

RESEARCH ARTICLE

The landscape of transcription initiation across latent and lytic KSHV genomes

Xiang Ye¹ , Yang Zhao¹ , John Karijovich^{1,2,3*} 

1 Department of Pathology, Microbiology, and Immunology, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, **2** Vanderbilt-Ingram Cancer Center, Nashville, Tennessee, United States of America, **3** Vanderbilt Institute for Infection, Immunology and Inflammation, Nashville, Tennessee, United States of America

 These authors contributed equally to this work.

* John.karijovich@vanderbilt.edu



Abstract

Precise promoter annotation is required for understanding the mechanistic basis of transcription initiation. In the context of complex genomes, such as herpesviruses where there is extensive genic overlap, identification of transcription start sites (TSSs) is particularly problematic and cannot be comprehensively accessed by standard RNA sequencing approaches. Kaposi's sarcoma-associated herpesvirus (KSHV) is an oncogenic gammaherpesvirus and the etiological agent of Kaposi's sarcoma and the B cell lymphoma primary effusion lymphoma (PEL). Here, we leverage RNA annotation and mapping of promoters for analysis of gene expression (RAMPAGE) and define KSHV TSSs transcriptome-wide and at nucleotide resolution in two widely used models of KSHV infection, namely iSLK.219 cells and the PEL cell line TREx-BCBL1-RTA. By mapping TSSs over a 96 h time course of reactivation we confirm 48 of 50 previously identified TSSs. Moreover, we identify over 100 novel transcription start site clusters (TSCs) in each cell line. Our analyses identified cell-type specific differences in TSC positions as well as promoter strength, and defined motifs within viral core promoters. Collectively, by defining TSSs at high resolution we have greatly expanded the transcriptional landscape of the KSHV genome and identified transcriptional control mechanisms at play during KSHV lytic reactivation.

OPEN ACCESS

Citation: Ye X, Zhao Y, Karijovich J (2019) The landscape of transcription initiation across latent and lytic KSHV genomes. *PLoS Pathog* 15(6): e1007852. <https://doi.org/10.1371/journal.ppat.1007852>

Editor: Paul M. Lieberman, Wistar Institute, UNITED STATES

Received: January 23, 2019

Accepted: May 20, 2019

Published: June 12, 2019

Copyright: © 2019 Ye et al. 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 relevant data are within the manuscript and its Supporting Information files. RAMPAGE sequencing data are available in GEO GSE129902

Funding: This work was supported by institutional startup funds (to J.K.), a grant from the Vanderbilt-Ingram Cancer Center Ambassadors program (to J.K.), American Cancer Society Institutional Research Grant (#IRG-58-009-58; to J.K.). J.K. is a Pew Biomedical Scholar. The funders had no role in study design, data collection and analysis,

Author summary

Kaposi's sarcoma-associated herpesvirus (KSHV) is an oncogenic gammaherpesvirus and the etiological agent of Kaposi's sarcoma and the B cell lymphoma primary effusion lymphoma (PEL). Despite identification of the virus over 20 years ago there is still an incomplete understanding of how many RNAs are transcribed from the viral genome and the location these RNAs are derived from. To fill this gap in knowledge we determined the landscape of transcription initiation on the KSHV genome. Our analyses more than tripled the number of known TSCs and thus viral-expressed RNAs. Furthermore, we identified key sequence features associated with the regulation of viral transcription start sites. This study provides the first transcriptome-wide characterization of KSHV transcription

decision to publish, or preparation of the manuscript.

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

initiation sites as well as a framework for future studies to define functions of novel viral transcripts and viral gene regulatory elements.

Introduction

Regulated gene expression is an essential process for all eukaryotic cells as well as the pathogens that infect them. A fundamental step in the regulation of gene expression is the initiation of RNA synthesis. RNA Polymerase II (RNAP II) transcription begins with the binding of gene-specific regulatory factors within the core promoter located near the transcription start site (TSS). Sequence elements found within core promoters include the TATA element, BRE (Transcription factor II B recognition element), Initiator (Inr), and downstream promoter element (DPE) [1]. These elements serve as binding sites for subunits of the transcription machinery and their nucleotide composition can impact the efficiency of transcription initiation.

Models of transcription initiation often depict RNA synthesis initiating from a single nucleotide. However, the application of high-throughput 5'-end sequencing technologies, such as Cap Analysis of Gene Expression (CAGE), have demonstrated that most RNAP II promoters have an array of closely spaced TSSs instead of the expected single TSS [2–7]. Thus, promoters can be more accurately described as a distribution of initiation events (transcription start site clusters, TSCs) on a stretch of given nucleotides [5]. Moreover, the nucleotide sequences that comprise the core promoter can influence the distribution of TSSs as well as the expression profile for a given gene. This has led to the classification of animal promoters into two major groups, narrow and broad [8, 9]. For example, more broadly distributed TSSs are correlated with CpG islands and ubiquitously expressed genes, whereas promoters harboring TATA and TATA-like sequences exhibit a narrow distribution of initiation sites and frequently drive expression of tissue-specific genes [4, 10–15].

Viruses employ a parasitic lifestyle thus they exploit many, and in some cases, all, of the host gene expression machineries. The co-option of cellular processes allows viruses to dramatically downsize their genomes, however, it necessitates the incorporation of host cell-specific gene regulatory mechanisms and features that enable the use of such mechanisms. Additionally, the compact nature of viral genomes requires the encoding of viral proteins and RNAs within a relatively small amount of genomic space.

Kaposi sarcoma-associated herpesvirus (KSHV) is a human oncogenic virus and a member of the gammaherpesvirus subfamily [16, 17]. While KSHV infection is generally unproblematic in the healthy adult population, in the context of immunosuppression such as iatrogenic immunosuppression or advanced HIV-infection resulting in AIDS, the virus is associated with several malignancies [18, 19]. In fact, it is estimated that approximately 1% of all human tumors are associated with KSHV infection and the World Health Organization (WHO) has classified the virus as a class I carcinogen [18, 20]. KSHV is the etiological agent of all forms of Kaposi sarcoma [21–23], a complex and highly vascularized solid tumor of endothelial origin [24], as well as the B cell lymphoproliferative disorders multicentric Castleman's disease (MCD) and primary effusion lymphoma (PEL) [25, 26].

As with other herpesviruses, KSHV displays two distinct phases of its viral lifecycle, latency and lytic infection. During latent infection KSHV exists in a dormant state in which viral gene expression is dramatically restricted such that few viral antigens and no viral particles are produced. In contrast, during the lytic cycle, which is driven by the viral-encoded transcription factor replication and transcription activator (RTA) [27, 28], the full repertoire of viral genes is

expressed and new infectious virions are produced. KSHV has a large DNA genome and similar to other DNA viruses has a complex gene organization including overlapping genes and polycistronic mRNAs. The current annotation of the KSHV genome has over 90 genes that encode for viral proteins and both long and small noncoding RNAs [29].

Despite identification of KSHV over two decades ago we still lack fundamental knowledge regarding how viral gene expression is controlled at the level of transcription initiation. Additionally, although several studies have employed RNA-seq to define the transcriptional output of the KSHV genome as well as the temporal kinetics of viral gene expression, it is often difficult to accurately identify TSSs from traditional RNA-seq data alone. Identification of TSSs is particularly important in complex transcriptomes, such as that of KSHV, where there is extensive overlapping transcription of viral genes and the appearance of polycistronic RNAs. To fill this gap in fundamental knowledge we have mapped all transcription initiation events on the viral genome using RNA annotation and mapping of promoters for analysis of gene expression (RAMPAGE) [30]. Leveraging RAMPAGE, we now provide transcriptome-wide nucleotide resolution TSC annotations for viral genes in two widely used models of KSHV infection, namely the iSLK.219 [31, 32] and TREx-BCBL1-RTA [33] systems. We confirm 48 of 50 previously known TSSs as well as identify over 100 novel TSCs in each cell line. Moreover, these analyses identify cell-type specific differences in TSC positions as well as promoter strength, and define sequence elements comprising viral core promoters. Collectively, we greatly expand the transcriptional landscape of the KSHV genome and identify transcriptional control mechanisms at play during KSHV lytic reactivation.

Results

Transcriptome-wide identification of KSHV transcription start sites at single-nucleotide resolution by RAMPAGE

Precise promoter annotation is required for understanding the mechanistic basis of condition- and tissue-specific gene regulation. While the transcriptional landscape of KSHV has been characterized by RNA-seq there is still a gap in fundamental knowledge regarding the location of viral TSSs, and thus it has not been feasible to comprehensively define viral promoters and the sequence motifs embedded within them. Moreover, whether there are condition (i.e. latent vs. lytic) or cell type specific differences in TSS usage is not known. To fill this gap in knowledge we took advantage of two well established models of KSHV infection, namely, iSLK.219 and TREx-BCBL1-RTA cells. While iSLK.219 cells are of clear cell renal cell carcinoma origin and thus of lesser biological relevance [34], the iSLK system has proven invaluable to KSHV research as it is routinely used in studies addressing viral gene function through bacterial artificial chromosome (BAC) mutagenesis. iSLK.219 cells are infected with a recombinant virus, KSHV.219, in which GFP, expressed from the elongation factor 1- α promoter, and RFP, expressed from the viral lytic gene PAN promoter, were inserted into the viral genome [31, 32]. In contrast, TREx-BCBL1-RTA cells are a genetically engineered derivative of KSHV-infected B cells isolated from a patient with PEL [33]. A key feature in both systems is the integration of a doxycycline (Dox)-inducible version of the major viral transcription activator RTA. Thus, upon introduction of Dox into the cell culture media KSHV enters the lytic cycle.

Leveraging iSLK.219 and TREx-BCBL1-RTA cells we mapped TSSs transcriptome-wide at nucleotide resolution across a 96 h Dox-induced reactivation time course using RAMPAGE. RAMPAGE, which combines the two orthogonal 5'-selection approaches of template-switching and cap-trapping, enables TSS identification with high specificity through the high-throughput sequencing of 5'-complete complementary DNAs (Fig 1A). Additionally, an important distinction from CAGE is that RAMPAGE allows for pair-ended sequencing and

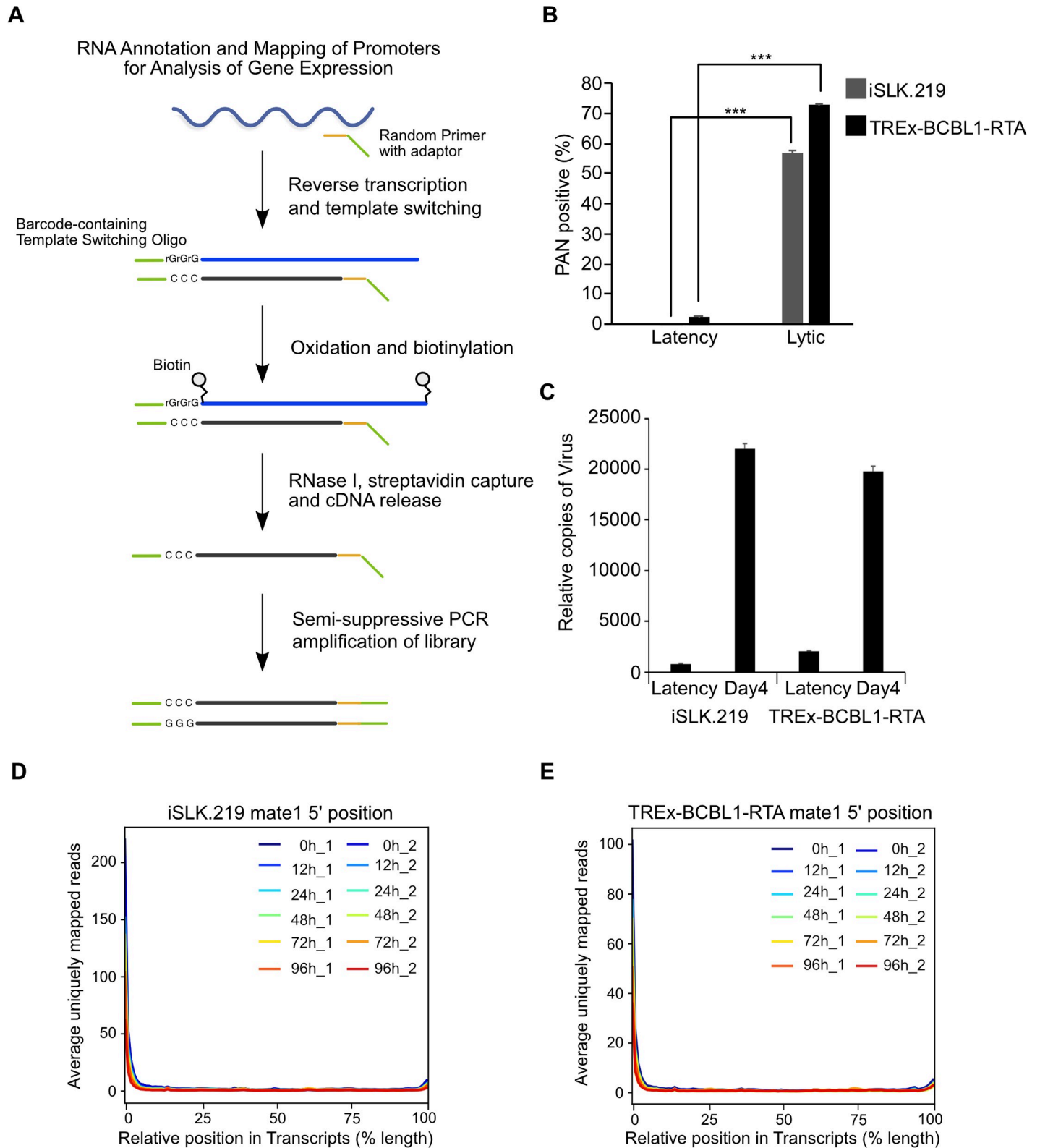


Fig 1. RAMPAGE enables transcription start site identification at nucleotide resolution. (A) Schematic illustration of the RAMPAGE method. (B) iSLK.219 and TREx-BCBL1-RTA cells were reactivated for 72 h and the percentage of lytic reactivated cells was quantified by PAN RNA FISH-FLOW. (C) Virions were isolated from supernatants of iSLK.219 and TREx-BCBL1-RTA cells induced for 96 h and KSHV genomes were quantified by qPCR against ORF52. (D and E) Average

uniquely mapped reads by position for endogenous transcripts, for libraries generated from (D) iSLK.219 and (E) TREx-BCBL1-RTA RNA. Error bars in all panels represent mean±SD from three independent experiments. p Values were determined by the Student's t test ***, $p < 0.001$.

<https://doi.org/10.1371/journal.ppat.1007852.g001>

thus yields extensive transcript connectivity information allowing the annotation of genes with their TSS [30]. Total RNA was isolated from cells in a latent state, or at 12 h, 24 h, 48 h, 72 h, and 96 h post-lytic reactivation and RAMPAGE libraries were prepared. These time points were chosen as we observed >70% of cells in the lytic stage by 72 h post-dox induction as well as a significant increase in virion associated DNA in the culture supernatant (Fig 1B and 1C).

The resultant RAMPAGE sequencing reads were aligned to the human (GRCh38) and KSHV (GQ994935.1) genomes and TSCs were identified by Paraclu [5] and quantified in tags per million reads mapped (TPM). To assess the specificity of our RAMPAGE data for 5' ends we examined the distribution of the raw TSS 5' signal over cellular annotated transcripts. Importantly, the metaprofile of signal density clearly demonstrates enrichment near the 5' end and confirms the specificity of our data (Fig 1D and 1E).

Within 12 h of Dox addition to the culture media there was an increase in TSSs that continued through the time course (Figs 2 and 3). In total, we identified 164 and 292 TSCs in iSLK.219 and TREx-BCBL1-RTA cells, respectively (Figs 2 and 3, S1 Table). TSSs for 50 viral transcripts have been previously mapped or annotated [27, 29, 35–53]. Within our RAMPAGE data we observed prominent TSSs at the exact position or within a few nucleotides of 48 of the 50 previously identified TSSs, with the exceptions being the TSSs driving ORF17 and ORF42 expression (S1 Fig). We surmise that the slight variations in TSS architecture observed between previous studies and our data is related to the use of histone deacetylase (HDAC) inhibitors to promote KSHV reactivation in the previous reports.

An advantage of RAMPAGE over other 5'-end sequencing technologies is that sequencing is performed in a pair-ended format and thus more information regarding transcript content is present in the data. Leveraging the pair-end information of RAMPAGE we were able to assign many TSSs to ORFs that were previously lacking TSS information (S1 Table). For example, ORF7, ORF10 and ORF7K were assigned TSSs. This high-resolution map of viral transcription initiation refines the current annotation of KSHV transcriptome, and reveals a much more complex transcriptional landscape than previously appreciated.

TSCs are enriched in regions of open chromatin

Transcriptional regulation is highly complex and chromatin structure and modifications directly contribute to transcription initiation. The open chromatin landscape of the KSHV genome in PEL has been investigated by formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq) [54]. Previous FAIRE-seq analyses found that regions of open chromatin were not restricted to transcriptionally active loci, but that transcriptionally inactive loci were also nucleosome depleted. Moreover, the transcriptional repressor CTCF occupied the majority of nucleosome depleted regions regardless of transcriptional activity [54]. Having mapped the transcription initiation landscape of PEL cells, we intersected FAIRE-seq and CTCF chromatin immunoprecipitation sequencing (ChIP-seq) data sets with the RAMPAGE defined TSCs from TREx-BCBL1-RTA cells (Fig 4A). Consistent with previous work, we observed that TSCs of both latent and lytic reactivated TREx-BCBL1-RTA cells were highly correlated ($p < 0.00001$) with regions of open chromatin. Furthermore, we similarly observed enrichment of CTCF at both latent and lytic TSCs ($p < 0.00001$). In fact, more than 50% of KSHV TSCs are located within 250 bp of a CTCF or FAIRE-seq peak in TREx-BCBL1-RTA cells (Fig 4B and 4C).

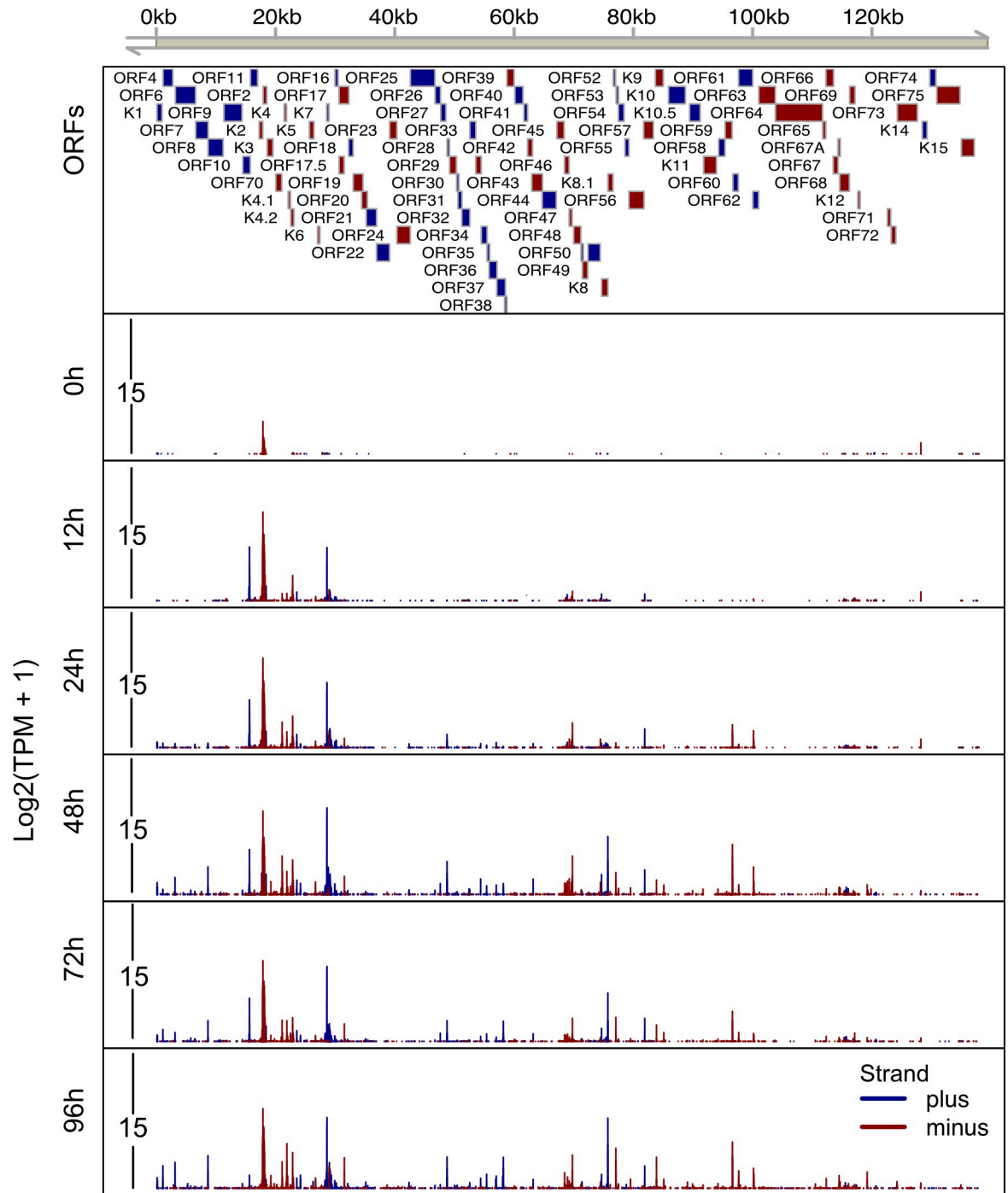


Fig 2. High resolution map of transcription start sites on the KSHV.219 genome in iSLK.219 cells. Total RNA was isolated from iSLK.219 cells at 0h, 12h, 24h, 48h, 72h and 96h post-dox induced lytic reactivation and RAMPAGE libraries were prepared and sequenced. The 5' signal of read 1, which corresponds to transcriptional start sites, were mapped on the KSHV genome (GQ994935.1). A schematic of the KSHV genome including genomic coordinates of annotated ORFs is depicted on the top horizontal axis. The lower 6 tracks depict the transcripts initiated from every nucleotide across the whole KSHV genome at each time point. The Y axis depicts the log2 transformed TPM value. Blue boxes/lines indicate ORFs/TSSs on the plus strand, red boxes/lines indicate ORFs/TSSs on the minus strand.

<https://doi.org/10.1371/journal.ppat.1007852.g002>

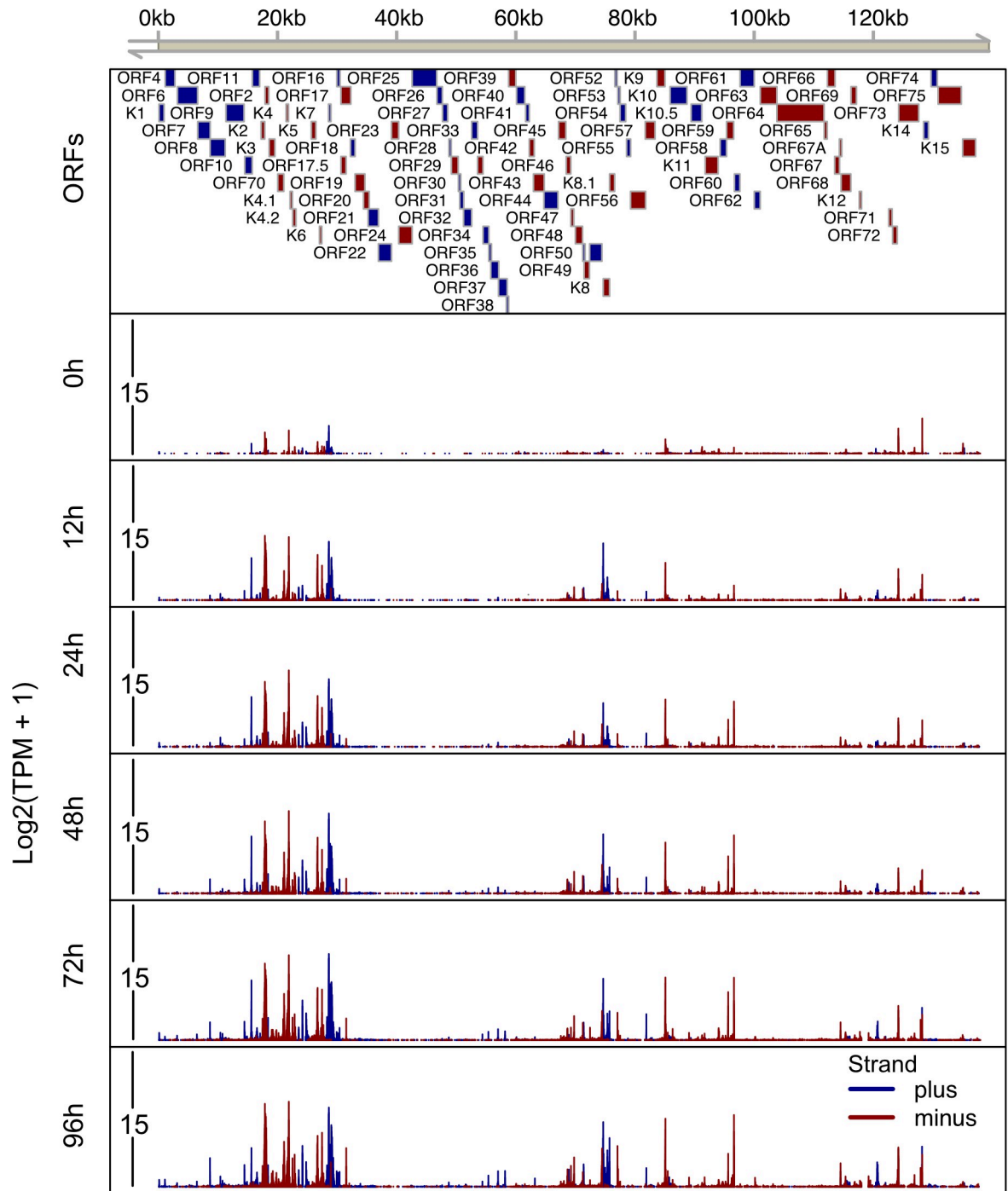


Fig 3. High resolution mapping of transcription start sites on the KSHV genome in TReX-BCBL1-RTA cells. Total RNA was isolated from TReX-BCBL1-RTA cells at 0h, 12h, 24h, 48h, 72h and 96h post-dox induced lytic reactivation and RAMPAGE libraries were prepared and sequenced. The 5' signal of read 1, which corresponds to transcriptional start sites, were mapped on the KSHV genome (GQ994935.1). A schematic of the KSHV genome including genomic coordinates of annotated ORFs is depicted on the top horizontal axis. The lower 6 tracks transcripts initiated from every nucleotide across the whole KSHV genome at each time point. The Y axis depicts the log₂ transformed TPM value. Blue boxes/lines indicate ORFs/TSSs on the plus strand, red boxes/lines indicate ORFs/TSSs on the minus strand.

<https://doi.org/10.1371/journal.ppat.1007852.g003>

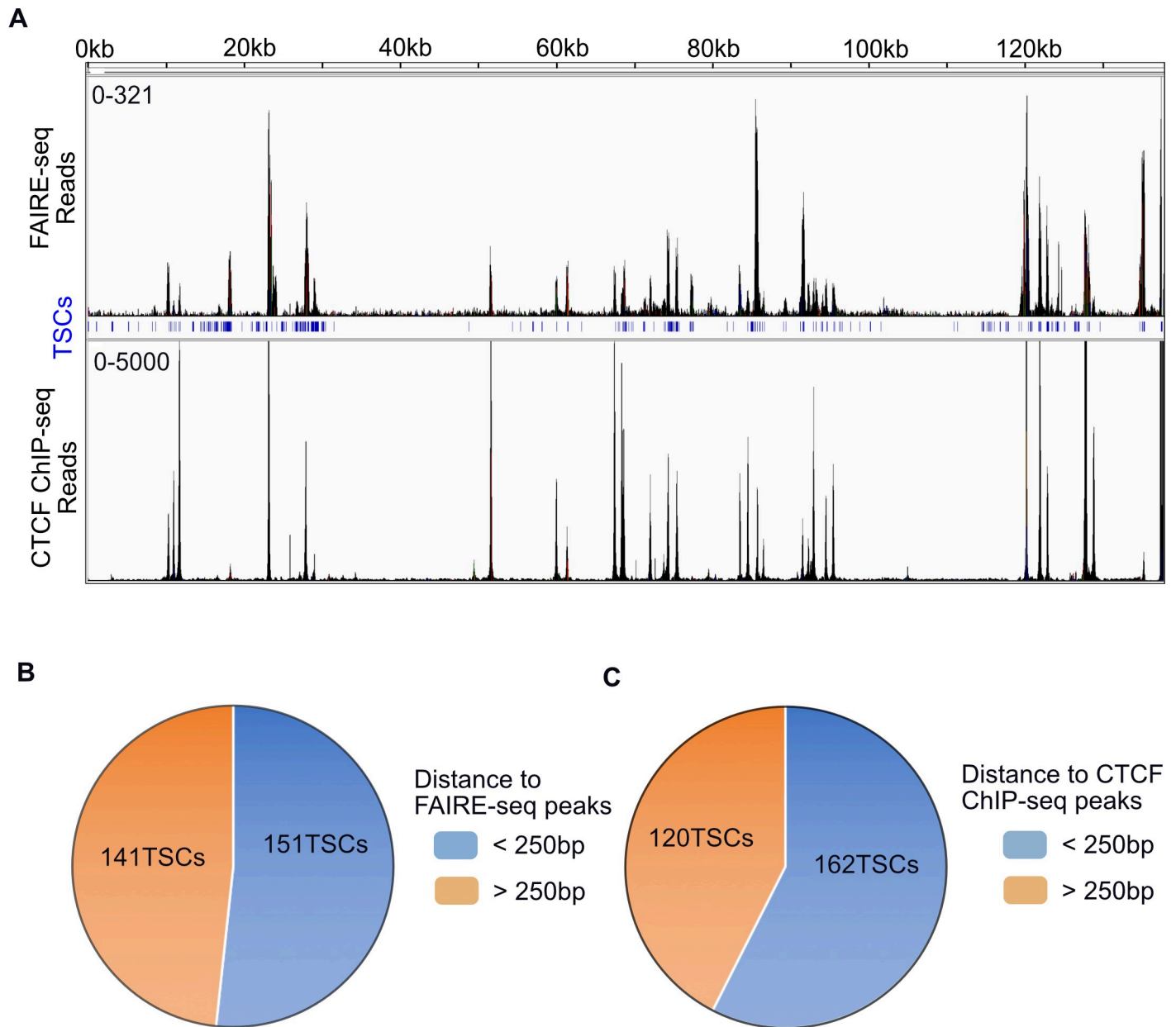


Fig 4. KSHV TSCs are located in regions of open chromatin and CTCF binding. (A) Overlap of RAMPAGE identified TSCs with FAIRE-seq and CTCF ChIP-seq signal in TREx-BCBL1-RTA cells. The top coordinates mark positions on KSHV genome, middle blue track shows the position of TSCs defined by RAMPAGE in TREx-BCBL1-RTA cell. (B) Pie chart depiction of the number of viral TSCs within 250bp of FAIRE-seq peaks in TREx-BCBL1-RTA. (C) Pie chart depiction of the number of viral TSCs within 250bp of CTCF ChIP-seq peaks in TREx-BCBL1-RTA. The enrichment of TSCs on FAIRE-seq ($p < 0.00001$) and CTCF ChIP-seq peaks ($p < 0.00001$) were tested by permutations.

<https://doi.org/10.1371/journal.ppat.1007852.g004>

TREx-BCBL1 cells have been the subject of extensive investigation regarding the presence and location of various chromatin modifications, including the activating mark H3K4me3 and the repressive mark H3K27me3, as well as the KSHV latent protein LANA and cellular RNAP II [55]. By intersection analysis we visualized the previous ChIP-seq data with our comprehensive TSC map (S2 Fig). Interestingly, although we analyzed H3K4me3 ChIP-seq data from latent TREx-BCBL1-RTA cells, TSCs belonging to both latent and lytic genes are located in close proximity to H3K4me3 signal. However, given that a small percentage of PEL cells are

continually undergoing spontaneous reactivation it is possible that the H3K4me3 signal over the lytic genes is derived from this population of cells.

Evidence for cell-type specific KSHV transcription start site selection

We next sought to determine whether KSHV promoter usage is regulated in a cell-type specific manner. Intersection analysis of TSCs identified in iSLK.219 and TREx-BCBL1-RTA cells revealed that viral gene expression is highly cell-type specific. We identified 129 TSCs were present in both cell lines, we identified 35 and 163 TSCs present exclusively in either iSLK.219 or TREx-BCBL1-RTA cells, respectively (Fig 5A). Importantly, the difference in numbers of TSCs identified is likely not due to sequencing depth as in both iSLK.219 and TREx-BCBL1-RTA a similar number of reads were mapped to the viral genome (S3 Fig). Interestingly, among the 129 TSCs present in both cell-types we identified three novel TSCs that have a corresponding RTA ChIP-seq peak immediately preceding the TSC, suggesting the transcripts are RTA-dependent (Fig 5B). Importantly, the RTA ChIP-seq peaks identified were present in two separate data sets analyzed. In support of a role for RTA in the expression of these TSCs, RTA was capable of activating the expression of a luciferase reporter harboring these genomic fragments as promoters (S4 Fig). While this data does not directly demonstrate the TSCs are RTA-dependent, these data do demonstrate that these genomic fragments harbor the potential to serve as promoters that drive productive transcription initiation.

Inspection of our data uncovered that many immediate early genes, such as K6, tend to be associated with multiple TSCs, all of which are shared between both iSLK.219 and TREx-BCBL1-RTA cells. Previous studies have demonstrated that immediate early genes can be expressed at a low basal level during latency, and are subsequently robustly induced upon lytic reactivation. Given that both multiple shared TSCs are present in both cell-types we hypothesized that basal and induced gene expression may be controlled by different promoters. Indeed, inspection of TSCs that drive expression of K6 uncovered multiple promoters, with a prominent switch in TSC usage upon lytic reactivation (Fig 5C). As shown in Fig 5C, K6 is expressed from three major promoters (P1, P2 and P3). While P3 was preferentially used during latency, usage of P1 and P2 increased substantially upon lytic induction. The conservation of regulated TSC selection for K6 between cell-types suggests this may be an underlying mechanism controlling the proper regulation of K6 expression.

As noted above we also identified cell-type specific TSCs. For example, in TREx-BCBL1-RTA cells we identified 11 ORFs expressed by TSCs only present within TREx-BCBL1-RTA cells. In contrast, we identified a distinct set of 11 ORFs that were expressed from TSCs only identified in iSLK.219 cells. For instance, we identified a TSC for LANA2 (K10.5) in TREx-BCBL1 cells while it was absent from iSLK.219 cells (S5 Fig). This is consistent with a previous report demonstrating LANA2 expression is restricted to PEL and MCD cells [56]. Moreover, in TREx-BCBL1-RTA cells we identified a prominent TSC initiating synthesis of the ORF58 mRNA (Fig 5D), however, in iSLK.219 cells a homologous TSC is not present and ORF58 would need to be expressed from a bicistronic mRNA ORF58/59 (Fig 5D). However, it is equally possible that ORF58 is not expressed in iSLK.219 cells, and we are unaware of any studies detecting ORF58 protein expression in iSLK.219 cells.

Lytic reactivation in iSLK.219 and TREx-BCBL1-RTA follows different kinetics

In highly complex genomes with extensive genic overlap, 5'-end RNA sequencing is more accurate in measuring transcript expression than traditional RNA-seq as it avoids ambiguous read assignment and transcript length normalization. We thus quantified viral gene expression

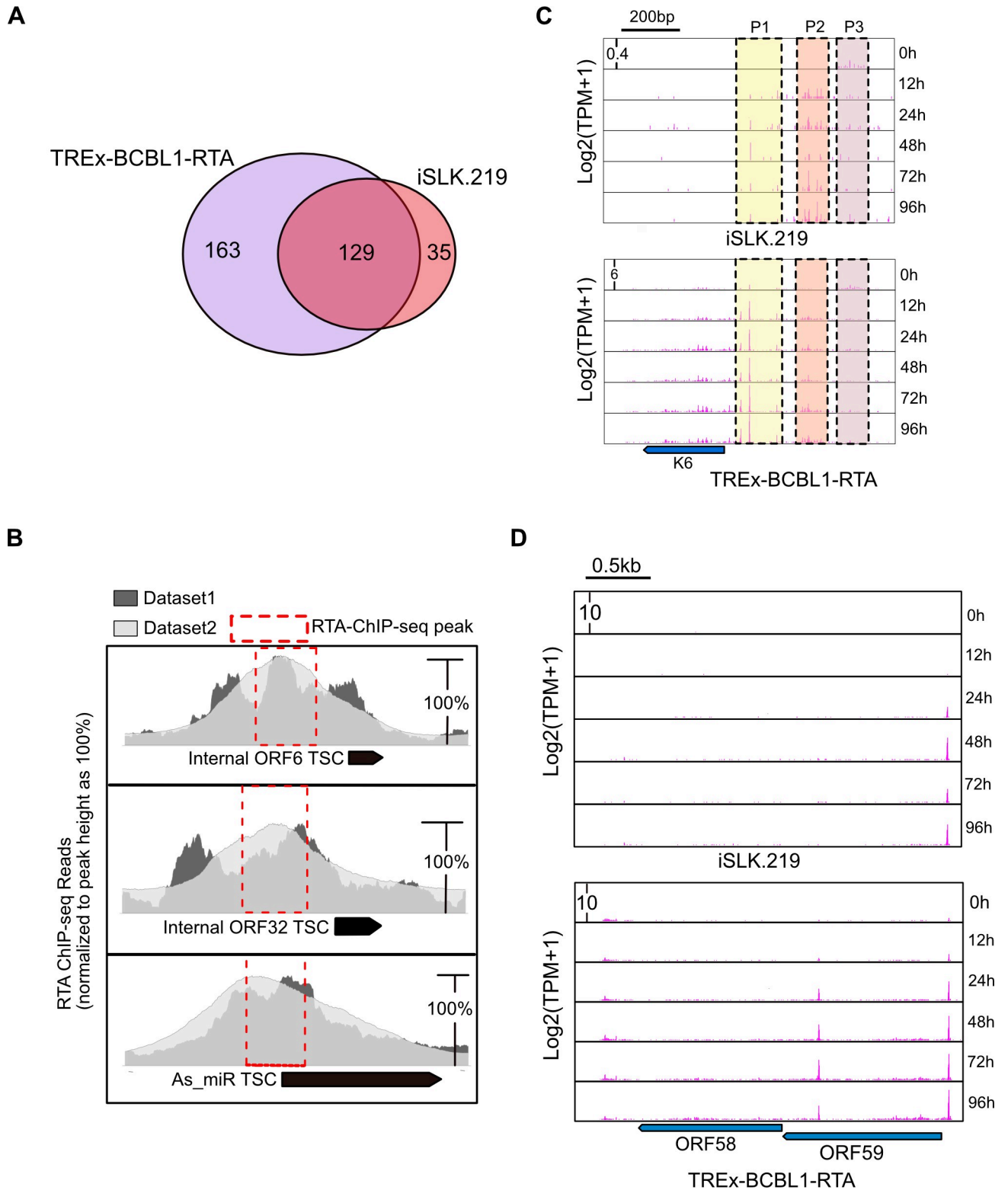


Fig 5. Cell-type specific TSC usage in KSHV-infected cells. (A) Venn diagram illustrating the overlap of viral TSCs in iSLK.219 and TREx-BCBL1-RTA cells. (B) RTA ChIP-seq read pile associated with three novel TSCs. The peak summit of each RTA ChIP-seq dataset was normalized to 100% for comparison. Black rectangles mark the position of RAMPAGE identified TSCs. Red-dashed boxes represent the location of RTA ChIP-seq peaks identified by MACs2. (C) RAMPAGE 5' signal at the K6 locus in iSLK.219 and TREx-BCBL1-RTA cells. The 5' signal on y axis is log transformed TPM value at single nucleotide resolution. The dash-line rectangle marks the primary TSCs (promoters), P1, P2, P3. (D) 5' nucleotide signal of RAMPAGE

read 1 associated with ORF58 and ORF59 in TREx-BCBL1-RTA, and iSLK.219 cells. The black bar on the top of the panel indicates the scale, the middle six tracks depict the six different time points post-dox induction. The location of the coding regions for ORF58 and ORF59, which is present on the minus strand, is shown below the panel.

<https://doi.org/10.1371/journal.ppat.1007852.g005>

in both cell lines based on the intensity of the 5' RAMPAGE signal (Fig 6A). Moreover, since we did not identify prominent TSCs for all ORFs we also quantified viral gene expression by mapping all of the reads directly to the ORFs (S6 Fig). Latency in iSLK.219 cells is extremely tight, and only RNAs belonging to known latent genes were expressed. In contrast, in latent TREx-BCBL1-RTA cells we observed a much broader transcriptional profile, with the promoters of several lytic genes, including ORF11, ORF50, ORFK8, and ORF57, producing mRNAs. The broader expression profile in TREx-BCBL1-RTA cells is consistent with low level spontaneous reactivation in PEL cells.

Unsupervised hierarchical clustering analysis further demonstrated that viral reactivation in iSLK.219 and TREx-BCBL1 cells follows different kinetic programs (Fig 6B). While the activity of most promoters in TREx-BCBL1-RTA cells gradually increases over the 96 h time course, expression from the majority of promoters in iSLK.219 cells peaked between 48 h–72 h (Fig 6A and 6B). Consistent with this analysis, in iSLK.219 cells the number of RAMPAGE reads mapping to the viral genome peak at 48 h, while in TREx-BCBL1-RTA cells this occurs at 96 h post-reactivation (S3 Fig).

Motif analysis of KSHV TSCs reveals time course related nucleotide usage bias

Given that unsupervised hierarchical clustering demonstrated distinct kinetic programs we next sought to investigate whether the individual programs were associated with unique core promoter compositions. It has previously been suggested that TSCs within a 50 bp window are under control of the same transcription initiation complex [57], thus we merged TSCs within a 50 bp window into a single cluster and performed K-means analysis on the resulting TSCs. K-means analysis separated the iSLK.219 and TREx-BCBL1-RTA expression profile in to three and four distinct classes, respectively (Fig 7A and 7B). The clusters defined revealed sets of genes with matched peak expression across the time points (Fig 7C and Fig 7D). The identification of variable cluster numbers between iSLK.219 and TREx-BCBL1-RTA transcription programs supports the notion of cell-type specific kinetic programs.

To identify sequence elements associated with the distinct clusters we extracted nucleotide sequences 50 bp upstream and downstream of the maximally expressed nucleotide within each cluster (MaxTSN) and searched for position-specific ultra-short motifs by kplogo analysis. All clusters, with the exception of TREx-BCBL1-RTA cluster 2, display a preference for a Py-Pu dinucleotide at the $-1/+1$ position and a similar preference has been observed at both mouse and human TSCs [4]. TREx-BCBL1-RTA cluster 2 is also unique in that a GG motif is present at the $+7/8$ position. Interestingly, when all cellular TSCs are ranked by expression the presence of a downstream GG motif is associated with higher expression (S7 Fig). Consistent with this analysis, cluster 2 displays the highest mean expression value across all time points. While this motif has been previously noted in CAGE data its association with more robust expression was not reported. The molecular basis for this observation is not known although its distance from the MaxTSN is approximately one helical turn of DNA and thus may be phased accordingly with the MaxTSN.

iSLK.219 clusters 1 and 3, which exhibit maximum expression 48 h and 96 h post-induction, respectively, and TREx-BCBL1-RTA cluster 1, which displays maximum expression at 96 h post-induction, all harbor an AT rich region 25–31 upstream of the MaxTSN (Fig 7E and

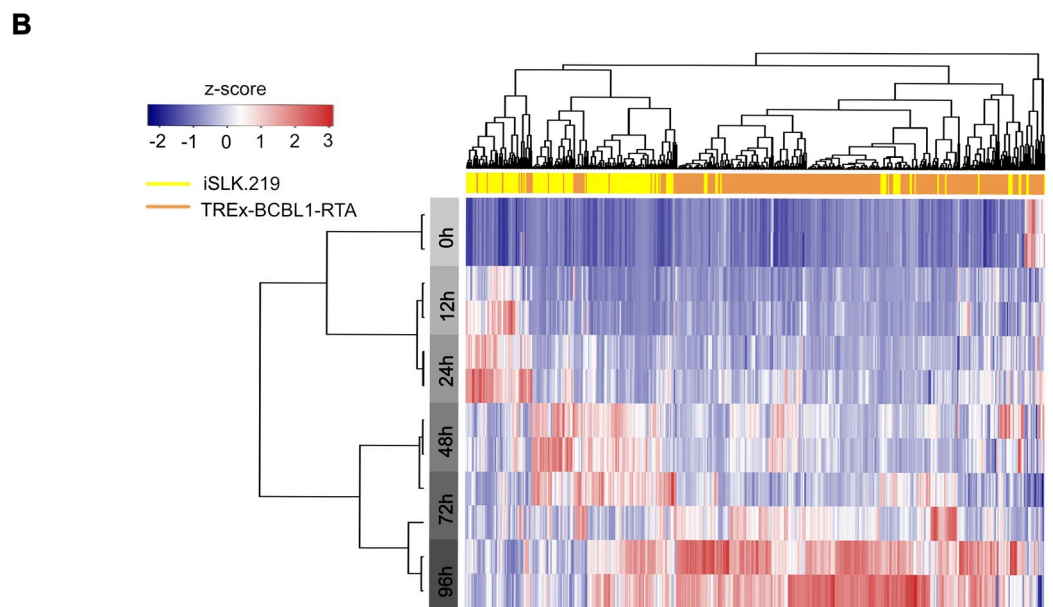
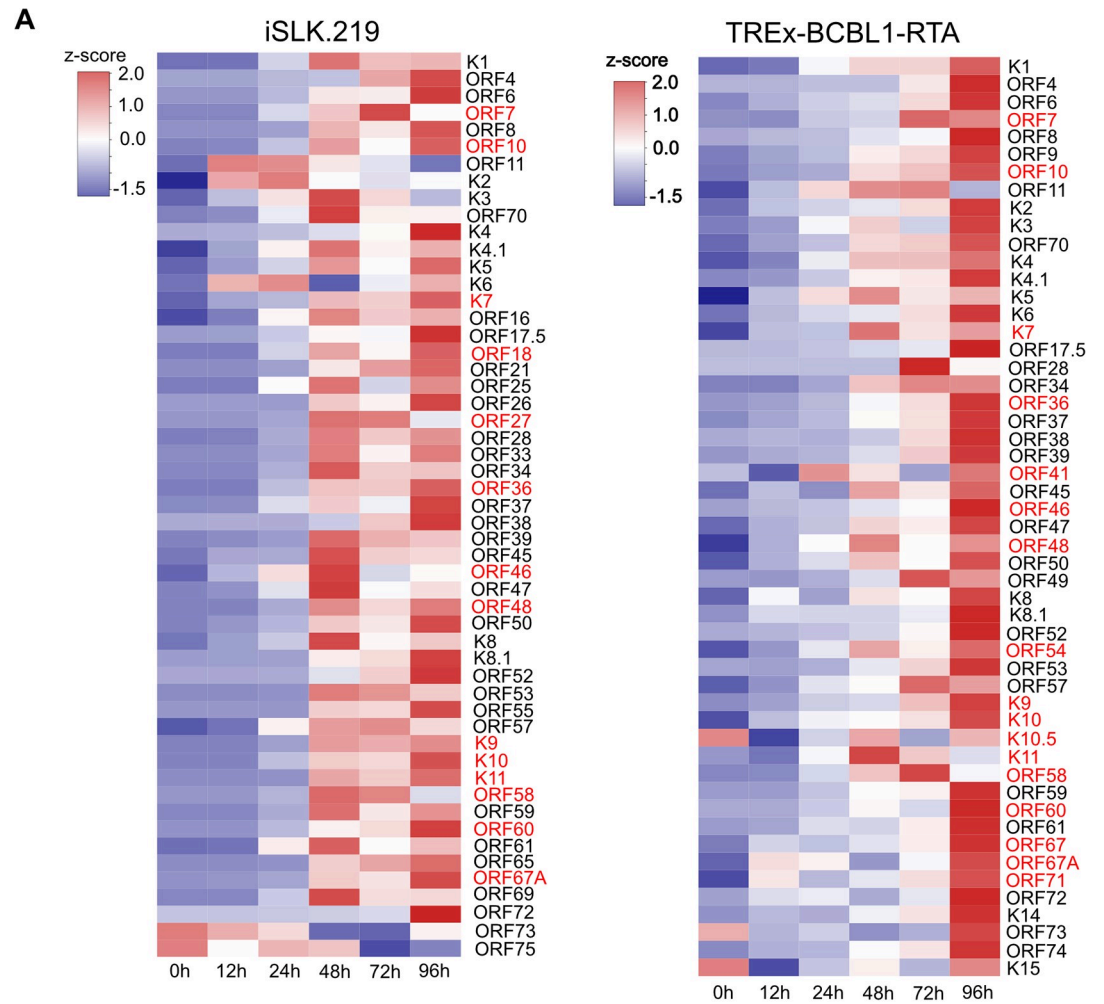


Fig 6. Kinetics of KSHV gene expression leveraging RAMPAGE. (A) Heat map illustrating the expression level of known ORFs based on promoter activity at the indicated time points. iSLK.219 cells (left panel) and TREx-BCBL1-RTA cells (right panel). The TSCs of ORFs in red are identified for the first time in this study. (B) Unsupervised hierarchical clustering analysis of transcriptional activity of all viral TSCs identified by RAMPAGE. The top horizontal bar indicates the cell lines: yellow and tan represent iSLK.219 and TREx-BCBL1-RTA cells, respectively. The left gray bar indicates the time point of sampling after doxycycline induced lytic reactivation. For all heat maps TPM expressions of TSCs were scaled to Z score before mapping.

<https://doi.org/10.1371/journal.ppat.1007852.g006>

7F). Moreover, motif analysis using HOMER identified an enrichment of TATA binding protein (TBP) motif within these clusters. However, we are cautious to interpret this as indicating that TBP is involved in the regulation of genes displaying maximum induction at later time points as viral late genes contain a variant TATA motif, TATT, that can also be defined by HOMER as a TBP motif. Thus, it is more likely that the enrichment of TBP motifs within these clusters reflects a combination of TSCs that are driven by TBP and the viral late gene initiation complex.

While neither an AT rich region or TBP motif was found in the other clusters, an initiator (Inr) motif (CA₊₁) is found enriched at TSCs within TREx-BCBL1-RTA cluster four. An Inr motif is similar to a TATA motif in that it can directly recruit the general transcription factor TFIID. However, the Inr recruits TFIID through interactions with TAF1 and TAF2 rather than TBP. The enrichment of Inr motifs in cellular genes that lack TATA motifs has been previously observed and it has been speculated that promoters with an Inr are less dependent on a TATA box and vice versa [58]. Thus, our observation mirrors what has been reported in humans and suggests preferential use of TATA or Inr motifs on viral promoters. Thus, we hypothesize that KSHV uses a combination of sequence elements to recruit TFIID and promote viral gene expression.

We also determined whether the presence of Inr motif was associated with a higher maximum of gene expression. When ranking TSC expression an expanded consensus Inr motif (BBCA₊₁BW) can be found dispersed among the moderately and highly expressed cellular TSCs within iSLK.219 cells (S7 Fig). In contrast, the expanded Inr motif is more commonly identified within higher expressed cellular TSCs in TREx-BCBL1-RTA cells. This suggests that distinct transcriptional control mechanisms and/or preferences may be at play within these two commonly used models of KSHV infection. With regard to viral TSCs, no clear enrichment of the Inr motif is found among genes with a distinct expression profile (S7 Fig).

We also investigated the relationship between the presence of a TATA motif with expression strength. Analysis of the TATA motif density plot clearly shows enrichment of this motif approximately 30 bp upstream of the MaxTSN in both iSLK.219 and TREx-BCBL1-RTA cells (Fig 8A). Interestingly, similar to what was observed for the Inr motif, a TATA motif can be found dispersed among the moderately and highly expressed cellular TSCs within iSLK.219 cells, while it is more prevalent among the highest expressed genes in TREx-BCBL1-RTA cells. With regard to viral TSCs, a TATA motif is not associated with expression.

The TATT motif is a noncanonical TBP-like motif that enables late gene expression via a specialized viral transcription complex [59, 60]. Interestingly, while we observed clear enrichment of this motif approximately 30 bp upstream of the viral MaxTSN in motif density plots, there also appears to be a preference for it upstream of the cellular MaxTSN (Fig 8B). To further investigate this, we first quantified the presence of TATA and TATT motifs within cellular and viral promoters (Fig 8C). Consistent with previous reports we observed TATA motifs within only a small percentage (2~3%) of cellular promoters. Moreover, a TATT motif was also observed in a similar portion of cellular promoters. In contrast, we observed a higher percentage of viral promoters harboring TBP-like motifs. Specifically, we observed a TATA motif within 16% of viral promoters and a TATT motif within 7%. We next calculated the relative

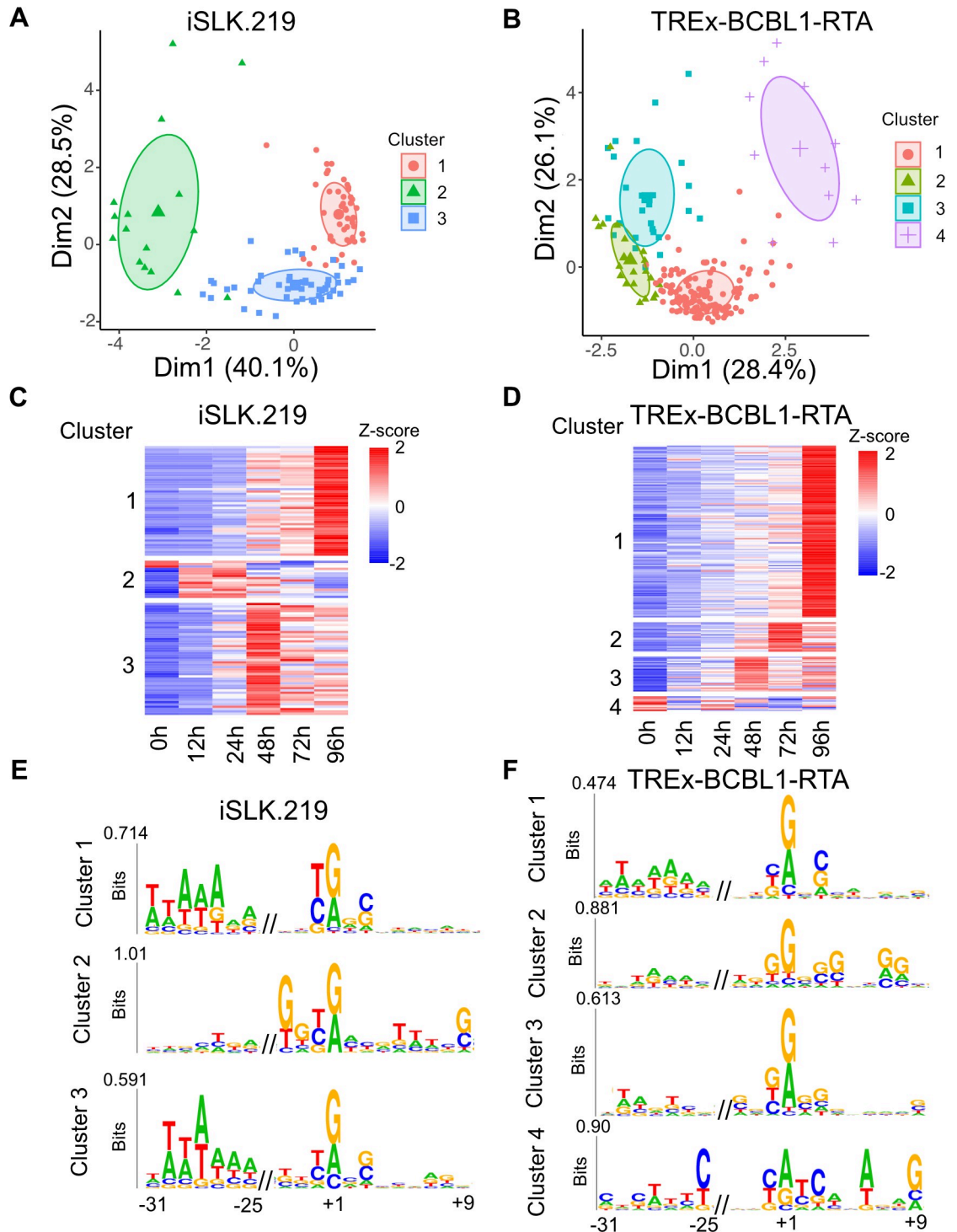


Fig 7. K-means and motif analysis of cis-elements within KSHV promoters. (A and B) PCA plot of all KSHV TSCs expression for (A) iSLK.219 and (B) TREx-BCBL1-RTA cells. (C and D) Corresponding heat map of the KSHV TSCs identified for (C) iSLK.219 and (D) TREx-BCBL1-RTA cells. Clusters identified by K-means are shown. (E and F) Sequences and up and downstream of the MaxTSN within each cluster were extracted and subjected to kplogo analysis. Ultra-short motifs identified in (E) iSLK.219 and (F) TREx-BCBL1-RTA cells are depicted.

<https://doi.org/10.1371/journal.ppat.1007852.g007>

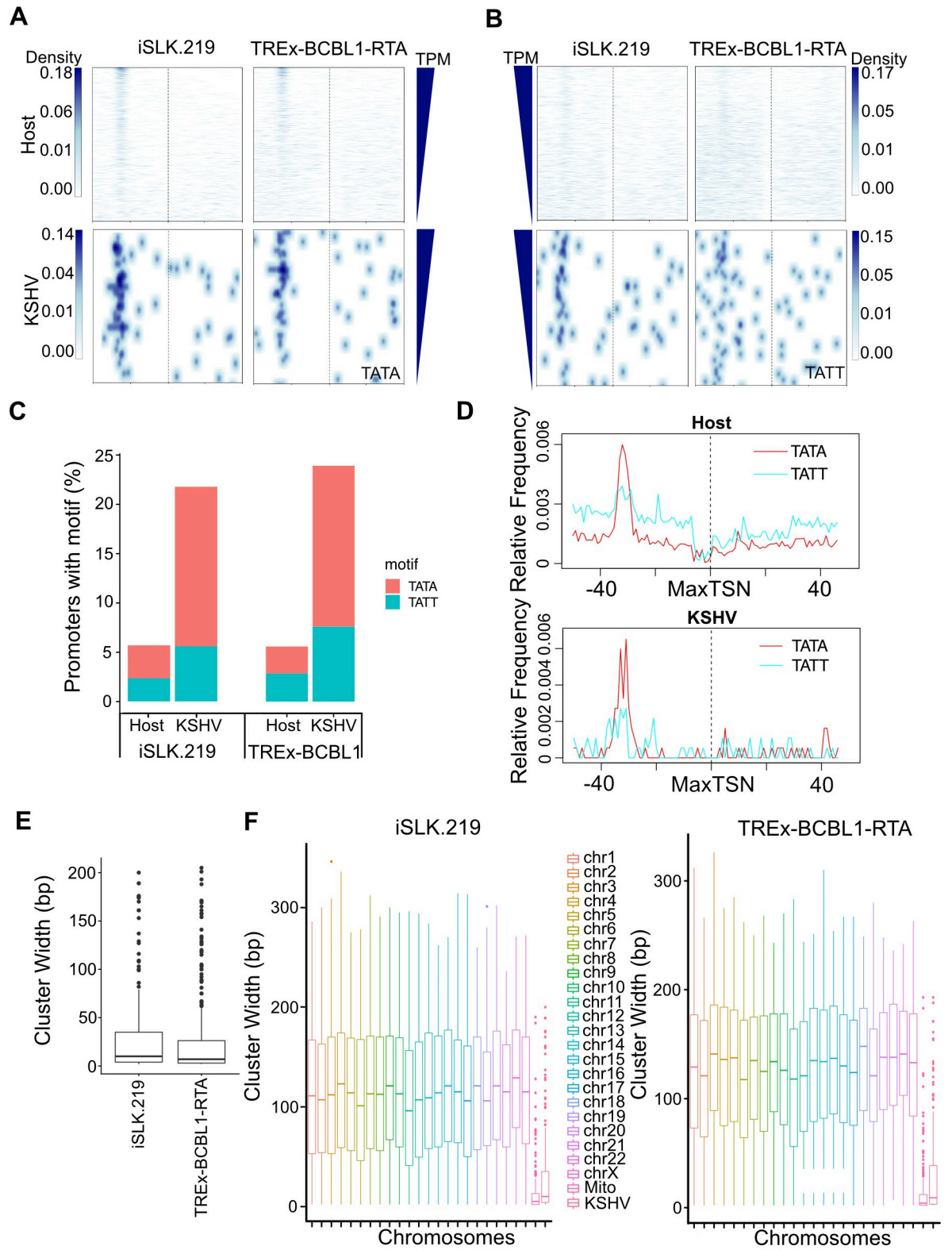


Fig 8. Comparison between host and KSHV cis-element use and TSC architecture. (A) Heatmap showing the TATA box motif density 50 bp up- and downstream of MaxTSN in iSLK.219 and TREx-BCBL1-RTA cells. (B) Heatmap showing the TATT box motif density 50bp up- and downstream of MaxTSN in iSLK.219 and TREx-BCBL1-RTA cells. In both A and B the TSCs were order by the expression value of MaxTSN. (C) Percentage of TSCs with TATA and TATT motifs between 20-31bp upstream of the MaxTSN for host and KSHV. (D)

Frequency of TATA and TATT motifs found relative to MaxTSN of TSCs from host and KSHV in TREx-BCBL1-RTA cell. (E) Cluster width of KSHV TSCs from iSLK.219 and TREx-BCBL1-RTA cells. (F) Cluster width of TSCs identified from host chromosomes and KSHV genome for iSLK.219 and TREx-BCBL1-RTA cells.

<https://doi.org/10.1371/journal.ppat.1007852.g008>

frequency of TATA and TATT motifs 50 bp upstream and downstream of cellular and viral MaxTSNs to search for positional enrichment (Fig 8D). Within viral promoters we identified a clear enrichment for these motifs 40–30 bp upstream of the MaxTSN. Interestingly, a second minor enrichment of the TATT motif can be found 25–20 bp upstream of the MaxTSN. The function significance of this second cluster is unclear. With regard to cellular TSSs, the TATA motif is clearly enriched 35–30 bp upstream of the MaxTSN. Unexpectedly, we also observe a peak of TATT motifs at a similar position within cellular promoters. Moreover, there is a clear depletion of this motif immediately upstream of the cellular MaxTSN. We anticipate this suggests that TATT motifs are engaged by cellular TBP for productive transcription and hypothesize that the viral transcription initiation complex may be similarly capable of regulating a subset of cellular genes due to the presence of its preferred binding motif, TATT.

Viral TSSs exhibit tighter spatial control

The presence of TATA- and Inr-like motifs is primarily associated with tissue-specific or terminally differentiated cell-specific gene expression. Moreover, the presence of these motifs tends to focus transcription initiation events into more well defined clusters. Thus, we compared the breadth of cellular and viral TSCs. The width distribution of viral TSCs is similar between iSLK.219 and TREx-BCBL1-RTA cells (Fig 8E). When compared with the host, we find that the spatial distribution of viral TSCs in both iSLK.219 and TREx-BCBL1-RTA cells is more confined than host TSCs, indicating viral gene expression is under a higher degree of spatial control (Fig 8F).

Discussion

Comprehensive genome annotations are critical for understanding the biology of all organisms, as well as viruses. Moreover, precise annotations of key gene regulatory features such as TSSs and core promoters, provides the opportunity to define mechanistic features of gene expression regulation. Our transcriptome-wide nucleotide resolution mapping of KSHV TSSs has greatly expanded the transcriptional landscape of KSHV and has identified core promoter sequences associated with TSC architecture and regulation. Moreover, the identification of over 100 novel TSCs in both iSLK.219 and TREx-BCBL1-RTA cells highlights the breadth of the KSHV transcriptome that remains uncharacterized. Future studies must be directed towards understanding the functional significance of the newly defined viral RNAs as they may function as noncoding RNAs or encode proteins that impact the KSHV lifecycle and thus KSHV-associated disease progression.

The high-resolution map of TSSs described here more than triples the number of known transcripts that are expressed from the KSHV genome. However, without full-length RNA-sequencing data it is difficult to fully assess whether the novel TSSs contribute to the expression of yet to be identified RNAs, or whether these extend the 5'-UTRs of known mRNAs. Though analyses here indicate that both are true as we observe novel splice junctions associated with newly discovered TSSs, as well as novel TSSs in which all mate pair reads are associated with a known ORF. Along this line, over 300 previously unknown transcripts were recently identified within the EBV transcriptome through the integration of multiple RNA-sequencing approaches, including CAGE-seq and full-length RNA sequencing [61]. While determining the function of novel transcripts will likely require a detailed description of

transcript structure coupled with biochemical assays, we surmise that extension of 5'-UTR sequences impacts gene expression regulation. Indeed, alternative TSS selection has been shown to lead to large differences in translation efficiency of RNA in both yeast and mouse embryonic fibroblasts [62, 63]. Moreover, RNA structures within 5'-UTRs can influence RNA half-life [64, 65]. Thus, in addition to extending the number of RNAs transcribed off the KSHV genome this study also uncovers additional points of entry for post-transcriptional control processes. Furthermore, the use of multiple TSSs to expand the landscape of viral 5'-UTRs appears to be common among viruses as recent studies leveraging CAGE-seq and precision nuclear run-on sequencing (PRO-Seq) have observed multiple TSSs for many EBV and cytomegalovirus (CMV) transcripts [61, 66, 67].

While transcription initiation is often viewed as originating from a single nucleotide, this is an antiquated model. Transcriptome-wide 5'-sequencing methodologies, such as CAGE-seq, PRO-seq, and RAMPAGE, have demonstrated that transcription initiation occurs over a range of nucleotides within promoters [2, 3, 30]. Additionally, low level transcription initiation is also frequently observed within the coding regions of human genes. Given that this has been observed with multiple technologies it is unlikely to be an experimental artifact. Indeed, RAMPAGE analysis on KSHV infected cells reveals that KSHV transcription initiation is similarly structured, with multiple TSSs driving expression of individual genes as well as transcription initiation within protein coding genes. The use of multiple TSSs for an individual protein coding gene enables additional mechanisms of post-transcriptional control of gene expression mediated through the various 5'-UTR sequences. The extent to which KSHV gene expression is regulated via 5'-UTR based mechanisms is unclear, however, our study highlights the need for consideration of these mechanisms when investigating the regulation of viral gene expression.

Using our experimentally defined TSSs we identified unknown similarities and differences in TSS usage between iSLK.219 and TReX-BCBL1-RTA cells. For example, leveraging previously published RTA ChIP-seq data sets we identified three novel TSSs present in both cell types that harbor prominent RTA binding peaks immediately upstream suggesting their expression is regulated by RTA. Moreover, when cloned upstream of a luciferase reporter these genomic fragments confer RTA-inducible luciferase expression. As we noted earlier these data do not conclude that the three novel TSSs are RTA-dependent. However, they do demonstrate the genomic fragments have the potential to drive productive RTA-dependent transcription initiation events. We also identified clear examples where a TSS was specifically used in only one cell-type. For example, in TReX-BCBL1-RTA cells a TSS for ORF58 is identified while in iSLK.219 cells this TSS is absent. While it is possible that that ORF58 is translated from a bicistronic ORF58/59 transcript in iSLK.219 cells, it is equally plausible to hypothesize that ORF58 protein is not expressed there.

Core promoter nucleotide composition is known to influence various aspects of transcription, including the mechanisms of transcription factor recruitment, expression level, and breadth of transcription initiation sites. Leveraging our RAMPAGE mapped TSSs we have defined the nucleotide composition of cellular and viral promoters and defined motifs associated with transcription breadth and strength (Fig 8A, 8B and 8F and S7 Fig). We find that viral promoters more heavily rely on TATA-like motifs than human promoters, and K-means clustering identified a distinct set of viral genes that harbor an Inr motif. TATA-like and Inr motifs are associated with more focused transcription initiation profiles and indeed we discovered that viral promoters drive more focused transcription initiation than cellular promoters. We hypothesize that the high gene density of the KSHV genome, and likely other viruses, necessitates a tighter control on the spatial distribution of transcription initiation events.

During KSHV infection the TATT motif is recognized by the viral TBP mimic ORF24 to facilitate late gene expression [59]. Interestingly, quantification of the TATT motif frequency

within cellular promoters identifies a minor enrichment of this sequence approximately 35–30 bp upstream of cellular promoters. TATT can serve to recruit TBP. For example, the mouse α 4 IFN promoter contains a well-defined TATT motif [68]. Whether cellular genes that contain this motif are bound by the viral initiation complex is not known although it would be very interesting to investigate this. While it would likely be unfavorable for the virus to activate the expression of an interferon via the viral initiation complex, we hypothesize that cellular promoters that harbor TATT motifs contain sequences that recruit additional factors that are required for their expression. Thus, in this scenario, the binding of the viral late gene complex would not necessarily be activating, but rather could sterically inhibit the recruit of the additional factors and inhibit gene expression.

Our study provides evidence for cell-type specific differences in both TSC usage as well as promoter strength. While the iSLK system has proven invaluable for KSHV community, the transcription initiation profile of the recombinant KSHV.219 virus is distinct from that within PEL cells. While we do identify a much boarder and complex transcriptional profile in iSLK.219 cells than previously known, we identify 130 fewer TSSs within iSLK.219 cells, including the lack of TSCs that facilitate expression of novel RNAs. It is important to note that we cannot ascertain with certainty whether the differences are due to the nature of the viruses, as one is a recombinant virus while the other is the virus that initially infected the BCBL1 cells, or whether KSHV gene expression in iSLK cells is fundamentally different than BCBL1 cells. Additional work is needed to resolve this issue. Moreover, it is formally possible that the bioinformatics thresholds for TSC classification result in a missed TSC in one system. However, if this were the case weakly expressed and spatially dispersed TSCs would be the most susceptible. Nonetheless, our data demonstrate that iSLK.219 and TREx-BCBL1-RTA gene expression programs are distinct.

KS is an endothelial cell derived lesion, and in the future, it will be imperative to determine the viral TSS landscape within endothelial cells. Infection of endothelial cells with virus derived from iSLK.219 cells versus virus prepared from TREx-BCBL1-RTA cells should resolve whether the TSCs landscapes observed here are virus or cell-type specific. Moreover, the KSHV gene expression program is also affected by the environment infected cells are cultured in. For example, during conditions of hypoxia a cluster of genes spanning ORF34–37 are expressed [69]. Application of RAMPAGE to KSHV infected cells cultured in different growth conditions will provide a comprehensive view of how KSHV gene expression is remodeled by the environment and may uncover other unknown viral transcripts.

Materials and methods

Cell culture

TREx-BCBL1-RTA (a gift from Dr. Britt Glaunsinger, University of California, Berkeley) was maintained in RPMI 1640 medium (Invitrogen) supplemented with 10% fetal bovine serum (FBS; Invitrogen), 200 μ M L-glutamine, 100 U/ml penicillin/streptomycin. iSLK.219 (a gift from Dr. Britt Glaunsinger, University of California, Berkeley) and Human embryonic kidney 293 T (HEK293T) cells (purchased from ATCC) were grown in Dulbecco's modified Eagle medium (DMEM; Invitrogen) supplemented with 10% FBS. Lytic reactivation of TREx-BCBL1-RTA and iSLK.219 cells was induced by the addition of 1 μ g/ml doxycycline (Fisher Scientific).

Fluorescence in situ hybridization and flow cytometry

Fluorescence in situ hybridization and flow cytometry (FISH-Flow) of PAN RNA in iSLK.219 and TREx-BCBL1-RTA cells was done as previously described [70, 71]. Briefly, cells with

indicated treatment were fixed in 4% (vol/vol) paraformaldehyde for 30 min at RT, wash with PBS-FISH buffer (1XPBS, 0.2 mg/ml RNase-free BSA) for twice, and then permeabilized with 1X PBS containing 0.2% (vol/vol) Tween-20 for another 30 min at RT. The permeabilized cells were then hybridized with Alexa-Fluor 647 labeled PAN anti-sense oligos (PAN-1, 5'-ACAA ATGCCACCTCACTTTGTGCGC-3'; PAN-2, 5'-CGCTGCTTTCCTTTCACATT-3'; PAN-3, 5'-GTGAAGCGGCAGCCAAGGTGACTGG-3') in HB 10% dx buffer (10% (wt/vol) dextran sulfate, 2× saline-sodium citrate (SSC), 10% (vol/vol) formamide, 1 mg/ml tRNA and 0.2 mg/ml BSA) at 37°C overnight. After extensive washing with HBW buffer (2× SSC, 10% (vol/vol) formamide and 0.2 mg/ml RNase-free BSA), cells were analyzed on BD LSR Fortessa instrument. Data were analyzed with FlowJo X software (TreeStar).

RAMPAGE

RAMPAGE was performed as previously described with minor modifications [30]. RNA was isolated from latent or dox-induced iSLK.219 and TREx-BCBL1-RTA cells by TRIzol (Thermo Scientific) according to manufacture instructions. 5 µg of total RNA from each sample was subjected to ribosomal RNA depletion using NEBNext rRNA depletion kit (NEB) followed by Terminator (Epicentre) treatment. The RNA was purified by RNAClean XP (Agencourt) according to the manufacturer's instructions and eluted in 7.5 µl H₂O. 1 µl of reverse-transcription primer (400 µM) and 1 µl of barcoded template-switching oligo (4 mM) were added before denaturation (65°C, 10 min). Ice-cold reverse-transcription mix (7.5 µl Invitrogen first strand buffer, 1.9 µl dNTPs (10 mM), 7.5 µl Sorbitol/Trehalose mix (3.3 M/0.66 M), 1.9 µl DTT (100 mM), 5.6 µl betaine (5 M), 4 µl Invitrogen SuperScript III) was added and the reaction was incubated in a thermal cycler (4°C 10 sec, 22°C 1 min, 42°C 30 min, 75°C 15 min). The RNAs and their associated 5'-complete cDNAs were purified by RNAClean XP, quantified by quantitative PCR and pooled. Pool volumes were adjusted to 40 µl by RNAClean XP precipitation. The oxidization buffer (2 µl 1 M NaOAc pH4.5, 2 µl 250 mM NaIO₄) was added and the samples were incubated on ice in the dark for 45 min. After quenching the reaction with 2 µl 40% glycerol, 14 µl of 1 M Tris-HCl pH8.5 was added prior to RNAClean XP purification. The samples were eluted in 40 µl H₂O. Biotinylation was performed by adding the biotinylation mix (4 µl 1 M NaCitrate pH6.0, 13.5 µl 15 mM biotin hydrazide long arm (Vector Labs)) and incubated at room temperature for 14–15 hours in the dark. For RNase I digest, we added the digestion mix (6 µl 1 M Tris-HCl pH8.5, 1 µl 0.5 M EDTA pH8.0, 5 µl RNase I (Thermo Scientific, 10U/µl)) and incubated the samples for 45 min at 37°C, and then 5 min at 65°C. After RNAClean XP purification, the samples were resuspended in 40 µl H₂O. The biotinylated capped RNAs and their associated 5'-complete cDNAs were captured by streptavidin-coated magnetic beads (Thermo Scientific) for 30 min at room temperature followed by extensive washing: once with wash buffer (WB) 1 (4.5 M NaCl, 50 mM EDTA), once with WB2 (0.3 M NaCl, 1 mM EDTA), twice with WB3 (20 mM Tris-HCl pH8.5, 1 mM EDTA, 0.5 M NaOAc, 0.4% SDS) and twice with WB4 (10 mM Tris-HCl pH8.5, 1 mM EDTA, 0.5 M NaOAc). The cDNAs were eluted by incubation with 65 µl 50 mM NaOH for 10 min at room temperature, and purified with AMPure XP (Agencourt). The cDNAs were then amplified with the KAPA HiFi HotStart PCR kit (KAPA Biosystems). The size selection (~300 to 1000 bp) was conducted with two rounds AMPure XP Cleanup and final libraries were eluted in 20 µl H₂O. The libraries were run on a DNA HS Bioanalyzer chip for quality control, quantified by quantitative PCR, and sequenced on a NextSeq 500 for PE75 run. Raw RAMPAGE sequencing files can be found in GEO accession # GSE129902. Oligonucleotides used in RAMPAGE analysis are in [S2 Table](#).

Bioinformatics analysis, sequence alignments

Data processing was performed according to the original RAMPAGE method with minor modifications. Briefly, raw reads in fastq files were demultiplexed and barcodes removed using cutadapt [72]. Trimmed reads were mapped to the human (hg38 assembly) and KSHV (NCBI accession number GQ994935.1) genomes using STAR (V2.6.1b) [73]. Six non-templated G bases on read 1, which were added by reverse transcriptase during the template switch reaction, and 15 bases of read 2, which come from primers used for reverse transcription, were soft clipped in the alignment process. Read pairs that share similar alignment coordinates and an identical reverse-transcription primer sequence were considered PCR duplications and removed. To assess RAMPAGE specificity we calculated 5'- versus 3'-end coverage of uniquely mapping mate 1 reads using the most 5' nucleotides by deeptools as described in [74]. Reads were mapped to reference transcripts extracted from hg38 and transcripts less than 500 bp were excluded from the analysis [75].

Identification of transcription start sites clusters and annotation

Start site clusters in mapped RAMPAGE data were extracted using Paraclu [5] with the following parameters (i) a minimum of 10 tags/cluster, (ii) (maximum density/baseline density) ≥ 2 and (iii) 1–200 base cluster length. Consensus clusters between replicates were identified using a custom R script, and subsequently merged into megaclusters by overlap. Coordinates of megaclusters were then applied to calculate the number of tags mapping to each consensus cluster in each sample with Rsubread and normalized for sequencing depth, converting tag counts to tags per million mapped reads (TPM). Assignment of megaclusters to promoters was done with NCBI annotation of KSHV genome GQ994935.1 using following metrics in order: I) megaclusters within 500bp upstream of an CDS were assigned to the gene; II) megaclusters overlapping CDS were annotated as CDS_internal; III) otherwise, megaclusters were considered intergenic. TSS clusters found in iSLK.219 and TReX-BCBL1-RTA cells were considered the same if their coordinates overlap.

Motif analysis

For motif analysis, TSCs within 50 bp were merged and the maximum expressed nucleotide (MaxTSN) was identified. Sequences 50 upstream and downstream of the MaxTSN were extracted and subjected to motif analysis and pattern density visualization. DNA motif pattern density and position were visualized by R package seqPattern (R package, <http://bioconductor.org/packages/release/bioc/html/seqPattern.html>), sequences were ordered by MaxTSN expression value (TPM). Transcripts initiating from each of the merged TSS were counted using Rsubread and normalized to the total uniquely mapped reads as TPM (tag per million). The mean expression value of duplicates at each time point were used in the K-means analysis with kmeans function from cluster (R package, <https://svn.r-project.org/R-packages/trunk/cluster>) and visualized with factoextra, the optimal cluster number K was determined by Average Silhouette Method (R package, <http://www.sthda.com/english/rpkgs/factoextra>). The K-means clustered heatmaps were generated with pheatmap (R package, <https://github.com/raivokolde/pheatmap>). Kplogo was used to search for enriched k-mers at each position [76]. De novo analysis was also performed using HOMER script findMotifsGenomics.pl on nucleotide sequences 100bp up- and downstream of MaxTSN using default parameters. hg38 and GQ994935.1 genomes were applied as background [77].

Data acquisition

CTCF ChIP-seq (SRR503420), FAIRE-seq (SRR965330, SRR965331), and RTA ChIP-seq (SRR5676584, SRR5676585, SRR8324529, SRR8324530), LANA (SRR1030378, SRR1030385,

SRR1030386, SRR1030387, SRR1030388), PolII (SRR1030375), H3K27me3 (SRR1030374) and H3K4me3 (SRR1030370, SRR1030371, SRR1030372, SRR1030373) were downloaded from Sequence Read Archive (SRA) use sra-tools (<https://www.ncbi.nlm.nih.gov/sra>). Processed fastq reads were mapped to GQ994935.1 and GRCh38 genome using default STAR parameters, and peaks identified using MACS2 [78].

Cloning

Novel RTA-associated TSSs were synthesized as gBlock gene fragments (IDT). gBlock fragments were amplified using KAPA HiFi DNA polymerase (Kapa Biosystems), digested with XhoI and HindIII and cloned into pGL4.28 (Promega). pFLAG-RTA was a kind gift from Dr. Melanie M. Brinkmann (Helmholtz Centre for Infection Research).

Luciferase assay

HEK293T cells, at 80% confluency, were transfected with 50 ng pGL4.28 luciferase reporter and 450 ng pFlag-RTA (kindly provided by Dr. Kendra Bussey and Dr. Melanie Brinkmann, Helmholtz Centre for Infection Research) plasmid using PolyJet In Vitro DNA Transfection Reagent (SignaGen Laboratories). Cells were collected and used to measure firefly luciferase activity using luciferase assay system (Promega) on a GloMax 20/20 Luminometer (Promega).

Supporting information

S1 Fig. RAMPAGE signal over 10 previously defined TSSs.
(TIFF)

S2 Fig. Comparison of RAMPAGE defined TSCs to other ChIP-seq datasets.
(TIFF)

S3 Fig. Number of reads mapping to the KSHV genome in iSLK.219 and TREx-BCBL1-RTA cells.
(TIFF)

S4 Fig. Luciferase assay for internal ORF6, internal ORF32 and As_miR RTA binding elements.
(TIFF)

S5 Fig. Comparison of LANA2 RAMPAGE 5' signal in iSLK.219 and TREx-BCBL1-RTA cells.
(TIFF)

S6 Fig. Heatmap based on reads mapping to known ORFs.
(TIFF)

S7 Fig. Pattern density of Inr and GG around MaxTSN (-50~+50).
(TIFF)

S1 Table. Transcription start sites clusters identified in iSLK.219 and TREx-BCBL1-RTA cells.
(XLSX)

S2 Table. Oligos used for RAMPAGE and cloning.
(XLSX)

Acknowledgments

We would like to thank members of the Karijovich lab for insightful discussion, and Dr. Melanie M. Brinkmann (Helmholtz Centre for Infection Research) for providing the pFLAG-RTA construct.

Author Contributions

Conceptualization: John Karijovich.

Data curation: Xiang Ye.

Formal analysis: Xiang Ye.

Funding acquisition: John Karijovich.

Investigation: Xiang Ye, Yang Zhao.

Methodology: Xiang Ye, Yang Zhao, John Karijovich.

Supervision: John Karijovich.

Validation: Xiang Ye, Yang Zhao.

Writing – original draft: Xiang Ye, Yang Zhao, John Karijovich.

Writing – review & editing: Xiang Ye, Yang Zhao, John Karijovich.

References

- Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol.* 2018; 19(10):621–37. <https://doi.org/10.1038/s41580-018-0028-8> PMID: 29946135; PubMed Central PMCID: PMC6205604.
- Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, et al. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A.* 2003; 100(26):15776–81. Epub 2003/12/08. <https://doi.org/10.1073/pnas.2136655100> PMID: 14663149; PubMed Central PMCID: PMC307644.
- Kawaji H, Kasukawa T, Forrest A, Carninci P, Hayashizaki Y. The FANTOM5 collection, a data series underpinning mammalian transcriptome atlases in diverse cell types. *Sci Data.* 2017; 4:170113. Epub 2017/08/29. <https://doi.org/10.1038/sdata.2017.113> PMID: 28850107; PubMed Central PMCID: PMC5574373.
- Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet.* 2006; 38(6):626–35. Epub 2006/04/28. <https://doi.org/10.1038/ng1789> PMID: 16645617.
- Frith MC, Valen E, Krogh A, Hayashizaki Y, Carninci P, Sandelin A. A code for transcription initiation in mammalian genomes. *Genome Res.* 2008; 18(1):1–12. Epub 2007/11/21. <https://doi.org/10.1101/gr.6831208> PMID: 18032727; PubMed Central PMCID: PMC2134772.
- Hoskins RA, Landolin JM, Brown JB, Sandler JE, Takahashi H, Lassmann T, et al. Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. *Genome Res.* 2011; 21(2):182–92. Epub 2010/12/22. <https://doi.org/10.1101/gr.112466.110> PMID: 21177961; PubMed Central PMCID: PMC3032922.
- Ni T, Corcoran DL, Rach EA, Song S, Spana EP, Gao Y, et al. A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat Methods.* 2010; 7(7):521–7. Epub 2010/05/23. <https://doi.org/10.1038/nmeth.1464> PMID: 20495556; PubMed Central PMCID: PMC3197272.
- Lenhard B, Sandelin A, Carninci P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet.* 2012; 13(4):233–45. Epub 2012/03/06. <https://doi.org/10.1038/nrg3163> PMID: 22392219.
- Akalin A, Fredman D, Arner E, Dong X, Bryne JC, Suzuki H, et al. Transcriptional features of genomic regulatory blocks. *Genome Biol.* 2009; 10(4):R38. Epub 2009/04/19. <https://doi.org/10.1186/gb-2009-10-4-r38> PMID: 19374772; PubMed Central PMCID: PMC2688929.

10. Schor IE, Degner JF, Harnett D, Cannavò E, Casale FP, Shim H, et al. Promoter shape varies across populations and affects promoter evolution and expression noise. *Nat Genet.* 2017; 49(4):550–8. Epub 2017/02/13. <https://doi.org/10.1038/ng.3791> PMID: 28191888.
11. FitzGerald PC, Shlyakhtenko A, Mir AA, Vinson C. Clustering of DNA sequences in human promoters. *Genome Res.* 2004; 14(8):1562–74. Epub 2004/07/15. <https://doi.org/10.1101/gr.1953904> PMID: 15256515; PubMed Central PMCID: PMCPMC509265.
12. Suzuki Y, Taira H, Tsunoda T, Mizushima-Sugano J, Sese J, Hata H, et al. Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites. *EMBO Rep.* 2001; 2(5):388–93. <https://doi.org/10.1093/embo-reports/kve085> PMID: 11375929; PubMed Central PMCID: PMCPMC1083880.
13. Cooper SJ, Trinklein ND, Anton ED, Nguyen L, Myers RM. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* 2006; 16(1):1–10. Epub 2005/12/12. <https://doi.org/10.1101/gr.4222606> PMID: 16344566; PubMed Central PMCID: PMCPMC1356123.
14. FitzGerald PC, Sturgill D, Shyakhtenko A, Oliver B, Vinson C. Comparative genomics of *Drosophila* and human core promoters. *Genome Biol.* 2006; 7(7):R53. <https://doi.org/10.1186/gb-2006-7-7-r53> PMID: 16827941; PubMed Central PMCID: PMCPMC1779564.
15. Ohler U. Identification of core promoter modules in *Drosophila* and their application in accurate transcription start site prediction. *Nucleic Acids Res.* 2006; 34(20):5943–50. Epub 2006/10/26. <https://doi.org/10.1093/nar/gkl608> PMID: 17068082; PubMed Central PMCID: PMCPMC1635271.
16. Boshoff C, Weiss R. AIDS-related malignancies. *Nat Rev Cancer.* 2002; 2(5):373–82. <https://doi.org/10.1038/nrc797> PMID: 12044013.
17. Dittmer DP, Damania B. Kaposi sarcoma-associated herpesvirus: immunobiology, oncogenesis, and therapy. *J Clin Invest.* 2016; 126(9):3165–75. Epub 2016/09/01. <https://doi.org/10.1172/JCI84418> PMID: 27584730; PubMed Central PMCID: PMCPMC5004954.
18. Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, et al. A review of human carcinogens—Part B: biological agents. *Lancet Oncol.* 2009; 10(4):321–2. PMID: 19350698.
19. Moore PS, Chang Y. Why do viruses cause cancer? Highlights of the first century of human tumour virology. *Nat Rev Cancer.* 2010; 10(12):878–89. Epub 2010/11/24. <https://doi.org/10.1038/nrc2961> PMID: 21102637; PubMed Central PMCID: PMCPMC3718018.
20. Parkin DM. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer.* 2006; 118(12):3030–44. <https://doi.org/10.1002/ijc.21731> PMID: 16404738.
21. Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science.* 1994; 266(5192):1865–9. PMID: 7997879.
22. Kedes DH, Operskalski E, Busch M, Kohn R, Flood J, Ganem D. The seroepidemiology of human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus): distribution of infection in KS risk groups and evidence for sexual transmission. *Nat Med.* 1996; 2(8):918–24. PMID: 8705863.
23. Moore PS, Chang Y. Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma in patients with and those without HIV infection. *N Engl J Med.* 1995; 332(18):1181–5. <https://doi.org/10.1056/NEJM199505043321801> PMID: 7700310.
24. Cancian L, Hansen A, Boshoff C. Cellular origin of Kaposi's sarcoma and Kaposi's sarcoma-associated herpesvirus-induced cell reprogramming. *Trends Cell Biol.* 2013; 23(9):421–32. Epub 2013/05/17. <https://doi.org/10.1016/j.tcb.2013.04.001> PMID: 23685018.
25. Cesarman E, Chang Y, Moore PS, Said JW, Knowles DM. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas. *N Engl J Med.* 1995; 332(18):1186–91. <https://doi.org/10.1056/NEJM199505043321802> PMID: 7700311.
26. Soulier J, Grollet L, Oksenhendler E, Cacoub P, Cazals-Hatem D, Babinet P, et al. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castlemann's disease. *Blood.* 1995; 86(4):1276–80. PMID: 7632932.
27. Lukac DM, Kirshner JR, Ganem D. Transcriptional activation by the product of open reading frame 50 of Kaposi's sarcoma-associated herpesvirus is required for lytic viral reactivation in B cells. *J Virol.* 1999; 73(11):9348–61. PMID: 10516043; PubMed Central PMCID: PMCPMC112969.
28. Sun R, Lin SF, Gradoville L, Yuan Y, Zhu F, Miller G. A viral gene that activates lytic cycle expression of Kaposi's sarcoma-associated herpesvirus. *Proc Natl Acad Sci U S A.* 1998; 95(18):10866–71. <https://doi.org/10.1073/pnas.95.18.10866> PMID: 9724796; PubMed Central PMCID: PMCPMC27987.
29. 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.* 2014; 10(1):e1003847. Epub

- 2014/01/16. <https://doi.org/10.1371/journal.ppat.1003847> PMID: 24453964; PubMed Central PMCID: PMC3894221.
30. Batut P, Dobin A, Plessy C, Carninci P, Gingeras TR. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome Res.* 2013; 23(1):169–80. Epub 2012/08/30. <https://doi.org/10.1101/gr.139618.112> PMID: 22936248; PubMed Central PMCID: PMC3530677.
 31. Myoung J, Ganem D. Generation of a doxycycline-inducible KSHV producer cell line of endothelial origin: maintenance of tight latency with efficient reactivation upon induction. *J Virol Methods.* 2011; 174(1–2):12–21. Epub 2011/03/17. <https://doi.org/10.1016/j.jviromet.2011.03.012> PMID: 21419799; PubMed Central PMCID: PMC3095772.
 32. Vieira J, O'Hearn PM. Use of the red fluorescent protein as a marker of Kaposi's sarcoma-associated herpesvirus lytic gene expression. *Virology.* 2004; 325(2):225–40. <https://doi.org/10.1016/j.virol.2004.03.049> PMID: 15246263.
 33. Nakamura H, Lu M, Gwack Y, Souvlis J, Zeichner SL, Jung JU. Global changes in Kaposi's sarcoma-associated virus gene expression patterns following expression of a tetracycline-inducible Rta transactivator. *J Virol.* 2003; 77(7):4205–20. <https://doi.org/10.1128/JVI.77.7.4205-4220.2003> PMID: 12634378; PubMed Central PMCID: PMC3095772.
 34. Stürzl M, Gaus D, Dirks WG, Ganem D, Jochmann R. Kaposi's sarcoma-derived cell line SLK is not of endothelial origin, but is a contaminant from a known renal carcinoma cell line. *Int J Cancer.* 2013; 132(8):1954–8. Epub 2012/10/12. <https://doi.org/10.1002/ijc.27849> PMID: 22987579.
 35. Tang S, Zheng ZM. Kaposi's sarcoma-associated herpesvirus K8 exon 3 contains three 5'-splice sites and harbors a K8.1 transcription start site. *J Biol Chem.* 2002; 277(17):14547–56. Epub 2002/02/08. <https://doi.org/10.1074/jbc.M111308200> PMID: 11832484.
 36. Lin SF, Robinson DR, Miller G, Kung HJ. Kaposi's sarcoma-associated herpesvirus encodes a bZIP protein with homology to BZLF1 of Epstein-Barr virus. *J Virol.* 1999; 73(3):1909–17. Epub 1999/02/11. PMID: 9971770; PubMed Central PMCID: PMC104432.
 37. Majerciak V, Yamanegi K, Zheng ZM. Gene structure and expression of Kaposi's sarcoma-associated herpesvirus ORF56, ORF57, ORF58, and ORF59. *J Virol.* 2006; 80(24):11968–81. Epub 2006/10/06. <https://doi.org/10.1128/JVI.01394-06> PMID: 17020939; PubMed Central PMCID: PMC1676266.
 38. Song MJ, Brown HJ, Wu TT, Sun R. Transcription activation of polyadenylated nuclear rna by rta in human herpesvirus 8/Kaposi's sarcoma-associated herpesvirus. *J Virol.* 2001; 75(7):3129–40. Epub 2001/03/10. <https://doi.org/10.1128/JVI.75.7.3129-3140.2001> PMID: 11238840; PubMed Central PMCID: PMC114107.
 39. Deng H, Song MJ, Chu JT, Sun R. Transcriptional regulation of the interleukin-6 gene of human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus). *J Virol.* 2002; 76(16):8252–64. Epub 2002/07/23. <https://doi.org/10.1128/JVI.76.16.8252-8264.2002> PMID: 12134031; PubMed Central PMCID: PMC155161.
 40. Pearce M, Matsumura S, Wilson AC. Transcripts encoding K12, v-FLIP, v-cyclin, and the microRNA cluster of Kaposi's sarcoma-associated herpesvirus originate from a common promoter. *J Virol.* 2005; 79(22):14457–64. Epub 2005/10/29. <https://doi.org/10.1128/JVI.79.22.14457-14464.2005> PMID: 16254382; PubMed Central PMCID: PMC1280212.
 41. Bowser BS, DeWire SM, Damania B. Transcriptional regulation of the K1 gene product of Kaposi's sarcoma-associated herpesvirus. *J Virol.* 2002; 76(24):12574–83. Epub 2002/11/20. <https://doi.org/10.1128/JVI.76.24.12574-12583.2002> PMID: 12438583; PubMed Central PMCID: PMC136681.
 42. Spiller OB, Robinson M, O'Donnell E, Milligan S, Morgan BP, Davison AJ, et al. Complement regulation by Kaposi's sarcoma-associated herpesvirus ORF4 protein. *J Virol.* 2003; 77(1):592–9. Epub 2002/12/13. <https://doi.org/10.1128/JVI.77.1.592-599.2003> PMID: 12477863; PubMed Central PMCID: PMC140610.
 43. Matsumura S, Fujita Y, Gomez E, Tanese N, Wilson AC. Activation of the Kaposi's sarcoma-associated herpesvirus major latency locus by the lytic switch protein RTA (ORF50). *J Virol.* 2005; 79(13):8493–505. Epub 2005/06/16. <https://doi.org/10.1128/JVI.79.13.8493-8505.2005> PMID: 15956592; PubMed Central PMCID: PMC1143749.
 44. Wong EL, Damania B. Transcriptional regulation of the Kaposi's sarcoma-associated herpesvirus K15 gene. *J Virol.* 2006; 80(3):1385–92. Epub 2006/01/18. <https://doi.org/10.1128/JVI.80.3.1385-1392.2006> PMID: 16415016; PubMed Central PMCID: PMC1346939.
 45. Taylor JL, Bennett HN, Snyder BA, Moore PS, Chang Y. Transcriptional analysis of latent and inducible Kaposi's sarcoma-associated herpesvirus transcripts in the K4 to K7 region. *J Virol.* 2005; 79(24):15099–106. Epub 2005/11/25. <https://doi.org/10.1128/JVI.79.24.15099-15106.2005> PMID: 16306581; PubMed Central PMCID: PMC1315995.

46. Saveliev A, Zhu F, Yuan Y. Transcription mapping and expression patterns of genes in the major immediate-early region of Kaposi's sarcoma-associated herpesvirus. *Virology*. 2002; 299(2):301–14. Epub 2002/08/31. PMID: [12202233](#).
47. Chen L, Park MS. Identification and characterization of the promoter region of Kaposi's sarcoma-associated herpesvirus ORF11. *Virus Res*. 2009; 142(1–2):160–8. Epub 2009/05/12. <https://doi.org/10.1016/j.virusres.2009.02.006> PMID: [19428749](#).
48. Chen J, Ueda K, Sakakibara S, Okuno T, Yamanishi K. Transcriptional regulation of the Kaposi's sarcoma-associated herpesvirus viral interferon regulatory factor gene. *J Virol*. 2000; 74(18):8623–34. Epub 2000/08/23. <https://doi.org/10.1128/jvi.74.18.8623-8634.2000> PMID: [10954564](#); PubMed Central PMCID: [PMC116375](#).
49. Masa SR, Lando R, Sarid R. Transcriptional regulation of the open reading frame 35 encoded by Kaposi's sarcoma-associated herpesvirus. *Virology*. 2008; 371(1):14–31. Epub 2007/10/30. <https://doi.org/10.1016/j.virol.2007.08.023> PMID: [17963810](#).
50. Wang SS, Chen LW, Chen LY, Tsai HH, Shih YC, Yang CT, et al. Transcriptional regulation of the ORF61 and ORF60 genes of Kaposi's sarcoma-associated herpesvirus. *Virology*. 2010; 397(2):311–21. Epub 2009/12/09. <https://doi.org/10.1016/j.virol.2009.11.031> PMID: [19969319](#).
51. Wang SS, Chang PJ, Chen LW, Chen LY, Hung CH, Liou JY, et al. Positive and negative regulation in the promoter of the ORF46 gene of Kaposi's sarcoma-associated herpesvirus. *Virus Res*. 2012; 165(2):157–69. Epub 2012/03/01. <https://doi.org/10.1016/j.virusres.2012.02.010> PMID: [22366521](#).
52. 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. *Pathogens*. 2017;6(1). Epub 2017/03/19. <https://doi.org/10.3390/pathogens6010011> PMID: [28335496](#); PubMed Central PMCID: [PMC5371899](#).
53. Cai X, Cullen BR. Transcriptional origin of Kaposi's sarcoma-associated herpesvirus microRNAs. *J Virol*. 2006; 80(5):2234–42. Epub 2006/02/14. <https://doi.org/10.1128/JVI.80.5.2234-2242.2006> PMID: [16474131](#); PubMed Central PMCID: [PMC1395403](#).
54. Hilton IB, Simon JM, Lieb JD, Davis IJ, Damania B, Dittmer DP. The open chromatin landscape of Kaposi's sarcoma-associated herpesvirus. *J Virol*. 2013; 87(21):11831–42. Epub 2013/08/28. <https://doi.org/10.1128/JVI.01685-13> PMID: [23986576](#); PubMed Central PMCID: [PMC3807352](#).
55. Toth Z, Maglinte DT, Lee SH, Lee HR, Wong LY, Brulois KF, et al. Epigenetic analysis of KSHV latent and lytic genomes. *PLoS Pathog*. 2010; 6(7):e1001013. Epub 2010/07/22. <https://doi.org/10.1371/journal.ppat.1001013> PMID: [20661424](#); PubMed Central PMCID: [PMC2908616](#).
56. Rivas C, Thlick AE, Parravicini C, Moore PS, Chang Y. Kaposi's sarcoma-associated herpesvirus LANA2 is a B-cell-specific latent viral protein that inhibits p53. *J Virol*. 2001; 75(1):429–38. <https://doi.org/10.1128/JVI.75.1.429-438.2001> PMID: [11119611](#); PubMed Central PMCID: [PMC113935](#).
57. Tome JM, Tippens ND, Lis JT. Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nat Genet*. 2018; 50(11):1533–41. Epub 2018/10/22. <https://doi.org/10.1038/s41588-018-0234-5> PMID: [30349116](#); PubMed Central PMCID: [PMC6422046](#).
58. Vo Ngoc L, Cassidy CJ, Huang CY, Duttke SH, Kadonaga JT. The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev*. 2017; 31(1):6–11. Epub 2017/01/20. <https://doi.org/10.1101/gad.293837.116> PMID: [28108474](#); PubMed Central PMCID: [PMC5287114](#).
59. 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*. 2015; 57(2):349–60. Epub 2014/12/24. <https://doi.org/10.1016/j.molcel.2014.11.026> PMID: [25544563](#); PubMed Central PMCID: [PMC4305015](#).
60. Wong-Ho E, Wu TT, Davis ZH, Zhang B, Huang J, Gong H, et al. Unconventional sequence requirement for viral late gene core promoters of murine gammaherpesvirus 68. *J Virol*. 2014; 88(6):3411–22. Epub 2014/01/08. <https://doi.org/10.1128/JVI.01374-13> PMID: [24403583](#); PubMed Central PMCID: [PMC3957950](#).
61. O'Grady T, Wang X, Höner Zu Bentrup K, Baddoo M, Concha M, Flemington EK. Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucleic Acids Res*. 2016; 44(18):e145. Epub 2016/07/12. <https://doi.org/10.1093/nar/gkw629> PMID: [27407110](#); PubMed Central PMCID: [PMC5062983](#).
62. Rojas-Duran MF, Gilbert WV. Alternative transcription start site selection leads to large differences in translation activity in yeast. *RNA*. 2012; 18(12):2299–305. Epub 2012/10/25. <https://doi.org/10.1261/ma.035865.112> PMID: [23105001](#); PubMed Central PMCID: [PMC3504680](#).
63. Wang X, Hou J, Quedenau C, Chen W. Pervasive isoform-specific translational regulation via alternative transcription start sites in mammals. *Mol Syst Biol*. 2016; 12(7):875. Epub 2016/07/18. <https://doi.org/10.1525/msb.20166941> PMID: [27430939](#); PubMed Central PMCID: [PMC4965872](#).

64. Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science*. 2016; 352(6292):1413–6. <https://doi.org/10.1126/science.aad9868> PMID: 27313038.
65. Ringnér M, Krogh M. Folding free energies of 5'-UTRs impact post-transcriptional regulation on a genomic scale in yeast. *PLoS Comput Biol*. 2005; 1(7):e72. Epub 2005/12/09. <https://doi.org/10.1371/journal.pcbi.0010072> PMID: 16355254; PubMed Central PMCID: PMCPMC1309706.
66. Djavadian R, Hayes M, Johannsen E. CAGE-seq analysis of Epstein-Barr virus lytic gene transcription: 3 kinetic classes from 2 mechanisms. *PLoS Pathog*. 2018; 14(6):e1007114. Epub 2018/06/04. <https://doi.org/10.1371/journal.ppat.1007114> PMID: 29864140; PubMed Central PMCID: PMCPMC6005644.
67. Parida M, Nilson KA, Li M, Ball CB, Fuchs HA, Lawson CK, et al. Nucleotide Resolution Comparison of Transcription of Human Cytomegalovirus and Host Genomes Reveals Universal Use of RNA Polymerase II Elongation Control Driven by Dissimilar Core Promoter Elements. *MBio*. 2019; 10(1). Epub 2019/02/12. <https://doi.org/10.1128/mBio.02047-18> PMID: 30755505; PubMed Central PMCID: PMCPMC6372792.
68. Perera LP. The TATA motif specifies the differential activation of minimal promoters by varicella zoster virus immediate-early regulatory protein IE62. *J Biol Chem*. 2000; 275(1):487–96. <https://doi.org/10.1074/jbc.275.1.487> PMID: 10617643.
69. Haque M, Wang V, Davis DA, Zheng ZM, Yarchoan R. Genetic organization and hypoxic activation of the Kaposi's sarcoma-associated herpesvirus ORF34-37 gene cluster. *J Virol*. 2006; 80(14):7037–51. <https://doi.org/10.1128/JVI.00553-06> PMID: 16809309; PubMed Central PMCID: PMCPMC1489055.
70. Arrigucci R, Bushkin Y, Radford F, Lakehal K, Vir P, Pine R, et al. FISH-Flow, a protocol for the concurrent detection of mRNA and protein in single cells using fluorescence in situ hybridization and flow cytometry. *Nat Protoc*. 2017; 12(6):1245–60. Epub 2017/05/19. <https://doi.org/10.1038/nprot.2017.039> PMID: 28518171; PubMed Central PMCID: PMCPMC5548662.
71. Zhao Y, Ye X, Dunker W, Song Y, Karijolich J. RIG-I like receptor sensing of host RNAs facilitates the cell-intrinsic immune response to KSHV infection. *Nat Commun*. 2018; 9(1):4841. Epub 2018/11/19. <https://doi.org/10.1038/s41467-018-07314-7> PMID: 30451863; PubMed Central PMCID: PMCPMC6242832.
72. Martin M. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet journal*. 2011.
73. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013; 29(1):15–21. Epub 2012/10/25. <https://doi.org/10.1093/bioinformatics/bts635> PMID: 23104886; PubMed Central PMCID: PMCPMC3530905.
74. Adiconis X, Haber AL, Simmons SK, Levy Moonshine A, Ji Z, Busby MA, et al. Comprehensive comparative analysis of 5'-end RNA-sequencing methods. *Nat Methods*. 2018; 15(7):505–11. Epub 2018/06/06. <https://doi.org/10.1038/s41592-018-0014-2> PMID: 29867192; PubMed Central PMCID: PMCPMC6075671.
75. Ramirez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014; 42(Web Server issue):W187–91. Epub 2014/05/05. <https://doi.org/10.1093/nar/gku365> PMID: 24799436; PubMed Central PMCID: PMCPMC4086134.
76. Wu X, Bartel DP. kpLogo: positional k-mer analysis reveals hidden specificity in biological sequences. *Nucleic Acids Res*. 2017; 45(W1):W534–W8. <https://doi.org/10.1093/nar/gkx323> PMID: 28460012; PubMed Central PMCID: PMCPMC5570168.
77. 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*. 2010; 38(4):576–89. <https://doi.org/10.1016/j.molcel.2010.05.004> PMID: 20513432; PubMed Central PMCID: PMCPMC2898526.
78. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008; 9(9):R137. Epub 2008/09/17. <https://doi.org/10.1186/gb-2008-9-9-r137> PMID: 18798982; PubMed Central PMCID: PMCPMC2592715.