Pseudouridine prevalence in Kaposi's sarcoma-associated herpesvirus transcriptome reveals an essential mechanism for viral replication.

Timothy J. Mottram[1,2], Katherine L. Harper[1,2], Elton J.R. Vasconcelos[1,3] and Adrian Whitehouse[1,2,4*]

1 School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, LS2 9JT, United Kingdom, **2** Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, United Kingdom, **3** LeedsOmics, University of Leeds, Leeds, LS2 9JT, United Kingdom, **4** Department of Biochemistry & Microbiology, Rhodes University, Grahamstown, 6140, South Africa.

*Correspondence to Adrian Whitehouse

Tel: +44 (0)113 3437096

Email: a.whitehouse@leeds.ac.uk

Significance

Resurging interest in RNA modifications has been driven by advancements in transcriptome-wide mapping and the identification of RNA modifications in all RNA species. Pseudouridine is the most abundant single nucleotide modification found in all functional RNA species, however its role within mRNA and ncRNA is largely unknown. Recent studies show that changes to pseudouridine modified mRNA status often occur when under cell stress, however how pseudouridine modifications are dynamically altered in pathogen RNAs during infection is currently unknown. Here we utilize single-site direct RNA nanopore sequencing to show the human DNA tumour virus, Kaposi sarcoma associated herpesvirus, transcriptome is functionally pseudouridylated. We further highlight that pseudouridylation of the KSHV PAN RNA affects its expression enhancing virus replication.

Abstract

Pseudouridylation is a prevalent RNA modification occurring in tRNAs, rRNAs, snoRNAs and most recently mRNAs and lncRNAs. Emerging evidence suggests that this dynamic RNA modification is implicated in altering gene expression by regulating RNA stability, modulating translation elongation and modifying amino acid substitution rates. However, the role of pseudouridylation in infection is poorly understood. Here we demonstrate that Kaposi's sarcoma-associated herpesvirus (KSHV) manipulates the pseudouridylation pathway to enhance replication. We show that the pseudouridine synthases (PUS), PUS1 and PUS7 are essential for efficient KSHV lytic replication, supported by their redistribution to viral replication and transcription complexes. We present a comprehensive analysis of KSHV RNA pseudouridylation, revealing hundreds of modified RNAs at single-nucleotide resolution. Notably, we demonstrate that pseudouridylation of the KSHV-encoded polyadenylated nuclear RNA (PAN) plays a significant role in the expression of PAN RNA. These findings reveal a novel and essential role of pseudouridine modification in the KSHV replication cycle.

Introduction

Post-transcriptional chemical modifications of RNAs are widely abundant across all forms of RNA, affecting up to 25% of all nucleotides present. There are over 170 modifications currently identified, exhibiting a plethora of functions including RNA stabilisation, localisation and the facilitation of intermolecular interactions. Resurging interest in RNA modifications has been driven by advancements in transcriptome-wide RNA modification mapping and the identification of RNA modifications in all RNA species, including mRNA and ncRNAs. Pseudouridine (Ψ) is the most abundant single nucleotide modification found in all functional RNA species (1, 2). Ψ is catalysed by two groups of enzymes, RNA dependent (H/ACA Box snoRNA-guided) such as Dyskerin (3-5), or RNA independent (direct) known as pseudouridine synthases (PUSs) (6, 7). PUS enzymes function to break the carbon-nitrogen bond found in uridine, then subsequently reform a carbon-carbon bond through the C5 position of the cleaved uridine to the ribose sugar. This function can be site dependent, driven by a specific motif binding or secondary RNA structure (6). The function of Ψ within RNA species such as tRNA, snoRNA and rRNAs are well characterised, however Ψ function within mRNA and lncRNA is largely unknown. Recent transcriptome-wide studies have revealed Ψ can be dynamically modified in response to cellular stress, which may unveil an analogous function to other modifications, such as m^6A. Ψ has been implicated in RNA folding, protein binding, protein translation, RNA-RNA interactions and RNA stability (8-11), however evidence suggests this is a transcript specific, rather than a global effect, as often when Ψ is removed, stability of the RNA molecule remains intact. Interestingly, changes to Ψ-modified mRNA status often occur under cell stress, conferring an enhancement to cell survivability (12, 13). Although, how Ψ sites function in host or pathogen mRNAs and lncRNAs during infection is currently understudied.

The ability of RNA modifications to regulate gene expression offers unique possibilities for viruses to modulate viral and host genes, but also for the host to regulate a response to infection (14). Recent studies demonstrate that m^6A is deposited upon the RNAs transcribed by a diverse range of both RNA and DNA viruses (15-18), and highlighted distinct pro and antiviral roles, indicating widespread regulatory control over viral life cycles (18, 19). Additionally, m^5C, a modification prevalent across many virus genomes, may play a role in modulating viral gene expression. For instance, m^5C knockdown in HIV-1 resulted in the dysregulation of alternative splicing within viral RNAs (20). Most recently, Ψ was identified within Epstein-Barr virus (EBV)-encoded non-coding RNA EBER2, proving essential for the stability of the RNA and required for efficient lytic viral replication (21). However, little is known

regarding the influence of Ψ on regulatory mechanisms controlling virus replication and a lack of global transcriptome-wide analysis has yet to reveal the global prevalence of Ψ within viral genomes.

Kaposi's sarcoma-associated herpesvirus (KSHV) is a large double stranded DNA virus associated with Kaposi's sarcoma and two lymphoproliferative disorders: primary effusion lymphoma and multicentric Castleman's disease. Like all herpesviruses, KSHV has a biphasic life cycle comprising of latent persistence and lytic replication cycles. KSHV establishes latency in B cells and in the tumour setting, where viral gene expression is limited to a small subset of viral genes allowing the viral genome to persist as a non-integrated episome. Upon reactivation through certain stimuli such as cell stress, KSHV expresses the key latent-lytic switch protein, RTA, which initiates entry into the lytic replication phase. Lytic replication leads to the orchestrated temporal expression of over 80 viral proteins necessary for the production of infectious virions (22). During the early stages of KSHV lytic replication, virus-induced nuclear structures are formed, termed virus replication and transcription compartments (vRTCs), which support viral transcription, DNA replication, genome packaging and capsid assembly (31, 32). These structures cause a dramatic remodeling of the nuclear architecture ultimately filling most of the nuclear space, compressing and marginalising the cellular chromatin to the nuclear periphery (23). In KS lesions, most infected cells harbour the virus in a latent state. Although, a small proportion of cells undergo lytic or abortive lytic replication. Here, lytic replication sustains KSHV episomes in latently-infected cells that would otherwise be lost during cell division. Moreover, expression of lytically expressed genes lead to the secretion of angiogenic, inflammatory and proliferative factors that act in a paracrine manner to enhance tumorigenesis (24-27) and manipulation of cellular RNA processing pathways required for efficient lytic replication drives genomic instability (28). Therefore, both the latent and lytic replication phases are implicated in KSHV-mediated tumourigenicity. Interestingly several recent studies have shown that m^6A modification is highly prevalent throughout the KSHV transcriptome and enhances the stability of the essential latent-lytic switch transcriptional protein RTA transcript (16, 29). This highlights the importance of RNA modifications in regulating KSHV gene expression.

Herein we use a CRISPR-Cas9 knockout approach to identify the importance of PUS1 and PUS7 for KSHV lytic replication. Subsequent single-site nanopore transcriptomic analysis demonstrates that the KSHV transcriptome is functionally pseudouridylated, with 162 candidate PUS1 Ψ sites and 49 candidate PUS7 Ψ sites. Notably, specific Ψ sites play an important functional role in the expression levels of polyadenylated nuclear RNA (PAN). These results describe a novel mechanism for KSHV to utilise the host cell post-transcriptional RNA

modification protein PUS1 to modify KSHV RNA transcripts allowing for efficient viral replication.

Results

PUS enzymes are essential for KSHV lytic replication.

To determine whether KSHV manipulated the host cell pseudouridylation machinery we first sought to identify any changes in localisation of the PUS enzymes during KSHV reactivation. We focused specifically on PUS1 and PUS7 has they have both been shown to be modulated under stress conditions and both readily target mRNA and ncRNA species (6, 13). TREx BCBL1-Rta cells, a KSHV-latently infected B-lymphocyte cell line containing a Myc-tagged version of the viral RTA under the control of a doxycycline-inducible promoter, remained latent or were reactivated for 24 hours prior to immunostaining with antibodies against KSHV early proteins and PUS1 or PUS7 proteins. PUS1 showed a diffuse staining throughout the nucleus and cytoplasm in latently-infected cells. In contrast, during lytic replication, the nuclear localized proportion of PUS1 was redistributed to the replication and transcription compartments (RTCs), a virally-induced intra-nuclear structure where viral transcription, viral DNA replication and capsid assembly occur (Fig. 1a). Similarly, PUS7 was redistributed from a diffuse nuclear localisation exclusively into KSHV RTCs (Fig. 1b). However, although changes in subcellular localization were observed, a direct comparison of latent and lytic cells show no changes in PUS1 or PUS7 expression using immunofluorescence (Supplementary Fig. 1a) or immunoblotting (Fig. 2a).

Figure 1. PUS1 and PUS7 are essential for KSHV viral replication and are relocalised into vRTCs upon viral reactivation.

(A, B) TREx BCBL1-Rta cells were reactivated using doxycycline and fixed at 24 hours post activation. Untreated TREx BCBL1-Rta cells were used as a control. Cells were stained using anti-PUS1 or anti-PUS7 antibodies and anti-ORF57 or anti-ORF59 antibodies. Cross section fluorescent intensity graph of representative cell identified by a white arrow. Representative image of n = 3 shown.

To determine whether PUS enzymes were essential for KSHV lytic replication, PUS1 (Fig. 2a) and PUS7 (Fig. 2b) were successfully knocked out in TREx BCBL1-Rta cells utilising the lentiCRISPRv2 CRISPR-cas9 system (Supplementary Fig. 1b-c). In addition, these knock out cell lines were used to confirm the specificity of PUS1 and PUS7 antibodies (Supplementary Fig. 1b-c). To ensure the CRISPR-cas9 single cell cloning did not result in large variations in viral episome count, levels of latent viral DNA was assessed (Supplementary Fig. 2a), which

confirmed single cell clones had no significant changes to KSHV episome load. Furthermore, as pseudouridylation can be important for cellular processes, we assessed the effect of PUS1 and PUS7 knock outs on TREx BCBL1-Rta cellular replication. Knockouts showed a limited effect on cellular replication as assessed by cell count over 72h (Supplementary Fig. 2b) and housekeeping gene expression both during latency (Supplementary Fig. 2c) and reactivation (Supplementary Fig. 2d). However, upon reactivation of PUS1ko and PUS7ko TREx BCBL1-Rta cells results showed a reduction in early KSHV ORF57 protein levels and a complete abolition of late ORF65 protein production compared to scrambled controls (Fig. 2a-b, Supplementary Fig. 3a-b), suggesting pseudouridylation may be important in the later stages of the KSHV lytic temporal cascade. Furthermore, viral RNA expression of immediate-early gene PAN (Fig. 2c), early gene ORF57 (Fig. 2d) and late gene ORF65 (Fig. 2e) (22) was impaired, with PAN and ORF57 showing a significant ~30% reduction and ORF65 ~80%, indicating that the effect seen is at a transcriptional level. To confirm that knockout of PUS1 and PUS7 affected infectious virion production, supernatants of reactivated PUS1ko or PUS7ko TREx BCBL1-Rta cells were used to re-infect naïve HEK-293T cells and KSHV ORF57 expression was determined by qRT-PCR (Fig. 2 f-g). Cells reinfected with supernatant from both PUS1ko and PUS7ko showed a significant reduction (>90%) in infectious virion production. As a further confirmation, PUS1 and PUS7 were rescued through lentiviral transduction of PUS1-eGFP and PUS7-eGFP (Fig 2h, Supplementary Fig. 4). Viral RNA expression of PAN was restored to control levels in PUS1-rescue TREx BCBL1-Rta cells while ORF57 saw an increase in expression in PUS1-rescue. ORF65 expression however, only saw a partial restoration of the control phenotype in PUS1-rescue, whereas PUS7-rescue failed to recapitulate PAN, ORF57 and ORF65 expression (Fig. 2h). Together this suggests that KSHV may directly utilise PUS1 and indirectly utilise PUS7, redistributing both PUS1 and PUS7 into RTCs, to enhance later stages of viral lytic replication.

Figure 2. Early and late lytic KSHV RNA and protein expression are disrupted in PUS1ko or PUS7ko cells.

(A, B) Early and late viral protein expression of reactivated KSHV in scrambled, PUS1ko or PUS7ko TREx BCBL1-Rta cells. Expression of viral early protein ORF57, late protein ORF65 and housekeeping protein GAPDH were determined by Western blot using anti-PUS1 or anti-PUS7, ORF57, ORF65 and GAPDH specific antibodies. Representative image of three biological repeats shown. **(C, D, E)** Gene expression of immediate-early gene non-coding RNA PAN early gene ORF57 and late gene ORF65 in scrambled control, PUS1ko or PUS7ko TREx BCBL1-Rta cells after 24 hours lytic reactivation. Viral RNA levels determined via qRT-PCR and normalised to GAPDH. Values were normalised to scrambled control KSHV infection Error bars represent SD, n = 4 for all experiments, $p \leq 0.05$ *, $p \leq 0.01$ **, $p \leq 0.0001$ **** using a two tailed Students unpaired t test. **(F, G)** Reduced production of infectious KSHV virions in PUS1ko or PUS7ko cells. Successful infection and replication of KSHV virions was determined by reinfection of naïve HEK-293Ts. After 72 hours of doxycycline induction in TREx BCBL1-

Rta cells, supernatant was used to infect HEK-293Ts, which were subsequently harvested at 48 hours post infection. Viral mRNA levels were determined via qRT-PCR of the ORF57 gene and normalised to GAPDH. Values were normalised to scrambled control KSHV infection. **(H)** Rescue of PUS1 and PUS7 in PUS1ko and PUS7ko TREx BCBL1-Rta respectively. Gene expression of PAN, ORF57 and ORF65 was then determined via qRT-PCR and normalised to GAPDH. Values were normalised to scrambled control. Error bars represent SD, n = ≥ 3 for all experiments, $p \leq 0.0001$ **** using a two tailed Students unpaired t-test.

Transcriptome-wide mapping of Ψ during KSHV lytic replication.

To elucidate the landscape of Ψ in the KSHV transcriptome, single site direct RNA nanopore sequencing was performed in TREx BCBL1-Rta cells undergoing lytic replication at 24 h post-induction. Recent studies have allowed for the identification of Ψ through distinct ion current differences in 5 nucleotide k-mer sequences as the RNA strand passes through the pore, while also identifying U-to-C base calling errors at sequence sites that indicate the presence of a Ψ (30, 31). Other current methods rely on chemical modification involving CMC or Bisulfite adducts which may introduce artifacts from labeling, such as incomplete adduct removal or RT read-through of modified sites (32, 33). Two distinct pseudouridine identification algorithms were applied in tandem to the nanopore reads, allowing for highly reproducible single base identification of Ψ sites. Here we utilised the custom bioinformatics pipelines of NanoPsiPy (30), which compares U-to-C base calling errors between control and PUSko reads, and NanoSPA, that was generated through machine learning of 12 distinct base calling features including U-to-C mismatch (34). Together, this identified 162 candidate PUS1 Ψ sites (Fig. 3a-b) and 49 candidate PUS7 Ψ sites (Supplementary Fig. 5a-b) shared across NanoPsiPy and NanoSPA within the KSHV transcriptome, along with an additional 274 candidate Ψ sites identified exclusively with NanoSPA using a higher threshold of >80% likelihood of modification.

Furthermore, mapping the Ψ sites to genome features previously identified in KSHV (22) showed that Ψ is found predominantly within the CDS of KSHV genes with the remaining Ψ sites identified in the UTR or in more rare uORF or sORF features of both PUS1 (Fig. 3c) and PUS7 sites (Supplementary Fig. 5c). Additionally, Ψ was identified in the stop codons for ORF34 and an ORF54 isoform. By performing STREME motif analysis of 15 nt sequences flanking each Ψ site, we observed varying consensus motifs within NanoPsiPy and NanoSPA datasets. Motif analysis of NanoPsiPy PUS1ko data (Fig. 3d), identified a CWU motif as the predominant motif at 71.2% of sites, which shows commonality to the recently identified short HRU PUS1 motif (35), however the primary mode of substrate recognition for PUS1 is assumed to be structural. Interestingly, a broad SDWAAAAD motif was the most prominently conserved motif within the NanoPsiPy PUS7 data (Supplementary Fig. 5d). STREME analysis

of NanoSPA identified a predominant UU motif indicating no common binding motif across all pseudouridine synthase enzymes. PUS7 has previously been identified to preferentially bind a UGUAR motif without a strong requirement for secondary RNA structure (6). Assessing this PUS7 motif in the KSHV transcriptome revealed 8 sites within the CDS that are strong candidates for Ψ modification (Supplementary. Fig. 5e). Together this data shows that the KSHV transcriptome is pseudouridylated during lytic reactivation, comprising of ~0.5% Ψ/U ratio, comparative to ~0.2-0.6% Ψ/U ratio within the human transcriptome.

Figure 3. The KSHV transcriptome is heavily ᵠ modified by PUS1.

(A) Proskee plot of individual Ψ sites identified by NanoPsiPy in TREx BCBL1-Rta PUS1ko cells, in addition to Ψ sites identified by NanoSPA with a >80% chance of modification in TREx BCBL1-Rta cells at 24 hours post reactivation. **(B)** Proskee plot of shared significance Ψ sites identified by NanoPsiPy in TREx BCBL1-Rta PUS1ko cells using NanoSPA sites with >65% probability of modification as a cut-off. **(C)** Distribution of Ψ sites identified by NanoPsiPy and NanoSPA across topological regions of viral genome features include CDS, UTR, miscellaneous RNA, regulatory RNA and repeat regions. **(D)** STREME analysis of Ψ motif sequences. Analysis of sequence motifs of within 15 nt surrounding Ψ site. * denotes position of ᵠ on either positive (T) or negative (A) strand.

Validation of specific Ψ sites in the KSHV transcriptome.

Following the identification of Ψ sites by NanoSPA and NanoPsiPy, validation of a subset of targets was performed to confirm the robustness of the dataset. Initially, RNA immunoprecipitations (RIP) were performed in reactivated TREx BCBL1-Rta cells using a previously characterised Ψ-specific antibody (36) to precipitate KSHV-encoded RNAs with Ψ sites (Supplementary Fig. 6a). This included the PAN and ORF4 RNAs, 28S as a control for a highly pseudouridylated RNA transcript and Myoglobin pre-mRNA (MB) as a negative control previously used as a control for no Ψ sites (Diogenode Cat# C15200247 Lot# 001). Results showed that both PAN and ORF4 RNAs were significantly enriched over MB, confirming they contain Ψ sites. The dose-dependent specificity of the Ψ-specific antibody was also assessed by precipitating in vitro transcribed PAN RNA spiked with varying proportions of Ψ (Supplementary Fig. 6b). PAN RNA undertakes multiple essential functions in KSHV lytic replication, specifically acting as a scaffold for early expressed viral proteins resulting in enhanced late gene expression (37, 38). Therefore, to determine the functional significance of the Ψ modification on PAN RNA, we mapped the four potential PUS1/PUS7 Ψ sites identified by NanoPsiPy onto a previously identified SHAPE structure of PAN RNA (Fig. 4a) (39). These sites, particularly site 344 and site 904, were noted to be in close proximity to known binding sites of a multitude of viral proteins, including the ORF57 protein. This is

particularly intriguing as the ORF57 protein has been previously reported to enhance the expression of PAN (40-42). This close proximity indicated that these Ψ sites may be involved in the expression of the PAN molecule allowing or disrupting binding of these important viral interactors (40). To investigate this possibility, we firstly experimentally confirmed each PAN Ψ modification site, using CMC-assisted RT-qPCR (Fig. 4b) (32, 43). CMC-assisted RT-qPCR relies on the addition of a CMC adjunct to the Ψ residue acting as a reverse transcriptase terminator, thus resulting in shortened DNA fragments reverse transcribed from Ψ modified RNA. Through the amplification of both a Ψ modified region and a negative control region of the RNA, in CMC or non-CMC treated samples, one can compare the relative read through of the qPCR and thus determine a site of Ψ modification. Levels of Ψ within an RNA population can be inferred from read-through %. CMC-assisted RT-qPCR confirmed that all four Ψ regions identified by nanopore sequencing on PAN are Ψ modified. Further analysis utilising PUS1ko and PUS7ko TREx BCBL1-Rta cells confirmed Ψ169 and Ψ994 sites as specifically modified by PUS1. However, the modifying enzyme for sites Ψ344 and Ψ904 could not be identified as neither PUS1ko nor PUS7ko ablated CMC read-through. Together, RIP and CMC-assisted RT-qPCR analysis confirm that the KSHV PAN transcriptome is Ψ modified during lytic replication.

Figure 4. KSHV PAN RNA is pseudouridylated in close proximity to known interaction motifs.

(A) Nanopore sequencing identified KSHV non-coding RNA PAN Ψ sites. Adapted from Sztuba-Solinska et al 2017 PAN SHAPE analysis with Ψ sites highlighted. **(B)** Individual PAN Ψ sites were confirmed via CMC RT-qPCR and their respective acting PUS enzyme identified through PUS1 and PUS7 knockouts. Read-through % was calculated by first normalising the amplification of the Ψ region to the Ψ negative control region, before further normalisation between CMC-treated and untreated samples. Error bars represent SEM, n = ≥ 3 for all experiments. $p \leq 0.05$ *, $p \leq 0.01$ ** $p \leq 0.001$ *** using a two tailed Students unpaired t-test.

Ψ plays a functional role in PAN expression and stability.

The ability of Ψ to influence the stability and expression of mRNA has been previously studied (44). To further examine the specific Ψ sites within PAN and how they contribute to PAN expression levels and function, pCMV-PAN mutants were generated at sites 169, 344, 904 and 994. A T-A substitution was chosen to maintain GC content and minimise any effects to overall RNA structure. Results demonstrated that all mutants showed significantly reduced PAN expression levels in comparison with WT PAN RNA (Fig. 5a). Previous research has shown that ORF57 expression can stabilize and increase PAN expression, therefore co-transfection assays were also performed for each mutant in the presence of a KSHV ORF57 expression construct (45). Interestingly, expression levels of mutants 344 and 944 were restored to WT levels in the presence of ORF57, however the expression levels of mutants

169 and 904 remained impaired (Fig. 5b). These levels were comparable with the negative ORF57 RNA binding mutant, RGG (46), which showed no enhancement of PAN RNA levels. Further, the mutants showed no significant reduction in expression levels in PUS1ko cells (Supplementary Fig7). These results suggested that the PAN Ψ mutants may affect the stability of PAN RNA or the positive effect ORF57 protein has on PAN expression.

Figure 5. PUS1 shows important functions for PAN stability and expression.

(A) Overexpression of PAN pseudouridine mutants, or **(B)** co-transfected with ORF57-eGFP. HEK-293T cells were transfected with pCMV-PAN (or mutant pCMV-PAN) or both pCMV-PAN (or mutant pCMV-PAN) and pCMV-ORF57-eGFP before harvesting after 48 hours. Samples were RNA extracted and analysed via qRT-PCR. Samples were first normalised to GAPDH before normalising to respective PAN only samples and WT PAN. Error bars represent SE, n = 4 for all experiments, $p \leq 0.05$ *, $p \leq 0.01$ ** using a one way ANOVA with Dunnett post-test. **(C)** The stability of PAN RNA during KSHV lytic reactivation was determined by assessing mRNA decay through Actinomycin D (AcD) treatment of TREx- BCBL1-Rta cells compared with PUS1ko. Cells were reactivated using doxycycline 24 hours prior to the addition of 2.5 µg / ml AcD. Cells were then collected at 0, 4 and 8 hours post transcription inhibition (Post-TI) by AcD treatment and total RNA was extracted followed by qRT-PCR. Values were first normalised to GAPDH before normalising to 0 hour time point. Error bars represent SD, n = 3 for all experiments. **(D)** Knockout of PUS1 in HEK-293T cells as confirmed via western blotting. PUS1ko clone 2 was used for subsequent experiments. **(E)** Overexpression of PAN and ORF57 in PUS1ko HEK-293T cells. HEK-293T cells were transfected with pCMV-PAN alone or pCMV-PAN with pCMV-ORF57-eGFP before harvesting after 48 hours. Samples were first normalised to GAPDH before normalising to respective PAN only sample. (F, G) RIP analysis of ORF57-eGFP in HEK-293T cells. HEK-293T cells were transfected with pCMV-PAN and pCMV-ORF57-eGFP and an RIP was performed at 24 hours post transfection using GFP-TRAP beads. RNA immunoprecipitation was determined by qPCR of BTG1 pre-mRNA positive control or PAN genes ($n \geq 3$). RIPs are expressed as fold change over GAPDH following normalisation to scrambled control. Error bars represent SEM, $n = \geq 3$ for all experiments. $p \leq 0.05$ *, $p \leq 0.01$ ** $p \leq 0.001$ *** using one way ANOVA with Dunnett post test.

Next we sought to investigate the effect of PUS1 and PUS7 knockouts on PAN stability. Here an Actinomycin D stability assay was performed in scrambled and PUS knockdown cell lines. PUS1ko in TREx-BCBL1-Rta cells led to a substantial reduction in PAN RNA abundance at both 4 and 8 hours post drug treatment (Fig. 5c), whereas PUS7ko resulted in no significant reduction by 8 hours post treatment (Supplementary Fig. 8). This suggests that the confirmed PUS1 Ψ169 and Ψ994 sites may be important for PAN stability. At present, the mechanism by which Ψ affects RNA stability has not been fully elucidated. Knockout of PUS1 or PUS7 has been shown to result in global reduction of pseudouridylation at PUS1 or PUS7 sites, respectively. Therefore, the effects on PAN stability observed in PUS1 knockouts cannot be directly attributed to Ψ within PAN. Thus, to isolate Ψ function in PAN RNA from other KSHV

Ψ modifications and their effects, we focused on PUS1 by perform a CRISPR-cas9 knockout (Fig. 5d) in HEK-293T cells.

To explore the functional implications of Ψ modification on PAN and its relation to ORF57, we next determined if the ORF57 protein was able to enhance PAN RNA levels in PUS1ko cells (Fig. 5e). Interestingly, PUS1ko showed a significant reduction in the enhancing effect of ORF57 on PAN expression. Furthermore, due to the close proximity of the Ψ site to local ORF57 binding sites, we wanted to assess if binding of the ORF57 protein to PAN RNA could be affected upon loss of PUS enzyme activity. Initially, we co-transfected PAN with ORF57-eGFP in our control HEK-293T or PUS1ko HEK-293T cell line and carried out a RIP using GFP-TRAP beads pulldown and bound PAN RNA levels were assessed (Fig. 5f-g, Supplementary Fig. 9a-b). PUS1ko resulted in reduced binding of PAN to ORF57-eGFP. The expression levels and pulldown efficiency of ORF57-eGFP and ORF57-RGG remained constant between HEK-293T and PUS1ko cells (Supplementary Fig. 9c-d). Pre-BTG1, a transcript that has previously been shown to be bound by ORF57 (47) was used as a positive control. Pre-BTG1 has not previously been shown to contain Ψ (7) and its binding levels remained constant, suggesting that general binding of ORF57 is not impaired. Together this suggests that the presence of Ψ within PAN RNA may be important for ORF57 binding, allowing enhanced PAN expression.

Discussion

This study is the first to investigate the transcriptome-wide role of Ψ within a DNA virus. KSHV was used as a model pathogen since it is known to manipulate and utilise a large number of host cell proteins and pathways during its replication cycle, including other RNA modifications (22, 25, 48, 49). Redistribution of PUS1 and PUS7 to KSHV RTCs during lytic reactivation reinforced the hypothesis that KSHV may utilise pseudouridylation to positively modify viral RNA transcripts. These virus-induced intranuclear structures, enable multiple processes required for KSHV lytic replication to occur, including viral transcription, viral DNA synthesis and capsid assembly. Sequestering of PUS1 and PUS7 into KSHV RTCs may serve the virus twofold; the utilisation of PUS1 and PUS7 pseudouridylation activity for KSHV lytic transcripts such as PAN and ORF4 occurring co-transcriptionally, and/or the reduction of Ψ on cellular genes.

Through the use of CRISPR-cas9 knockouts, we show that both PUS1 and PUS7 are essential for KSHV lytic replication. The ~30% reduction in expression of the immediate-early PAN RNA and early ORF57 RNA and protein, compared to the significant 80% reduction of late ORF65 protein in both PUS1ko and PUS7ko cells indicates that Ψ is not directly involved in the

latent/lytic switch but more likely affecting one or more viral processes that occur downstream in the lytic temporal cascade. While PUS1 showed partial rescue of ORF65 expression, the failure of PUS7 rescue to restore ORF65 expression indicates potentially lasting changes to the viral transcriptome that require further study to elucidate. Continual passage of PUS7ko may have irrevocably changed the cellular transcriptomes functional landscape, through possible compensation of Ψ by other PUS enzymes. Further, overexpression may not recapitulate the subcellular localisation observed with native PUS7 during lytic reactivation, thus may not appropriately mimic the regulation observed. To further understand the effect of the PUS enzymes on viral replication, we sought to determine if the virus transcriptome was Ψ modified. It has been previously shown through transcriptome wide analysis that the human pathogens *Trypanosoma brucei*, Influenza A and HIV have Ψ modified transcripts (44, 50, 51) however there is limited research on the functional role of Ψ in the context of viral infection. Our direct RNA nanopore sequencing and application of NanoPsiPy and NanoSPA algorithms revealed that the KSHV transcriptome is heavily Ψ modified. Mapping the Ψ sites to genome features revealed the distribution of Ψ to be preferentially located on the CDS of the KSHV transcriptome. While the majority of traditional Ψ found in human or yeast cellular transcripts are also within the CDS, there are proportionally more sites found in UTR regions, particularly 3' UTR than within the KSHV transcriptome (34). Examining Ψ sites in alternative topological features showed sites in uORF, sORF and stop codon regions which may have implications on downstream viral protein expression. Motif analysis of nucleotides surrounding Ψ sites revealed differential motifs between NanoSPA, NanoPsiPy PUS1ko and NanoPsiPy PUS7ko datasets. Recent research has shown that motif binding alone is not sufficient to predict Ψ sites. When analysing Ψ sites identified by NanoSPA, the broad WT motif represents the pseudouridylation by all PUS enzymes, indicating that pseudouridylation of the viral transcriptome is facilitated by a range of pseudouridine synthases. The PUS1 HRU binding motif would contribute a significant proportion to the motif analysis and thus, is likely to pseudouridylate the most Ψ sites in the KSHV transcriptome. This is reflected in NanoPsiPy PUS1ko motif analysis displaying a similar HWU motif, though PUS1 pseudouridylation specificity is thought to be determined by secondary structure rather than specific binding motifs. Comparatively, NanoPsiPy PUS7ko motif analysis did not broadly correspond to the UGUAR motif previously identified, with that specific binding motif being found in just 7 Ψ sites. This indicates PUS7 may have a reduced role of direct pseudouridylation of the viral transcriptome and the effect on KSHV replication is more likely to relate to changes in pseudouridylation of host cell RNA. A limitation of this study is the reduction in Ψ identified during reactivation of PUS7ko cells, which may be a result of changes in localisation or expression of other pseudouridine synthase enzymes not captured during this study. A larger panel of pseudouridine synthase knockouts in addition to comparsions with latent knockout

cells would provide additional insights into specific PUS1 and PUS7 function. Ψ calling must be approached cautiously, particularly in RNA transcripts with potentially low Ψ level, as there can be a high level of false positive calls (52). To mitigate this, we opted to cross reference both NanoSPA and NanoPsiPy, and to focus validation strategies on highly abundant RNA transcripts. Whilst outside the scope of this study, the nanopore sequencing dataset generated here can be further examined to identify Ψ within cellular genes and any changes that may occur during KSHV lytic reactivation that could influence the overall landscape of anti-viral or pro-viral cellular genes.

Validation of the non-coding viral PAN and ORF4 RNAs was performed via an RNA-immunoprecipitation using a Ψ-specific antibody to immunoprecipitate Ψ modified RNA. PAN and ORF4 were selected as two hits for their importance during viral replication (53). Furthermore, PAN was selected for further study due to its huge abundance during viral reactivation, accounting for >90% of viral reads within a cell, and thus we surmised that PAN may lead to significant changes in virus replication in the absence of Ψ. There has been growing evidence that Ψ can directly affect levels of protein translation, (11, 54) thus we wished to examine further the lesser known effects of Ψ on a viral non-coding RNA. Within PAN, four Ψ sites were detected at nucleotides 169, 344, 904 and 994. These regions were validated as containing Ψ via CMC RT-qPCR. While PAN regions surrounding 169 and 994 were confidently identified as PUS1 modified, PAN 344 and 904 may be modified by other PUS enzymes, or can be modified by both PUS1 and PUS7, as previous studies have identified single Ψ sites written by multiple writers (55). Further study through the generation of double knockouts or knocking out a larger panel of PUS enzymes would provide more specific writer data. Due to the proximal nature of multiple viral protein binding sites surrounding the Ψ sites, we hypothesised these sites may be important in influencing the binding of viral proteins essential for viral replication (39). The identification of PUS1 modified Ψ sites led us to focus on the role of PUS1 during KSHV reactivation specifically.

During reactivation, there are thousands of Ψ modifications occurring across cellular and viral transcripts and thus focusing on single transcript modifications in PUS knockouts is a significant challenge. To isolate PAN from the multitude of associating viral factors during lytic reactivation, we generated PUS1ko HEK-293T cells with which we could assess PAN function through transfection. Single Ψ sites can affect the dynamics of an RNA molecule. A recent study utilising an NMR based approach suggests that single Ψ modifications can have profound effects on stabilising or destabilising the RNA molecule based on structural elements and alternative hydrogen bonding interactions (56). By generating Ψ PAN mutants, we showed that all four Ψ sites had an effect on the PAN molecule, with all mutants displaying a reduced expression phenotype. Interestingly, when examined further, ORF57 was able to restore PAN

457 expression in mutants 344 and 994, while mutants 169 and 904 expression remained
458 impaired. While this reduction was no longer significant when mutant PAN were transfected
459 into PUS1ko cells, indicating a Ψ specific effect, indirect structural effects of the mutations
460 cannot be ruled out. Interestingly, PAN showed reduced stability in PUS1ko TREx BCBL1-Rta
461 cells, while PUS7ko had a minimal effect.

462 Previous work has shown that PAN has a number of interaction motifs enabling ORF57 protein
463 recruitment, including an Mta responsive element (MRE) and expression and nuclear retention
464 element (ENE). These regions are important for achieving high levels of expression, up to 20
465 fold increase and also nuclear retention (41). Here we show that the expression enhancing
466 effect of ORF57 is impaired in PUS1ko HEK-293T cells, consistent with the reduction seen in
467 TREx BCBL1-RTA PUS1ko cells. Ψ has been shown to reduce RNA-protein binding affinity
468 when the modification is located directly within the binding motif (8) however in this case, the
469 344 modification is 4-6 bases from the binding site. Sites 169 and 904 are not in close proximity
470 to any known binding sites however 904 is found within the ENE region on an unpaired bulge
471 within the secondary structure. Interestingly, PUS1ko reduces the overall binding of ORF57 to
472 PAN as shown by RIP analysis. While PUS1 was confirmed to be an exclusive PAN Ψ writer
473 of sites 169 and 994, PUS7 was not confirmed as the sole writer for any sites. The effect on
474 binding by Ψ modification may affect protein-protein interactions through changes to
475 secondary structure rather than immediate changes to binding dynamics within the motif.
476 Additionally, sites 169, 344 and 994 are not located in the MRE or ENE elements that have
477 been shown to improve PAN stability. However, this impaired binding is likely a contributor to
478 the reduction in PAN expression highlighting the importance of all multiple binding regions on
479 PAN. Further work may identify Ψ importance in the impaired binding of other important viral
480 factors such as ORF59 or cellular factors such as PABPC1 and Aly/REF (41, 57).

481 The landscape of Ψ is broad as to affect many cellular and viral genes (58), of which there
482 can be multiple interactions affect PAN molecular dynamics. Thus, while clearly essential for
483 PAN function, Ψ knockdown can affect numerous RNAs across the host and virus
484 transcriptome making examining direct function challenging. Further work making use of
485 investigating multiple PUS knockouts in both latent and lytic KSHV reactivation, examining
486 both the viral and host transcriptome, would allow a broader overview of how the landscape
487 of Ψ changes, and may provide more elucidation of Ψ function with KSHV replication.

488

489 **Methods**

490 **Reagents tables.** Plasmids (Supplementary Table 1, antibodies (Supplementary Table 2) and
491 primers (Supplementary Table 3).

Cells lines and reagents. TREx BCBL1-Rta cells are a genetically engineered BCBL-1 primary effusion lymphoma (PEL) B cell line that expresses Myc-tagged RTA under a doxycycline inducible promoter, a gift from Professor Jae U. Jung (University of Southern California, USA). TREx BCBL1-Rta cells were cultured in RPMI1640 growth media with glutamine (Gibco®) supplemented with 10% (v/v) fetal bovine serum (FBS, Gibco®), 1% (v/v) penicillin-streptomycin (P/S, Gibco®) and 100 µg/ul hygromycin B (Thermo Scientific). For virus reactivation, RTA expression was induced through the addition of 2 µg/mL doxycycline hyclate (Sigma-Aldrich). HEK-293Ts were purchased from ATCC (American Type Culture Collection) and cultured in Dulbecco's modified Eagle's medium with glutamine (DMEM, Lonza) supplemented with 10% (v/v) FBS and 1% P/S. The antibodies used throughout this study include anti-ORF57 (Santa Cruz, sc-135746 1:1,000), anti-ORF65 (CRB crb2005223, 1:100), anti-GAPDH (Abcam, ab8245 1:5,000), anti-PUS1 (SIGMA-ALDRICH, SAB1411457 1:1000) and anti-PUS7 (Invitrogen™, PA5-54983, 1:1000).

CRISPR stable cell lines. HEK-293T cells were transfected with the 3 plasmid lentiCRISPR v2 system. In 12-well plates, 4 ul of lipofectamine 2000 (Invitrogen™) was combined with 1 ug lenti CRISPR V2 plasmid (a gift from Feng Zhang, Addgene plasmid #52961) expressing the guide RNA (gRNA) targeting the protein of interest, 0.65 µg of pVSV.G and 0.65 µg psPAX2. pVSV.G and psPAX2 were gifts from Dr. Edwin Chen at the University of Leeds. Two days post transfection the viral supernatant was harvested, filtered (0.45 um pore, Merck Millipore) and used to transduce TREx BCBL1-Rta cells in the presence of 8 µg/mL of polybrene (Merck Millipore). Virus supernatant was removed 6 hours post transduction and cells were maintained for 48 hours before puromycin (Sigma-Aldrich) selection. Stable mixed population cell lines were maintained until confluent before single cell selection. Single cell populations were generated through serial dilution of ~100 cells in 96 well plates. Positive wells were cultured for 3-5 weeks and maintained with fresh media before transferal into 6 well plates. Upon confluence, clones were tested via western blot for expression of target protein of interest. TREx BCBL1-Rta rescues were generated through lentiviral transduction of TREx BCBL1-Rta PUS1 or PUS7ko cells generated from a modified plentiCMV containing a zeomycin resistance gene and PUS1-eGFP or PUS7-eGFP genes respectively. PUS1ko HEK-293Ts were generated as above.

Nanopore direct RNA sequencing. Library preparation of direct RNA sequencing samples was performed as described by Oxford Nanopore Technology using their Direct RNA

Sequencing Kit (SQK-RNA004). Briefly, 500ng of total RNA per sample was used to perform a sequencing run. The RT adaptor was ligated using T4 DNA ligase (NEB) followed by reverse transcription with SuperScript III Reverse Transcriptase (ThermoFisher). The RT product was then purified using RNAClean XP beads (Beckamn Coulter). The second RNA adaptor (RLA) was ligated by T4 DNA ligase. The RNA product was purified with RNAClean XP beads and eluted with elution buffer. Sample was loaded onto a FLO-MIN004RA flowcell.

Nanopore data processing.

Two independent ONT sequencing-driven algorithms were employed in order to map PSU sites onto the KSHV genome (NC_009333.1) *i)* NanoSPA (https://github.com/sihaohuanguc/NanoSPA) for *de novo* PSU identification of TREx BCBL1-Rta (pooled triplicates), and *ii)* NanoPsiPy (https://github.com/vetmohit89/NanoPsiPy) of both TREx BCBL1-Rta PUS1ko-vs-WT (pooled duplicates) and TREx BCBL1-Rta PUS7ko-vs-WT (pooled triplicates) sample groups. The latter tool, as described by developers, is a more recent workflow built around Nanopore_psU (31). Both NanoSPA and NanoPsiPy workflows were run as described on their tutorials, using the guppy base caller (v.3.2.2+9fe0a78). The default chi-square p-value threshold of 0.01 was kept for PSU calling significance by NanoPsiPy, whereas a PSU confidence score above 80% was arbitrarily chosen regarding NanoSPA results. A combination of KSHV 2.0 (22) and more recently annotated novel transcripts and ORFs (59) were used for assigning PSU sites to KSHV annotated features through bedtools (intersect -s -split -loj) (60) followed by *ad-hoc* PERL script for filtering and summarizing of results.

Viral reinfection assay. TREx BCBL1-Rta, TREx BCBL-Rta PUS1ko or PUS7ko cells were reactivated and harvested after 72 h as previously described (61). Cellular supernatant was filtered with a 0.45 μm pore filter (Merck Millipore) and subsequently used to inoculate HEK-293T cells at a 1:1 ratio with DMEM tissue culture media. Active KSHV transcription was quantified at 48 h post-infection by RT qRT-PCR. Total RNA was extracted from cell lysates using RNeasy Mini Kit as described by the manufacturer. cDNA synthesis was carried out with 1 μg total RNA using LunaScript™ RT SuperMix Kit according to the manufacturers protocol. Subsequent qPCR reactions were carried out using ORF57 and GAPDH specific primers as described in the qRT-PCR method.

Viral episome count assay. TREx BCBL1-Rta, TREx BCBL-Rta PUS1ko or PUS7ko cells were serially passaged and cells harvested after 14 days. Total DNA was extracted from cell pellets using Monarch Genomic DNA Purification Kits (New England Biolabs) and viral episome copies quantified by qPCR of the viral gene ORF57 as described in the qRT-PCR method.

Two-step reverse transcription quantitative PCR (qRT-PCR). Total RNA from cell pellets was extracted using a Monarch Total RNA Miniprep kit (New England Biolabs) according to the manufacturer's instructions. Reverse transcription was performed on 500 ng of total RNA using a LunaScript™ RT SuperMix Kit (New England Biolabs) as according to the manufacturer's instructions. Quantitative PCR (qPCR) reactions included 10 µl 1 X GoTaq® qPCR Master Mix (Promega), 0.5 µM of each primer and 5 µl template cDNA. Cycling was performed in a RotorGene Q 2plex machine (Qiagen). The cycling programme used was; a 10 minute initial preincubation at 95 °C, followed by 40 cycles of 95 °C for 15 s, 60 °C for 30 s and 72 °C for 20 s. A melt curve step was performed post qPCR to confirm single product amplification. Gene expression analysis was performed with normalisation to the housekeeping gene GAPDH (ΔC_T) and reference sample ($\Delta\Delta C_T$).

Overexpression assays. Scrambled HEK 293T or HEK 293T PUS1ko cells were seeded at $1x10^5$ in a 24 well plate. After 24 hours, cells were transfected with 250 ng PAN, or 250 ng PAN and 250 ng ORF57-eGFP or ORF57-RGG using Lipofectamine 2000 (Thermo Fischer Scientific) according to manufacturer's instructions. Cells were harvested at 24 hours post transfection before total RNA extraction using Monarch® Total RNA Miniprep Kits (New England Biolabs) according to the manufacturer instructions. cDNA synthesis was carried out using LunaScript™ RT SuperMix Kit. qRT-PCR was performed as described above. Normalisation was carried out using GAPDH, and data was further normalised to WT samples samples.

RNA stability assay. TREx BCBL1-Rta cells were reactivated with doxycycline hyclate. At 24 hours post reactivation, cells were treated with 2.5 µg/ml of actinomycin D (Thermo Scientific) and samples were collected at 0, 4 and 8 hours post treatment. Total RNA was extracted using Monarch® Total RNA Miniprep Kits (New England Biolabs) according to the manufacturer instructions. cDNA synthesis was carried out using LunaScript™ RT SuperMix Kit. qRT-PCR

was performed as described above. Normalisation was carried out using GAPDH and data was further normalised to 0 hour sample.

RNA immunoprecipitations.

For Ψ RIPs, TREx BCBL1-Rta cells were reactivated with doxycycline hyclate and harvested 24 hours post reactivation. Cells were lysed and RNA extracted with TRIzol LS (Invitrogen™) as per manufacturer's instructions. 10 µg total RNA was incubated overnight at 4 °C with Dynabeads™ Protein G magnetic beads pre-bound with anti- Ψ (Diagenode) according to manufacturer's instructions. Following pulldown, RNA samples were incubated with Proteinase K buffer (10 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 10% SDS, proteinase K) for 30 minutes at 55 °C minutes before further RNA extraction with TRIzol LS (Invitrogen™). cDNA was synthsised using using LunaScript™ RT SuperMix Kit (New England Biolabs) before analysis via qRT-PCR. Human ChIP-seq grade Myoglobin Exon 2 primer pair (diagenode) were used as a negative control according to diagenode manufacturer instructions. For Ψ RIP of varying Ψ% in PAN, PAN RNA was in-vitro transcribed using the MegaScript T7 kit (Invitrogen) according to manufacturer's instructions, substituting the U nucleotides through proportionally supplying Ψ at 0, 10, 50 and 100% increments. 5 µg PAN RNA was then used to perform Ψ RIPs as previously described. Samples were analysed using fold enrichment over GAPDH before further normalisation with scrambled samples.

For GFP RIPs, scrambled or PUS1ko HEK-293Ts were transfected with 2 µg PAN and 2 µg ORF57-eGFP plasmids using Lipofectamine 2000 (Thermo Fisher Scientific) according to manufacturer's instructions. Cells were lysed before incubation with GFP-Trap Agarose beads (Chromotek) overnight at 4 °C using manufacturer's instructions. Samples were incubated with Proteinase K buffer for 30 minutes at 55 °C before RNA extraction with TRIzol LS (Invitrogen™) as per manufacturer's instructions. cDNA was synthesised using LunaScript™ RT SuperMix Kit (New England Biolabs) before analysis via qRT-PCR. Samples were analysed using fold enrichment over GAPDH before further normalisation with scrambled samples.

Immunoblotting. Protein samples were run on 12% polyacrylamide gels and transferred to nitrocellulose membranes (GE Healthcare) via semi-dry transfer in a Bio-Rad Trans-blot Turbo transfer machine. Membranes were blocked with TBS + 0.1% (v/v) Tween 20 (TBS-T) and 5% (w/v) dried milk powder (Marvel) for 1 hour. The membrane was then incubated with relevant primary, followed by secondary antibodies for 1 hour incubations diluted in 5% (w/v) milk TBS-

T. Membranes incubated with secondary fluorescent antibodies were dried and imaged on an Odyssey® CLx (LI-COR).

Immunofluorescence. Sterilised glass coverslips were treated with Poly-L-Lysine for 15 mins before seeding TREx BCBL1-Rta cells. After 8 hours post seeding, TREx cells were reactivated with doxycycline hyclate. Cells were fixed at 24, 48 or 72 hours post reactivation with 4% (v/v) paraformaldehyde in PBS for 15 minutes. Subsequently, wells were washed twice in PBS and permeabilised in PBS containing 1% Triton X-100 for 10 minutes. Coverslips were blocked with 5% (v/v) BSA in PBS for 1 hour before subsequent incubation with primary and secondary antibodies, both for 1 hour at 37 °C. Glass coverslips were then mounted onto microscope slides using Vectashield® Hardset with DAPI. Slides were visualised on a Zeiss LSM 880 laser scanning confocal microscope and images analysed using Zen® 2011 (Zeiss).

CMC-assisted RT-qPCR assay

CMC-assisted RT-qPCR assay was carried out as previously described. Briefly, 5 µg total RNA was subjected to CMC-treatment or control treatment followed by CMC reversal and ethanol precipitation. For each RT reaction, RNA was incubated with random hexamers (Thermo Scientific) at 70 °C for 2 min followed by 2 µl 5x first strand buffer, 0.5 µl dithiothreitol (100 mM), 1 µl dNTP (10 mM), 0.5 µl RNaseOut and 1 µl SuperScript III RT were added with a final volume of 20 µl. RT-qPCR primers were designed for the Ψ containing region and a control region without a Ψ site. Samples were subjected to RT-qPCR quantification as previously described.

Mapping identified PSU sites onto KSHV annotated features. KSHV 2.0 and 3.0 annotations (22) were merged to NC_009333.1 reference genome through BlastN (62) (-FF -W7 -v1 -b1) of all novel features from the former against the latter whole genome. An a*d hoc* PERL script was used to convert Blast tabular format to gff, allowing the integration of both annotations. Gff was converted to bed6 through regular Unix “cut” commands, setting “feature_type” from gff (3rd column) as “feature_name” for bed6 (4th column). Bed files containing identified PSU sites from the NanoSPA, NanoPsiPy PUS1ko and NanoPsiPy PUS7ko were compared against the recently created NC_009333.1 + KSHV 2.0 bed6 through a “bedtools intersect -loj” execution (60).

PSU motif analysis. STREME tool from the MEME suite (63) (--dna --minw 3 --maxw 10 --thresh 0.01) was employed for screening short motifs on the KSHV genome within a 15 nt-long PSU site surrounding area (7 nt upstream - PSU site - 7 nt downstream). PSU sites' bed files (described above) were combined into a single non-redundant tabular format used as input for an *ad hoc* PERL script aimed at generating the 15 nt-long sequences (with a central PSU site) fasta file used as input for STREME. PSU sites mapped on the minus strand of the genome had their 15 nt-long sequences reversed-complemented. A KSHV whole genome-derived background control file for STREME was created with pyfasta: pyfasta split -n 1 -o 7 -k 15 NC_009333.1.fasta.

Statistical analysis. Except where otherwise stated, graphical data shown represent mean plus/minus standard error of mean or standard deviation (SD) using at least 3 independent experiments. Differences between means were analysed by Students t-test or one-way ANOVA as described in the figure legends. Statistics were considered significant at $p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***, $p < 0.0001$ **** between groups.

Data availability. Source data for Nanopore sequencing has been deposited to NCBI SRA. Reviewer login details provided below:

https://dataview.ncbi.nlm.nih.gov/object/PRJNA1248188?reviewer=9jpikuqub0o41jrhpichql2543

References

1. J. Ge, Y. T. Yu, RNA pseudouridylation: new insights into an old modification. *Trends Biochem Sci* **38**, 210-218 (2013).
2. F. F. Davis, F. W. Allen, Ribonucleic acids from yeast which contain a fifth nucleotide. *J Biol Chem* **227**, 907-915 (1957).
3. P. Schattner, S. Barberan-Soler, T. M. Lowe, A computational screen for mammalian pseudouridylation guide H/ACA RNAs. *RNA* **12**, 15-25 (2006).
4. M. Xiao, C. Yang, P. Schattner, Y. T. Yu, Functionality and substrate specificity of human box H/ACA guide RNAs. *RNA* **15**, 176-186 (2009).
5. A. Garus, C. Autexier, Dyskerin: an essential pseudouridine synthase with multifaceted roles in ribosome biogenesis, splicing, and telomere maintenance. *RNA* **27**, 1441-1458 (2021).
6. M. K. Purchal *et al.*, Pseudouridine synthase 7 is an opportunistic enzyme that binds and modifies substrates with diverse sequences and structures. *Proc Natl Acad Sci U S A* **119** (2022).
7. N. M. Martinez *et al.*, Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect pre-mRNA processing. *Mol Cell* **82**, 645-659 e649 (2022).
8. P. P. Vaidyanathan, I. AlSadhan, D. K. Merriman, H. M. Al-Hashimi, D. Herschlag, Pseudouridine and N(6)-methyladenosine modifications weaken PUF protein/RNA interactions. *RNA* **23**, 611-618 (2017).
9. E. Kierzek *et al.*, The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res* **42**, 3492-3501 (2014).
10. E. deLorimier *et al.*, Pseudouridine Modification Inhibits Muscleblind-like 1 (MBNL1) Binding to CCUG Repeats and Minimally Structured RNA through Reduced RNA Flexibility. *J Biol Chem* **292**, 4350-4357 (2017).
11. D. E. Eyler *et al.*, Pseudouridinylation of mRNA coding sequences alters translation. *Proc Natl Acad Sci U S A* **116**, 23068-23074 (2019).
12. K. Ishida *et al.*, Pseudouridine at position 55 in tRNA controls the contents of other modified nucleotides for low-temperature adaptation in the extreme-thermophilic eubacterium Thermus thermophilus. *Nucleic Acids Res* **39**, 2304-2318 (2011).
13. C. van der Feltz, A. C. DeHaven, A. A. Hoskins, Stress-induced Pseudouridylation Alters the Structural Equilibrium of Yeast U2 snRNA Stem II. *J Mol Biol* **430**, 524-536 (2018).
14. M. Brocard, A. Ruggieri, N. Locker, m6A RNA methylation, a new hallmark in virus-host interactions. *J Gen Virol* **98**, 2207-2214 (2017).
15. G. Lichinchi *et al.*, Dynamics of the human and viral m(6)A RNA methylomes during HIV-1 infection of T cells. *Nat Microbiol* **1**, 16011 (2016).
16. B. Baquero-Perez *et al.*, The Tudor SND1 protein is an m(6)A RNA reader essential for replication of Kaposi's sarcoma-associated herpesvirus. *Elife* **8** (2019).
17. N. S. Gokhale *et al.*, N6-Methyladenosine in Flaviviridae Viral RNA Genomes Regulates Infection. *Cell Host Microbe* **20**, 654-665 (2016).
18. O. Manners, B. Baquero-Perez, A. Whitehouse, m(6)A: Widespread regulatory control in virus replication. *Biochim Biophys Acta Gene Regul Mech* **1862**, 370-381 (2019).
19. B. Baquero-Perez, D. Geers, J. Diez, From A to m(6)A: The Emerging Viral Epitranscriptome. *Viruses* **13** (2021).
20. D. G. Courtney *et al.*, Epitranscriptomic Addition of m(5)C to HIV-1 Transcripts Regulates Viral Gene Expression. *Cell Host Microbe* **26**, 217-227 e216 (2019).
21. B. A. Henry *et al.*, Pseudouridylation of Epstein-Barr Virus Noncoding RNA EBER2 Facilitates Lytic Replication. *RNA* 10.1261/rna.079219.122 (2022).
22. C. Arias *et al.*, 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* **10**, e1003847 (2014).

23. B. Baquero-Perez, A. Whitehouse, Hsp70 Isoforms Are Essential for the Formation of Kaposi's Sarcoma-Associated Herpesvirus Replication and Transcription Compartments. *PLoS Pathog* **11**, e1005274 (2015).
24. J. Nicholas, Human herpesvirus 8-encoded proteins with potential roles in virus-associated neoplasia. *Front Biosci* **12**, 265-281 (2007).
25. K. L. Harper *et al.*, Virus-modified paraspeckle-like condensates are hubs for viral RNA processing and their formation drives genomic instability. *Nat Commun* **15**, 10240 (2024).
26. A. Grundhoff, D. Ganem, Inefficient establishment of KSHV latency suggests an additional role for continued lytic replication in Kaposi sarcoma pathogenesis. *J Clin Invest* **113**, 124-136 (2004).
27. L. Giffin, F. Yan, M. Ben Major, B. Damania, Modulation of Kaposi's sarcoma-associated herpesvirus interleukin-6 function by hypoxia-upregulated protein 1. *J Virol* **88**, 9429-9441 (2014).
28. B. R. Jackson, M. Noerenberg, A. Whitehouse, A Novel Mechanism Inducing Genome Instability in Kaposi's Sarcoma-Associated Herpesvirus Infected Cells. *PLoS Pathog* **10**, e1004098 (2014).
29. D. Macveigh-Fierro, A. Cicerchia, A. Cadorette, V. Sharma, M. Muller, The m(6)A reader YTHDC2 is essential for escape from KSHV SOX-induced RNA decay. *Proc Natl Acad Sci U S A* **119** (2022).
30. M. Bansal *et al.*, Integrative analysis of nanopore direct RNA sequencing data reveals a role of PUS7-dependent pseudouridylation in regulation of m (6) A and m (5) C modifications. *bioRxiv* 10.1101/2024.01.31.578250 (2024).
31. S. Huang *et al.*, Interferon inducible pseudouridine modification in human mRNA by quantitative nanopore profiling. *Genome Biol* **22**, 330 (2021).
32. W. Zhang, T. Pan, Pseudouridine RNA modification detection and quantification by RT-PCR. *Methods* **203**, 1-4 (2022).
33. L. S. Zhang, Q. Dai, C. He, BID-seq: The Quantitative and Base-Resolution Sequencing Method for RNA Pseudouridine. *ACS Chem Biol* **18**, 4-6 (2023).
34. V. Khoddami *et al.*, Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc Natl Acad Sci U S A* **116**, 6784-6789 (2019).
35. T. M. Carlile *et al.*, Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143-146 (2014).
36. X. Wang *et al.*, Dysregulation of pseudouridylation in small RNAs contributes to papillary thyroid carcinoma metastasis. *Cancer Cell Int* **24**, 337 (2024).
37. M. Campbell, Y. Izumiya, PAN RNA: transcriptional exhaust from a viral engine. *J Biomed Sci* **27**, 41 (2020).
38. K. Hiura *et al.*, KSHV ORF59 and PAN RNA Recruit Histone Demethylases to the Viral Chromatin during Lytic Reactivation. *Viruses* **12** (2020).
39. J. Sztuba-Solinska *et al.*, Kaposi's sarcoma-associated herpesvirus polyadenylated nuclear RNA: a structural scaffold for nuclear, cytoplasmic and viral proteins. *Nucleic Acids Res* **45**, 6805-6821 (2017).
40. R. B. Tunnicliffe, C. Levy, H. D. Ruiz Nivia, R. M. Sandri-Goldin, A. P. Golovanov, Structural identification of conserved RNA binding sites in herpesvirus ORF57 homologs: implications for PAN RNA recognition. *Nucleic Acids Res* **47**, 1987-2001 (2019).
41. M. J. Massimelli *et al.*, Stability of a long noncoding viral RNA depends on a 9-nt core element at the RNA 5' end to interact with viral ORF57 and cellular PABPC1. *Int J Biol Sci* **7**, 1145-1160 (2011).
42. M. J. Massimelli, V. Majerciak, M. Kruhlak, Z. M. Zheng, Interplay between polyadenylate-binding protein 1 and Kaposi's sarcoma-associated herpesvirus ORF57 in accumulation of polyadenylated nuclear RNA, a viral long noncoding RNA. *J Virol* **87**, 243-256 (2013).

43. W. Zhang, M. J. Eckwahl, K. I. Zhou, T. Pan, Sensitive and quantitative probing of pseudouridine modification in mRNA and long noncoding RNA. *RNA* **25**, 1218-1225 (2019).
44. K. S. Rajan *et al.*, Pseudouridines on Trypanosoma brucei mRNAs are developmentally regulated: Implications to mRNA stability and protein binding. *Mol Microbiol* **116**, 808-826 (2021).
45. S. Schumann *et al.*, Targeting the ATP-dependent formation of herpesvirus ribonucleoprotein particle assembly as an antiviral approach. *Nat Microbiol* **2**, 16201 (2016).
46. A. Taylor *et al.*, Mutation of a C-terminal motif affects Kaposi's sarcoma-associated herpesvirus ORF57 RNA binding, nuclear trafficking, and multimerization. *J Virol* **85**, 7881-7891 (2011).
47. E. Sei, T. Wang, O. V. Hunter, Y. Xie, N. K. Conrad, HITS-CLIP analysis uncovers a link between the Kaposi's sarcoma-associated herpesvirus ORF57 protein and host pre-mRNA metabolism. *PLoS Pathog* **11**, e1004652 (2015).
48. J. C. Murphy *et al.*, Kaposi's sarcoma-associated herpesvirus induces specialised ribosomes to efficiently translate viral lytic mRNAs. *Nat Commun* **14**, 300 (2023).
49. H. Carden *et al.*, K(v)1.3-induced hyperpolarization is required for efficient Kaposi's sarcoma-associated herpesvirus lytic replication. *Sci Signal* **17**, eadg4124 (2024).
50. Y. Furuse, RNA Modifications in Genomic RNA of Influenza A Virus and the Relationship between RNA Modifications and Viral Infection. *Int J Mol Sci* **22** (2021).
51. C. Martinez Campos *et al.*, Mapping of pseudouridine residues on cellular and viral transcripts using a novel antibody-based technique. *RNA* **27**, 1400-1411 (2021).
52. Y. Zou *et al.*, A Comparative Evaluation of Computational Models for RNA modification detection using Nanopore sequencing with RNA004 Chemistry. *bioRxiv* 10.1101/2025.02.03.636352 (2025).
53. O. B. Spiller *et al.*, Complement regulation by Kaposi's sarcoma-associated herpesvirus ORF4 protein. *J Virol* **77**, 592-599 (2003).
54. B. R. Anderson *et al.*, Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. *Nucleic Acids Res* **38**, 5884-5892 (2010).
55. A. C. Rintala-Dempsey, U. Kothe, Eukaryotic stand-alone pseudouridine synthases - RNA modifying enzymes and emerging regulators of gene expression? *RNA Biol* **14**, 1185-1196 (2017).
56. J. Vogele *et al.*, Structural and dynamic effects of pseudouridine modifications on noncanonical interactions in RNA. *RNA* **29**, 790-807 (2023).
57. S. H. Stubbs, O. V. Hunter, A. Hoover, N. K. Conrad, Viral factors reveal a role for REF/Aly in nuclear RNA stability. *Mol Cell Biol* **32**, 1260-1270 (2012).
58. S. Schwartz *et al.*, Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148-162 (2014).
59. R. Shekhar *et al.*, High-density resolution of the Kaposi's sarcoma associated herpesvirus transcriptome identifies novel transcript isoforms generated by long-range transcription and alternative splicing. *Nucleic Acids Res* **52**, 7720-7739 (2024).
60. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
61. K. L. Harper *et al.*, Dysregulation of the miR-30c/DLL4 axis by circHIPK3 is essential for KSHV lytic replication. *EMBO Rep* **23**, e54117 (2022).
62. S. F. Altschul *et al.*, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402 (1997).
63. T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME Suite. *Nucleic Acids Res* **43**, W39-49 (2015).

Acknowledgements

We are very thankful to Professor Jae Jung (UCLA) for the TREx BCBL1-Rta cell line and Dr Edwin Chen (University of Westminster) for the psPAX2 and pVSV.G plasmids. Part of the bioinformatics experiments were undertaken on ARC4, a cluster from the High Performance Computing facility at the University of Leeds, UK. This work was supported in parts by Worldwide Cancer Research (16-1025) and BBSRC (BB/T00021X/1).

Author Contributions

AW conceived the study and acquired project funding. TJM performed the experiments, TJM, KLH and AW analysed the resulting data. TJM and EJRV analysed transcriptomics datasets. The manuscript draft was written by TJM and AW. All authors reviewed and edited the final version.

Competing interests

The authors declare no competing financial or non-financial interests.