Kiesewetter K, Wakabayashi Y, Ohler U, Peng W, Zhu J. Distinct polyadenylation landscapes of diverse human tissues revealed by a modified PA-seq strategy. *BMC. Genomics.* 2013; 14:615. [PubMed: 24025092]
- Ozbun MA, Meyers C. Characterization of late gene transcripts expressed during vegetative replication of human papillomavirus type 31b. *J. Virol.* 1997; 71:5161–5172. [PubMed: 9188583]
- Palermo-Dilts DA, Broker TR, Chow LT. Human papillomavirus type 1 produces redundant as well as polycistronic mRNAs in plantar warts. *J. Virol.* 1990; 64:3144–3149. [PubMed: 2159571]
- Pauws E, van Kampen AH, van de Graaf SA, de Vijlder JJ, Ris-Stalpers C. Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis. *Nucleic Acids Res.* 2001; 29:1690–1694. [PubMed: 11292841]

- Renaud KJ, Cowser LM. Characterization of human papillomavirus-11 mRNAs expressed in the context of autonomously replicating viral genomes. *Virology*. 1996; 220:177–185. [PubMed: 8659110]
- Sankovski E, Mannik A, Geimanen J, Ustav E, Ustav M. Mapping of betapapillomavirus human papillomavirus 5 transcription and characterization of viral-genome replication function. *J. Virol*. 2014; 88:961–973. [PubMed: 24198410]
- Tang S, Zheng ZM. Kaposi's sarcoma-associated herpesvirus K8 exon 3 contains three 5'-splice sites and harbors a K8.1 transcription start site. *J. Biol. Chem*. 2002; 277:14547–14556. [PubMed: 11832484]
- Wang HK, Duffy AA, Broker TR, Chow LT. Robust production and passaging of infectious HPV in squamous epithelium of primary human keratinocytes. *Genes Dev*. 2009a; 23:181–194. [PubMed: 19131434]
- Wang X, Meyers C, Wang HK, Chow LT, Zheng ZM. Construction of a full transcription map of human papillomavirus type 18 during productive viral infection. *J. Virol*. 2011; 85:8080–8092. [PubMed: 21680515]
- Wang X, Tang S, Le SY, Lu R, Rader JS, Meyers C, Zheng ZM. Aberrant expression of oncogenic and tumor-suppressive microRNAs in cervical cancer is required for cancer cell growth. *PLoS. ONE*. 2008; 3:e2557. [PubMed: 18596939]
- Wang X, Wang HK, Li Y, Hafner M, Banerjee NS, Tang S, Briskin D, Meyers C, Chow LT, Xie X, Tuschl T, Zheng ZM. microRNAs are biomarkers of oncogenic human papillomavirus infections. *Proc. Natl. Acad. Sci. U. S. A*. 2014; 111:4262–4267. [PubMed: 24591631]
- Wang X, Wang H-K, McCoy JP, Banerjee NS, Rader JS, Broker TR, Meyers C, Chow LT, Zheng ZM. Oncogenic HPV infection interrupts the expression of tumor-suppressive miR-34a through viral oncoprotein E6. *RNA*. 2009b; 15:637–647. [PubMed: 19258450]
- Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene*. 2007; 389:52–65. [PubMed: 17123746]
- Zheng ZM, Baker CC. Papillomavirus genome structure, expression, and post-transcriptional regulation. *Front Biosci*. 2006; 11:2286–2302. [PubMed: 16720315]
- Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques*. 2001; 30:892–897. [PubMed: 11314272]
- Zhumabayeva B, Diatchenko L, Chenchik A, Siebert PD. Use of SMART-generated cDNA for gene expression studies in multiple human tumors. *Biotechniques*. 2001; 30:158–163. [PubMed: 11196307]

Figure 1. Identification of HPV18 splice junctions by primer walking RT-PCR
(A) The primers designed for primer walking RT-PCRs to map RNA splicing of HPV18 late transcripts. Arrows below the line diagram for the linearized HPV18 genome are primer positions and orientations. (B) Primer walking RT-PCR products. Poly(A)⁺ total RNA isolated from 16-day-old, HPV18-infected rafts was amplified using three different pairs of HPV18-specific primers (Pr3599 plus Pr5793, Pr5939, or Pr5628). Products 1 to 4 were gel purified, cloned, and sequenced. Shown below the gel are the corresponding splice junctions identified by sequencing. The products 2 and 3 in the gel are major late transcripts (L1)

spliced from nt 3696 to 5613. This splice junction could be also verified by using a backward splice junction primer, Pr5628, of which the 3' end has 2 nt identical to nt 3696 to 3695, giving the product 4 in the gel. The product 1 in the gel, detected with a primer pair of Pr3599 and Pr5793, is a product spliced from nt 3786 to 5776. This minor transcript is detectable only with a primer pair of Pr3599 and Pr5793 because the 3'-end 3 nt of Pr5793 are identical to the sequences from nt 3786 to 3784 and thereby the Pr5793 primer can function as a splice junction 3786/5776 primer. This figure is modified with permission from a reference (Wang, et al., 2011).

Table 1
Primers used for 5' RACE to map an HPV18 TSS and 3' RACE to map an HPV18 pA cleavage site*

Forward primer		
nt 3976-3996	3'RACE mapping HPV18 pA cleavage sites for early transcripts	5'- TGTATGTGTGCTGCCATGTCC-3'
nt 7038-7056	3'RACE mapping HPV18 pA cleavage sites for late transcripts	5'- CGTCGCAAGCCCACCATAG-3'
Reverse primer		
nt 233-209	5'RACE mapping HPV18 TSS for early transcripts	5'- CTCTGTAAGTTCCAATACTGTCTTG-3'
nt 850-833	5'RACE mapping HPV18 TSS for late transcripts	5'- CTGGAATGCTCGAAGGTC-3'
nt 904-886	5'RACE mapping HPV18 TSS for late transcripts	5'- CACTGAGGTAC/CTGCTGGGATGCACACCAC-3'
nt 3517-3500	5'RACE mapping HPV18 splice junctions	5'- ACGGACACGGTGCTGGAA-3'

* HPV18-specific primers (Wang, et al. 2011) being designed to meet regular primer criteria and not complementary to the 3'-end of the Universal Primer A Mix. The primers in the similar positions for other papillomaviruses can be designed accordingly.