HHS Public Access

Author manuscript

Curr Protoc Microbiol. Author manuscript; available in PMC 2017 May 06.

Published in final edited form as:

Curr Protoc Microbiol. ; 41: 14E.7.1–14E.7.18. doi:10.1002/cpmc.1.

PA-seq for global identification of RNA polyadenylation sites of Kaposi sarcoma-associated herpesvirus transcripts

Ting Ni[1,a], **Vladimir Majerciak**[2,a], **Zhi-Ming Zheng**[2], and **Jun Zhu**[3,#]

[1] Ministry of Education (MOE) Key Laboratory of Contemporary Anthropology & State Key Laboratory of Genetics Engineering, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai, P.R. China

[2] Tumor Virus RNA Biology Section, Gene Regulation and Chromosome Biology Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, MD, USA

[3] Systems Biology Center, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA

Abstract

Kaposi sarcoma-associated herpesvirus (KSHV) is a human oncovirus linked to the development of several malignancies in immunocompromised patients. Like other herpesviruses, KSHV has a large DNA genome encoding more than 100 distinct gene products. Despite being transcribed and processed by cellular machinery, the structure and organization of KSHV genes in the virus genome differs from that observed in cellular genes from the human genome. A typical feature of KSHV expression is the production of polycistronic transcripts initiated from different promoters but sharing the same polyadenylation site (pA site). This represents a challenge in determination of the 3' end of individual viral transcripts. Such information is critical for generation of a virus transcriptional map for genetic studies. Here we present PA-seq, a high-throughput method for genome-wide analysis of pA sites of KSHV transcripts in B lymphocyte with latent or lytic KSHV infection. Besides identification of all viral pA sites, PA-seq also provides quantitative information about the levels of viral transcripts associated with each pA site enabling to determine the relative expression levels of viral genes at various stages of infection. Due to the indiscriminate nature of PA-seq, the pA sites of host transcripts are also concurrently mapped in the testing samples. Therefore, this technology can simultaneously estimate the expression changes of host genes and RNA polyadenylation upon KSHV infection.

Keywords

polyadenylation; PA-seq; KSHV; herpesvirus; transcript

corresponding author.

[a]co-first authors

INTRODUCTION

Herpesviruses consist of a large group of animal viruses found in hosts ranging from mollusks to humans, and many herpesviruses are important pathogens. Kaposi sarcoma-associated herpesvirus (KSHV), also known as human herpesvirus 8 (HHV-8), is a *gamma*-herpesvirus, evolutionarily related to Epstein-Barr virus (EBV) (Chang et al., 1994; Russo et al., 1996), and like other herpesviruses, it establishes a life-long infection in the infected host. While asymptomatic in healthy individuals, KSHV infection in immunocompromised patients leads to development of several life-threatening malignancies including all forms of Kaposi sarcoma (a skin tumor of endothelial origin) and two rare B-cell lymphomas (primary effusion lymphoma and multicentric Castleman disease) (Chang and Moore, 1996; Cesarman et al., 1995; Soulier et al., 1995).

KSHV has a large double stranded DNA (dsDNA) genome of ~165 kb that encodes more than 100 gene products including protein-coding transcripts, several long non-coding RNAs and a set of viral microRNAs (Russo et al., 1996; Arias et al., 2014; Cai et al., 2005). KSHV exhibits two distinct life cycles; latent and lytic. During latent infection, KSHV expresses only a limited number of viral genes required for maintenance of the viral genome in the infected cells. Upon various cellular and environmental stresses, the latent virus can be reactivated into the lytic cycle to express all viral genes for active replication of viral genome and assembly of infectious virions.

All KSHV transcripts are transcribed by cellular RNA polymerase II (pol II) and their maturation requires multiple co-transcriptional and post-transcriptional processing steps including capping at the 5' end, splicing of introns and addition of a poly-A tail at the 3' end of primary transcripts. RNA polyadenylation is triggered when RNA pol II reads through a polyadenylation signal (PAS), usually an AAUAAA hexamer towards the 3' end of a gene, during ongoing transcription. Recruitment of CPSF (Cleavage and Polyadenylation Specificity Factor) to the PAS catalyzes the cleavage of the newly synthesized RNA, which occurs at a CA (cytosine-adenine) dinucleotide 10-30 nucleotides (nts) downstream of the PAS. Poly-A polymerase then synthesizes a poly-A tail at the cleavage site, also called the pA site. The newly synthesized poly-A tail recruits PABPs (polyadenylate binding proteins), which protect the mRNAs from degradation by cellular exonucleases. PABPs also interacts with eIF4G, a translational initiation factor bound to the RNA 5' Cap, to stimulate mRNA translation (Kahvejian et al., 2005).

In general, each eukaryotic gene contains one promoter and one (or often more) PASs downstream of the gene body. Herpesviruses and some other viruses often use a single PAS for polyadenylation of multiple collinear transcripts that originated from different promoters. The gene clusters KSHV ORF50/K8/K8.1 (Tang and Zheng, 2002), ORF73/ORF72/K13 (Ajiro and Zheng, 2014), ORF56/ORF57 and ORF58/ORF59/ORF60/ORF61/ORF62 (Majerciak et al., 2006) are examples of such a gene structure. The use of a single pA site for polyadenylation of multiple transcripts is also found in hepatitis B virus, adenoviruses, human papillomaviruses and others (Majerciak and Zheng, 2013; Wang et al., 2011). As a result, transcripts derived from the upstream genes in a given cluster span over several downstream genes in order to reach a PAS further downstream. This gene structure

represents a challenge in transcription mapping and causes a hurdle for mutational and knock-down studies because targeting a single gene locus may inadvertently affect the expression of other genes (Majerciak et al., 2007). Therefore, identification of the 3' end of individual viral transcripts is critical for understanding the expression, regulation and function of viral genes. Here we present a detailed step-by-step PA-seq protocol that has been successfully applied to identify both viral and host polyadenylation sites in three primary effusion B cell lines (JSC-1, BCBL-1 and BCBL1-TREx) undergoing latent or lytic infection. This method has been used to reconstruct a genome-wide landscape of all functional polyadenylation sites in the KSHV genome at single nucleotide resolution (Majerciak et al., 2013), and could be extended to monitor pA site usages of poly-A^+ transcripts in any organism of interest.

PA-SEQ PROTOCOL

Library construction of PA-seq

Historically, pA sites are determined in a gene-by-gene fashion using cDNA construction, 3' rapid amplification of cDNA ends (3' RACE), RNase protection assay (RPA) and other methods. These assays are relatively low throughput and often biased toward the most distal or prevalent pA site. In recent years, several high-throughput methods have been developed to determine pA sites at the transcriptome level (Jan et al., 2011; Mangone et al., 2010; Ozsolak et al., 2010; Hafez et al., 2013). The protocol presented here describes genome-wide analysis of pA sites based on paired-end sequence tags mapping to transcript 3' ends (Ni et al., 2013). Briefly, the poly-A^+ mRNA is randomly fragmented and reversely transcribed with a biotinylated oligo(dT) primer containing a deoxyuracil (dU) at the fourth position. The resulting cDNA are captured by streptavidin-coated magnetic beads and digested with USER enzyme that cleaves at dU, removing most of the poly-A tail. After end repair and A-tailing the Illumina-compatible Y linkers (or adaptors) are ligated to both ends of cDNA. The resulting library is PCR-amplified and subjected to paired-end sequencing. Raw sequence reads are then aligned to the reference genomes and pA sites are determined based on mapped reads containing a signature “TTT” sequence at their 5' ends. This method allows for strand-specific identification of polyadenylation cleavage site at single nucleotide resolution (see Figure 1A).

Materials

Primary effusion lymphoma cell lines- JSC-1 (Cannon et al., 2000), BCBL-1 (Renne et al., 1996), BCBL1-TREx-vector/Rta (Nakamura et al., 2003) cultivated in RPMI 1640 (ThermoFisher Scientific, cat. no. 22400-089) supplemented with 10% fetal bovine serum (FBS, Hyclone-GE Healthcare, cat. no. SH30070.03) and 1 × Penicillin-Streptomycin-Glutamine (PSG, ThermoFisher Scientific, cat. no. 10378-016). Cell lines could be obtained upon request from the corresponding authors.

Activators of KSHV lytic cycle: valproic acid, sodium salt (Sigma-Aldrich, cat. no. P4543), n-butyric acid, sodium salt (Sigma-Aldrich, cat. no. B-5887), doxycycline hyclate (Sigma-Aldrich, cat. no. D9891)

Phosphate-Buffered Saline, pH 7.4 (PBS, ThermoFisher Scientific, cat. no. 10010-023)

Nuclease-free 1.7 mL graduated microcentrifuge tubes (GeneMates, cat. no. C-3262-1)

MicroAmp Reaction Tube with Cap, 0.2 mL, autoclaved (ThermoFisher Scientific, cat. no. N8010612)

RNAlater RNA Stabilization Reagent (Qiagen, cat. no. 76104)

TRIzol reagent (Life Technologies, cat. no. 15596-026)

Chloroform (Sigma-Aldrich, cat. no. C2432)

Isopropanol (Sigma-Aldrich, cat. no. I9516)

GlycoBlue Coprecipitant (15 mg/mL, Ambion, cat. no. AM9516)

Nuclease-free water (Ambion, cat. no. AM9932)

RNeasy Mini kit (Qiagen, cat. no. 74104)

RNase-Free DNase Set (contains DNase I, RNase-free RDD buffer and RNase-free water, Qiagen, cat. no. 79254)

5 × fragmentation buffer (200 mM Tris acetate, pH 8.2, 500 mM potassium acetate and 150 mM magnesium acetate). Filter with 0.2 μm filter unit and shine the solution with UV (energy 9600 × 100 μJ/cm^2) using a UV crosslinker (Stratalinker model 1800, Stratagene) to make it RNase-free.

3 M sodium acetate (NaOAc, pH 5.2, RNase-free, ThermoFisher Scientific, cat. no. R1181)

DNA Clean & Concentrator kit (ZYMO Research, cat. no. D4014)

SuperScript II Reverse Transcriptase (ThermoFisher Scientific, the kit contains 5 × first strand buffer and 0.5 M DTT, cat. no. 18064-014)

Recombinant RNasin Ribonuclease Inhibitor (Promega, cat. no. N2515)

BdUT3VN primer: 5′-bio-T$_{16}$dUTTTVN-3′, ‘bio’denotes duo biotin group, ‘dU’ stands for deoxyuridine, ‘V’ represents any nucleotide except T and ‘N’ denotes any nucleotide. The primer was synthesized by Integrated DNA Technologies (IDT).

Agencourt RNAClean XP magnetic beads (Beckman Coulter, cat. no. A63987)

10 × second strand buffer (500 mM Tris-HCl, pH 7.8, 50 mM $MgCl_2$ and 10 mM DTT). The buffer is filtered with 0.2 μm filter unit and UV-treated without DTT (energy 9600 × 100 μJ/cm^2 with a Stratalinker model 1800 (Stratagene). Add DTT from a concentrated stock (Invitrogen, cat. no. R0861) to make final concentration of 10 mM.

10 mM dNTP mix, each (Bioline, cat. no. BIO-39029)

DNA Polymerase I (*E. coli*) (NEB, cat. no. M0209L)

RNase H (NEB, cat. no. M0297L)

10 mM Tris-HCl (pH 7.4)

100%, 80% and 70% ethanol

Dynabeads MyOne Streptavidin C1 magnetic beads (ThermoFisher Scientific, cat. no. 65001)

2 × Binding and Wash (B&W) Buffer for MyOne beads (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 2 M NaCl)

1 × Binding and Wash (B&W) Buffer for MyOne beads (5 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl)

APex Heat-Labile Alkaline Phosphatase (Epicentre, cat. no. AP49100)

TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0)

TE1 buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0)

USER (Uracil-Specific Excision Reagent) Enzyme (NEB, cat. no. M5505L)

MagneSphere Technology Magnetic Separation Stand, 1.5 ml (twelve-position) (Promega, cat. no. Z5342)

10 × NEB buffer 2 (NEB, cat. no. B7002S)

T4 DNA Polymerase (NEB, cat. no. M0203L)

dATP, 25 μM (Bioline, cat. no. BIO-39036)

Exo-Minus Klenow DNA Polymerase (Epicentre, cat. no. KL11101K)

Illumina-compatible indexed adapters (NEXTflex DNA Barcodes, BIOO Scientific, cat. no. 514104):

TruSeq Universal Adapter

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'

TruSeq Indexed Adapter

5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG-3' ('NNNNNN' denotes the index, of which each index has a unique sequence according to the vendor's manual)

T4 DNA ligase, high concentration (NEB, cat. no. M0202M)

Tween-20, for molecular biology (Sigma-Aldrich, cat. no. P9416-50ML)

E-Gel EX Agarose Gels, 2% (Invitrogen, cat. no. G4020-02)

Zymoclean Gel DNA Recovery Kit (ZYMO Research, cat. no. D4008)

Phusion High-Fidelity DNA Polymerase (NEB, cat. no. M0530L)

Illumina-compatible NEXTflex PCR Primer 1: 5′-AATGATACGGCGACCACCGAGATCTACAC-3′ (IDT, HPLC grade)

Illumina-compatible NEXTflex PCR Primer 2: 5′-CAAGCAGAAGACGGCATACGAGAT-3′ (IDT, HPLC grade)

TaqMan probe for β-actin (ACTB) (ThermoFisher Scientific, cat. no. 4331182, probe set Hs01060665_g1)

TaqMan Gene Expression Master Mix (ThermoFisher Scientific, cat. no. 4369016)

Instruments

Nanodrop (ThermoFisher Scientific, cat. no. A30221) or equivalent Qubit 3.0 Fluorometer (ThermoFisher Scientific, cat. no. Q33216)

TissueLyser LT (Qiagen, cat. no. 85600)

Eppendorf Thermomixer (Eppendorf, cat. no. 22670107)

Veriti PCR Thermocycler (ThermoFisher Scientific, cat. no. 4375786)

StepOnePlus Real-Time PCR System (ThermoFisher Scientific, cat. no. 4376600)

HiSeq 2500 Ultra-High-Throughput Sequencing System (Illumina)

Recommended software

Burrows-Wheeler Aligner, free download from http://bio-bwa.sourceforge.net/

SAMtools, free download from http://samtools.sourceforge.net/

F-Seq, free download from http://fureylab.web.unc.edu/software/fseq/

Integrative Genomics Viewer (IGV), free download from https://www.broadinstitute.org/igv/

Isolation of total RNA from cells or tissues

1. Culture PEL cells in complete RPMI 1640 medium (RPMI 1640 supplemented with 10% FBS and 1 × PSG). Induce KSHV lytic cycle by sodium butyrate for JSC-1 cells at a final concentration of 3 mM, by 1 mM sodium valproate for BCBL-1 cells, or by 1 μg/ml of doxycycline for BCBL1-TREX-Rta cells. After

48-hour (h) incubation, harvest the treated cells and wash one time with cold PBS followed by 5 min centrifugation at 1000 × g. Harvest the cells without chemical induction or doxycycline-treated control BCBL1-TREx-vector cells representing latently infected cells in parallel. Transfer the cells resuspended in PBS to a 2-ml Eppendorf tube and wash again by 5 min centrifugation at 1000 × g. Discard the supernatant, and resuspend the cells with residual buffer left over the pellet.

This protocol can work with any uninfected and infected primary cells, cultured primary cells, cultured cell lines or fresh tissues that intact total RNA can be isolated. Animal tissues should be quickly stored in liquid nitrogen or the RNAlater reagent until RNA isolation.

2. Add 1 ml TRIzol reagent directly to the pelleted cells, and isolate total RNA according to the manufacturer's protocol.

If tissues are used for total RNA purification, add 1 ml TRIzol to ~20 mg (but less than 100 mg) of tissues in a 2-ml Eppendorf tube and then homogenize the tissues with one steel bead using Qiagen TissueLyser for 2 min at 50 oscillations per second. Transfer the homogenized tissue suspension into another fresh 2-ml Eppendorf tube and perform RNA isolation according to the manufacturer's manual.

3. If more than 5 μg of total RNA is obtained, RNeasy MinElute cleanup kit with an on-column DNase I digestion step is recommended to remove potential genomic DNA contamination according to the manufacturer's protocol. However, if less than 5 μg of total RNA is obtained, in-solution DNase I digestion coupled with magnetic beads purification protocol is recommended as below:

- **a.** For a 40-μl reaction, mix 35 μl of total RNA, 4 μl of RDD buffer and 1 μl of DNase I (both DNase I and RDD buffer were from Qiagen RNase-Free DNase Set) and incubate at room temperature (RT) for 15 min.
- **b.** Add 80 μl of Agencourt RNAClean XP magnetic beads (abbreviate as 'RNAClean beads' hereafter) to the reaction and mix thoroughly. Incubate the mixture at RT for 12 min and put it on magnetic stand for 5 min.
- **c.** Carefully remove all liquid followed by two consecutive wash with 300 μl of freshly-made 70% ethanol.
- **d.** After removing the 70% ethanol, dry the beads in a clean hood for 3-5 min until no visible liquid remaining.
- **e.** Add 20 μl of RNase-free water to the magnetic beads and mix thoroughly. Incubate the tube at room temperature for 1 min without magnetic rack, followed by another minute on a magnetic

stand. Transfer the clear solution into an RNase-free tube and quantify the RNA concentration by NanoDrop Spectrophotometer.

Fragmentation of total RNA

4. Add the following components to a PCR tube:	
Purified total RNA from step 3e	2-10 μg
5 × fragmentation buffer	6 μl
Nuclease-free water	bring up to 30 μl (total reaction volume)
Mix by gentle pipetting.	

5. Incubate at 94 °C for 3 min in a thermocycler. Immediately transfer the tube to ice after the incubation.

Purification of fragmented RNA

6. The fragmented RNA is subject to either ethanol precipitation or bead purification.

7. **I. Ethanol precipitation protocol:**

a. After placing the PCR tube on ice for 2 min, directly add 70 μl of RNase-free water to bring the total volume to 100 μl.

b. Add 10 μl of 3 M NaOAc (pH 5.2), 2 μl of GlycoBlue and 250 μl of 100% ethanol to the mixture.

c. Mix the solution and place the tube on dry ice for 30 min.

d. Spin the tube at 10,000 × g for 1 h at 4 °C.

e. Wash the pellet with 500 μl of 70% ethanol.

f. Dry the pellet in a clean hood and resuspend in 10 μl of nuclease-free water.

II. Beads purification protocol (Alternative for automation):

a. Add 1.8 × RNAClean beads to the 30 μl reaction. Mix well and incubate at RT for at least 5 min.

b. Place the tube on a magnetic stand for 5 min or until the solution becomes clear.

c. Remove the supernatant and wash the beads twice with 100 μl of freshly made 70% ethanol, which is added while the tube remains on magnetic stand. During each wash step, incubate at room temperature for 30 seconds (sec) before removing the liquid.

d. After the second wash, briefly spin the tube and put it back to magnetic stand. Remove residual ethanol with a 10-μl pipette, and dry the beads at room temperature for 10 min.

e. Add 10 μl of nuclease-free water and mix well by pipetting up and down ten times. Put the tube back to magnetic stand.

f. Transfer the supernatant to a fresh PCR tube after the solution becomes clear.

Note: All PCR tubes should be RNase-free.

Reverse transcription

8. Add the following components to a PCR tube:	
RNA from step 7	10 μl
10 mM dNTP (each)	1 μ l
10 μM BdUT3VN primer	1 μ l

9. Incubate the mixture at 65 °C for 5min. Immediately transfer the tube to ice and keep on ice for at least 1 min.

10. Add additional components to the reaction:	
5 × first strand buffer	4 μl
0.1 M DTT	2 μl
RNasin Ribonuclease Inhibitor	1 μl

11. Incubate at 42 °C for 2 min. Add 1 μl of SuperScript II (SSII) reverse transcriptase and mix well when the reaction is on the thermocycler. Close the cap of the PCR tube and incubate at 42 °C for 60 min, 70 °C for 15 min. Hold at 10 °C.

Purification of reversely transcribed cDNA

12. Isolate the reversely transcribed cDNA by bead purification following the same protocol in step 7II except the RNAClean bead to reaction ratio is 1.5 × (v/v).

13. Add 41 μl of nuclease-free water to beads after the final wash to elute first-stand cDNA.

Second strand synthesis

14. Add the following components to a PCR tube:	
First strand cDNA from step 13	40.5 μl
10 × second strand buffer	5 μl
10 mM dNTP (each)	1.5 μ l

15. Mix well and incubate on ice for 5 min.

16. Transfer the PCR tube to a PCR cooler and add following enzymes to the tube:	
DNA polymerase I (*E.coli*)	2.5 μl
RNase H	0.5 μl

17. Mix well and incubate at 15 °C for 2.5 h in a thermocycler. Disable the heated lid function and do not let the temperature rise above 15 °C.

Note: Transfer the reaction into a 4 °C refrigerator if one wants to hold at this step.

Pulling down of double stranded cDNA with MyOne magnetic beads

18. Aliquot 25 μl of Dynabeads MyOne Streptavidin C1 magnetic beads ('MyOne Beads' for abbreviation) into a fresh 1.7-ml microcentrifuge tube.

19. Wash MyOne Beads twice with 100 μl of 2 × Binding & Wash buffer ('B&W' buffer for abbreviation); Resuspend the beads in 50 μl of 2 × B&W buffer to a final concentration of 5 μg/μl (twice the original volume) in a 1.7-ml microcentrifuge tube. [Binding capacity: 100 μl beads bind 20 μg DNA]

20. Add 50 μl of Second Strand Synthesis reaction from step 17 to 50 μl of resuspended MyOne beads.

The resulting final NaCl concentration is 1M, for optimal binding.

21. Incubate the mixture for 15 min at room temperature with gentle rotation.

22. Separate the biotinylated cDNA coated beads with a magnetic stand for 2-3 min.

23. Wash 2-3 times with 100 μl of 1 × B&W buffer.

24. Wash once with 100 μl of 10 mM Tris-HCl (pH 7.4), and resuspend the MyOne beads in 44 μl of 10 mM Tris-HCl (pH 7.4).

Removal of phosphate group with APex Heat-Labile alkaline phosphatase

25. Add 5 μl of 10 × APex buffer and 1 μl of APex heat-labile alkaline phosphatase to the 44-μl sample from step 24.

26. Incubate at 37 °C for 10 min.

27. Heat at 70 °C for 5 min to inactivate APex heat-labile alkaline phosphatase.

28. Use magnetic rack to pull down the MyOne beads.

29. Wash twice with 1 × B&W buffer and once with 10 mM Tris-HCl (pH 7.4).

30. Resuspend the beads in 28 μl of TE1 buffer.

USER enzyme digestion to release double stranded cDNA

31. Add 2 μl of USER enzyme to the reaction tube from step 30, which contains both beads and solution.

32. Incubate at 37 °C for 1 h with constant shaking in an Eppendorf ThermoMixer to release double-stranded cDNA from the MyOne beads.

33. Put the tube back to magnetic stand. The solution becomes clear within 1 min. Transfer the supernatant to a fresh 1.7-ml microcentrifuge tube.

34. Purify the eluted DNA by ZYMO DNA clean & concentrator-5 kit and elute the purified products with 10 μl of nuclease-free water according to the manufacturer's protocols. Alternatively, column purification can also be replaced by RNAClean bead purification according to step 7II.

Quality control with qPCR

35. A quality control (QC) step is recommended to confirm the release of cDNA products from MyOne beads (and to estimate the overall cDNA yield). Take 1 μl of the eluted DNA and check the expression level of β-actin (or other house-keeping genes) by quantitative real-time PCR. Follow the procedure provided by ThermoFisher where the qPCR probe set and master mixes can be ordered. The typical C_t value we obtained is ~17 cycles and the negative control (water) shows a C_t value of around 35 cycles.

Blunt double stranded cDNA with T4 DNA polymerase

36. Add the following components to a new PCR tube:	
DNA from step 34	9 μl
10 × NEB buffer 2	5 μl
10 mM dNTP (each)	1 μl
T4 DNA polymerase	1 μl
Nuclease-free water	33.5 μl

37. Mix well and incubate at 15 °C for 15 min.

38. Purify the reaction by ZYMO DNA clean & concentrator-5 kit and elute with 40 μl of nuclease-free water according to the manufacturer's protocol. Alternatively, column purification can also be replaced by RNAClean bead purification according to step 7II.

A-tailing

39. Add the following components in a 1.7-ml tube:	
DNA from step 38	40 μl
10 × Exo-Minus Klenow buffer	5 μl
2.5 mM dATP	4 μl
Exo-Minus Klenow DNA polymerase	1 μl

40. Incubate at 37 °C for 30 min.

41. Purify the reaction by ZYMO DNA clean & concentrator-5 kit and elute with 8 μl of nuclease-free water according to the manufacturer's protocol. Alternatively, column purification can also be replaced by RNAClean bead purification according to step 7II.

Ligation with Illumina-compatible indexed adapter

42. Add the following components to a PCR tube:	
DNA from step 41	7 µl
1.5 µM Illumina-compatible indexed adapter (BIOO Scientific)	1 µl
10 × T4 DNA ligase buffer	1 µl
T4 DNA ligase, high concentration	1 µl

43. Incubate at room temperature for 30 min.

44. Purify the ligation products by ZYMO DNA clean & concentrator-5 kit and elute with 20 µl of nuclease-free water according to the manufacturer's protocol. Run the products on a 2% E-Gel EX agarose gel to select the band at 250-450 bp range. Purify ligated products from gel slices with Zymoclean Gel DNA Recovery Kit and elute with 20 µl of nuclease-free water according to the manufacturer's protocol.

(1) Use TE1 buffer to dilute Illumina-compatible indexed adapters to 1.5 µM.

(2) If multiple PA-seq libraries are sequenced in one lane of an Illumina flowcell, add different index primers to different samples. The resulting PA-seq libraries can then be pooled together for multiplex sequencing. Use different indexed adaptors for different samples to allow for sequencing multiple PA-seq libraries in the same lane.

(3) A 3-step bead purification procedure can be used to replace gel purification as following:

- **a.** Add equal volume (10 µl) of RNAClean beads to the ligation reaction. Incubate at room temperature for 2-3 min.
- **b.** Collect the beads on a magnetic stand. Carefully remove and discard the supernatant.
- **c.** Elute with 10 µl of TE buffer.
- **d.** Purify the ligation products again using 1.4 volumes of RNAClean beads (14 µl) and elute with 10 µl of TE buffer.
- **e.** Add exactly 1 volume of RNAClean beads (10 µl) to the eluted DNA and incubate at RT for 12 min.
- **f.** Pull the beads to the side of the plate/tube on a magnetic stand.
- **g.** Wash twice with freshly made 70% ethanol without disturbing the beads.
- **h.** Elute the DNA with 20 µl of nuclease-free water.

  Note 1. This is a possible stopping point. The purified DNA can be stored at − 20 °C.

Note 2. Instead of using one gel-cutting step to remove adaptor contamination, the bead purification strategy needs 3 rounds of size selection to eliminate most of the Y-shaped adaptors.

PCR amplification of PA-seq library

45. Add the following components to a PCR tube:	
DNA from step 44	20 μl
5 × Phusion HF buffer	10 μl
10 mM dNTP (each)	1 μl
10 μM NEXTflex PCR Primer 1	1 μl
10 μM NEXTflex PCR Primer 2	1 μl
Phusion High-Fidelity DNA polymerase	0.5 μl
Nuclease-free water	16.5 μl

46. Perform initial denaturation at 98 °C for 30 sec, followed by 14-16 cycles of amplification (98 °C for 10 sec, 65 °C for 30 sec, 72 °C for 30 sec) and one cycle of extension step (72 °C for 5 min). Hold the reaction at 10 °C.

Final PA-seq library purification

47. Load 3 μl of PCR reaction onto a 2% agarose gel for a quick quality check. If a smear of 300-500 bp can be detected, load 20 μl of PCR reaction to a 2% agarose EX E-gel and excise 300-500 bp smear (see Figure 1B). Purify with ZYMO gel recovery kit and elute the final library in 20 μl of nuclease-free water. Add 2 μl of 1% Tween-20 to the final library (final concentration is ~ 0.1% Tween-20) to prevent DNA loss during prolonged (or long-term) storage at −20 °C.

Note: If no band can be seen after 16 cycles of amplification, add 2-4 more cycles and check again.

Illumina sequencing and data collection

48. Quantify the PA-seq libraries using a Qubit fluorometer and sequence using the Illumina HiSeq2000 platform with a 2 × 50 bp modality following the manufacturer's procedure. Other instruments can also be used based on availability. Pre-process the raw data to generate sequence reads in fastq format. An average sequencing depth is ~10-20 million paired reads for individual PA-seq libraries.

The sequencing step is typically handled by core facilities or commercial service providers.

Mapping and pA site determination

49. Map the raw sequencing reads to the corresponding genome using the BWA aligner (Li and Durbin, 2009) allowing a maximum of two mismatches. For our study, we sequentially aligned the reads to the KSHV genome (GenBank acc. no. U75698.1), the EBV wt strain genome (GenBank acc. no. NC007605.1) and the human genome (UCSC version hg19). All uniquely mapped KSHV-specific reads

were used for further analyses, including peak calling and expression estimation. The IGV genome browser was used for data visualization (see Figure 2).

50. Perform peak calling with the F-Seq method (parameters: feature length = 30 and threshold = 10, and we recommend resizing the clusters to the shortest distance that contained 95% of the reads) to determine individual pA sites for transcripts from both KSHV and host cells on combined libraries (see Figure 3). Two additional filters should be applied to remove unreliable pA sites. (1) Peaks with less than 50 tags in combined libraries are considered as low expression and subsequently removed. (2) Filter out pseudo pA sites resulting from "internal priming" due to a stretch of continuous "A"s in the corresponding genome.

51. Bona fide peaks that survive the stringent filtering steps serve as a "guide" to assign sequence reads to corresponding peaks in individual libraries. To estimate the relative expression level of each poly-A^+ transcript of KSHV, the read counts of individual peaks should be normalized to overall reads mapped to the KSHV and human genomes, and the results are presented as count per million reads (CPM).

Validation

Our previous studies have demonstrated the high accuracy, sensitivity and reliability of the PA-seq for detecting pA sites (Ni et al., 2013; Majerciak et al., 2013) (see Figure 4). If necessary, the identified pA sites can be validated at the gene level by several methods, including but not limited to 3' RACE and RNase protection assays (RPA).

RACE was originally designed to determine the ends of cDNA generated by reverse transcription. 3' RACE is a modified version allowing for specific amplification of cDNA 3' ends. It consists of two major steps: (1) reverse transcription; (2) PCR amplification. Reverse transcription is performed using a hybrid primer consisting of an oligo dT sequence, which binds to a mRNA's poly-A tail, fused with a generic adaptor sequence. The resulting cDNA is used for PCR amplification using a gene-specific forward primer and a reverse primer complementary to the adaptor sequence introduced to cDNA by reverse transcription. If no specific product is detected after the first amplification, a second PCR step can be performed using a nested primer set. The resulting PCR products are then cloned into a vector by TA cloning or similar methods. Multiple colonies are then picked for sequencing. As a result, pA sites are determined based on the sequence immediately upstream of the poly-A tail in each clone. The same strategy can be used to determine multiple pA sites of a given gene by targeting different gene-specific regions. Versatility and high sensitivity are among the benefits of 3' RACE. However, due to non-linear PCR amplification it provides only semi-quantitative information about the abundance of the RNA tested in the sample. Several 3' RACE kits are commercially available, such as SMARTer 5'/3' RACE Kit (Takara/Clontech, cat. no. 634858). Also see Wang and Zheng, 2016 for detailed protocol.

RPA (RNase protection assay) detects RNA or its fragment(s) using antisense riboprobe(s) complementary to a transcript of interest. The riboprobe is synthesized by *in vitro* transcription from a PCR or linearized plasmid template. During the probe synthesis, radioactive nucleotides are typically incorporated for subsequent detection. Several non-radioactive labeling alternatives are also available. The purified probe is mixed and

hybridized with the RNA sample, followed by RNase digestion to remove unprotected probes. The remaining probes are protected either entirely or partially along their length by complementary RNAs, and are then visualized and quantified on a denaturing PAGE (polyacrylamide gel electrophoresis). For pA site mapping, it is recommended to design probes that extend beyond the predicted pA site allowing distinction of shorter cleavage product from full-length undigested probe. The concomitant running of sequence markers prepared by Sanger sequencing with radiolabeled primer on the same gel allows determining pA site at single nucleotide resolution (Majerciak et al., 2006: Wang and Zheng, 2016; see UNIT 14B.6). In contrast to 3' RACE, RPA represents a non-amplification method that provides quantification of specific transcripts in the RNA sample. In addition to the use of radioisotope, RPA is labor intensive and time consuming, thereby limiting its broad usage for global pA site mapping. Nontheless, the use of Riboprobe System-T7 (Promega, cat. no. P1440) and RPAIII Ribonuclease Protection Assay (Ambion, cat. no. AM1415) provides satisfactory results.

COMMENTARY

RNA polyadenylation consists of two major enzymatic reactions: an endonucleolytic cleavage of primary transcripts followed by covalent addition of a poly-A tail at the 3' end of cleavage products (Proudfoot, 2011). Polyadenylation at a specific site is determined by its flanking sequences, which are recognized by a multiprotein complex consisting of more than 90 different factors in human (Mandel et al., 2008). First, an A/U-rich element (AWUAAA) upstream of pA sites, also called PAS, is recognized by CPSF (Cleavage and Polyadenylation Specificity Factor), initiating the binding of other factors including CStF (Cleavage Stimulatory Factor) that interacts with the GU- or U-rich sequences downstream of the cleavage site or downstream sequence element, DSE (Wilusz et al., 1990; Takagaki and Manley, 1997; Murthy and Manley, 1995). Then, canonical PAP (Poly A-Polymerase) is recruited to mediate non-templated synthesis of the poly-A tail in a size of up to ~250 nts.

The efficiency of polyadenylation is primarily regulated by sequence conservation of PAS and DSE (Hafez et al., 2013; Majerciak et al., 2013). Approximately 70% of the human transcripts employ canonical PAS AAUAAA followed by less frequent variants. However ~4% of transcripts do not have an identifiable PAS (Beaudoing et al., 2000; Tian et al., 2005). DSE is less conserved both in sequence and length. Other distal elements, such as U-rich element located immediately upstream of PAS (USE, upstream sequence element), can also enhance polyadenylation efficiency (Gil and Proudfoot, 1987). The distance between PAS and DSE that dictates the cleavage site varies from a few to several tens of nucleotides. The sequence at the cleavage site is not conserved, but the cleavage preferably occurs at CA dinucleotide (Chen et al., 1995).

We have applied the aforementioned PA-seq method to determine a genome-wide landscape of the pA sites of KSHV transcripts. Our results showed that the KSHV genome utilizes 64 distinct pA sites to express various genes (Majerciak et al., 2013). The pA sites identified by PA-seq have been successfully validated by 3' RACE and agreed well with previously reported pA sites, confirming the high accuracy of PA-seq method. In addition, we identified several unannotated pA sites, which help us uncover several new viral transcripts. Further

analysis of the pA sites identified in the context of viral genome reaffirmed the notion of prevalent usage of a single pA site by multiple genes. Indeed, more than half of all identified pA sites are located downstream of the gene clusters containing two or more viral genes. Interestingly, with the exception of few transcripts such as lncRNA T1.5, most viral transcripts are terminated at a single pA site. These results indicate that, in contrast to cellular transcripts, alternative polyadenylation (APA) is rather rare in KSHV expression.

Quantitative analysis of the sequence tags associated with individual pA site provides further insight into the regulation of KSHV polyadenylation. The pA site usage was found to closely reflect the relative expression of corresponding transcripts, allowing the detection of expression changes of viral genes associated with virus reactivation among individual PEL cell lines. We found that the expression ratio of each viral gene from the latent to lytic stage is a much better measurement than the overall expression measurement for the latent genes. This is because a few cells in the latent stage could be spontaneously reactivated to express the viral lytic genes.

Because PA-seq library constructed from each cell/tissue sample contains all polyadenylated transcripts. In the study mentioned above, non-KSHV transcripts and KSHV transcripts can be interrogated in parallel (Majerciak et al., 2013). For example, by analyzing sequence reads obtained from KSHV/EBV double-infected JSC-1 cells, we were able to determine pA sites and expression profile of EBV transcripts. Therefore, PA-seq is powerful for simultaneous analysis of multiple pathogens during co-infection. Moreover, by mapping sequence reads to the human genome, we can carry out a genome-wide analysis of cellular gene expression and pA site usage from viral latent to lytic infection. Since PA-seq uses an oligo(dT) primer for cDNA synthesis, it is limited to detection of polyadenylated transcripts. Although polyadenylation is ubiquitous in mRNAs generated by RNA polymerase II, other functional transcripts including ribosomal RNAs synthesized by RNA polymerase I and III, small RNAs produced by RNA pol III, replication-dependent histone mRNAs, and some long non-coding RNAs lack a poly-A tail (Yang et al., 2011). Some viral transcripts also lack a poly-A tail, such as EBV EBER RNAs. Thus, they are not detected by our PA-seq method.

Critical Parameters

RNA quality—Although RNA quality is essential for regular RNA-seq protocols, it's not as important for PA-seq since total RNA will be fragmented at the first step of the library construction and only the 3′ end of the fragment of the RNA will eventually be sequenced. However, PA-seq can globally identify polyadenylation sites not only qualitatively but also quantitatively (Ni et al., 2013; Majerciak et al., 2013). PA-seq library constructed with high quality RNA would reflect the expression level more accurately and in particular provide more insights in the regulation of alternative polyadenylation (APA).

QC steps before deep sequencing—We recommend two QC steps during PA-seq library preparation. The first QC step is to estimate the yield of double stranded cDNA after USER digestion (see Step 35.). Such information is critical for assessing the relative efficiency of all steps previously performed. The second QC step is devised after the final

library PCR (See Step 47.). Typically, 14 cycles of PCR is required for library amplification if starting with 10 μg of total RNA. If less total RNA (e.g. 2 μg) is used, 16 PCR cycles might be required. A uniform smear is a good indicator of high quality PA-seq library (see Figure 1B). In addition, further validation can be achieved by sequencing a few clones by Sanger's sequencing, especially for the beginners who do not have prior experience. Alternatively, one can use Illumina MiSeq sequencer (or equivalent platforms) to determine library quality before large-scale data acquisition.

Troubleshooting

1. Low amount of double stranded cDNA at the first QC step: Either RNA is lost or the reverse transcription efficiency is low. Although the concentration of total RNA is quantified, the concentration of RNA after DNase I treatment should also be evaluated. If RNase was contaminated in the DNase I, the RNA will be significantly reduced after removal of genomic DNA. One can also perform qPCR right after the reverse transcription step to determine whether enough cDNA is generated or estimate reverse transcription efficiency.

2. Strong adaptor ligation band and weak 300-500 bp smear when checking final PCR products by agarose gel electrophoresis: A strong adaptor ligation band indicates low cDNA input and/or low efficiency of removing excessive adaptors after ligation. Make sure 70% and 80% ethanol is freshly made and the trace amount of liquid is completely removed during each bead purification step. To avoid adaptor contamination, larger size PCR fragments can be excised during the gel purification step. If possible, increasing the amount of input RNA is recommended. Using 10 μg total RNA is expected to have a higher signal-to-noise ratio than that of 2 μg of total RNA.

Anticipated Results—In order to map the 3′ ends of poly A^+ RNA, the most important step is to construct a library suitable for sequencing the very end of individual transcripts. We simplified the protocol by skipping the step of poly A^+ RNA enrichment, either using oligo(dT) or the Ribozero kit to remove rRNA. All column purification steps can be replaced by bead purification; thus, it is possible to automate the entire library construction procedure. For a typical high-quality PA-seq library, approximately 80% of the raw reads are expected to overlap with or in close proximity to known PA sites.

Time Considerations—It typically requires 2-3 days to complete the entire PA-seq procedures. Isolation of total RNA from cells or tissues requires 2-3 hours. Fragmentation of total RNA only takes a few minutes; however, purification of fragmented RNA and reverse transcription takes approximately 2.5 hours. Second strand synthesis requires 3 hours. These steps can be accomplished in the first day. In the second day, it requires 1 hour to pull down double-stranded cDNA with MyOne magnetic beads. Removal phosphate group with APex heat-labile alkaline phosphatase takes 1 hour, followed by USER enzyme digestion, which takes 1.5 hours. Quality control with qPCR requires another 1.5 hours. It is expected that 2 hours are required to complete blunt double stranded cDNA with T4 DNA polymerase, A-tailing and ligation with Illumina-compatible indexed adapter. Final PCR and purification

take 1 hour. Two days are likely sufficient if 1-8 samples are prepared. To prepare more samples in parallel, one additional day is usually required.

Acknowledgements

This work is supported by the Intramural Research Program of National Heart, Lung, and Blood Institute and National Cancer Institute of NIH (USA), National Key Basic Research Program of China (973 program: 2013CB530700; 2015CB943000), National Science Foundation of China (31271348; 31471192), the Thousand Talents Plan of China and the 111 Project of China (B13016).

Literature Cited

Ajiro M, Zheng ZM. Oncogenes and RNA splicing of human tumor viruses. Emerg.Microbes.Infect. 2014; 3:e63. [PubMed: 26038756]

Arias C, Weisburd B, Stern-Ginossar N, Mercier A, Madrid AS, Bellare P, Holdorf M, Weissman JS, Ganem D. KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. PLoS.Pathog. 2014; 10:e1003847. [PubMed: 24453964]

Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. Genome Res. 2000; 10:1001–1010. [PubMed: 10899149]

Cai X, Lu S, Zhang Z, Gonzalez CM, Damania B, Cullen BR. Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells. Proc.Natl.Acad.Sci.U.S.A. 2005; 102:5570–5575. [PubMed: 15800047]

Cannon JS, Ciufo D, Hawkins AL, Griffin CA, Borowitz MJ, Hayward GS, Ambinder RF. A new primary effusion lymphoma-derived cell line yields a highly infectious Kaposi's sarcoma herpesvirus-containing supernatant. J. Virol. 2000; 74:10187–10193. [PubMed: 11024147]

Cesarman E, Chang Y, Moore PS, Said JW, Knowles DM. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas. N.Engl.J.Med. 1995; 332:1186–1191. [PubMed: 7700311]

Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, Moore PS. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science. 1994; 266:1865–1869. [PubMed: 7997879]

Chang Y, Moore PS. Kaposi's Sarcoma (KS)-associated herpesvirus and its role in KS. Infect.Agents Dis. 1996; 5:215–222. [PubMed: 8884366]

Chen F, MacDonald CC, Wilusz J. Cleavage site determinants in the mammalian polyadenylation signal. Nucleic Acids Res. 1995; 23:2614–2620. [PubMed: 7651822]

Gil A, Proudfoot NJ. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. Cell. 1987; 49:399–406. [PubMed: 3568131]

Hafez D, Ni T, Mukherjee S, Zhu J, Ohler U. Genome-wide identification and predictive modeling of tissue-specific alternative polyadenylation. Bioinformatics. 2013; 29:i108–i116. [PubMed: 23812974]

Jan CH, Friedman RC, Ruby JG, Bartel DP. Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs. Nature. 2011; 469:97–101. [PubMed: 21085120]

Kahvejian A, Svitkin YV, Sukarieh R, M'Boutchou MN, Sonenberg N. Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. Genes Dev. 2005; 19:104–113. [PubMed: 15630022]

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]

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, Pripuzova N, McCoy JP, Gao SJ, Zheng ZM. Targeted disruption of Kaposi's sarcoma-associated herpesvirus ORF57 in the viral genome is detrimental for the expression of ORF59,

K8alpha, and K8.1 and the production of infectious virus. J.Virol. 2007; 81:1062–1071. [PubMed: 17108026]

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]

Majerciak, V.; Zheng, ZM. Detection of Viral RNA Splicing in Diagnostic Virology. In: Tang, Y-W.; Stratton, CW., editors. In Advanced Techniques in Diagnostic Microbiology. Springer; New York: 2013. p. 693-748.

Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. Cell.Mol.Life Sci. 2008; 65:1099–1122. [PubMed: 18158581]

Mangone M, Manoharan AP, Thierry-Mieg D, Thierry-Mieg J, Han T, Mackowiak SD, Mis E, Zegar C, Gutwein MR, Khivansara V, Attie O, Chen K, Salehi-Ashtiani K, Vidal M, Harkins TT, Bouffard P, Suzuki Y, Sugano S, Kohara Y, Rajewsky N, Piano F, Gunsalus KC, Kim JK. The landscape of C. elegans 3'UTRs. Science. 2010; 329:432–435. [PubMed: 20522740]

Murthy KG, Manley JL. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. Genes Dev. 1995; 9:2672–2683. [PubMed: 7590244]

Nakamura H, Lu M, Gwack Y, Souvlis J, Zeichner SL, Jung JU. Global changes in Kaposi's sarcoma-associated virus gene expression patterns following expression of a tetracycline-inducible Rta transactivator. J.Virol. 2003; 77:4205–4220. [PubMed: 12634378]

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]

Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B, Milos PM. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell. 2010; 143:1018–1029. [PubMed: 21145465]

Proudfoot NJ. Ending the message: poly(A) signals then and now. Genes Dev. 2011; 25:1770–1782. [PubMed: 21896654]

Renne R, Zhong W, Herndier B, McGrath M, Abbey N, Kedes D, Ganem D. Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. Nat.Med. 1996; 2:342–346. [PubMed: 8612236]

Russo JJ, Bohenzky RA, Chien MC, Chen J, Yan M, Maddalena D, Parry JP, Peruzzi D, Edelman IS, Chang Y, Moore PS. Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). Proc.Natl.Acad.Sci.U.S.A. 1996; 93:14862–14867. [PubMed: 8962146]

Soulier J, Grollet L, Oksenhendler E, Cacoub P, Cazals-Hatem D, Babinet P, d'Agay MF, Clauvel JP, Raphael M, Degos L, Sigaux F. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease. Blood. 1995; 86:1276–1280. [PubMed: 7632932]

Takagaki Y, Manley JL. RNA recognition by the human polyadenylation factor CstF. Mol.Cell Biol. 1997; 17:3907–3914. [PubMed: 9199325]

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]

Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res. 2005; 33:201–212. [PubMed: 15647503]

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, Zheng ZM. Construction of a Transcription Map for Papillomaviruses. Curr. Protoc. Microbiol. 2016 Unit14B.6.

Wilusz J, Shenk T, Takagaki Y, Manley JL. A multicomponent complex is required for the AAUAAA-dependent cross-linking of a 64-kilodalton protein to polyadenylation substrates. Mol.Cell Biol. 1990; 10:1244–1248. [PubMed: 2304466]

Yang L, Duff MO, Graveley BR, Carmichael GG, Chen LL. Genomewide characterization of non-polyadenylated RNAs. Genome Biol. 2011; 12:R16. [PubMed: 21324177]

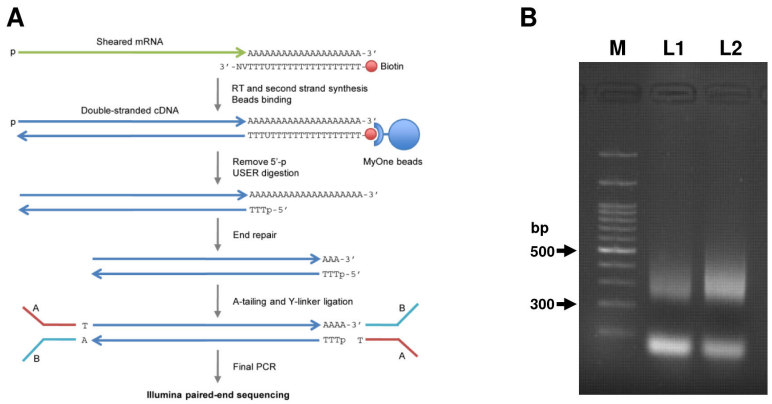

Figure 1. Generation of PA-seq libraries

(A) Diagram of PA-seq protocol *(with permission from Ni et al., 2013)*. (B) A typical agarose gel for PA-seq libraries. The Left lane shows DNA ladder (M) while the two right lanes (L1 and L2) are final PCR products from two samples.

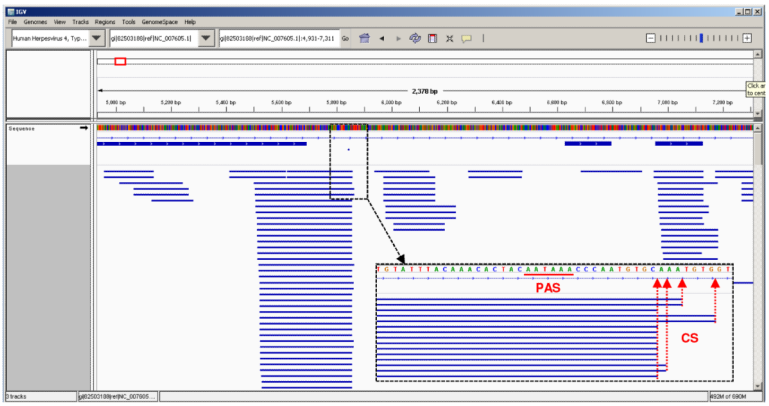

Figure 2. Mapping of PA-seq reads

Visualization viral-specific reads in IGV browser using ".bed" files. The inset zooms in the 3' end of reads corresponding to polyadenylation cleavage sites (CS) marked by dashed red arrows. The upstream sequence underlined in red is a canonical "AATAAA" polyadenylation signal (PAS).

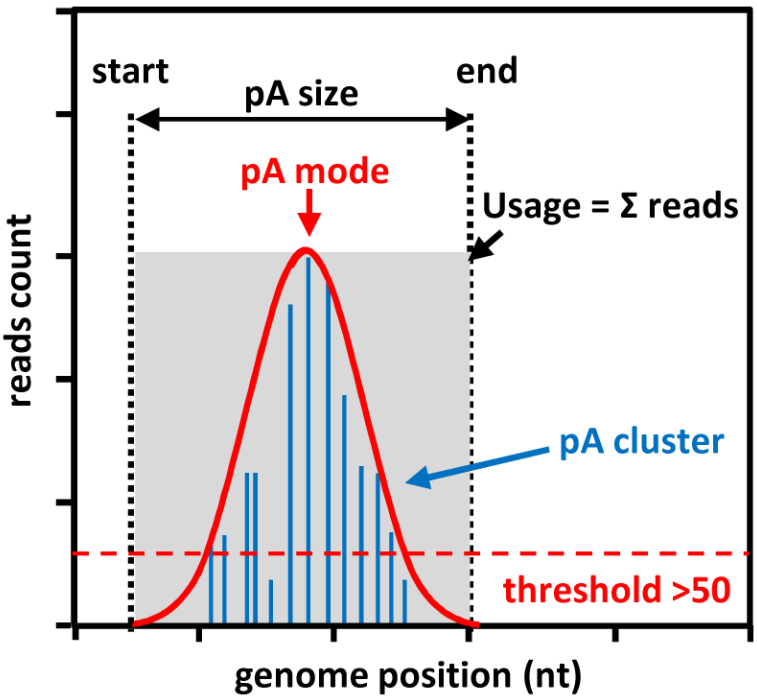

Figure 3. Determination of KSHV pA sites by F-seq analysis

Diagram shows PA peak (red line) identified by F-seq analysis of viral sequence reads (blue bars) aligned to the KSHV genome. The PA mode, the nucleotide position with the highest read count in each peak, is designated as a pA site. The peak size is a distance from nucleotide position of the beginning to the end of the peak within which a pA site is assigned. The total number of all reads within the peak represents usage of the pA site *(with permission from Majerciak et al., 2013).*

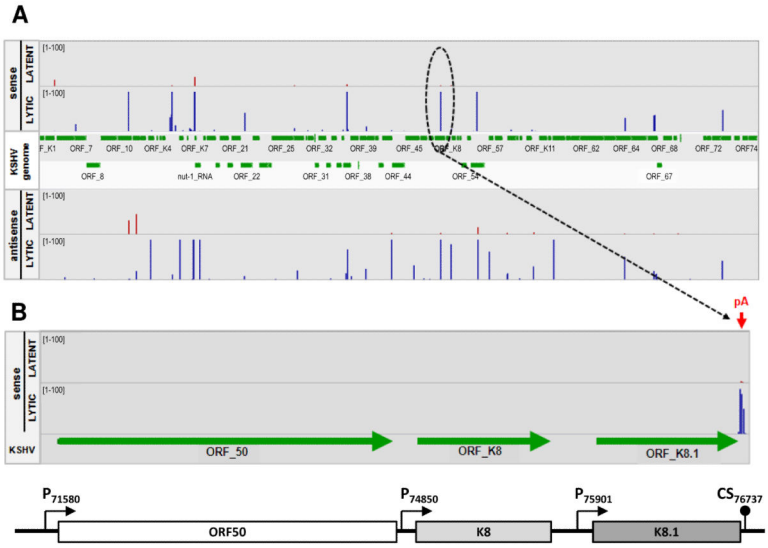

Figure 4. Validation of the pA sites identified

(A) Genome-wide distribution of viral specific reads in the KSHV genome using ".wig" files in IGV browser. The red and blue vertical lines represents reads obtained from samples undergoing latent or lytic infection, respectively. The height of the line indicates the read counts in RPKM (Reads Per Kilobase per Million mapped reads). (B) The read distribution in KSHV ORF50/K8/K8.1 region overlaps the polyadenylation cleavage site (CS) previously identified from the transcripts originated from all three independent promoters (P) upstream of each ORF (Modified with permission from Majerciak et al., 2013).