

RESEARCH ARTICLE

# An integrative approach identifies direct targets of the late viral transcription complex and an expanded promoter recognition motif in Kaposi's sarcoma-associated herpesvirus

Divya Nandakumar<sup>1,2</sup>, Britt Glaunsinger<sup>1,2\*</sup>

**1** Department of Plant & Microbial Biology, University of California Berkeley, CA, United States of America, **2** Howard Hughes Medical Institute, Berkeley, CA, United States of America

\* [glaunsinger@berkeley.edu](mailto:glaunsinger@berkeley.edu)



**OPEN ACCESS**

**Citation:** Nandakumar D, Glaunsinger B (2019) An integrative approach identifies direct targets of the late viral transcription complex and an expanded promoter recognition motif in Kaposi's sarcoma-associated herpesvirus. *PLoS Pathog* 15(5): e1007774. <https://doi.org/10.1371/journal.ppat.1007774>

**Editor:** Eric Johannsen, University of Wisconsin, UNITED STATES

**Received:** February 15, 2019

**Accepted:** April 22, 2019

**Published:** May 16, 2019

**Copyright:** © 2019 Nandakumar, Glaunsinger. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All raw data from RNA sequencing and chromatin immunoprecipitation sequencing experiments have been deposited on the NCBI Gene Expression Omnibus database (GEO accession number GSE126603).

**Funding:** This work was supported by NIH R01 AI122528 (<https://www.nih.gov/>) to B.G., who is also a Howard Hughes Medical Institute

## Abstract

The structural proteins of DNA viruses are generally encoded by late genes, whose expression relies on recruitment of the host transcriptional machinery only after the onset of viral genome replication.  $\beta$  and  $\gamma$ -herpesviruses encode a unique six-member viral pre-initiation complex (vPIC) for this purpose, although how the vPIC directs specific activation of late genes remains largely unknown. The specificity underlying late transcription is particularly notable given that late gene promoters are unusually small, with a modified TATA-box being the only recognizable element. Here, we explored the basis for this specificity using an integrative approach to evaluate vPIC-dependent gene expression combined with promoter occupancy during Kaposi's sarcoma-associated herpesvirus (KSHV) infection. This approach distinguished the direct and indirect targets of the vPIC, ultimately revealing a novel promoter motif critical for KSHV vPIC binding. Additionally, we found that the KSHV vPIC component ORF24 is required for efficient viral DNA replication and identified a ORF24 binding element in the origin of replication that is necessary for late gene promoter activation. Together, these results identify an elusive element that contributes to vPIC specificity and suggest novel links between KSHV DNA replication and late transcription.

## Author summary

Gene expression in DNA viruses often occurs in temporal waves, with expression of essential structural proteins occurring late in infection, after viral genome replication has begun. Strategies underlying expression of these viral late genes are often sophisticated; for example, the  $\beta$ - and  $\gamma$ -herpesviruses encode a six-component viral complex that directs late gene transcription, largely by unknown mechanisms. Here, we evaluated how this complex specifically recognizes late promoters during infection with the oncogenic human  $\gamma$ -herpesvirus Kaposi's sarcoma-associated herpesvirus (KSHV). We found that one of the components of the late transcription complex was required for robust viral DNA replication and binds at the origin of replication, suggesting new links between

Investigator (<https://www.hhmi.org>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

KSHV replication and transcription. Combined measurements of late gene expression and promoter occupancy then revealed which KSHV genes are directly controlled by the late gene transcription complex, leading to identification of a key new regulatory element in KSHV late promoters. Together, these data help explain how the late gene transcription complex is able to bind seemingly minimal promoters with high specificity, ensuring robust expression of viral factors necessary for assembly of progeny virions.

## Introduction

A conserved feature of double-stranded (ds) DNA viruses is that they possess a set of genes that become strongly activated only following viral DNA replication [1]. These “late genes” generally encode factors involved in viral assembly that are essential for virion production. The mechanisms underlying their transcriptional regulation are often non-canonical or distinct from those involved in activating either cellular genes or early viral genes. For example, baculoviruses encode a multisubunit,  $\alpha$ -amanitin-resistant DNA-directed RNA polymerase dedicated specifically to the activation of compact late promoters [2]. Similarly, bacteriophage T7 encodes its own late gene specific RNA polymerase that recognizes unique promoter elements to transcribe late genes [3,4]. Adenoviruses regulate late gene expression using a complex combination of viral and cellular factors; beyond promoter activation, these also regulate transcription termination and contribute to differential posttranscriptional processing of late viral mRNA [5].

Several studies suggest that the expression of late genes is also uniquely regulated in  $\beta$ - and  $\gamma$ -herpesviruses, which include human cytomegalovirus (HCMV), Epstein-Barr virus (EBV), and Kaposi’s sarcoma-associated herpesvirus (KSHV) [6]. Unlike other eukaryotic dsDNA viruses,  $\beta$  and  $\gamma$  herpesviruses do not engage cellular TATA binding protein (TBP) during late gene pre-initiation complex formation [7]. Their core late gene promoters are unusually small (12–15 base pairs (bp)), with the only recognizable conserved element being a modified TATA box (TATT) ~30 bp upstream of the transcription start site (TSS) [8–10]. This is in contrast to eukaryotic and most early herpesviral promoters, which have multiple promoter elements that are important for promoter recognition and activation by RNA polymerase II (RNA pol II) [11]. Another notable feature is that activation of these minimal promoters requires a virally-encoded six-protein complex called the viral Pre-Initiation Complex (vPIC) that coordinates mammalian RNA pol II [12,13]. The vPIC is presumed to be uniquely adapted to activate minimal promoters, yet the mechanisms underlying its activity and specificity remain largely unknown.

The core component of the KSHV vPIC is ORF24 (BcRF1 in EBV and UL87 in HCMV), a viral TBP mimic that displays putative structural similarity to cellular TBP [7,14,15]. Unlike cellular TBP, the viral TBP mimics have expanded functionality in that they bind RNA pol II in addition to promoter DNA, making them unique among transcription factors that function in eukaryotic cells. Beyond ORF24, the other essential components of the KSHV vPIC are ORFs 18, 30, 31, 34, and 66, but little is known about how each of these factors contributes to vPIC activity [7,16–18]. ORF34 interacts with multiple components of the vPIC, and is thus thought to function as a scaffolding factor [13]. The ORF18 homolog in HCMV (pUL79) has been shown to function as an elongation factor, although whether it plays a similar role in KSHV and EBV is unclear [19]. Despite the paucity of functional information on the individual vPIC components, protein-protein interaction mapping has provided insight into the overall complex organization [13,18]. Furthermore, point mutants that disrupt individual inter-

subunit vPIC interactions abrogate late gene transcription, highlighting the fact that vPIC complex integrity is critical for function [13,20,21].

The mechanistic basis for promoter selectivity has also been challenging to grasp in the context of the minimal, largely nondescript nature of late gene promoters. While a modified TATA box with a TATT sequence is enriched in several late promoters of KSHV, MHV68, and EBV [8–10,22], the ORF24 TBP mimic shows partial binding to a mutant promoter in which the TATT sequence is modified to resemble a canonical TATA box [8]. However, ORF24 does not bind promoters that inherently have a canonical TATA box [8], which suggests that other DNA elements surrounding the TATT sequence in late gene promoters contribute to ORF24 binding or vPIC assembly on the DNA.

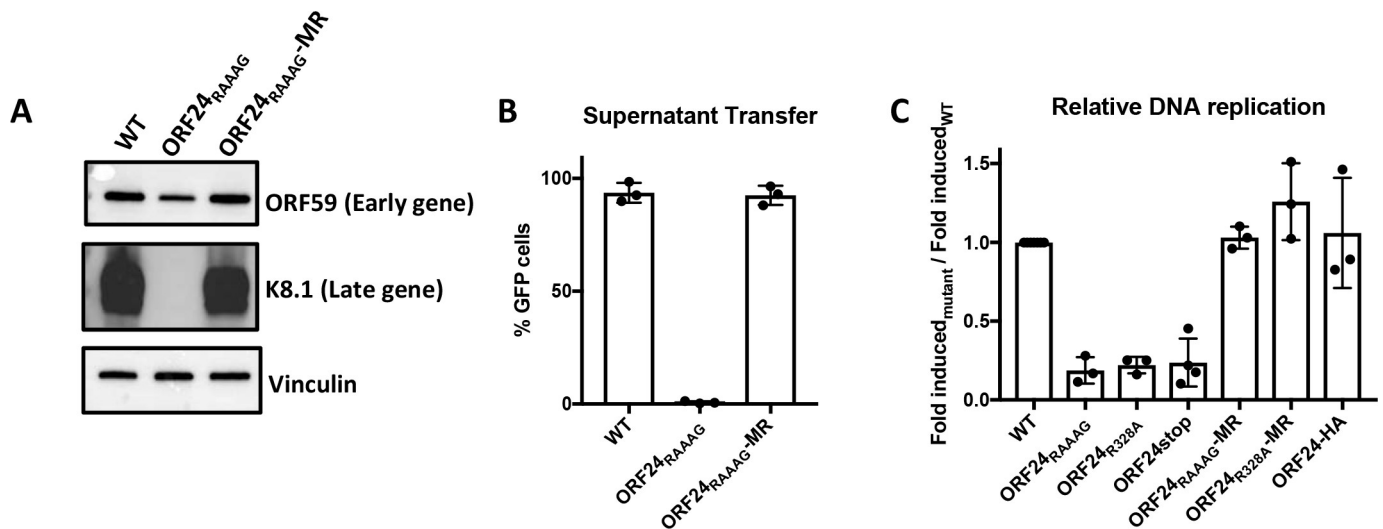
One possible reason that additional DNA elements involved in vPIC promoter recognition have not been identified is that we do not yet know the subset of viral promoters under the direct influence of the vPIC. Techniques such as RNA Sequencing (RNA Seq) and microarrays have helped define the set of genes that are expressed with late kinetics and in a manner dependent on viral DNA replication [23–26]. Furthermore, viral mutagenesis and knockdown studies have revealed viral genes whose expression is impaired in the absence of vPIC components [17,18,22,27,28]. However, these assays cannot distinguish direct vPIC targets from indirect targets, such as genes regulated by factors that are themselves dependent on the vPIC. Inclusion of such indirect targets in the pool of vPIC-regulated promoters could confound the identification of conserved elements important for direct vPIC regulation.

Here, we integrated RNA Seq with chromatin immunoprecipitation high-throughput sequencing (ChIP Seq) for vPIC components during KSHV infection to distinguish genes whose expression is directly versus indirectly controlled by the vPIC. In doing so we made four key findings that provide insight into late gene promoter activation in KSHV. First, we defined the set of direct vPIC promoter targets within the KSHV genome and showed that other promoters are indeed nonproductively bound or only indirectly regulated by the vPIC. Second, by restricting our analysis to these direct targets, we revealed a novel promoter motif that contributes significantly to vPIC binding and thus provides insight into features governing vPIC specificity. Notably, this motif did not emerge when the indirect vPIC targets were included in the analysis. Third, we found that the KSHV TBP mimic ORF24 is required for efficient viral DNA replication. This was unexpected given that the other studied vPIC components have been shown to be dispensable for this process. Finally, we identified a binding site for ORF24-34 at the lytic origin of replication that is essential for activating late gene expression, providing a new connection between viral DNA replication and late gene transcription.

## Results

### KSHV ORF24 is essential for late gene expression and contributes to viral DNA replication

Three leucine residues in the N-terminal domain of KSHV ORF24 (amino acids 73–75) are essential for its interaction with RNA Pol II and its ability to promote late gene transcription [7]. Given that most viral proteins are multifunctional, we engineered the ORF24<sub>RLLLG->RAAAG</sub> mutation (ORF24<sub>RAAAG</sub>) into KSHV BAC16 to precisely evaluate the impact of the transcriptional function of ORF24 during lytic infection. We also generated the corresponding mutant rescue (ORF24<sub>RAAAG</sub>-MR) to ensure that any observed phenotypes were not the effect of secondary mutations in the BAC. Identical banding patterns from *Rsr*II digestion of WT and mutant BAC16 DNAs verified that there were no gross rearrangements of the viral genome during mutagenesis (S1 Fig-lanes1-3). We established stable cell lines latently infected with WT, ORF24<sub>RAAAG</sub>, and ORF24<sub>RAAAG</sub>-MR KSHV in the iSLK background, which contains a



**Fig 1. The ORF24-RNA pol II interaction is essential for late gene expression and contributes to viral DNA replication.** iSLK cells harboring WT BAC16, the ORF24<sub>RAAAG</sub> mutant or the ORF24<sub>RAAAG</sub> mutant rescue (MR) were reactivated with doxycycline and sodium butyrate for 72 hours. (A) Western blot showing expression of an early protein (ORF59) and a late protein (K8.1) in WT, ORF24<sub>RAAAG</sub> and ORF24<sub>RAAAG</sub>-MR. Vinculin was used as a loading control. (B) Progeny virion production was measured by supernatant transfer on to 293T cells and quantified by flow cytometry. (C) DNA replication of different mutants and MRs was measured by qPCR and normalized to the respective uninduced sample. The fold induction of all mutants was normalized to the WT to get the relative change in DNA replication compared to WT.

<https://doi.org/10.1371/journal.ppat.1007774.g001>

doxycycline-inducible version of the KSHV lytic transactivator protein ORF50 (RTA) to enable lytic reactivation [29]. During lytic reactivation, these three viruses expressed similar levels of the early protein ORF59, while the ORF24<sub>RAAAG</sub> mutant was selectively unable to express the K8.1 late protein or produce infectious virions (Fig 1A and 1B). These findings are consistent with previous complementation data [7], and confirm the importance of the ORF24-Pol II interaction for late gene expression.

Surprisingly, we also observed ~6-fold defect in DNA replication in the ORF24<sub>RAAAG</sub> mutant (Fig 1C), which was unexpected given that previous studies with mutations in ORF24 or other components of the vPIC did not show this defect [7,13,16–18]. This defect was rescued in the ORF24<sub>RAAAG</sub>-MR cell line, indicating that it was not due to secondary mutations in the BAC (Fig 1C). However, to independently confirm this observation, we engineered two additional ORF24 viral mutants: the point mutant ORF24<sub>R328A</sub>, which abrogates its essential interaction with the ORF34 component of the vPIC [13], and a stop mutant (ORF24<sub>Stop</sub>) that has a premature stop codon in the ORF24 coding sequence. Inclusion of the ORF24<sub>Stop</sub> virus allowed us to evaluate whether the point mutants were functioning in a dominant negative manner, for example through nonproductive DNA binding. Finally, we also N-terminally tagged endogenous WT ORF24 with HA (HA-ORF24) to assess whether any alterations to this locus impacted viral DNA replication. Notably, we observed a similar ~5-6-fold defect in DNA replication in the ORF24<sub>RAAAG</sub>, ORF24<sub>R328A</sub>, and ORF24<sub>Stop</sub> viruses, which was restored to WT levels in the ORF24<sub>RAAAG</sub> and ORF24<sub>R328A</sub> MR viruses. Moreover, there was no DNA replication defect in the HA-ORF24 virus (Fig 1C), demonstrating that the defect in DNA replication is not a consequence of manipulating the ORF24 locus but is specific to disrupting its role in transcription.

### Specific gene clusters show differential dependence on the ORF24-Pol II interaction

To more comprehensively analyze how the ORF24-Pol II interaction impacts gene expression during infection, we deep sequenced RNA from three replicates each of iSLK cells infected

with the ORF24<sub>RAAAG</sub> or the ORF24<sub>RAAAG</sub>-MR virus. Samples were sequenced at 24 and 48 h post lytic reactivation with doxycycline and sodium butyrate, as these times captured predominantly early (24 h) versus both early and late (48 h) viral gene expression (S2 Fig). Given that overlapping transcripts are pervasive in the KSHV genome [22,30], we calculated the primary transcript counts as described in Bruce et al [30] to accurately quantify transcript levels.

Differential expression analysis of genes in infected cells containing WT ORF24 (ORF24<sub>RAAAG</sub> MR) confirmed that the majority of viral genes are upregulated between 24 and 48 h post reactivation (Fig 2A), and those that are most strongly upregulated are predominantly classified as late genes (shown in blue). Conversely, there was a generalized reduction in transcript levels at the 48 h time point in cells containing KSHV ORF24<sub>RAAAG</sub> compared to MR, with the most strongly downregulated genes being those classified as late genes (Fig 2B).

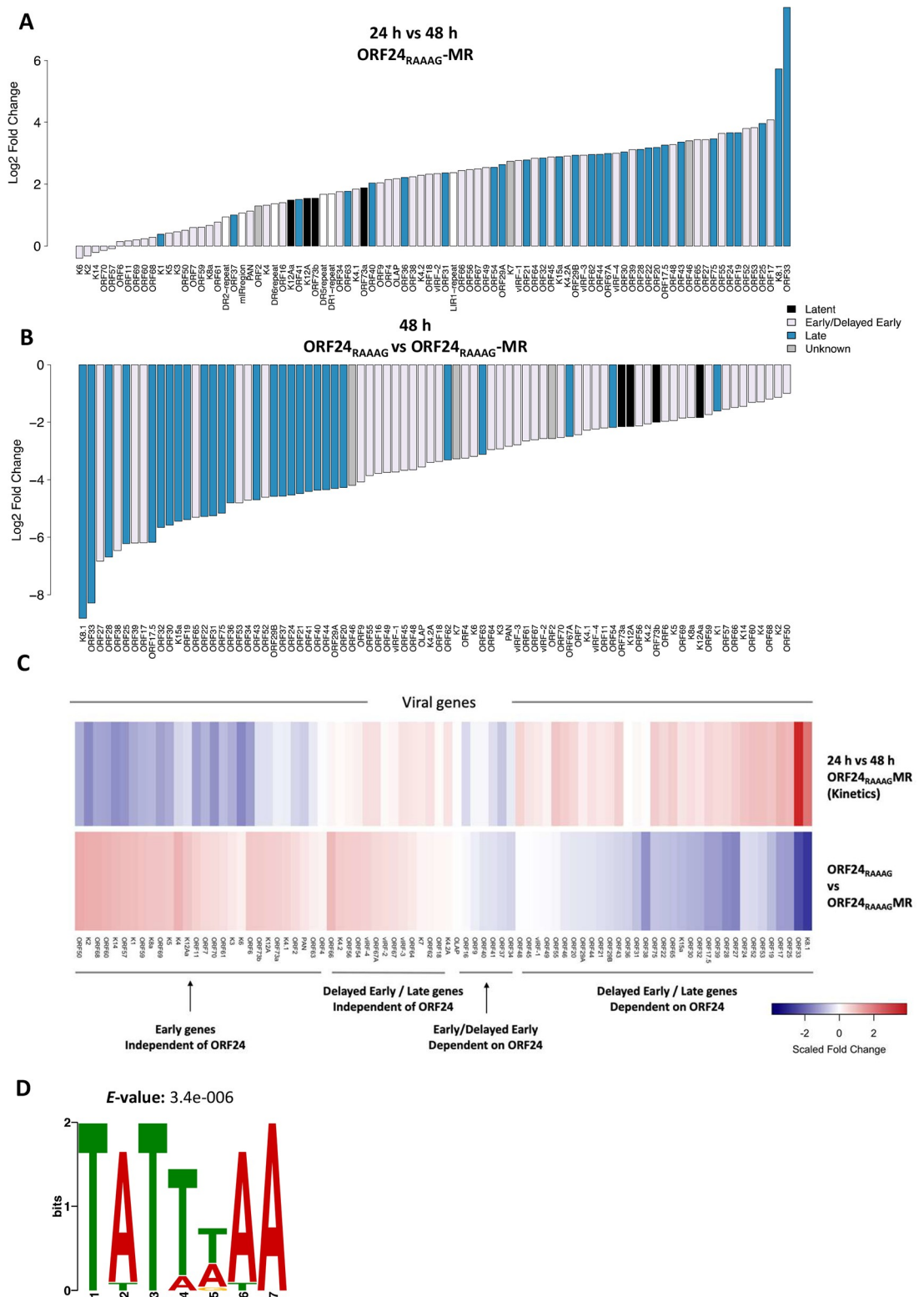
Combining the above ORF24<sub>RAAAG</sub> mutant and MR induction data into a heatmap confirmed that in general, genes with later expression kinetics tended to be more severely impacted by the ORF24<sub>RAAAG</sub> mutation compared to genes with early expression kinetics (Fig 2C). (The fold change data was scaled to allow for easy visualization). However, not all genes fit this model. Some genes categorized as early based on their induction pattern showed a moderate dependence on the ORF24-Pol II interaction (e.g. ORFs 9, 40, 34, 41, 16 and 37). Similarly, a subset of genes that show late kinetics were less dependent on the ORF24-Pol II interaction (e.g. ORFs 67A, 62, K4.2A, 64, K7, 54, 67, 56, 66, K4.2 and vIRFs-2, 3 and 4). Classification as early or late and dependence on the ORF24-Pol II interaction was determined using mean log<sub>2</sub> fold change. Together, these data suggest that not all kinetic late genes are regulated by the vPIC and there may be subset of early genes that are direct or indirect targets of the vPIC.

### The TATTWAA motif is enriched in the promoters of genes whose expression is linked to the ORF24-Pol II interaction

We then looked for motifs enriched in the promoters of genes most strongly downregulated in the ORF24<sub>RAAAG</sub> mutant using Multiple EM for Motif Elicitation (MEME) [31], a program that discovers un-gapped motifs that are enriched in a given set of sequences. We defined the promoter sequence as 100 bp upstream from the transcription start site (TSS) or 300 bp upstream from the translation start site when the TSS was unknown. Analysis of promoter sequences from 14 of the most strongly down-regulated genes (log<sub>2</sub> fold change greater than one standard deviation from the mean) showed that TATTWAA is the only motif that is enriched in this set of sequences (Fig 2D).

We next monitored for the occurrence of this motif across all promoters in the viral genome using Find Individual Motif Occurrences (FIMO) [32], which scans a set of sequences for matches to motifs provided by the user. 23 of 83 viral promoters were found to have the TATTWAA motif with a high score, of which 16 promoters had the motif ~30 bp upstream from the TSS (Table 1), consistent with the location of the canonical TATA box in eukaryotic promoters. Of the remaining seven promoter sequences, the TSS of six of the promoters is not known (shown in red in Table 1); hence, the exact position of the TATTWAA motif from the TSS cannot be estimated. ORF25 is the only exception, where the TATTWAA sequence is not present ~30 bp upstream from the TSS, although it does have a canonical TATA box (TATAAAA) at that position.

Notably, a subset of the genes containing the TATTWAA motif showed less than average repression by the ORF24<sub>RAAAG</sub> mutation (ORF55, ORF64, ORF66, ORF59, K4.1, ORF4, ORF8, ORF58, ORF26 and ORF62). One explanation for this observation is that these genes could be regulated by cellular transcription factors and other *cis* elements which are absent in



**Fig 2. Differential expression analysis of ORF24-Pol II interaction mutant and mutant rescue.** iSLK cells harboring ORF24<sub>RAAAG</sub> and ORF24<sub>RAAAG</sub>-MR were reactivated with doxycycline and sodium butyrate for 24 and 48 h, whereupon total RNA was extracted and subjected to library preparation and RNA sequencing. (A) Bar plot showing log<sub>2</sub> fold change of the viral genes in the ORF24<sub>RAAAG</sub>-MR at 48 h relative to 24 h. Genes were color coded by kinetic class as defined in (30). (B) Bar plot showing log<sub>2</sub> fold change of the viral genes in the mutant relative to the MR. Genes were color coded by kinetic class as defined in

(30). (C) Heatmap of the scaled  $\log_2$  fold change of the genes induced in the MR between 24 to 48 h and genes repressed in the mutant compared to MR at 48 h. Gene clusters with different kinetics and dependence on ORF24-Pol II interaction are indicated. (D) Consensus motif identified by MEME analysis of promoters from 14 genes that showed greater than one standard deviation repression in the mutant relative to the MR.

<https://doi.org/10.1371/journal.ppat.1007774.g002>

the genes that are downregulated in the ORF24<sub>RAAAG</sub> mutant. The data also suggest that the motif may not be sufficient to mark regulation by the vPIC, which is consistent with a similar observation in MHV68 [8]. Furthermore, several genes (ORF28, ORF30, K15, ORF31 and ORF75) were strongly affected by the mutation ( $\log_2$  fold change greater than one standard deviation from the mean) but lack the TATTWAA motif. The number of genes in this set increases to 26 if we broaden the criteria to include all the genes that show greater than average repression in the presence of the mutation. These genes could either be down regulated by an indirect mechanism in the absence of the ORF24-Pol II interaction or could be bound by the vPIC in a manner independent of the TATT motif.

### Chromatin immunoprecipitation reveals vPIC occupancy across the KSHV genome

Our RNA Seq experiments identified KSHV genes whose expression is influenced by the transcription regulatory function of ORF24. However, RNA Seq does not distinguish genes that are directly bound by the vPIC from those that are indirect transcriptional targets (e.g. controlled by other factors whose expression is induced by ORF24). We therefore sought to identify promoters bound by the KSHV vPIC complex using ChIP Seq. Previous studies suggest

**Table 1. Occurrence of TATTWAA motif in KSHV.** ORFs shown in red do not have a known Transcription Start Site (TSS).

Promoter	Start	Stop	Strand	Score	p-value	q-value	Matched Sequence	Position
ORF17	68	74	+	11.8614	0.000114	0.143	TATTTAA	32
ORF62	68	74	+	11.8614	0.000114	0.143	TATTTAA	32
ORF55	69	75	+	11.8614	0.000114	0.143	TATTTAA	31
ORF39	70	76	+	11.8614	0.000114	0.143	TATTTAA	30
ORF8	71	77	+	11.8614	0.000114	0.143	TATTTAA	29
ORF52	71	77	+	11.8614	0.000114	0.143	TATTTAA	29
ORF17.5	71	77	+	11.8614	0.000114	0.143	TATTTAA	29
ORF59	71	77	+	11.8614	0.000114	0.143	TATTTAA	29
ORF58	71	77	+	11.8614	0.000114	0.143	TATTTAA	29
ORF19	187	193	+	11.8614	0.000114	0.143	TATTTAA	113
ORF27	197	203	+	11.8614	0.000114	0.143	TATTTAA	103
ORF32	243	249	+	11.8614	0.000114	0.143	TATTTAA	57
ORF25	11	17	+	11.198	0.000235	0.153	TATTAAA	89
K8.1	68	74	+	11.198	0.000235	0.153	TATTAAA	32
K4.1	68	74	+	11.198	0.000235	0.153	TATTAAA	32
ORF65	69	75	+	11.198	0.000235	0.153	TATTAAA	31
ORF33	70	76	+	11.198	0.000235	0.153	TATTAAA	30
ORF26	71	77	+	11.198	0.000235	0.153	TATTAAA	29
ORF38	71	77	+	11.198	0.000235	0.153	TATTAAA	29
ORF4	71	77	+	11.198	0.000235	0.153	TATTAAA	29
ORF64	145	151	+	11.198	0.000235	0.153	TATTAAA	155
ORF66	218	224	+	11.198	0.000235	0.153	TATTAAA	82
ORF22	229	235	+	11.198	0.000235	0.153	TATTAAA	71

<https://doi.org/10.1371/journal.ppat.1007774.t001>

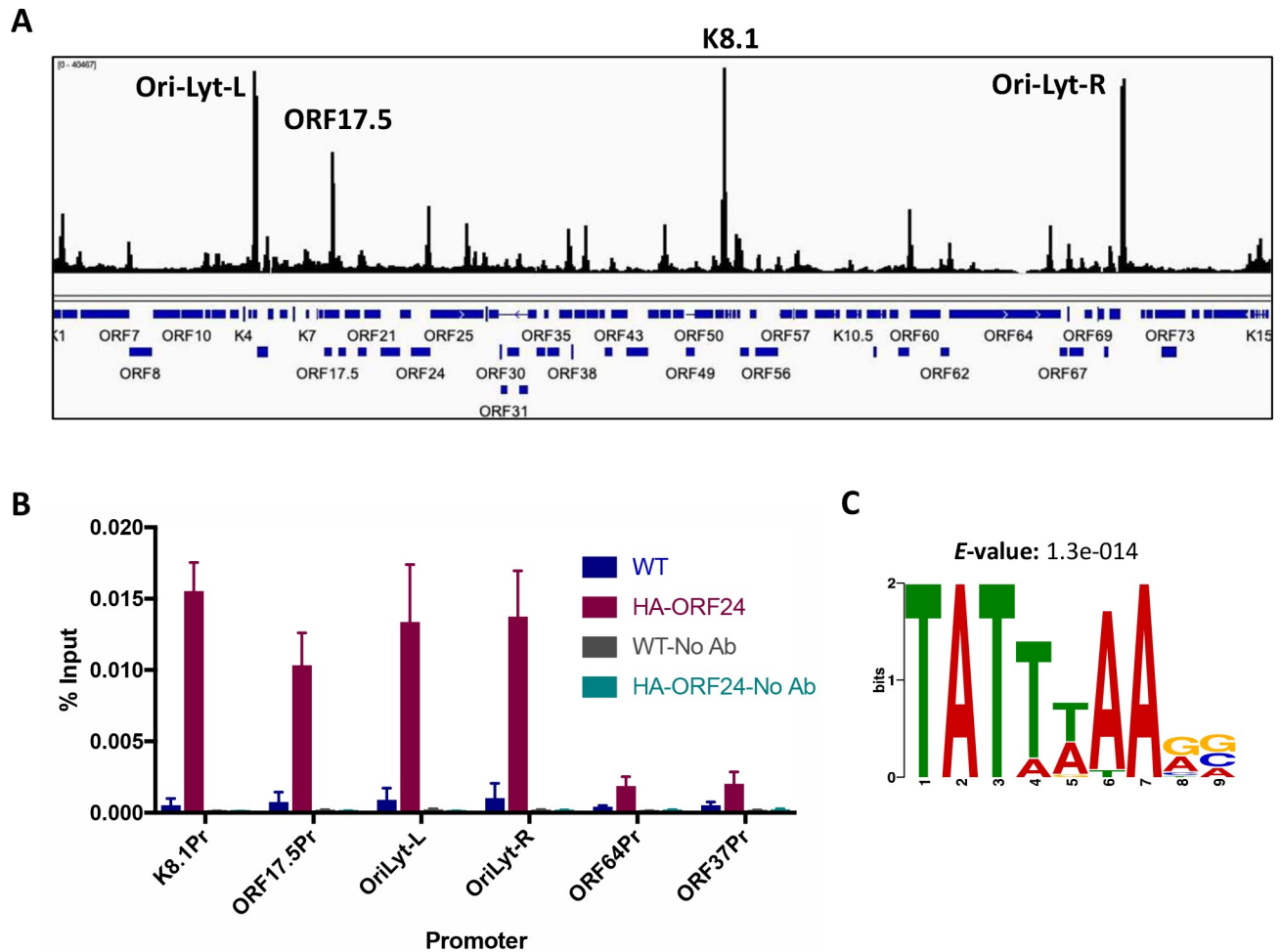
that ORF34 acts as a scaffolding factor that brings together all components of the vPIC [13] including ORF24, which binds both RNA Pol II and TATT-containing DNA. We reasoned that ORF34 promoter occupancy would therefore be a good proxy for vPIC assembly, and additionally would avoid potential pitfalls associated with any non-productive or independent interactions between ORF24 and DNA.

To facilitate ChIP Seq, we added an N-terminal HA tag to endogenous ORF34 in KSHV BAC16 (HA-ORF34) (S3A Fig), and generated a stable, doxycycline-inducible iSLK cell line containing this tagged virus as previously described. HA-ORF34 produced WT levels of infectious virions in a supernatant transfer assay (S3B Fig), indicating that the tag does not disrupt ORF34 activity. The integrity of the BAC DNA was verified by *RsrII* digestion (S1 Fig-Lane7). We performed the ChIP Seq experiment in duplicate at 48 h post reactivation to ensure active late gene expression. Reactivated iSLK cells with untagged BAC16 served as a negative control for the immunoprecipitation. The individual ChIP peaks for the two replicates as well as the respective inputs and controls are shown in S3C Fig and S1 File. The ChIP Seq data showed that HA-ORF34 bound several locations across the viral genome (Fig 3A) but was not detectably associated with the host genome. A peak calling algorithm (HOMER) [33] identified 25 common peaks in the viral genome across two replicates that were significantly enriched in HA-ORF34 relative to untagged WT or input DNA (S1 File and S3C Fig). Notably, 20 of the 25 peaks were found upstream of ORFs, with ~80% of the peaks centered less than 200 bp from the closest TSS (S1 File). Given the link between DNA replication and late transcription in herpesviruses, it was particularly interesting that two of the most prominent peaks are upstream of the two minimal lytic origins of replication (Ori) in the KSHV genome (Fig 3A). One of the peaks could not be uniquely assigned to a single ORF and was not considered further. Of the remaining 19 ORFs associated with peaks, 13 are involved in packaging, capsid assembly (or form part of capsid itself), 4 are involved in immune evasion and 2 in DNA replication (Table 2). While several of them have also been previously classified as late genes (12/20), it is notable that 6 of them have been classified as early genes, again suggesting that not all targets of the vPIC are late genes.

To confirm that the observed binding pattern was representative of vPIC assembly, we independently validated several prominent ChIP Seq peaks by ChIP-qPCR from cells infected with KSHV containing HA-tagged ORF24 (HA-ORF24). Similar to HA-ORF34, the HA-ORF24 virus produced WT levels of infectious virions (S3B Fig). Indeed, the HA-ORF24 ChIP-qPCR data were consistent with the HA-ORF34 ChIP-Seq data (Fig 3B), with peaks showing higher percent input values compared to control regions (ORF64 and ORF37 promoters).

A MEME analysis of the promoters associated with the peaks again showed that the only motif enriched in these promoters is TATTWAARV (Fig 3C). 15 of the 19 promoters have the TATTWAA motif and in a majority of the sequences, it is present ~30 bp from the TSS (Table 2). Of the four exceptions, two of the promoters (ORF53 and ORF6) have a TATATAA motif with an A at position 4 in place of the T. Indeed, mORF53 has previously shown to be a target of the vPIC in MHV68 despite the deviation from consensus [8]. Two key points can be drawn from the ChIP seq data. First, several genes whose promoters contain a TATTWAA motif do not show ORF34 binding (e.g. ORFs 17, 32, 22, 55, 58, 64 and 66). Second, while most promoters bound by ORF34 have a TATTWAA motif, there are some exceptions (ORF45, ORF68, ORF6 and ORF53). These observations indicate that while the TATTWAA motif is enriched in promoters regulated by the vPIC, it is not sufficient to ensure binding of this specialized transcription complex.





**Fig 3. ChIP-Seq with HA-ORF34 shows multiple binding sites across the viral genome.** (A) ChIP-Seq was performed using an anti-HA antibody to immunoprecipitate DNA prepared from iSLK HA-ORF34 cells reactivated for 48 h with doxycycline and sodium butyrate. iSLK-WT cells lacking the HA tag and input DNA were used as negative controls. The alignment files were converted to the tiled data file (tdf) format and the ChIP background from the untagged cells was subtracted and visualized on the Integrated Genome Viewer (IGV) (B) ChIP-qPCR was performed using an anti-HA antibody to immunoprecipitate DNA prepared from iSLK HA-ORF24 cells reactivated for 48 h with doxycycline and sodium butyrate. The DNA was quantified by qPCR with primers specific to the promoter regions of the indicated genes. Data were normalized to the level of input DNA and are presented as percentages of the input. Data shown are from three biological replicates. (C) Consensus motif identified by MEME analysis of promoters from 19 genes whose promoters had a ChIP peak.

<https://doi.org/10.1371/journal.ppat.1007774.g003>

### Thirteen KSHV genes are direct targets of the vPIC

Combining the RNA Seq and ChIP Seq data, we defined the direct targets of the vPIC as genes that are both downregulated in the absence of the ORF24-Pol II interaction and show ORF34 binding in the promoter region. By this measure, a total of 13 genes are direct targets of the vPIC (Fig 4A and Table 2 –gray rows). Of the 13 direct targets, 11 code for structural proteins or proteins involved in viral egress while two are tegument proteins involved in immune evasion. Genes whose expression was impaired in the KSHV ORF24<sub>RAAAG</sub> mutant but lacked ORF34 binding at their promoters are likely to be indirect targets of the vPIC (Fig 4A). Finally, it is interesting to note that not all genes that have a ChIP peak are down regulated in the RNA Seq data (e.g. ORFs 4, 6, K4.1, 59, 62 and 68). These could be false positive ChIP peaks and/or have alternate mechanisms of regulation that ensure their transcription even in the absence of the vPIC.

Table 2. Features of ChIP-Seq targets.

ChIP targets	Function	Kinetics	Strand	Position of motif from TSS	Motif
K8.1	Glycoprotein	L	+	32	TATTAAAGG
ORF52	Immune Evasion	L	-	29	TATTTAAAG
ORF27	Glycoprotein	L	+	103	TATTTAATC
ORF53	Glycoprotein	L	-	29	*TATATAAGA
ORF25	Major capsid protein	L	+	89 (29 <sup>#</sup> )	*TATAAAAGG TATTAACC
ORF62	Assembly/DNA maturation	L	-	32	TATTTAAAA
ORF33	Capsid binding	L	+	30	TATTAAGG
ORF8	Glycoprotein	L	+	29	TATTTAAAG
ORF17.5	Viral capsid assembly	L	-	29 (73 <sup>#</sup> )	TATTTAAAG TATTGAAAA
ORF19	Capsid vertex		-	113	TATTTAAGA
ORF26	Minor capsid protein	L	+	29	TATTAAGC
ORF45	Immune evasion (Tegument)	E	-	No motif	No motif
ORF6	Single strand DNA binding protein	E	+	89 (31 <sup>#</sup> )	*TATATAAAC TATTTTAGG
K4.1	Homologs of cellular chemokines	E	-	32	TATTAACC
ORF59	DNA polymerase processivity factor	E	-	29	TATTTAAGG
ORF68	Packaging protein	E	+	No motif	No motif
ORF4	Immune evasion	E	+	29	TATAAAAG
ORF65	Minor capsid protein	L	-	31	TATTAAGC
ORF38	Viral maturation and egress (Tegument protein)	L	+	29	TATTAAGC

# Second site with lower consensus to TATTWAARV motif

Direct targets are shown in gray

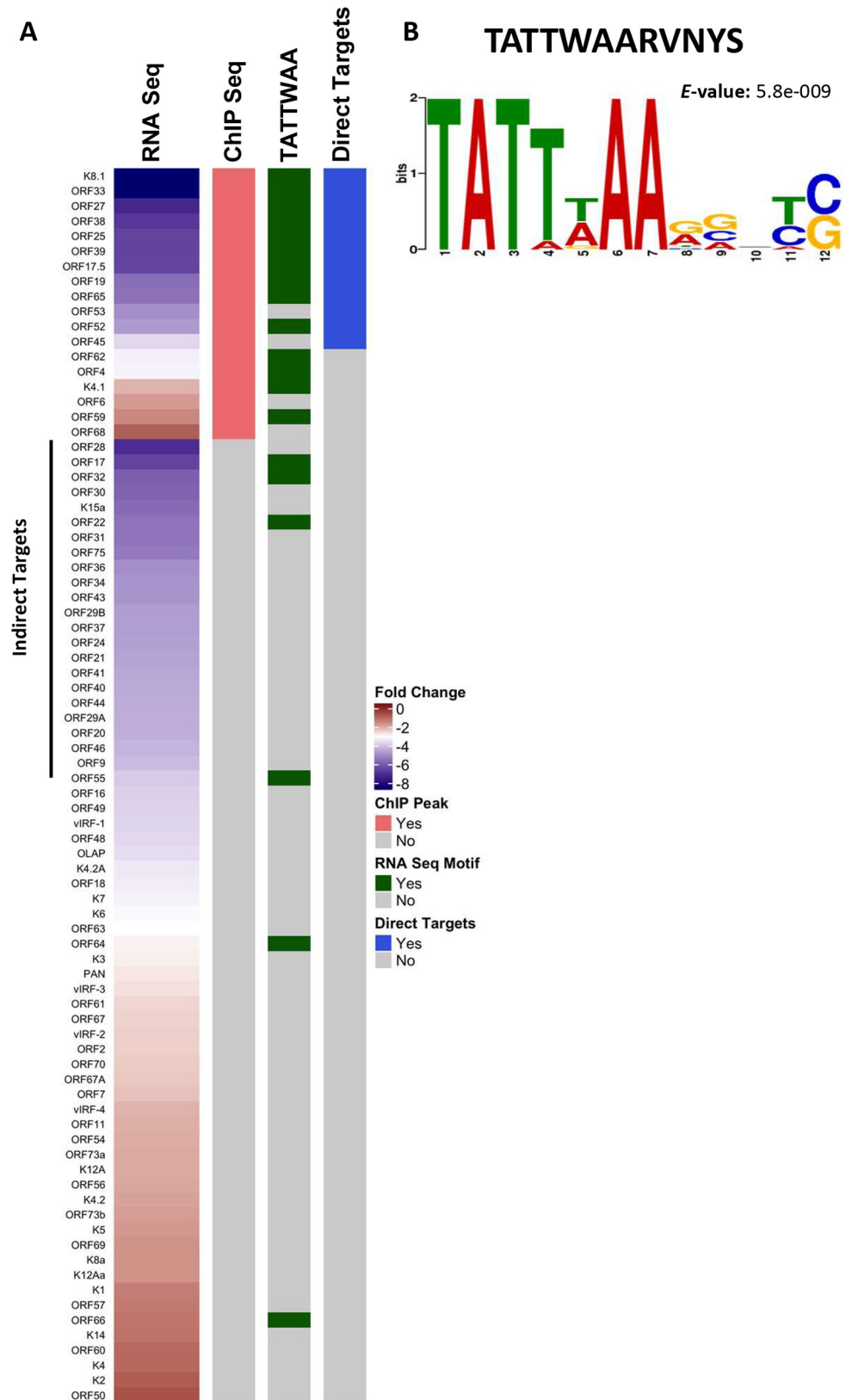
\* Promoters with TATA instead of the TATT sequence

<https://doi.org/10.1371/journal.ppat.1007774.t002>

### The direct vPIC targets reveal an expanded promoter motif

A defining feature of characterized late gene promoters is that they are minimalistic, with no conserved motifs beyond the TATTWAA signature recognized by ORF24. Yet, our data indicate that the TATTWAA motif is insufficient to confer activation by the late gene transcription complex. We hypothesized that the inclusion of potentially indirect vPIC targets into prior motif searches may have hindered identification of other regulatory sequences within these promoters. Indeed, repeating the MEME analysis on the 13 direct targets revealed that an additional 5 bp motif (RVNYS) downstream of the TATTWAA motif was enriched in these promoter sequences (Fig 4B).

To explore the contribution of the 5 bp region downstream of the TATTWAA sequence for regulation by the vPIC, we used a plasmid-based promoter activation assay in which expression of a luciferase reporter was driven by either the K8.1 late or ORF57 early promoter (Fig 5A). The KSHV left lytic origin of replication (ori-lyt) was included on the plasmids, as data from MHV68, KSHV and EBV indicate that the presence of ori-lyt in cis significantly boosts late gene promoter activation during infection [8,10,34]. Each plasmid was transfected together with a plasmid constitutively expressing renilla luciferase (to normalize for transfection efficiency) into lytically reactivated KSHV-positive iSLK cells and the luciferase signal was measured 48 h post transfection and reactivation. As expected, the presence of ori-lyt enhanced K8.1 promoter-driven luciferase expression by 15–20-fold (Fig 5B). Furthermore, unlike the consistently robust ORF57 promoter activity, K8.1 promoter activity on the plasmid



**Fig 4. Identifying direct targets of the vPIC.** (A) Heatmap showing genes repressed in the ORF24<sub>RAAAG</sub> mutant compared to the MR adjacent to the list of genes that show ORF34 binding at their promoter by ChIP-Seq analysis.

Also indicated are viral genes that were identified to have the TATTWAA motif by FIMO analysis. The direct targets were identified by their combined dependence on the ORF24-Pol II interaction and the presence of a ChIP signal at the promoter. (B) Consensus motif identified by MEME analysis of promoters from 13 direct targets.

<https://doi.org/10.1371/journal.ppat.1007774.g004>

was completely repressed in the presence of DNA replication inhibitor PAA (Fig 5C) or in cells lytically infected with the ORF24<sub>RAAAG</sub> mutant (Fig 5D). Taken together, these data confirm that the plasmid assay accurately recapitulates late gene promoter activity.

Notably, mutating the 5 bp RVNYS motif in the K8.1 promoter decreased transcription by > 50% (Fig 5E). This decrease was comparable to mutating the T at position four to an A in the TATTWAA sequence, which has been shown to be important (although not sufficient) for recognition and binding by ORF24 [10]. Replacement of the 5 bp RVNYS motif with either synonymous (-S) or nonsynonymous (-NS) bases yielded a similar decrease in promoter activity (Fig 5E-K8.1<sub>GGGAC->TACGA-S</sub>, K8.1<sub>GGGAC->ACTCG-NS</sub>), suggesting a role for the overall architecture of this motif rather than the primary sequence. In support of this hypothesis, a preliminary analysis of four structural features of the expanded motif in the WT and mutant predict that the average roll of the DNA in the expanded motif is lower in the mutant promoter compared to the WT promoter (S4 Fig). Combining the TATT->TATA mutation with the mutation of the 5 bp motif completely abolished transcription, indicating that these elements are absolutely essential for promoter activation (Fig 5E). Furthermore, with the exception of GG89TT-NS, systematic single and double base substitutions in the RVNYS motif were well tolerated, arguing that this region may not be contributing to base specific contacts with the vPIC (Fig 5F). In summary, while the expanded RVNYS motif appears important for transcriptional activation, we hypothesize it may contribute to vPIC specificity by impacting the local topology of the promoter DNA rather than through base specific contacts.

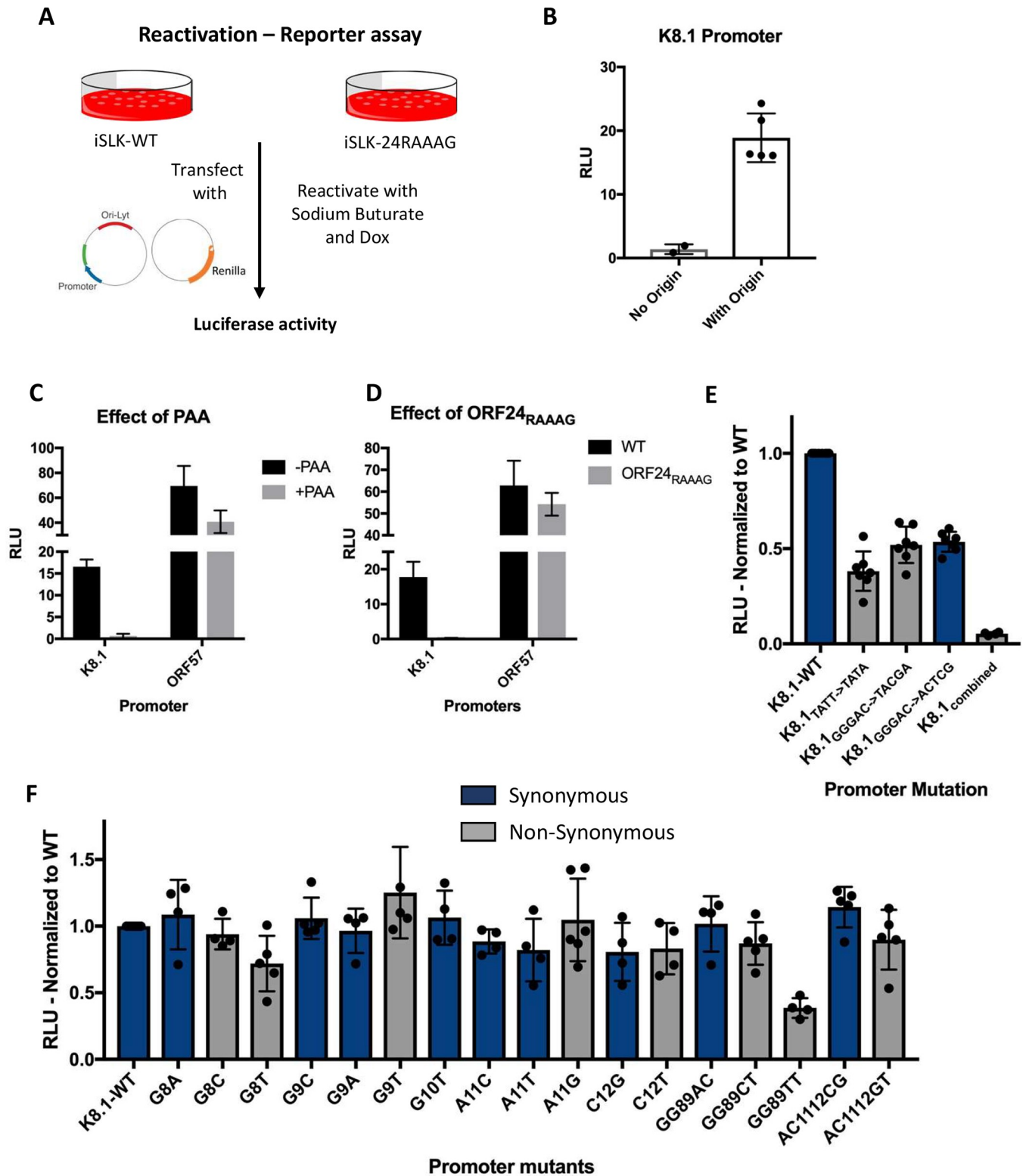
### The expanded promoter motif is important for DNA binding by vPIC

The RVNYS motif could either contribute to vPIC binding to its target promoters, as is the case for the TATTWAA sequence, or it could modulate downstream events involved in transcription initiation at vPIC bound promoters. To distinguish these possibilities, we used ChIP to measure HA-ORF34 occupancy on the luciferase plasmids containing WT or mutant K8.1 promoter in lytically reactivated HA-ORF34 iSLK cells. We engineered qPCR primers that enabled distinction of the K8.1 promoter on the plasmid DNA from the endogenous viral locus (Fig 6A).

Notably, both the K8.1<sub>TATT->TATA</sub> and K8.1<sub>GGGAC->ACTCG</sub> mutations similarly impaired HA-ORF34 binding to the plasmid promoter DNA (Fig 6B- left panel). The ChIP signal was specific for HA-ORF34, as we observed only a background signal when the assay was performed in the WT iSLK cells containing an untagged ORF34 locus. Furthermore, there was comparable vPIC assembly at the genomic K8.1 locus in cells transfected with WT or the mutant plasmids, confirming similar immunoprecipitation efficiencies across samples (Fig 6B—right panel). Together, these data indicate that the expanded promoter motif is important for vPIC assembly on the promoter DNA.

### The ORF24-34 binding element in the origin is necessary for late gene activation

Given that late gene transcription requires the presence of the viral lytic origin of replication (Fig 5B), binding of ORF34 and ORF24 at the KSHV Ori was particularly notable (see Fig 3A and 3B). The binding site mapped to a region of the Ori just outside of the core minimal region

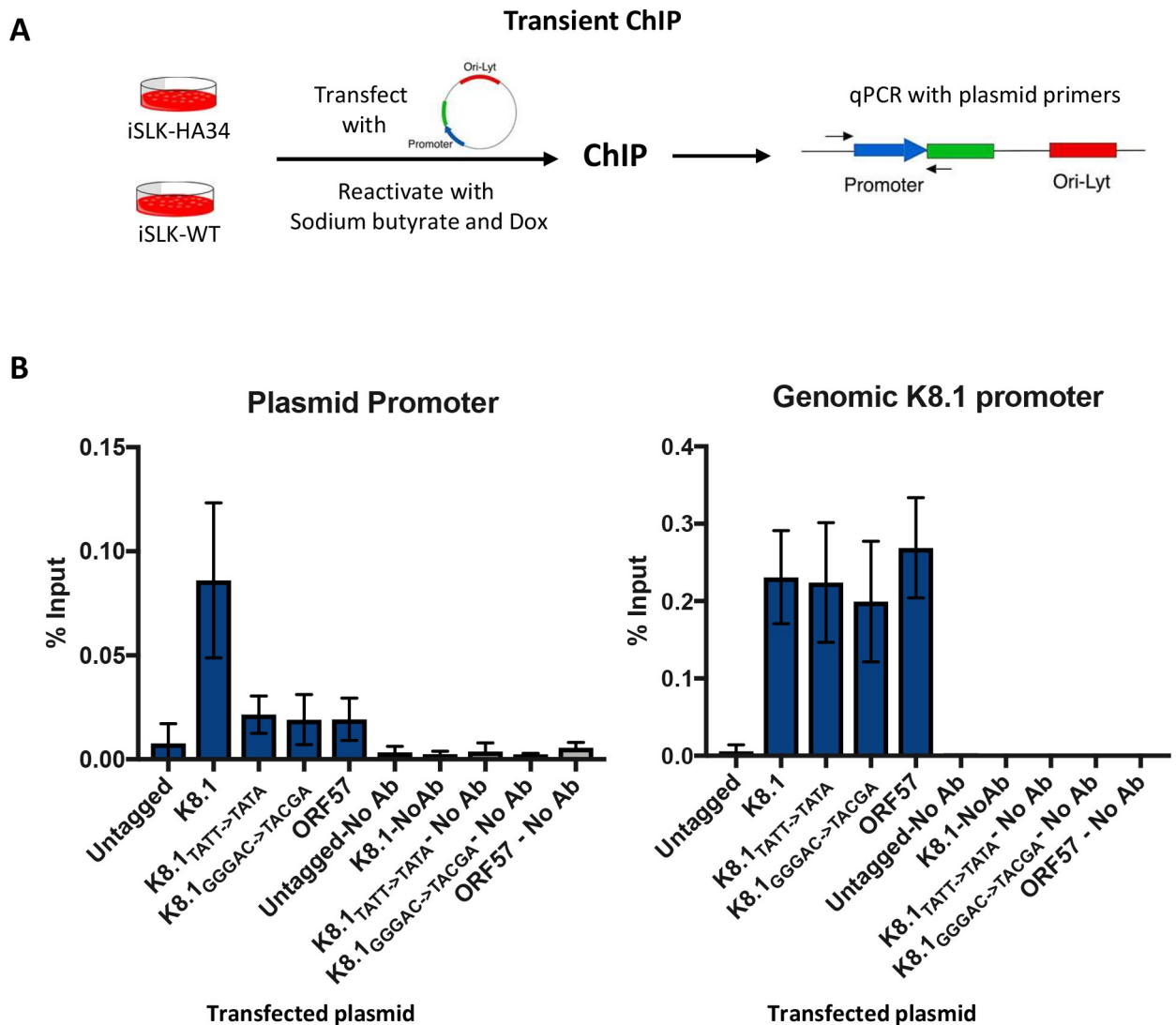


**Fig 5. The motif identified from direct targets is important for transcription from the promoter.** (A) Schematic of plasmid-based promoter activation assay. iSLK-WT or ORF24<sub>RAAAG</sub> mutant cells reactivated with doxycycline and sodium butyrate were transfected with reporter plasmid carrying the indicated promoter sequence upstream of the firefly luciferase gene and a control renilla luciferase expressing plasmid. 48 h post transfection and reactivation, the cells were assessed for luciferase activity. (B) Reactivated iSLK-WT cells were transfected with the K8.1 promoter reporter plasmid with or without the KSHV left lytic origin of replication cloned in *cis* and assessed for luciferase activity. (C) Reactivated iSLK-WT cells were transfected with the K8.1 and ORF57 promoter

reporter plasmids in the presence or absence of DNA replication inhibitor (PAA) and assessed for luciferase activity. (D) Reactivated iSLK-WT and iSLK-ORF24<sub>RAAAG</sub> mutant cells were transfected with K8.1 or ORF57 promoter reporter plasmids and assessed for luciferase activity. (E-F) iSLK-WT cells were transfected with plasmids containing the WT K8.1 promoter or the indicated single or double point mutant promoters and assessed for luciferase activity. The data for each mutant was normalized to the WT K8.1 promoter. All data shown are average of 3–5 biological replicates.

<https://doi.org/10.1371/journal.ppat.1007774.g005>

that is required for viral DNA replication (Fig 7A). To examine the role of vPIC Ori binding on K8.1 promoter activation, we engineered a mutant version of the K8.1 promoter-driven luciferase plasmid lacking the ORF24-34 binding element (BE). A deletion encompassing this region has been shown to only modestly reduce (~30%) viral DNA replication [35,36].



**Fig 6. The expanded motif is important for DNA binding by vPIC.** (A) Schematic of the transient ChIP assay. iSLK-WT or iSLK-HAORF34 cells reactivated with doxycycline and sodium butyrate were transfected with plasmid carrying the indicated promoter sequence. 48 h post transfection and reactivation, cells were subjected to ChIP with an anti-HA antibody, and the associated DNA was quantified using primers specific to the promoter on the plasmid and the genomic K8.1 promoter locus. (B) iSLK-HAORF34 cells were transfected with K8.1 promoter containing plasmid, mutant plasmids or ORF57 promoter plasmid (as a control). iSLK-WT cells without the HA tag transfected with K8.1 plasmid served as a control for the IP. The ChIP DNA was quantified using plasmid specific primers (left panel) and K8.1 promoter specific primers (right panel), normalized to the level of input DNA and presented as percent input. Data shown are an average of 3–6 biological replicates.

<https://doi.org/10.1371/journal.ppat.1007774.g006>

However, we observed a marked reduction in K8.1 promoter activation in the absence of the ORF24-34 Ori binding element (Fig 7B). The element by itself was insufficient to activate transcription in the absence of core minimal Ori, in agreement with the requirement for DNA replication to license late gene activation.

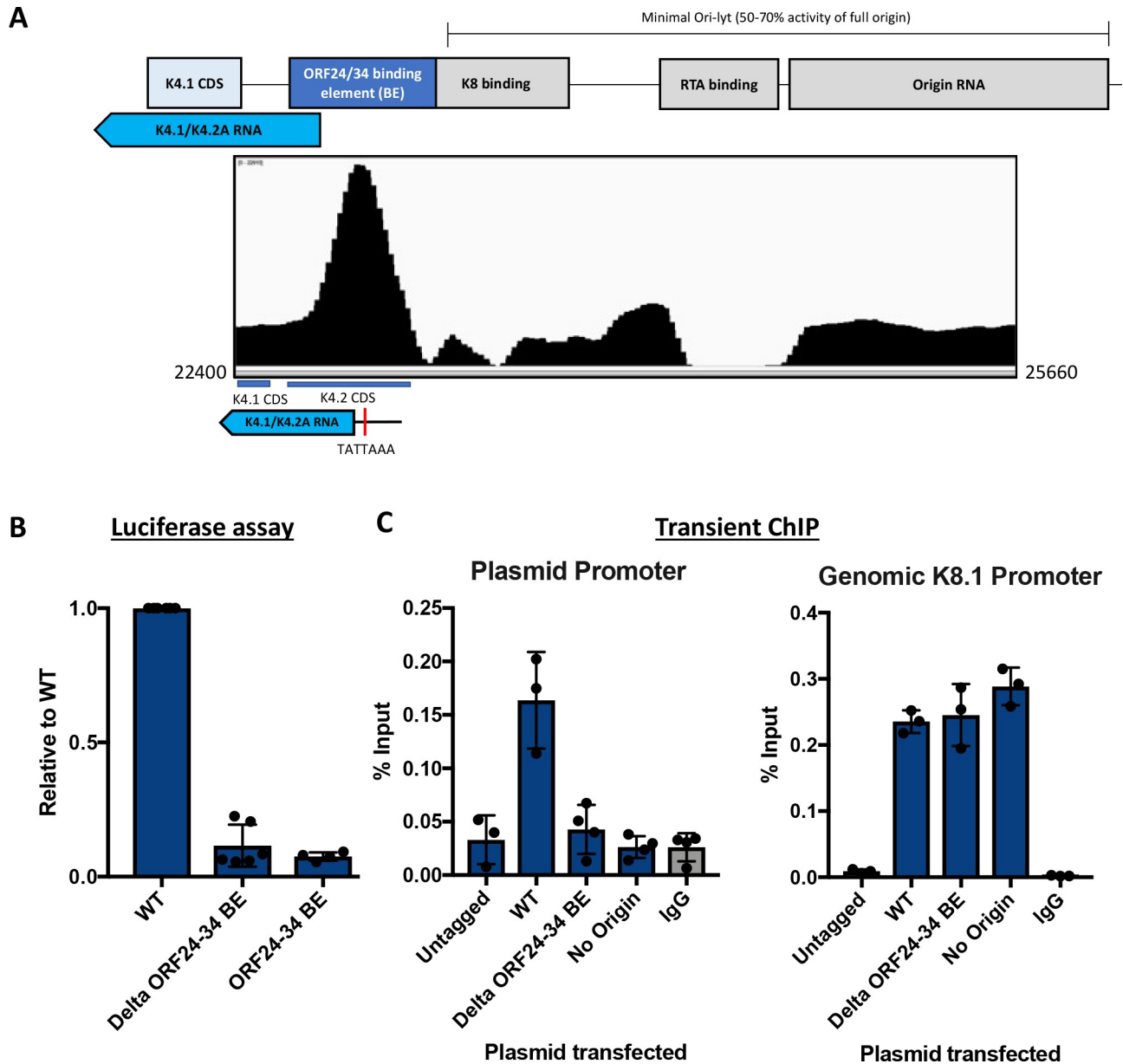
Ori binding by ORF24 and ORF34 could impact late gene transcription either by enabling vPIC promoter binding or by facilitating a downstream stage of transcription. To distinguish between these possibilities, we used our transient ChIP assay to compare HA-ORF34 binding to the K8.1 promoter on the plasmid containing either the WT Ori or the Ori lacking the ORF24-34 binding element. Removing the binding element within the Ori resulted in near complete loss of HA-ORF34 binding to the K8.1 promoter on the plasmid (Fig 7C), indicating that ORF24-34 occupancy at the lytic origin of replication is essential for vPIC assembly at the late gene promoter.

## Discussion

Late gene expression is uniquely regulated in KSHV by a set of viral proteins collectively known as the vPIC. In this study, we generated a comprehensive list of direct targets of the vPIC by combining transcriptome and genome-wide binding data of the complex. Among the direct targets, we identified an expanded promoter recognition motif that contributes to stable binding of the vPIC on the promoter. This is particularly useful to understanding alternate mechanisms of transcription initiation by RNA pol II because of the unique ability of the vPIC to directly recruit RNA Pol II to highly compact promoters. We also identified a vPIC binding element in the lytic origin of replication that licenses late gene activation at a distal promoter, suggesting a novel mechanism may underlie the link between DNA replication and late gene transcription.

Several key points emerge from independent analysis of the RNA Seq and ChIP Seq data. First, the previously identified TATTWAA motif is insufficient to mark binding and regulation by the vPIC, as demonstrated by the multiple ORFs that have the motif in their promoter (23/83 ORFs) but are not repressed in the ORF24<sub>RAAAG</sub> mutant (10/23 ORFs) or bound by ORF34 (7/23 ORFs). Second, the TATTWAA motif is not a pre-requisite for regulation or binding by the vPIC. For example, although ORF53 expression is dependent on the ORF24-Pol II interaction and its promoter is bound by ORF34, its promoter does not have the conserved TATTWAA sequence but a TATATAA sequence in its place. Third, several genes that exhibit late kinetics escape repression in the mutant, indicating that not all late genes are direct or indirect targets of vPIC. It is notable that several of the kinetic late genes that escape repression are involved in immune regulation, consistent with a similar observation in EBV [28]. In particular, EBV BPLF1 is a late gene whose expression is independent of the vPIC, and we observe that its homolog ORF64 also shows a similar pattern in KSHV. Finally, while there are several early genes that are indirect targets of the vPIC, only 1 of 13 direct targets exhibited early kinetics (ORF45). It is notable that ORF45 is a tegument protein, which might explain the requirement for its continued regulation at later times of infection.

Our integrative analysis revealed that a distinct 5 bp motif is integral to vPIC promoter recognition. Rather than engaging in base specific contacts (as is the case for the TATTWAA sequence) [7], we instead hypothesize that the importance of this motif lies in its shape or structure. In this regard, a preliminary analysis of basic structural properties of the expanded motif suggest that the average roll (representing the rotational flexibility) of the DNA in the expanded motif is lower in the 5 bp mutant promoter compared to WT K8.1 promoter. Structure-based recognition and a contribution of base pairs flanking the core binding site to binding affinity is a common feature among transcription factors, although the specific structural



**Fig 7. The ORF24-34 binding element at the origin is essential for late gene activation.** (A) (Top panel) Schematic of the left lytic origin of replication of the KSHV genome. (Bottom panel) IGV image of the ChIP peak for HA-ORF34 at the origin of replication. (B) iSLK-WT cells were transfected with the K8.1 promoter plasmid containing either the full length Ori (WT), the Ori lacking the ORF24-34 binding element (Delta ORF24-34 BE), or the Ori lacking the minimal Ori-lyt and retaining only the ORF24-34 binding element (ORF24-34 BE). 48 h post transfection and reactivation, the cells were assessed for luciferase activity. (C) iSLK-HAORF34 cells were transfected with a K8.1 promoter containing plasmid containing either the full-length Ori (WT), the Ori lacking the ORF24-34 binding element (Delta ORF24-34 BE), or no Ori. Untagged iSLK-WT cells transfected with K8.1 plasmid served as a control for the IP. The ChIP DNA was quantified using plasmid specific primers (left panel) and genomic K8.1 promoter specific primers (right panel), normalized to the level of input DNA and presented as percent input. Data shown are an average of 3–4 biological replicates.

<https://doi.org/10.1371/journal.ppat.1007774.g007>

properties that influence binding differ for various classes of factors [37]. A previous study of the bZIP family of transcription factors identified a correlation between the roll of DNA sequences that flank the core recognition element and binding affinity [37]. Consistent with that, we observed a defect in the assembly of the vPIC when the expanded motif was mutated, suggesting a stabilizing role for the flanking sequence.



The location of the expanded motif is reminiscent of the downstream recognition element of the general transcription factor TFIIB (BRE<sup>d</sup>) [38]. Interestingly, the BRE<sup>d</sup> element also shows low sequence conservation and does not require an active promoter sequence to conform with the consensus sequence at all positions [38], as seen with the expanded motif. Interaction of TFIIB with the BRE<sup>d</sup> element is thought to stabilize the TFIIB-TBP interaction on the promoter DNA, leading to activation or repression of the promoter in a context-dependent manner. While there is some evidence to suggest that TFIIB is present at late gene promoters [7], whether the interaction between the vPIC and the expanded motif is mediated by TFIIB or any other factors remains to be elucidated.

Notably, we found that KSHV DNA replication was reduced in the absence of functional ORF24, including with an ORF24 null mutant or ORF24 point mutants that fail to bind either RNA Pol II or the vPIC component ORF34. This was surprising, given that deletion or mutation of several other vPIC components has been shown not to affect DNA replication [7,13,17,18]. Furthermore, we previously failed to observe a replication defect in an ORF24 null virus [7]. We hypothesize that the discrepancy for the ORF24 null mutant relates to the methods used to establish the infected cell lines. In our earlier study, the viral BAC DNA was directly transfected into iSLK cells, resulting in only a ~5-7-fold increase in genome copy number upon reactivation. It is likely that the majority of viral DNA was present in a non-activatable state, providing a level of background that significantly reduced the dynamic range for DNA replication measurements. Here, the iSLK cells were established through co-culture with reactivated 293T cells harboring the BAC DNA, i.e. through infection rather than transfection. This reproducibly yields ~50-70-fold increases in genome copy number upon lytic reactivation, more closely representing actual infection. That said, the role of ORF24 in viral DNA replication may be unique to KSHV, as this link has not been observed upon depletion of its homologs in EBV and MHV68 [28,39].

One possibility for how KSHV ORF24 might contribute to DNA replication is through its indirect transcriptional targets. We observed that components of the helicase/primase complex (ORFs 40, 41), which are involved in DNA replication, show more than average repression in the absence of functional ORF24 (Fig 2). Importantly, these genes were not repressed in previous studies using stop mutants of the vPIC components ORF18, ORF30 and ORF31 [17,18], raising the possibility that ORF24 has additional gene regulatory functions outside the context of the vPIC. For example, our observed binding of ORF24 and ORF34 at the lytic origins of replication could facilitate DNA replication. Consistent with this hypothesis, two previous studies that defined the boundaries of the lytic origin of replication observed a modest decrease in replicated DNA upon deletion of the ORF24-34 binding site identified here [35,36].

Interestingly, we also observed that binding of ORF24-34 at the origin is critical for origin dependent late gene expression, as deletion of the ORF24-34 binding element at the origin prevented vPIC assembly and transcription at the distal K8.1 late promoter. We note that although the binding element overlaps the promoter of K4.1/K4.2A, we observe only average or less than average reduction in these transcript levels in the mutant compared to the mutant rescue. We hypothesize that alternate mechanisms of maintaining steady-state K4.1/K4.2A RNA levels are adopted in the absence of a functional vPIC and binding of ORF24-34 at the origin is likely related to the link between DNA replication and late gene activation. The proximity of the ORF24-34 binding element to the other origin elements also raises the possibility of vPIC recruitment by other origin binding proteins. Exploring the mechanism underlying this link between origin binding and late gene activation will be an exciting area for future study.

## Methods

### Cell lines

HEK293T cells (ATCC CRL-3216) were maintained in DMEM (Invitrogen) + 10% Fetal Bovine Serum (FBS). iSLK BAC16 cells [29] (kindly provided by Jae Jung's lab) were maintained in DMEM + 10% FBS and 0.5 mg/mL hygromycin B. HEK293T cells stably expressing ORF24-3xFlag were maintained in DMEM (Invitrogen) + 10% Fetal Bovine Serum (FBS) and 325 ug/ml of zeocin. iSLK-puro cells were maintained in DMEM + 10% FBS.

### Plasmids

A 100 bp fragment of the K8.1 promoter or ORF57 promoter was generated by PCR from the BAC16 genome and cloned into the KpnI and HindIII sites of the pGL4.16 reporter plasmid using Infusion cloning (Clontech). The left origin of replication was similarly generated by PCR from the BAC16 genome and cloned into the NotI and BstXI sites of the K8.1Pr-pGL4.16 plasmid. The promoter mutants were generated by site-directed mutagenesis using KAPA HiFi DNA polymerase using the primers shown in [S1 Table](#). To make the ORF24-3xFlag pJLM1 plasmid for lentiviral transduction, the ORF24-3xFlag fragment was generated by PCR using ORF24-3xFlag-pCDNA4 as the template and cloned in to pJLM1 plasmid digested with AgeI and EcoRI. All primers used for cloning and mutagenesis are listed in [S1 Table](#).

### Generation of BAC mutants and establishing latent cell lines

The ORF24<sub>RAAAG</sub>, ORF24<sub>R328A</sub> and ORF24<sub>stop</sub> mutants, the corresponding mutant rescues (MR) for ORF24<sub>RAAAG</sub>, ORF24<sub>R328A</sub> and N terminal HA tagged ORF24 and ORF34 were engineered using the scarless Red recombination system in BAC16 GS1783 *E. coli* [40] as described previously. The modified BACs were purified using the NucleoBond BAC 100 kit (Clontech).

To establish iSLK cell lines latently infected with the modified virus, HEK293T cells were transfected with 5–10 ug of the modified BAC16 using linear polyethylenimine (PEI, MW ~25,000) at a 1:3 DNA:PEI ratio. The following day, the 293T cells were mixed 1:1 with iSLK-puro cells (which contain inducible RTA but lack KSHV) and treated with 30 nM 12-O-Tetradecanoylphorbol-13-acetate (TPA) and 300 nM sodium butyrate for 4 days to induce lytic replication. Cells were then grown in selection media containing 300 ug/ml hygromycin B, 1 ug/ml puromycin and 250 ug/ml G418. The hygromycin concentration was gradually increased to 500 ug/ml and then 1 mg/ml until all the HEK293T cells died. To generate the ORF24<sub>RAAAG</sub>, ORF24<sub>R328A</sub>, and ORF24<sub>stop</sub> (which are defective in infectious virion production), HEK293T cells that stably express ORF24-3xFlag were used for the initial transfection and co-culture.

### Supernatant transfer assay

For supernatant transfer assays to assess virion production, BAC16-containing iSLK cells were plated at a density of  $1 \times 10^6$  cells in 10 cm dishes and treated with 1 ug/mL doxycycline and 1 mM sodium butyrate the following day. 72 h post induction, the supernatant was filtered through a 0.45 uM syringe filter and 2 ml of the supernatant was mixed with  $1 \times 10^6$  freshly trypsinized HEK293T cells in a 6 well plate and centrifuged at 1,500xg for 2 h at 37°C. The following day, cells were trypsinized, fixed in 4% paraformaldehyde and the percentage of cells expressing GFP was determined by flow cytometry (BD Csampller, BD Biosciences).

## Viral genome replication

To measure viral genome replication, iSLK-BAC16 cells were reactivated for 72 h as described above. The cells were then scraped into the media and the media + cells were digested with proteinase K (80 µg/mL) (Promega) in 5x proteinase K digestion buffer (50mM Tris-HCl pH 7.4, 500mM NaCl, 5mM EDTA, 2.5% SDS) overnight at 55 °C. The gDNA was isolated using Zymo Quick gDNA Miniprep Kit according to the manufacturer's instructions. Quantitative PCR (qPCR) was performed on the isolated DNA using iTaq Universal SYBR Green Supermix on a QuantStudio3 Real-Time PCR machine. DNA levels were quantified using relative standard curves with primers specific for KSHV ORF59 promoter and human CPSF6 promoter (S1 Table). The relative genome numbers were normalized to CPSF6 to account for loading differences and to uninduced samples to account for differences in starting genome copy number.

## RT-qPCR

To look at viral transcript levels, total RNA was isolated from reactivated iSLK cells at the indicated time points using the Zymo Direct-Zol RNA kit following the manufacturer's protocol. The purified RNA was treated with DNase and cDNAs were synthesized using AMV reverse transcriptase (Promega). The cDNAs were directly used for qPCR analysis using the iTaq Universal SYBR Green Supermix. The qPCR signal for each ORF was normalized to 18s rRNA.

## Western blotting

Cells were lysed in 1x lysis buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 0.5% NP-40, cOmplete EDTA-free Protease Inhibitors [Roche]) by rotating for 30 min at 4 °C. The lysate was clarified by centrifugation at 21,000 x g for 15 min at 4 °C. The supernatant was quantified by Bradford assay, and equivalent amounts of each sample were resolved by SDS-PAGE and western blotted with rabbit polyclonal antibodies against vinculin (Abcam, 1:1000), ORF59, or K8.1. The ORF59 and K8.1 antibodies were generated by injecting rabbits with purified MBP-ORF59 or MBP-K8.1 (gifts from Denise Whitby [41]) (Pocono Rabbit Farm and Laboratory).

## RNA Seq sample preparation and analysis

Total RNA was isolated from reactivated iSLK cells at indicated time points using Zymo Direct-Zol RNA kit following the manufacturer's protocol. The samples were ribo-depleted and stranded libraries were prepped by the QB3 Berkeley core sequencing facility. Multiplex sequencing was performed using the Illumina HiSeq 4000 to generate 100 bp paired end reads. The quality of the raw reads was checked using FastQC [42] and adapters were trimmed using TrimGalore [43] using default parameters. The trimmed reads were aligned to both the viral (NC\_009333.1) and host genomes (hg19) using STAR aligner (ver2.5.3a) [44]. The aligned reads were counted using htseq-counts [45] and differential expression was assessed using DESeq2 [46].

To compensate for the problem of quantitative assessment of transcripts due to the overlapping transcripts in the KSHV genome, the primary counts of the transcripts were calculated as recommended in Bruce et al [30] with the following small modifications. Length correction was not applied to allow for differential expression analysis by DESeq2. Transcripts for which the counts became negative in all 3 replicates when calculating primary counts were removed before running DESeq2. Bar plots showing Log<sub>2</sub> Fold Change were generated using R. The heatmap was generated using the heatmap.2 package in R. The fold change values were scaled

by column before plotting to ease visualization of the data. Raw and processed data files are available from the GEO database (accession number: GSE126602).

### Chromatin immunoprecipitation (ChIP)

ChIP was performed on 3\*15-cm plates of iSLK-cells (WT KSHV, KSHV<sub>HA-ORF34</sub> and KSHV<sub>ORF24-HA</sub>) reactivated for 48 h with 1 ug/ml dox and 1 mM sodium butyrate. Cells were crosslinked in 2% formaldehyde for 15 min at room temperature, quenched in 0.125 M glycine, and washed twice with ice-cold PBS. Crosslinked cell pellets were mixed with 1 ml ice-cold ChIP lysis buffer (5mM PIPES pH 8.0, 85mM KCl, 0.5% NP-40) and incubated on ice for 10 min, whereupon the lysate was dounce homogenized to release nuclei and spun at 1700 g for 5 min at 4°C. Nuclei were then resuspended in 800 µl of nuclei lysis buffer (50mM Tris-HCl pH 8.0, 0.3% SDS, 10mM EDTA) and rotated for 10 min at 4°C followed by sonication using a QSonica Ultrasonicator with a cup horn set to 25 amps for 20 min total (5 min on, 5 min off). Chromatin was spun at 16,000 x g for 10 min at 4°C and the pellet was discarded. The chromatin was precleared with protein A + protein G beads blocked with 5 ug/ml glycogen, 50 ug/ml BSA, 100 ug/ml Ecoli tRNA in 1ml dilution buffer (1.1% Triton-X-100, 1.2mM EDTA, 16.7mM Tris-HCl pH 8, 167mM NaCl) for 2 hours.

350 µl of chromatin was diluted in ChIP dilution buffer (16.7 mM Tris-HCl pH 8.0, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl) to 500 µl and incubated with 10 µg anti-HA antibody (Cell Signaling C29F4) overnight, whereupon samples were rotated with 20 µl pre-blocked protein A + G beads (ThermoFisher) for 2 h at 4°C. Beads were washed with low salt immune complex (20 mM Tris pH 8.0, 1% Triton-x-100, 2 mM EDTA, 150 mM NaCl, 0.1% SDS), high salt immune complex (20 mM Tris pH 8.0, 1% Triton-x-100, 2 mM EDTA, 500 mM NaCl, 0.1% SDS), lithium chloride immune complex (10mM Tris pH 8.0, 0.25 M LiCl, 1% NP-40, 1% Deoxycholic acid, 1 mM EDTA), and Tris-EDTA for 10 min each at 4°C with rotation. DNA was eluted from the beads using 100 µl of elution buffer (150 mM NaCl, 50 µg/ml Proteinase K) and incubated at 50°C for 2 h, then 65°C overnight. DNA was purified using the Zymo PCR purification kit. For ChIP-qPCR, purified DNA was quantified by qPCR using iTaq Universal SYBR Mastermix (BioRad) and the indicated primers (S1 Table). Each sample was normalized to its own input.

For ChIP-Seq, the concentration of the eluted DNA was measured using Qubit. 10–15 ng of eluted DNA was used to generate libraries using the Accel-NGS 2S Plus DNA Library Kit (Swift Biosciences). The libraries were sequenced on a Illumina HiSeq 4000 sequencer at the QB3 Berkeley core facility to generate 100 bp single end reads.

### ChIP Seq analysis

The quality of the raw reads was checked using FastQC [42] and adapters were trimmed using TrimGalore [43] using default parameters. The trimmed reads were aligned to both the viral and host genomes (hg19) using Bowtie2 aligner using default parameters. The aligned reads were filtered for uniquely mapped reads using SAMTools. The filtered reads were converted to tdf format using IGV Tools and visualized on Integrated Genome Viewer [47,48]. We used HOMER [33] with the following minor adjustments to default parameters to call peaks for the immunoprecipitated and input samples from cells that have HA tagged ORF34 and untagged ORF34 (WT) as a control. The fold over input and local fold change were set to 1.3 as recommended for smaller and denser viral genomes (personal communication with Chris Benner, UCSD). The peak size and fragment length were also set at 200 and 150 respectively. From the resulting peaks, only the ones that were present in both conditions (HA-ORF34 vs HA-Input, HA-ORF34 vs WT) and both replicates were retained for further analysis. The common peaks

identified were verified visually on the Integrated Genome Viewer and were manually annotated.

To identify the direct targets, transcripts that had a  $\text{Log}_2$  fold change lower than the mean  $\text{Log}_2$  fold change in the ORF24<sub>RAAAG</sub> mutant vs MR were considered to be down regulated in the mutant. Although ORF8 and ORF26 were not included in the differential expression analysis, they were considered to be downregulated because of negative values for their levels in the mutant and not in the mutant rescue on calculation of primary transcript levels. Raw and processed data files are available from the GEO database (accession number: GSE126601).

### Promoter analyses

We defined the promoter sequence as 100 base pairs (bp) upstream from the transcription start site (TSS) or 300 bp upstream from the translation start site when the TSS was unknown. The TSS data was obtained from Table 1 in [23] and a custom python script was used to extract the sequences. These were then run through MEME [31] allowing for multiple occurrences of a motif in a sequence and only searching the given strand to identify enriched motifs. The identified motif was then used to search for similar motifs across all 83 promoter sequences from the viral genome, again only allowing for searching the given strand.

### Reporter assays

$5 \times 10^5$  iSLK-BAC16 (WT or ORF24<sub>RAAAG</sub>) cells were plated in a 6 well plate 24 h before transfection and reactivation. The cells were reactivated with 1  $\mu\text{g}/\text{ml}$  doxycycline and 1 mM Sodium Butyrate and simultaneously transfected with 0.95  $\mu\text{g}$  luciferase reporter plasmid and 0.05  $\mu\text{g}$  renilla plasmid using polyjet transfection reagent. 48 h post transfection and reactivation, the cells were rinsed twice with room temperature PBS, lysed by rocking for 15 min at room temperature in 300  $\mu\text{l}$  of Passive Lysis Buffer (Promega), and clarified by centrifuging at 14,000 g for 1 min. 20  $\mu\text{l}$  of the clarified lysate was added in triplicate to a white chimney well microplate (Greiner bio-one) to measure luminescence on a Tecan M1000 using a Dual Luciferase Assay Kit (Promega). The firefly luminescence was normalized to the internal Renilla luciferase control. The induced sample was then normalized to the uninduced (but transfected) control to look at the fold induction during reactivation.

### Transient ChIP-qPCR

$3.5 \times 10^6$  iSLK-BAC16 were plated in a 15 cm dish 24 h before transfection and reactivation. The cells were reactivated with 1  $\mu\text{g}/\text{ml}$  doxycycline and 1 mM Sodium Butyrate and simultaneously transfected with 8  $\mu\text{g}$  of reporter plasmid using polyjet transfection reagent. 48 h post reactivation and transfection, cells were crosslinked and lysed as mentioned above. Nuclei were sonicated using a Covaris S220 (Covaris) for 5 min with the recommended sonication settings (Power: 140 W, Bursts per cycle: 200, Duty cycle: 5%). The chromatin was pre-cleared as mentioned above and the concentration measured using a qubit. 25  $\mu\text{g}$  of pre-cleared chromatin was used in the IP and the following steps were same as mentioned in the ChIP protocol above.

### Supporting information

**S1 Fig. Validation of mutant BAC DNA integrity.** The integrity of the various BAC mutants used in this study was verified by *RsrII* digestion.

(TIF)

**S2 Fig. Kinetics of early and late gene transcripts in ORF24<sub>RAAAG</sub> mutant and MR.** iSLK--ORF24<sub>RAAAG</sub> and MR were reactivated with doxycycline and sodium butyrate for the indicated time points, whereupon total RNA was extracted, converted to cDNA, and quantified by qPCR using primers specific to the CDS of K8.1 (late gene) or ORF57 (early gene). The fold induction was measured relative to uninduced sample.

(TIF)

**S3 Fig. ChIP-Seq data with HA-ORF34.** (A) Western blot showing expression of HA-ORF24 and HA-ORF34. iSLK cells harboring WT BAC16, HA-ORF34 BAC or HA-ORF24 BAC were reactivated with doxycycline and sodium butyrate for 72 hours. Protein lysate was immunoprecipitated with HA beads and visualized by SDS-PAGE-Western blot. (B) iSLK cells harboring WT BAC16, HA-ORF34 BAC or HA-ORF24 BAC were reactivated with doxycycline and sodium butyrate for 72 hours. Progeny virion production was measured by supernatant transfer on to 293T cells and quantified by flow cytometry. (B) IGV image showing ChIP-Seq data from iSLK-HA-ORF34 cells reactivated with doxycycline and sodium butyrate for 48 h and immunoprecipitated with anti-HA antibody. Also shown are plots for control data from untagged iSLK cells and input DNA for two biological replicates.

(TIF)

**S4 Fig. Effect of mutations on DNA structural parameters.** Four structural parameters for 40 bp of the WT and expanded motif mutants were calculated using the DNAShape program in R [49]. Average of each structural parameter for the mutated base pairs and the two adjacent base pairs at either end (a total of 7 bp) was calculated for each DNA sequence analyzed. The two adjacent base pairs were included in the averaging as some of the structural parameters refer to a base step rather than a specific base. The average for each mutant was normalized to the WT K8.1 promoter.

(TIF)

**S1 Table. List of all DNA sequences used in the study.**

(PDF)

**S1 File. List of peaks identified by HOMER from the HA-ORF34 ChIP-Seq data.**

(XLSX)

## Acknowledgments

We thank all members of the Glaunsinger and Coscoy labs for their helpful suggestions. In particular, we thank Angelica Casteñeda for providing the HEK293T cells stably expressing ORF24-3xFlag and Chris Benner (UCSD) for suggestions on proper usage of HOMER for organisms with small genomes.

## Author Contributions

**Conceptualization:** Divya Nandakumar, Britt Glaunsinger.

**Data curation:** Divya Nandakumar.

**Formal analysis:** Divya Nandakumar, Britt Glaunsinger.

**Investigation:** Divya Nandakumar.

**Supervision:** Britt Glaunsinger.

**Writing – original draft:** Divya Nandakumar.

**Writing – review & editing:** Britt Glaunsinger.

## References

1. Payne S. Chapter 28—Introduction to DNA Viruses. In: Payne S, editor. *Viruses* [Internet]. Academic Press; 2017. p. 231–6. Available from: <http://www.sciencedirect.com/science/article/pii/B9780128031094000283>
2. Grula MA, Buller PL, Weaver RF.  $\alpha$ -Amanitin-Resistant Viral RNA Synthesis in Nuclei Isolated from Nuclear Polyhedrosis Virus-Infected *Heliothis zea* Larvae and *Spodoptera frugiperda* Cells. *J Virol* [Internet]. 1981 Jun 1; 38(3):916. Available from: <http://jvi.asm.org/content/38/3/916.abstract> PMID: 16789208
3. Kochetkov SN, Rusakova EE, Tunitskaya VL. Recent studies of T7 RNA polymerase mechanism. *FEBS Lett* [Internet]. 1998; 440(3):264–7. Available from: <http://www.sciencedirect.com/science/article/pii/S0014579398014847> PMID: 9872383
4. Rong M, He B, McAllister WT, Durbin RK. Promoter specificity determinants of T7 RNA polymerase. *Proc Natl Acad Sci U S A* [Internet]. The National Academy of Sciences; 1998 Jan 20; 95(2):515–9. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/9435223> <https://doi.org/10.1073/pnas.95.2.515> PMID: 9435223
5. Young CSH. The Structure and Function of the Adenovirus Major Late Promoter. In: Doerfler W, Böhm P, editors. *Adenoviruses: Model and Vectors in Virus-Host Interactions: Virion-Structure, Viral Replication and Host-Cell Interactions* [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2003. p. 213–49. Available from: [https://doi.org/10.1007/978-3-662-05597-7\\_8](https://doi.org/10.1007/978-3-662-05597-7_8)
6. Gruffat H, Marchione R, Manet E. Herpesvirus Late Gene Expression: A Viral-Specific Pre-initiation Complex Is Key. *Front Microbiol* [Internet]. Frontiers Media S.A.; 2016 Jun 6; 7:869. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/27375590> <https://doi.org/10.3389/fmicb.2016.00869> PMID: 27375590
7. Davis ZH, Verschueren E, Jang GM, Kleffman K, Johnson JR, Park J, et al. Global Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for Late Genes. *Mol Cell* [Internet]. 2015; 57(2):349–60. Available from: <http://www.sciencedirect.com/science/article/pii/S1097276514009216> <https://doi.org/10.1016/j.molcel.2014.11.026> PMID: 25544563
8. Wong-Ho E, Wu T-T, Davis ZH, Zhang B, Huang J, Gong H, et al. Unconventional Sequence Requirement for Viral Late Gene Core Promoters of Murine Gammaherpesvirus 68. Frueh K, editor. *J Virol* [Internet]. 2014 Mar 15; 88(6):3411. Available from: <http://jvi.asm.org/content/88/6/3411.abstract> <https://doi.org/10.1128/JVI.01374-13> PMID: 24403583
9. Serio TR, Cahill N, Prout ME, Miller G. A Functionally Distinct TATA Box Required for Late Progression through the Epstein-Barr Virus Life Cycle. *J Virol* [Internet]. 1998 Oct 1; 72(10):8338. Available from: <http://jvi.asm.org/content/72/10/8338.abstract> PMID: 9733880
10. Tang S, Yamanegi K, Zheng Z-M. Requirement of a 12-Base-Pair TATT-Containing Sequence and Viral Lytic DNA Replication in Activation of the Kaposi's Sarcoma-Associated Herpesvirus K8.1 Late Promoter. *J Virol* [Internet]. 2004 Mar 1; 78(5):2609. Available from: <http://jvi.asm.org/content/78/5/2609.abstract> <https://doi.org/10.1128/JVI.78.5.2609-2614.2004> PMID: 14963167
11. Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol* [Internet]. 2018; 19(10):621–37. Available from: <https://doi.org/10.1038/s41580-018-0028-8> PMID: 29946135
12. Aubry V, Mure F, Mariamé B, Deschamps T, Wyrwicz LS, Manet E, et al. Epstein-Barr Virus Late Gene Transcription Depends on the Assembly of a Virus-Specific Preinitiation Complex. Longnecker RM, editor. *J Virol* [Internet]. 2014 Nov 1; 88(21):12825. Available from: <http://jvi.asm.org/content/88/21/12825.abstract> <https://doi.org/10.1128/JVI.02139-14> PMID: 25165108
13. Davis ZH, Hesser CR, Park J, Glaunsinger BA. Interaction between ORF24 and ORF34 in the Kaposi's Sarcoma-Associated Herpesvirus Late Gene Transcription Factor Complex Is Essential for Viral Late Gene Expression. Sandri-Goldin RM, editor. *J Virol* [Internet]. 2016 Jan 1; 90(1):599. Available from: <http://jvi.asm.org/content/90/1/599.abstract> <https://doi.org/10.1128/JVI.02157-15> PMID: 26468530
14. Wyrwicz LS, Rychlewski L. Identification of Herpes TATT-binding protein. *Antiviral Res* [Internet]. 2007; 75(2):167–72. Available from: <http://www.sciencedirect.com/science/article/pii/S0166354207002367> <https://doi.org/10.1016/j.antiviral.2007.03.002> PMID: 17400302
15. Gruffat H, Kadjouf F, Mariamé B, Manet E. The Epstein-Barr Virus BcRF1 Gene Product Is a TBP-Like Protein with an Essential Role in Late Gene Expression. *J Virol* [Internet]. 2012 Jun 1; 86(11):6023. Available from: <http://jvi.asm.org/content/86/11/6023.abstract> <https://doi.org/10.1128/JVI.00159-12> PMID: 22457524

16. Nishimura M, Watanabe T, Yagi S, Yamanaka T, Fujimuro M. Kaposi's sarcoma-associated herpesvirus ORF34 is essential for late gene expression and virus production. *Sci Rep* [Internet]. 2017; 7(1):329. Available from: <https://doi.org/10.1038/s41598-017-00401-7> PMID: 28336944
17. Gong D, Wu NC, Xie Y, Feng J, Tong L, Brulois KF, et al. Kaposi's Sarcoma-Associated Herpesvirus ORF18 and ORF30 Are Essential for Late Gene Expression during Lytic Replication. Frueh K, editor. *J Virol* [Internet]. 2014 Oct 1; 88(19):11369. Available from: <http://jvi.asm.org/content/88/19/11369.abstract> <https://doi.org/10.1128/JVI.00793-14> PMID: 25056896
18. Brulois K, Wong L-Y, Lee H-R, Sivadas P, Ensser A, Feng P, et al. Association of Kaposi's Sarcoma-Associated Herpesvirus ORF31 with ORF34 and ORF24 Is Critical for Late Gene Expression. Longnecker RM, editor. *J Virol* [Internet]. 2015 Jun 1; 89(11):6148. Available from: <http://jvi.asm.org/content/89/11/6148.abstract> <https://doi.org/10.1128/JVI.00272-15> PMID: 25810551
19. Perng Y-C, Campbell JA, Lenschow DJ, Yu D. Human Cytomegalovirus pUL79 Is an Elongation Factor of RNA Polymerase II for Viral Gene Transcription. *PLOS Pathog* [Internet]. Public Library of Science; 2014 Aug 28; 10(8):e1004350-. Available from: <https://doi.org/10.1371/journal.ppat.1004350>
20. Castañeda AF, Glaunsinger BA. The Interaction between ORF18 and ORF30 Is Required for Late Gene Expression in Kaposi's Sarcoma-Associated Herpesvirus. Sandri-Goldin RM, editor. *J Virol* [Internet]. 2019 Jan 1; 93(1):e01488-18. Available from: <http://jvi.asm.org/content/93/1/e01488-18.abstract> <https://doi.org/10.1128/JVI.01488-18> PMID: 30305361
21. Pan D, Han T, Tang S, Xu W, Bao Q, Sun Y, et al. Murine Cytomegalovirus Protein pM91 Interacts with pM79 and Is Critical for Viral Late Gene Expression. Sandri-Goldin RM, editor. *J Virol* [Internet]. 2018 Sep 15; 92(18):e00675-18. Available from: <http://jvi.asm.org/content/92/18/e00675-18.abstract> <https://doi.org/10.1128/JVI.00675-18> PMID: 29997217
22. Djavadian R, Hayes M, Johannsen E. CAGE-seq analysis of Epstein-Barr virus lytic gene transcription: 3 kinetic classes from 2 mechanisms. *PLOS Pathog* [Internet]. Public Library of Science; 2018 Jun 4; 14(6):e1007114-. Available from: <https://doi.org/10.1371/journal.ppat.1007114> PMID: 29864140
23. Arias C, Weisburd B, Stern-Ginossar N, Mercier A, Madrid AS, Bellare P, 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* [Internet]. Public Library of Science; 2014 Jan 16; 10(1):e1003847-. Available from: <https://doi.org/10.1371/journal.ppat.1003847> PMID: 24453964
24. 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* [Internet]. 2003 Apr 1; 77(7):4205. Available from: <http://jvi.asm.org/content/77/7/4205.abstract> <https://doi.org/10.1128/JVI.77.7.4205-4220.2003> PMID: 12634378
25. Jenner RG, Albà MM, Boshoff C, Kellam P. Kaposi's Sarcoma-Associated Herpesvirus Latent and Lytic Gene Expression as Revealed by DNA Arrays. *J Virol* [Internet]. 2001 Jan 15; 75(2):891. Available from: <http://jvi.asm.org/content/75/2/891.abstract> <https://doi.org/10.1128/JVI.75.2.891-902.2001> PMID: 11134302
26. Paulose-Murphy M, Ha N-K, Xiang C, Chen Y, Gillim L, Yarchoan R, et al. Transcription Program of Human Herpesvirus 8 (Kaposi's Sarcoma-Associated Herpesvirus). *J Virol* [Internet]. 2001 May 15; 75(10):4843. Available from: <http://jvi.asm.org/content/75/10/4843.abstract> <https://doi.org/10.1128/JVI.75.10.4843-4853.2001> PMID: 11312356
27. Arumugaswami V, Wu T-T, Martinez-Guzman D, Jia Q, Deng H, Reyes N, et al. ORF18 Is a Transfactor That Is Essential for Late Gene Transcription of a Gammaherpesvirus. *J Virol* [Internet]. 2006 Oct 1; 80(19):9730. Available from: <http://jvi.asm.org/content/80/19/9730.abstract> <https://doi.org/10.1128/JVI.00246-06> PMID: 16973577
28. McKenzie J, Lopez-Giraldez F, Delecluse H-J, Walsh A, El-Guindy A. The Epstein-Barr Virus Immunoevasins BCRF1 and BPLF1 Are Expressed by a Mechanism Independent of the Canonical Late Pre-initiation Complex. *PLOS Pathog* [Internet]. Public Library of Science; 2016 Nov 17; 12(11):e1006008-. Available from: <https://doi.org/10.1371/journal.ppat.1006008> PMID: 27855219
29. Brulois KF, Chang H, Lee AS-Y, Ensser A, Wong L-Y, Toth Z, et al. Construction and Manipulation of a New Kaposi's Sarcoma-Associated Herpesvirus Bacterial Artificial Chromosome Clone. *J Virol* [Internet]. 2012 Sep 15; 86(18):9708. Available from: <http://jvi.asm.org/content/86/18/9708.abstract> <https://doi.org/10.1128/JVI.01019-12> PMID: 22740391
30. Bruce AG, Barcy S, DiMaio T, Gan E, Garrigues HJ, Lagunoff M, et al. Quantitative Analysis of the KSHV Transcriptome Following Primary Infection of Blood and Lymphatic Endothelial Cells. *Pathog* (Basel, Switzerland) [Internet]. MDPI; 2017 Mar 19; 6(1):11. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/28335496>
31. Bailey L. T, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: *Proceedings of the Second International Conference on Intelligent Systems for Molecular*



- Biology [Internet]. Menlo Park; 1994. p. 28–36. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/?term=Fitting+a+mixture+model+by+expectation+maximization+to+discover+motifs+in+biopolymers>
32. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics* [Internet]. 2011 Feb 16; 27(7):1017–8. Available from: <https://doi.org/10.1093/bioinformatics/btr064> PMID: 21330290
  33. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* [Internet]. 2010; 38(4):576–89. Available from: <http://www.sciencedirect.com/science/article/pii/S1097276510003667> <https://doi.org/10.1016/j.molcel.2010.05.004> PMID: 20513432
  34. Djavadian R, Chiu Y-F, Johannsen E. An Epstein-Barr Virus-Encoded Protein Complex Requires an Origin of Lytic Replication In Cis to Mediate Late Gene Transcription. *PLOS Pathog* [Internet]. Public Library of Science; 2016 Jun 27; 12(6):e1005718–. Available from: <https://doi.org/10.1371/journal.ppat.1005718> PMID: 27348612
  35. Lin CL, Li H, Wang Y, Zhu FX, Kudchodkar S, Yuan Y. Kaposi's sarcoma-associated herpesvirus lytic origin (ori-Lyt)-dependent DNA replication: identification of the ori-Lyt and association of K8 bZip protein with the origin. *J Virol* [Internet]. American Society for Microbiology; 2003 May; 77(10):5578–88. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/12719550> <https://doi.org/10.1128/JVI.77.10.5578-5588.2003> PMID: 12719550
  36. AuCoin DP, Colletti KS, Xu Y, Cei SA, Pari GS. Kaposi's Sarcoma-Associated Herpesvirus (Human Herpesvirus 8) Contains Two Functional Lytic Origins of DNA Replication. *J Virol* [Internet]. 2002 Aug 1; 76(15):7890. Available from: <http://jvi.asm.org/content/76/15/7890.abstract> <https://doi.org/10.1128/JVI.76.15.7890-7896.2002> PMID: 12097603
  37. Yella VR, Bhimsaria D, Ghoshdastidar D, Rodríguez-Martínez JA, Ansari AZ, Bansal M. Flexibility and structure of flanking DNA impact transcription factor affinity for its core motif. *Nucleic Acids Res* [Internet]. 2018/11/05. Oxford University Press; 2018 Dec 14; 46(22):11883–97. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/30395339> <https://doi.org/10.1093/nar/gky1057> PMID: 30395339
  38. Deng W, Roberts SGE. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev* [Internet]. Cold Spring Harbor Laboratory Press; 2005 Oct 15; 19(20):2418–23. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/16230532> <https://doi.org/10.1101/gad.342405> PMID: 16230532
  39. Wong E, Wu T-T, Reyes N, Deng H, Sun R. Murine Gammaherpesvirus 68 Open Reading Frame 24 Is Required for Late Gene Expression after DNA Replication. *J Virol* [Internet]. 2007 Jun 15; 81(12):6761. Available from: <http://jvi.asm.org/content/81/12/6761.abstract> <https://doi.org/10.1128/JVI.02726-06> PMID: 17392360
  40. Gardner MR, Glaunsinger BA. Kaposi's Sarcoma-Associated Herpesvirus ORF68 Is a DNA Binding Protein Required for Viral Genome Cleavage and Packaging. Sandri-Goldin RM, editor. *J Virol* [Internet]. 2018 Aug 15; 92(16):e00840–18. Available from: <http://jvi.asm.org/content/92/16/e00840-18.abstract> <https://doi.org/10.1128/JVI.00840-18> PMID: 29875246
  41. Labo N, Miley W, Marshall V, Gillette W, Esposito D, Bess M, et al. Heterogeneity and Breadth of Host Antibody Response to KSHV Infection Demonstrated by Systematic Analysis of the KSHV Proteome. *PLOS Pathog* [Internet]. Public Library of Science; 2014 Mar 27; 10(3):e1004046. Available from: <https://doi.org/10.1371/journal.ppat.1004046> PMID: 24675986
  42. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
  43. Trim Galore - [http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Available from: [http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)
  44. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* [Internet]. 2012/10/25. Oxford University Press; 2013 Jan; 29(1):15–21. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/23104886> <https://doi.org/10.1093/bioinformatics/bts635> PMID: 23104886
  45. Pyl PT, Anders S, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* [Internet]. 2014 Sep 25; 31(2):166–9. Available from: <https://doi.org/10.1093/bioinformatics/btu638> PMID: 25260700
  46. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* [Internet]. 2014; 15(12):550. Available from: <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
  47. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol* [Internet]. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2011 Jan 10; 29:24. Available from: <https://doi.org/10.1038/nbt.1754> PMID: 21221095

48. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* [Internet]. 2012 Apr 19; 14(2):178–92. Available from: <https://doi.org/10.1093/bib/bbs017> PMID: 22517427
49. Zhou T, Yang L, Lu Y, Dror I, Dantas Machado AC, Ghane T, et al. DNASHape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res* [Internet]. 2013/05/22. Oxford University Press; 2013 Jul; 41(Web Server issue):W56–62. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/23703209> <https://doi.org/10.1093/nar/gkt437> PMID: 23703209