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 [a, b] and Britt Glaunsinger [a, b]

[a] Department of Plant & Microbial Biology, University of California Berkeley, CA, USA

[b] Howard Hughes Medical Institute, Berkeley, CA, USA

Corresponding Author: glaunsinger@berkeley.edu

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. β and γ-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.

24

25 **Author summary**

26 Gene expression in DNA viruses often occurs in temporal waves, with expression of
27 essential structural proteins occurring late in infection, after viral genome replication has begun.
28 Strategies underlying expression of these viral late genes are often sophisticated; for example, the
29 β- and γ-herpesviruses encode a six-component viral complex that directs late gene transcription,
30 largely by unknown mechanisms. Here, we evaluated how this complex specifically recognizes
31 late promoters during infection with the oncogenic human γ-herpesvirus Kaposi's sarcoma-
32 associated herpesvirus (KSHV). We found that one of the components of the late transcription
33 complex was required for robust viral DNA replication and binds at the origin of replication,
34 suggesting new links between KSHV replication and transcription. Combined measurements of
35 late gene expression and promoter occupancy then revealed which KHSV genes are directly
36 controlled by the late gene transcription complex, leading to identification of a key new regulatory
37 element in KSHV late promoters. Together, these data help explain how the late gene transcription
38 complex is able to bind seemingly minimal promoters with high specificity, ensuring robust
39 expression of viral factors necessary for assembly of progeny virions.

40

41 **Introduction**

42 A conserved feature of double-stranded DNA viruses is that they possess a set of genes that
43 become strongly activated only following viral DNA replication (1). These "late genes" generally
44 encode factors involved in viral assembly that are essential for virion production. The mechanisms
45 underlying their transcriptional regulation are often non-canonical or distinct from those involved
46 in activating either cellular genes or early viral genes. For example, baculoviruses encode a

47 multisubunit, α-amanitin-resistant DNA-directed RNA polymerase dedicated specifically to the
48 activation of compact late promoters (2). Similarly, bacteriophage T7 encodes its own late gene
49 specific RNA polymerase that recognizes unique promoter elements to transcribe late genes (3,4).
50 Adenoviruses regulate late gene expression using a complex combination of viral and cellular
51 factors; beyond promoter activation, these also regulate transcription termination and contribute
52 to differential posttranscriptional processing of late viral mRNA (5).

53 Several studies suggest that the expression of late genes is also uniquely regulated in β- and γ-
54 herpesviruses, which include human cytomegalovirus (HCMV), Epstein-Barr virus (EBV), and
55 Kaposi's sarcoma-associated herpesvirus (KSHV) (6). Unlike other eukaryotic dsDNA viruses, β
56 and γ herpesviruses do not engage cellular TATA binding protein (TBP) during late gene pre-
57 initiation complex formation (7). Their core late gene promoters are unusually small (12-15 base
58 pairs (bp)), with the only recognizable conserved element being a modified TATA box (TATT)
59 ~30 bp upstream of the transcription start site (TSS) (8–10). This is in contrast to eukaryotic and
60 most early herpesviral promoters, which have multiple promoter elements that are important for
61 promoter recognition and activation by RNA polymerase II (RNA pol II) (11). Another notable
62 feature is that activation of these minimal promoters requires a virally-encoded six-protein
63 complex called the viral Pre-Initiation Complex (vPIC) that coordinates mammalian RNA pol II
64 (12,13). The vPIC is presumed to be uniquely adapted to activate minimal promoters, yet the
65 mechanisms underlying its activity and specificity remain largely unknown.

66 The core component of the KSHV vPIC is ORF24 (BcRF1 in EBV and UL87 in HCMV), a
67 viral TBP mimic that displays putative structural similarity to cellular TBP (7,14,15). Unlike
68 cellular TBP, the viral TBP mimics have expanded functionality in that they bind RNA pol II in
69 addition to promoter DNA, making them unique among transcription factors that function in

70 eukaryotic cells. Beyond ORF24, the other essential components of the KSHV vPIC are ORFs 18,
71 30, 31, 34, and 66, but little is known about how each of these factors contributes to vPIC activity
72 (7,16–18). ORF34 interacts with multiple components of the vPIC, and is thus thought to function
73 as a scaffolding factor (13). The ORF18 homolog in HCMV (pUL79) has been shown to function
74 as an elongation factor, although whether it plays a similar role in KSHV and EBV is unclear (19).
75 Despite the paucity of functional information on the individual vPIC components, protein-protein
76 interaction mapping has provided insight into the overall complex organization (13,18).
77 Furthermore, point mutants that disrupt individual inter-subunit vPIC interactions abrogate late
78 gene transcription, highlighting the fact that vPIC complex integrity is critical for function
79 (13,20,21).

80 The mechanistic basis for promoter selectivity has also been challenging to grasp in the context
81 of the minimal, largely nondescript nature of late gene promoters. While a modified TATA box
82 with a TATT sequence is enriched in several late promoters of KSHV, MHV68, and EBV (8–
83 10,22), the ORF24 TBP mimic shows partial binding to a mutant promoter in which the TATT
84 sequence is modified to resemble a canonical TATA box (8). However, ORF24 does not bind
85 promoters that inherently have a canonical TATA box (8), which suggests that other DNA
86 elements surrounding the TATT sequence in late gene promoters contribute to ORF24 binding or
87 vPIC assembly on the DNA.

88 One possible reason that additional DNA elements involved in vPIC promoter recognition have
89 not been identified is that we do not yet know the subset of viral promoters under the direct
90 influence of the vPIC. Techniques such as RNA Sequencing (RNA Seq) and microarrays have
91 helped define the set of genes that are expressed with late kinetics and in a manner dependent on
92 viral DNA replication (23–26). Furthermore, viral mutagenesis and knockdown studies have

93 revealed viral genes whose expression is impaired in the absence of vPIC components
94 (17,18,22,27,28). However, these assays cannot distinguish direct vPIC targets from indirect
95 targets, such as genes regulated by factors that are themselves dependent on the vPIC. Inclusion
96 of such indirect targets in the pool of vPIC-regulated promoters could confound the identification
97 of conserved elements important for direct vPIC regulation.

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

112

113 **Results**

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

115 Three leucine residues in the N-terminal domain of KSHV ORF24 (amino acids 73-75) are
116 essential for its interaction with RNA Pol II and its ability to promote late gene transcription (7).
117 Given that most viral proteins are multifunctional, we engineered the ORF24$_{RLLLG->RAAAG}$
118 mutation (ORF24$_{RAAAG}$) into KSHV BAC16 to precisely evaluate the impact of the transcriptional
119 function of ORF24 during lytic infection. We also generated the corresponding mutant rescue
120 (ORF24$_{RAAAG}$-MR) to ensure that any observed phenotypes were not the effect of secondary
121 mutations in the BAC. Identical banding patterns from *Rsr*II digestion of WT and mutant BAC16
122 DNAs verified that there were no gross rearrangements of the viral genome during mutagenesis
123 (Figure S1-lanes1-3). We established stable cell lines latently infected with WT, ORF24$_{RAAAG}$,
124 and ORF24$_{RAAAG}$-MR KSHV in the iSLK background, which contains a doxycycline-inducible
125 version of the KSHV lytic transactivator protein ORF50 (RTA) to enable lytic reactivation (29).
126 During lytic reactivation, these three viruses expressed similar levels of the early protein ORF59,
127 while the ORF24$_{RAAAG}$ mutant was selectively unable to express the K8.1 late protein or produce
128 infectious virions (Fig 1A, 1B). These findings are consistent with previous complementation data
129 (7), and confirm the importance of the ORF24-Pol II interaction for late gene expression.
130 Surprisingly, we also observed ~6-fold defect in DNA replication in the ORF24$_{RAAAG}$ mutant
131 (Fig 1C), which was unexpected given that previous studies with mutations in ORF24 or other
132 components of the vPIC did not show this defect (7,13,16–18). This defect was rescued in the
133 ORF24$_{RAAAG}$-MR cell line, indicating that it was not due to secondary mutations in the BAC (Fig.
134 1C). However, to independently confirm this observation, we engineered two additional ORF24
135 viral mutants: the point mutant ORF24$_{R328A}$, which abrogates its essential interaction with the
136 ORF34 component of the vPIC (13), and a stop mutant (ORF24$_{Stop}$) that has a premature stop
137 codon in the ORF24 coding sequence. Inclusion of the ORF24$_{Stop}$ virus allowed us to evaluate

138 whether the point mutants were functioning in a dominant negative manner, for example through
139 nonproductive DNA binding. Finally, we also N-terminally tagged endogenous WT ORF24 with
140 HA (HA-ORF24) to assess whether any alterations to this locus impacted viral DNA replication.
141 Notably, we observed a similar ~5-6-fold defect in DNA replication in the $ORF24_{RAAAG}$,
142 $ORF24_{R328A}$, and $ORF24_{Stop}$ viruses, which was restored to WT levels in the $ORF24_{RAAAG}$ and
143 $ORF24_{R328A}$ MR viruses. Moreover, there was no DNA replication defect in the HA-ORF24 virus
144 (Fig 1C), demonstrating that the defect in DNA replication is not a consequence of manipulating
145 the ORF24 locus but is specific to disrupting its role in transcription.
146
147 **Specific gene clusters show differential dependence on the ORF24-Pol II interaction**
148 To more comprehensively analyze how the ORF24-Pol II interaction impacts gene
149 expression during infection, we deep sequenced RNA from three replicates each of iSLK cells
150 infected with the $ORF24_{RAAAG}$ or the $ORF24_{RAAAG}$-MR virus. Samples were sequenced at 24 and
151 48 h post lytic reactivation with doxycycline and sodium butyrate, as these times captured
152 predominantly early (24 h) versus both early and late (48 h) viral gene expression (Figure S2).
153 Given that overlapping transcripts are pervasive in the KSHV genome (22,30), we calculated the
154 primary transcript counts as described in Bruce et al (30) to accurately quantify transcript levels.
155 Differential expression analysis of genes in infected cells containing WT ORF24
156 ($ORF24_{RAAAG}$ MR) confirmed that the majority of viral genes are upregulated between 24 and 48
157 h post reactivation (Figure 2A), and those that are most strongly upregulated are predominantly
158 classified as late genes (shown in blue). Conversely, there was a generalized reduction in transcript
159 levels at the 48 h time point in cells containing KSHV $ORF24_{RAAAG}$ compared to MR, with the
160 most strongly downregulated genes being those classified as late genes (Figure 2B).

161 Combining the above ORF24$_{RAAAG}$ mutant and MR induction data into a heatmap confirmed
162 that in general, genes with later expression kinetics tended to be more severely impacted by the
163 ORF24$_{RAAAG}$ mutation compared to genes with early expression kinetics (Figure 2C). (The fold
164 change data was scaled to allow for easy visualization). However, not all genes fit this model.
165 Some genes categorized as early based on their induction pattern showed a moderate dependence
166 on the ORF24-Pol II interaction (e.g. ORFs 9, 40, 34, 41, 16 and 37). Similarly, a subset of genes
167 that show late kinetics were less dependent on the ORF24-Pol II interaction (e.g. ORFs 67A, 62,
168 K4.2A, 64, K7, 54, 67, 56, 66, K4.2 and vIRFs-2, 3 and 4). Classification as early or late and
169 dependence on the ORF24-Pol II interaction was determined using mean log_2 fold change.
170 Together, these data suggest that not all kinetic late genes are regulated by the vPIC and there may
171 be subset of early genes that are direct or indirect targets of the vPIC.
172

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

175 We then looked for motifs enriched in the promoters of genes most strongly downregulated
176 in the ORF24$_{RAAAG}$ mutant using Multiple EM for Motif Elicitation (MEME) (31), a program that
177 discovers un-gapped motifs that are enriched in a given set of sequences. We defined the promoter
178 sequence as 100 bp upstream from the transcription start site (TSS) or 300 bp upstream from the
179 translation start site when the TSS was unknown. Analysis of promoter sequences from 14 of the
180 most strongly down-regulated genes (log_2 fold change greater than one standard deviation from
181 the mean) showed that TATTWAA is the only motif that is enriched in this set of sequences
182 (Figure 2D).

183 We next monitored for the occurrence of this motif across all promoters in the viral genome
184 using Find Individual Motif Occurrences (FIMO) (32), which scans a set of sequences for matches
185 to motifs provided by the user. 23 of 83 viral promoters were found to have the TATTWAA motif
186 with a high score, of which 16 promoters had the motif ~30 bp upstream from the TSS (Figure 4
187 and Table 1), consistent with the location of the canonical TATA box in eukaryotic promoters. Of
188 the remaining seven promoter sequences, the TSS of six of the promoters is not known (shown in
189 red in Table 1); hence, the exact position of the TATTWAA motif from the TSS cannot be
190 estimated. ORF25 is the only exception, where the TATTWAA sequence is not present ~30 bp
191 upstream from the TSS, although it does have a canonical TATA box (TATAAAA) at that
192 position.

193 Notably, a subset of the genes containing the TATTWAA motif showed less than average
194 repression by the $ORF24_{RAAAG}$ mutation (ORF55, ORF64, ORF66, ORF59, K4.1, ORF4, ORF8,
195 ORF58, ORF26 and ORF62). One explanation for this observation is that these genes could be
196 regulated by cellular transcription factors and other *cis* elements which are absent in the genes that
197 are downregulated in the $ORF24_{RAAAG}$ mutant. The data also suggest that the motif may not be
198 sufficient to mark regulation by the vPIC, which is consistent with a similar observation in MHV68
199 (8). Furthermore, several genes (ORF28, ORF30, K15, ORF31 and ORF75) were strongly affected
200 by the mutation ($\log_2$ fold change greater than one standard deviation from the mean) but lack the
201 TATTWAA motif. The number of genes in this set increases to 26 if we broaden the criteria to
202 include all the genes that show greater than average repression in the presence of the mutation.
203 These genes could either be down regulated by an indirect mechanism in the absence of the
204 ORF24-Pol II interaction or could be bound by the vPIC in a manner independent of the TATT
205 motif.

206

207 **Chromatin immunoprecipitation reveals vPIC occupancy across the KSHV genome**

208 Our RNA Seq experiments identified KSHV genes whose expression is influenced by the
209 transcription regulatory function of ORF24. However, RNA Seq does not distinguish genes that
210 are directly bound by the vPIC from those that are indirect transcriptional targets (e.g. controlled
211 by other factors whose expression is induced by ORF24). We therefore sought to identify
212 promoters bound by the KSHV vPIC complex using ChIP Seq. Previous studies suggest that
213 ORF34 acts as a scaffolding factor that brings together all components of the vPIC (13) including
214 ORF24, which binds both RNA Pol II and TATT-containing DNA. We reasoned that ORF34
215 promoter occupancy would therefore be a good proxy for vPIC assembly, and additionally would
216 avoid potential pitfalls associated with any non-productive or independent interactions between
217 ORF24 and DNA.

218 To facilitate ChIP Seq, we added an N-terminal HA tag to endogenous ORF34 in KSHV
219 BAC16 (HA-ORF34) (Figure S3A), and generated a stable, doxycycline-inducible iSLK cell line
220 containing this tagged virus as previously described. HA-ORF34 produced WT levels of infectious
221 virions in a supernatant transfer assay (Figure S3B), indicating that the tag does not disrupt ORF34
222 activity. The integrity of the BAC DNA was verified by *Rsr*II digestion (Figure S1-Lane7). We
223 performed the ChIP Seq experiment in duplicate at 48 h post reactivation to ensure active late gene
224 expression. Reactivated iSLK cells with untagged BAC16 served as a negative control for the
225 immunoprecipitation. The individual ChIP peaks for the two replicates as well as the respective
226 inputs and controls are shown in Figure S3C and supplementary file 1. The ChIP Seq data showed
227 that HA-ORF34 bound several locations across the viral genome (Figure 3A) but was not
228 detectably associated with the host genome. A peak calling algorithm (HOMER) (33) identified

229 25 common peaks in the viral genome across two replicates that were significantly enriched in
230 HA-ORF34 relative to untagged WT or input DNA (Supplementary File 1 and Figure S3C).
231 Notably, 20 of the 25 peaks were found upstream of ORFs, with ~80% of the peaks centered less
232 than 200 bp from the closest TSS (Supplementary File 1). Given the link between DNA replication
233 and late transcription in herpesviruses, it was particularly interesting that two of the most
234 prominent peaks are upstream of the two minimal lytic origins of replication (Ori) in the KSHV
235 genome (Figure 3A). One of the peaks could not be uniquely assigned to a single ORF and was
236 not considered further. Of the remaining 19 ORFs associated with peaks, 13 are involved in
237 packaging, capsid assembly (or form part of capsid itself), 4 are involved in immune evasion and
238 2 in DNA replication (Table 2). While several of them have also been previously classified as late
239 genes (12/20), it is notable that 6 of them have been classified as early genes, again suggesting that
240 not all targets of the vPIC are late genes.

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

247 A MEME analysis of the promoters associated with the peaks again showed that the only
248 motif enriched in these promoters is TATTWAARV (Figure 3C). 15 of the 19 promoters have the
249 TATTWAA motif and in a majority of the sequences, it is present ~30 bp from the TSS (Table 2).
250 Of the four exceptions, two of the promoters (ORF53 and ORF6) have a TATATAA motif with
251 an A at position 4 in place of the T. Indeed, mORF53 has previously shown to be a target of the

252 vPIC in MHV68 despite the deviation from consensus (8). Two key points can be drawn from the
253 ChIP seq data. First, several genes whose promoters contain a TATTWAA motif do not show
254 ORF34 binding (e.g. ORFs 17, 32, 22, 55, 58, 64 and 66). Second, while most promoters bound
255 by ORF34 have a TATTWAA motif, there are some exceptions (ORF45, ORF68, ORF6 and
256 ORF53). These observations indicate that while the TATTWAA motif is enriched in promoters
257 regulated by the vPIC, it is not sufficient to ensure binding of this specialized transcription
258 complex.
259

260 **Thirteen KSHV genes are direct targets of the vPIC**

261 Combining the RNA Seq and ChIP Seq data, we defined the direct targets of the vPIC as
262 genes that are both downregulated in the absence of the ORF24-Pol II interaction and show ORF34
263 binding in the promoter region. By this measure, a total of 13 genes are direct targets of the vPIC
264 (Figure 4 and Table 2 – gray rows). Of the 13 direct targets, 11 code for structural proteins or
265 proteins involved in viral egress while two are tegument proteins involved in immune evasion.
266 Genes whose expression was impaired in the KSHV ORF24$_{RAAAG}$ mutant but lacked ORF34
267 binding at their promoters are likely to be indirect targets of the vPIC (Figure 4A). Finally, it is
268 interesting to note that not all genes that have a ChIP peak are down regulated in the RNA Seq
269 data (e.g. ORFs 4, 6, K4.1, 59, 62 and 68). These could be false positive ChIP peaks and/or have
270 alternate mechanisms of regulation that ensure their transcription even in the absence of the vPIC.
271

272 **The direct vPIC targets reveal an expanded promoter motif**

273 A defining feature of characterized late gene promoters is that they are minimalistic, with
274 no conserved motifs beyond the TATTWAA signature recognized by ORF24. Yet, our data

275 indicate that the TATTWAA motif is insufficient to confer activation by the late gene transcription
276 complex. We hypothesized that the inclusion of potentially indirect vPIC targets into prior motif
277 searches may have hindered identification of other regulatory sequences within these promoters.
278 Indeed, repeating the MEME analysis on the 13 direct targets revealed that an additional 5 bp motif
279 (RVNYS) downstream of the TATTWAA motif was enriched in these promoter sequences (Figure
280 4B).

281 To explore the contribution of the 5 bp region downstream of the TATTWAA sequence
282 for regulation by the vPIC, we used a plasmid-based promoter activation assay in which expression
283 of a luciferase reporter was driven by either the K8.1 late or ORF57 early promoter (Figure 5A).
284 The KSHV left lytic origin of replication (ori-lyt) was included on the plasmids, as data from
285 MHV68, KSHV and EBV indicate that the presence of ori-lyt in *cis* significantly boosts late gene
286 promoter activation during infection (8,10,34). Each plasmid was transfected together with a
287 plasmid constitutively expressing renilla luciferase (to normalize for transfection efficiency) into
288 lytically reactivated KSHV-positive iSLK cells and the luciferase signal was measured 48 h post
289 transfection and reactivation. As expected, the presence of ori-lyt enhanced K8.1 promoter-driven
290 luciferase expression by 15 – 20-fold (Figure 5B). Furthermore, unlike the consistently robust
291 ORF57 promoter activity, K8.1 promoter activity on the plasmid was completely repressed in the
292 presence of DNA replication inhibitor PAA (Figure 5C) or in cells lytically infected with the
293 $ORF24_{RAAAG}$ mutant (Figure 5D). Taken together, these data confirm that the plasmid assay
294 accurately recapitulates late gene promoter activity.

295 Notably, mutating the 5 bp RVNYS motif in the K8.1 promoter decreased transcription by
296 > 50% (Figure 5E). This decrease was comparable to mutating the T at position four to an A in the
297 TATTWAA sequence, which has been shown to be important (although not sufficient) for

298 recognition and binding by ORF24 (10). Replacement of the 5 bp RVNYS motif with either
299 synonymous (-S) or nonsynonymous (-NS) bases yielded a similar decrease in promoter activity
300 (Figure 5E-K8.1$_{GGGAC\text{->}TACGA\text{-}S}$, K8.1$_{GGGAC\text{->}ACTCG\text{-}NS}$), suggesting a role for the overall
301 architecture of this motif rather than the primary sequence. In support of this hypothesis, a
302 preliminary analysis of four structural features of the expanded motif in the WT and mutant predict
303 that the average roll of the DNA in the expanded motif is lower in the mutant promoter compared
304 to the WT promoter (Figure S4). Combining the TATT->TATA mutation with the mutation of the
305 5 bp motif completely abolished transcription, indicating that these elements are absolutely
306 essential for promoter activation (Figure 5E). Furthermore, with the exception of GG89TT-NS,
307 systematic single and double base substitutions in the RVNYS motif were well tolerated, arguing
308 that this region may not be contributing to base specific contacts with the vPIC (Figure 5F). In
309 summary, while the expanded RVNYS motif appears important for transcriptional activation, we
310 hypothesize it may contribute to vPIC specificity by impacting the local topology of the promoter
311 DNA rather than through base specific contacts.
312
313 **The expanded promoter motif is important for DNA binding by vPIC**
314 The RVNYS motif could either contribute to vPIC binding to its target promoters, as is the
315 case for the TATTWAA sequence, or it could modulate downstream events involved in
316 transcription initiation at vPIC bound promoters. To distinguish these possibilities, we used ChIP
317 to measure HA-ORF34 occupancy on the luciferase plasmids containing WT or mutant K8.1
318 promoter in lytically reactivated HA-ORF34 iSLK cells. We engineered qPCR primers that
319 enabled distinction of the K8.1 promoter on the plasmid DNA from the endogenous viral locus
320 (Figure 6A).

321 Notably, both the K8.1$_{TATT->TATA}$ and K8.1$_{GGGAC->ACTCG}$ mutations similarly impaired HA-
322 ORF34 binding to the plasmid promoter DNA (Figure 6B- left panel). The ChIP signal was
323 specific for HA-ORF34, as we observed only a background signal when the assay was performed
324 in the WT iSLK cells containing an untagged ORF34 locus. Furthermore, there was comparable
325 vPIC assembly at the genomic K8.1 locus in cells transfected with WT or the mutant plasmids,
326 confirming similar immunoprecipitation efficiencies across samples (Figure 6B - right panel).
327 Together, these data indicate that the expanded promoter motif is important for vPIC assembly on
328 the promoter DNA.
329
330 **The ORF24-34 binding element in the origin is necessary for late gene activation**
331 Given that late gene transcription requires the presence of the viral lytic origin of
332 replication (Figure 5B), binding of ORF34 and ORF24 at the KSHV Ori was particularly notable
333 (see Figures 3A-B). The binding site mapped to a region of the Ori just outside of the core minimal
334 region that is required for viral DNA replication (Figure 7A). To examine the role of vPIC Ori
335 binding on K8.1 promoter activation, we engineered a mutant version of the K8.1 promoter-driven
336 luciferase plasmid lacking the ORF24-34 binding element (BE). A deletion encompassing this
337 region has been shown to only modestly reduce (~30%) viral DNA replication (35,36) . However,
338 we observed a marked reduction in K8.1 promoter activation in the absence of the ORF24-34 Ori
339 binding element (Figure 7B). The element by itself was insufficient to activate transcription in the
340 absence of core minimal Ori, in agreement with the requirement for DNA replication to license
341 late gene activation.
342 Ori binding by ORF24 and ORF34 could impact late gene transcription either by enabling
343 vPIC promoter binding or by facilitating a downstream stage of transcription. To distinguish

344 between these possibilities, we used our transient ChIP assay to compare HA-ORF34 binding to
345 the K8.1 promoter on the plasmid containing either the WT Ori or the Ori lacking the ORF24-34
346 binding element. Removing the binding element within the Ori resulted in near complete loss of
347 HA-ORF34 binding to the K8.1 promoter on the plasmid (Figure 7C), indicating that ORF24-34
348 occupancy at the lytic origin of replication is essential for vPIC assembly at the late gene promoter.
349

350 **Discussion**
351
352 Late gene expression is uniquely regulated in KSHV by a set of viral proteins collectively
353 known as the vPIC. In this study, we generated a comprehensive list of direct targets of the vPIC
354 by combining transcriptome and genome-wide binding data of the complex. Among the direct
355 targets, we identified an expanded promoter recognition motif that contributes to stable binding of
356 the vPIC on the promoter. This is particularly useful to understanding alternate mechanisms of
357 transcription initiation by RNA pol II because of the unique ability of the vPIC to directly recruit
358 RNA Pol II to highly compact promoters. We also identified a vPIC binding element in the lytic
359 origin of replication that licenses late gene activation at a distal promoter, suggesting a novel
360 mechanism may underlie the link between DNA replication and late gene transcription.
361 Several key points emerge from independent analysis of the RNA Seq and ChIP Seq data.
362 First, the previously identified TATTWAA motif is insufficient to mark binding and regulation by
363 the vPIC, as demonstrated by the multiple ORFs that have the motif in their promoter (23/83 ORFs)
364 but are not repressed in the $ORF24_{RAAAG}$ mutant (10/23 ORFs) or bound by ORF34 (7/23 ORFs).
365 Second, the TATTWAA motif is not a pre-requisite for regulation or binding by the vPIC. For
366 example, although ORF53 expression is dependent on the ORF24-Pol II interaction and its
367 promoter is bound by ORF34, its promoter does not have the conserved TATTWAA sequence but

368 a TATATAA sequence in its place. Third, several genes that exhibit late kinetics escape repression
369 in the mutant, indicating that not all late genes are direct or indirect targets of vPIC. It is notable
370 that several of the kinetic late genes that escape repression are involved in immune regulation,
371 consistent with a similar observation in EBV (28). In particular, EBV BPLF1 is a late gene whose
372 expression is independent of the vPIC, and we observe that its homolog ORF64 also shows a
373 similar pattern in KSHV. Finally, while there are several early genes that are indirect targets of the
374 vPIC, only 1 of 13 direct targets exhibited early kinetics (ORF45). It is notable that ORF45 is a
375 tegument protein, which might explain the requirement for its continued regulation at later times
376 of infection.

377 Our integrative analysis revealed that a distinct 5 bp motif is integral to vPIC promoter
378 recognition. Rather than engaging in base specific contacts (as is the case for the TATTWAA
379 sequence) (7), we instead hypothesize that the importance of this motif lies in its shape or structure.
380 In this regard, a preliminary analysis of basic structural properties of the expanded motif suggest
381 that the average roll (representing the rotational flexibility) of the DNA in the expanded motif is
382 lower in the 5 bp mutant promoter compared to WT K8.1 promoter. Structure-based recognition
383 and a contribution of base pairs flanking the core binding site to binding affinity is a common
384 feature among transcription factors, although the specific structural properties that influence
385 binding differ for various classes of factors (37). A previous study of the bZIP family of
386 transcription factors identified a correlation between the roll of DNA sequences that flank the core
387 recognition element and binding affinity (37). Consistent with that, we observed a defect in the
388 assembly of the vPIC when the expanded motif was mutated, suggesting a stabilizing role for the
389 flanking sequence.

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

399 Notably, we found that KSHV DNA replication was reduced in the absence of functional
400 ORF24, including with an ORF24 null mutant or ORF24 point mutants that fail to bind either RNA
401 Pol II or the vPIC component ORF34. This was surprising, given that deletion or mutation of
402 several other vPIC components has been shown not to affect DNA replication (7,13,17,18).
403 Furthermore, we previously failed to observe a replication defect in an ORF24 null virus (7). We
404 hypothesize that the discrepancy for the ORF24 null mutant relates to the methods used to establish
405 the infected cell lines. In our earlier study, the viral BAC DNA was directly transfected into iSLK
406 cells, resulting in only a ~5-7-fold increase in genome copy number upon reactivation. It is likely
407 that the majority of viral DNA was present in a non-reactivatable state, providing a level of
408 background that significantly reduced the dynamic range for DNA replication measurements.
409 Here, the iSLK cells were established through co-culture with reactivated 293T cells harboring the
410 BAC DNA, i.e. through infection rather than transfection. This reproducibly yields ~50-70-fold
411 increases in genome copy number upon lytic reactivation, more closely representing actual

412 infection. That said, the role of ORF24 in viral DNA replication may be unique to KSHV, as this
413 link has not been observed upon depletion of its homologs in EBV and MHV68 (28,39).

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

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

435

436 **Methods**

437 **Cell lines**

438 HEK293T cells were maintained in DMEM (Invitrogen) + 10% Fetal Bovine Serum (FBS).
439 iSLK BAC16 cells (29) were maintained in DMEM + 10% FBS and 0.5 mg/mL hygromycin B.
440 HEK293T cells stably expressing ORF24-3xFlag were maintained in DMEM (Invitrogen) + 10%
441 Fetal Bovine Serum (FBS) and 325 ug/ml of zeocin. iSLK-puro cells were maintained in DMEM
442 + 10% FBS.

443

444 **Plasmids**

445 A 100 bp fragment of the K8.1 promoter or ORF57 promoter was generated by PCR from
446 the BAC16 genome and cloned into the KpnI and HindIII sites of the pGL4.16 reporter plasmid
447 using Infusion cloning (Clontech). The left origin of replication was similarly generated by PCR
448 from the BAC16 genome and cloned into the NotI and BstXI sites of the K8.1Pr-pGL4.16 plasmid.
449 The promoter mutants were generated by site-directed mutagenesis using KAPA HiFi DNA
450 polymerase using the primers shown in Supplementary table 1. To make the ORF24-3xFlag
451 pJLM1 plasmid for lentiviral transduction, the ORF24-3xFlag fragment was generated by PCR
452 using ORF24-3xFlag-pCDNA4 as the template and cloned in to pLJM1 plasmid digested with
453 AgeI and EcoRI. All primers used for cloning and mutagenesis are listed in Supplementary table
454 1.

455

456 **Generation of BAC mutants and establishing latent cell lines**

457 The $ORF24_{RAAAG}$, $ORF24_{R328A}$ and $ORF24_{Stop}$ mutants, the corresponding mutant rescues
458 (MR) for $ORF24_{RAAAG}$, $ORF24_{R328A}$ and N terminal HA tagged ORF24 and ORF34 were
459 engineered using the scarless Red recombination system in BAC16 GS1783 *E. coli* (40) as
460 described previously. The modified BACs were purified using the NucleoBond BAC 100 kit
461 (Clontech).

462 To establish iSLK cell lines latently infected with the modified virus, HEK293T cells were
463 transfected with 5-10 ug of the modified BAC16 using linear polyethylenimine (PEI, MW
464 ~25,000) at a 1:3 DNA:PEI ratio. The following day, the 293T cells were mixed 1:1 with iSLK-
465 puro cells (which contain inducible RTA but lack KSHV) and treated with 30 nM 12-O-
466 Tetradecanoylphorbol-13-acetate (TPA) and 300 nM sodium butyrate for 4 days to induce lytic
467 replication. Cells were then grown in selection media containing 300 ug/ml hygromycin B, 1 ug/ml
468 puromycin and 250 ug/ml G418. The hygromycin concentration was gradually increased to 500
469 ug/ml and then 1 mg/ ml until all the HEK293Ts cells died. To generate the $ORF24_{RAAAG}$,
470 $ORF24_{R328A}$, and $ORF24_{Stop}$ (which are defective in infectious virion production), HEK293T cells
471 that stably express ORF24-3xFlag were used for the initial transfection and co-culture.

472

473 **Supernatant transfer assay**

474 For supernatant transfer assays to assess virion production, BAC16-containing iSLK cells
475 were plated at a density of $1x10^6$ cells in 10 cm dishes and treated with 1 ug/mL doxycycline and
476 1 mM sodium butyrate the following day. 72 h post induction, the supernatant was filtered through
477 a 0.45 uM syringe filter and 2 ml of the supernatant was mixed with $1x10^6$ freshly trypsinized
478 HEK293T cells in a 6 well plate and centrifuged at 1,500x*g* for 2 h at 37˚C. The following day,

479 cells were trypsinized, fixed in 4% paraformaldehyde and the percentage of cells expressing GFP
480 was determined by flow cytometry (BD Csampler, BD Biosciences).
481
482 **Viral genome replication**
483 To measure viral genome replication, iSLK-BAC16 cells were reactivated for 72 h as
484 described above. The cells were then scraped into the media and the media + cells were digested
485 with proteinase K (80µg/mL) (Promega) in 5x proteinase K digestion buffer (50mM Tris-HCl pH
486 7.4, 500mM NaCl, 5mM EDTA, 2.5% SDS) overnight at 55˚C. The gDNA was isolated using
487 Zymo Quick gDNA Miniprep Kit according to the manufacturer's instructions. Quantitative PCR
488 (qPCR) was performed on the isolated DNA using iTaq Universal SYBR Green Supermix on a
489 QuantStudio3 Real-Time PCR machine. DNA levels were quantified using relative standard
490 curves with primers specific for KSHV ORF59 promoter and human CPSF6 promoter
491 (Supplementary table 1). The relative genome numbers were normalized to CPSF6 to account for
492 loading differences and to uninduced samples to account for differences in starting genome copy
493 number.
494
495 **RT-qPCR**
496 To look at viral transcript levels, total RNA was isolated from reactivated iSLK cells at the
497 indicated time points using the Zymo Direct-Zol RNA kit following the manufacturer's protocol.
498 The purified RNA was treated with DNase and cDNAs were synthesized using AMV reverse
499 transcriptase (Promega). The cDNAs were directly used for qPCR analysis using the iTaq
500 Universal SYBR Green Supermix. The qPCR signal for each ORF was normalized to 18s rRNA.
501

502 **Western blotting**

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

510

511 **RNA Seq sample preparation and analysis**

512 Total RNA was isolated from reactivated iSLK cells at indicated time points using Zymo
513 Direct-Zol RNA kit following the manufacturer's protocol. The samples were ribo-depleted and
514 stranded libraries were prepped by the QB3 Berkeley core sequencing facility. Multiplex
515 sequencing was performed using the Illumina HiSeq 4000 to generate 100 bp paired end reads.
516 The quality of the raw reads was checked using FastQC (42) and adapters were trimmed using
517 TrimGalore (43) using default parameters. The trimmed reads were aligned to both the viral
518 (NC_009333.1) and host genomes (hg19) using STAR aligner (ver2.5.3a) (44). The aligned reads
519 were counted using htseq-counts (45) and differential expression was assessed using DESeq2 (46).

520 To compensate for the problem of quantitative assessment of transcripts due to the
521 overlapping transcripts in the KSHV genome, the primary counts of the transcripts were calculated
522 as recommended in Bruce et al (30) with the following small modifications. Length correction was
523 not applied to allow for differential expression analysis by DESeq2. Transcripts for which the
524 counts became negative in all 3 replicates when calculating primary counts were removed before

525 running DESeq2. Bar plots showing Log_2 Fold Change were generated using R. The heatmap was
526 generated using the heatmap.2 package in R. The fold change values were scaled by column before
527 plotting to ease visualization of the data. Raw and processed data files are available from the GEO
528 database (accession number: GSE126602).
529
530 **Chromatin Immunoprecipitation (ChIP)**
531 ChIP was performed on 3*15-cm plates of iSLK-cells (WT KSHV, $KSHV_{HA\text{-}ORF34}$ and
532 $KSHV_{ORF24\text{-}HA}$) reactivated for 48 h with 1 ug/ml dox and 1 mM sodium butyrate. Cells were
533 crosslinked in 2% formaldehyde for 15 min at room temperature, quenched in 0.125 M glycine,
534 and washed twice with ice-cold PBS. Crosslinked cell pellets were mixed with 1 ml ice-cold ChIP
535 lysis buffer (5mM PIPES pH 8.0, 85mM KCl, 0.5% NP-40) and incubated on ice for 10 min,
536 whereupon the lysate was dounce homogenized to release nuclei and spun at 1700 g for 5 min at
537 4°C. Nuclei were then resuspended in 800 µl of nuclei lysis buffer (50mM Tris-HCl pH 8.0, 0.3%
538 SDS, 10mM EDTA) and rotated for 10 min at 4°C followed by sonication using a QSonica
539 Ultrasonicator with a cup horn set to 25 amps for 20 min total (5 min on, 5 min off). Chromatin
540 was spun at 16,000 x *g* for 10 min at 4°C and the pellet was discarded. The chromatin was
541 precleared with protein A + protein G beads blocked with 5 ug/ml glycogen, 50 ug/ml BSA, 100
542 ug/ml Ecoli tRNA in 1ml dilution buffer (1.1% Triton-X-100, 1.2mM EDTA, 16.7mM Tris-HCl
543 pH 8, 167mM NaCl) for 2 hours.
544 350 µl of chromatin was diluted in ChIP dilution buffer (16.7 mM Tris-HCl pH 8.0, 1.1%
545 Triton X-100, 1.2 mM EDTA, 167 mM NaCl) to 500 µl and incubated with 10 µg anti-HA
546 antibody (Cell Signaling C29F4) overnight, whereupon samples were rotated with 20 µl pre-
547 blocked protein A + G beads (Thermofisher) for 2 h at 4°C. Beads were washed with low salt

548 immune complex (20 mM Tris pH 8.0, 1% Triton-x-100, 2 mM EDTA, 150 mM NaCl, 0.1% SDS),
549 high salt immune complex (20 mM Tris pH 8.0, 1% Triton-x-100, 2 mM EDTA, 500 mM NaCl,
550 0.1% SDS), lithium chloride immune complex (10mM Tris pH 8.0, 0.25 M LiCl, 1% NP-40, 1%
551 Deoxycholic acid, 1 mM EDTA), and Tris-EDTA for 10 min each at 4°C with rotation. DNA was
552 eluted from the beads using 100 µl of elution buffer (150 mM NaCl, 50 µg/ml Proteinase K) and
553 incubated at 50 °C for 2 h, then 65 °C overnight. DNA was purified using the Zymo PCR
554 purification kit. For ChIP-qPCR, purified DNA was quantified by qPCR using iTaq Universal
555 SYBR Mastermix (BioRad) and the indicated primers (Supplementary table 1). Each sample was
556 normalized to its own input.
557 For ChIP-Seq, the concentration of the eluted DNA was measured using Qubit. 10-15 ng
558 of eluted DNA was used to generate libraries using the Accel-NGS 2S Plus DNA Library Kit
559 (Swift Biosciences). The libraries were sequenced on a Illumina HiSeq 4000 sequencer at the QB3
560 Berkeley core facility to generate 100 bp single end reads.
561
562 **ChIP Seq Analysis**
563 The quality of the raw reads was checked using FastQC (42) and adapters were trimmed
564 using TrimGalore (43) using default parameters. The trimmed reads were aligned to both the viral
565 and host genomes (hg19) using Bowtie2 aligner using default parameters. The aligned reads were
566 filtered for uniquely mapped reads using SAMTools. The filtered reads were converted to tdf
567 format using IGV Tools and visualized on Integrated Genome Viewer (47,48). We used HOMER
568 (33) with the following minor adjustments to default parameters to call peaks for the
569 immunoprecipitated and input samples from cells that have HA tagged ORF34 and untagged
570 ORF34 (WT) as a control. The fold over input and local fold change were set to 1.3 as

571 recommended for smaller and denser viral genomes (personal communication with Chris Benner,
572 UCSD). The peak size and fragment length were also set at 200 and 150 respectively. From the
573 resulting peaks, only the ones that were present in both conditions (HA-ORF34 vs HA-Input, HA-
574 ORF34 vs WT) and both replicates were retained for further analysis. The common peaks
575 identified were verified visually on the Integrated Genome Viewer and were manually annotated.
576 To identify the direct targets, transcripts that had a Log_2 fold change lower than the mean
577 Log_2 fold change in the $ORF24_{RAAAG}$ mutant vs MR were considered to be down regulated in the
578 mutant. Although ORF8 and ORF26 were not included in the differential expression analysis, they
579 were considered to be downregulated because of negative values for their levels in the mutant and
580 not in the mutant rescue on calculation of primary transcript levels. Raw and processed data files
581 are available from the GEO database (accession number: GSE126601).
582
583 **Promoter Analyses**
584 We defined the promoter sequence as 100 base pairs (bp) upstream from the transcription
585 start site (TSS) or 300 bp upstream from the translation start site when the TSS was unknown. The
586 TSS data was obtained from Table 1 in (23) and a custom python script was used to extract the
587 sequences. These were then run through MEME (31) allowing for multiple occurrences of a motif
588 in a sequence and only searching the given strand to identify enriched motifs. The identified motif
589 was then used to search for similar motifs across all 83 promoter sequences from the viral genome,
590 again only allowing for searching the given strand.
591
592 **Reporter assays**

593 $5*10^5$ iSLK-BAC16 (WT or $ORF24_{RAAAG}$) cells were plated in a 6 well plate 24 h before
594 transfection and reactivation. The cells were reactivated with 1 μg/ml doxycycline and 1 mM
595 Sodium Butyrate and simultaneously transfected with 0.95 μg luciferase reporter plasmid and 0.05
596 μg renilla plasmid using polyjet transfection reagent. 48 h post transfection and reactivation, the
597 cells were rinsed twice with room temperature PBS, lysed by rocking for 15 min at room
598 temperature in 300 μl of Passive Lysis Buffer (Promega), and clarified by centrifuging at 14,000
599 g for 1 min. 20 μl of the clarified lysate was added in triplicate to a white chimney well microplate
600 (Greiner bio-one) to measure luminescence on a Tecan M1000 using a Dual Luciferase Assay Kit
601 (Promega). The firefly luminescence was normalized to the internal Renilla luciferase control. The
602 induced sample was then normalized to the uninduced (but transfected) control to look at the fold
603 induction during reactivation.
604
605 **Transient ChIP-qPCR**
606 $3.5x10^6$ iSLK-BAC16 were plated in a 15 cm dish 24 h before transfection and reactivation.
607 The cells were reactivated with 1 μg/ml doxycycline and 1 mM Sodium Butyrate and
608 simultaneously transfected with 8 μg of reporter plasmid using polyjet transfection reagent. 48 h
609 post reactivation and transfection, cells were crosslinked and lysed as mentioned above. Nuclei
610 were sonicated using a Covaris S220 (Covaris) for 5 min with the recommended sonication settings
611 (Power: 140 W, Bursts per cycle: 200, Duty cycle: 5%). The chromatin was pre-cleared as
612 mentioned above and the concentration measured using a qubit. 25 μg of pre-cleared chromatin
613 was used in the IP and the following steps were same as mentioned in the ChIP protocol above.
614
615 **Acknowledgements**

616 We thank all members of the Glaunsinger and Coscoy labs for their helpful suggestions. In
617 particular, we thank Angelica Casteñeda for providing the HEK293T cells stably expressing
618 ORF24-3xFlag and Chris Benner (UCSD) for suggestions on proper usage of HOMER for
619 organisms with small genomes. This work was supported by NIH R01 AI122528 to B.G., who is
620 also a Howard Hughes Medical Institute Investigator.

621

622 **References**

623 1. Payne S. Chapter 28 - Introduction to DNA Viruses. In: Payne S, editor. Viruses
624 [Internet]. Academic Press; 2017. p. 231–6. Available from:
625 http://www.sciencedirect.com/science/article/pii/B9780128031094000283
626 2. Grula MA, Buller PL, Weaver RF. α-Amanitin-Resistant Viral RNA Synthesis in Nuclei
627 Isolated from Nuclear Polyhedrosis Virus-Infected Heliothis zea
628 Larvae and Spodoptera frugiperda Cells. J Virol [Internet]. 1981
629 Jun 1;38(3):916. Available from: http://jvi.asm.org/content/38/3/916.abstract
630 3. Kochetkov SN, Rusakova EE, Tunitskaya VL. Recent studies of T7 RNA polymerase
631 mechanism. FEBS Lett [Internet]. 1998;440(3):264–7. Available from:
632 http://www.sciencedirect.com/science/article/pii/S0014579398014847
633 4. Rong M, He B, McAllister WT, Durbin RK. Promoter specificity determinants of T7
634 RNA polymerase. Proc Natl Acad Sci U S A [Internet]. The National Academy of
635 Sciences; 1998 Jan 20;95(2):515–9. Available from:
636 https://www.ncbi.nlm.nih.gov/pubmed/9435223
637 5. Young CSH. The Structure and Function of the Adenovirus Major Late Promoter. In:
638 Doerfler W, Böhm P, editors. Adenoviruses: Model and Vectors in Virus-Host

639 Interactions: Virion-Structure, Viral Replication and Host-Cell Interactions [Internet].
640 Berlin, Heidelberg: Springer Berlin Heidelberg; 2003. p. 213–49. Available from:
641 https://doi.org/10.1007/978-3-662-05597-7_8
642 6. Gruffat H, Marchione R, Manet E. Herpesvirus Late Gene Expression: A Viral-Specific
643 Pre-initiation Complex Is Key. Front Microbiol [Internet]. Frontiers Media S.A.; 2016 Jun
644 6;7:869. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27375590
645 7. Davis ZH, Verschueren E, Jang GM, Kleffman K, Johnson JR, Park J, et al. Global
646 Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for
647 Late Genes. Mol Cell [Internet]. 2015;57(2):349–60. Available from:
648 http://www.sciencedirect.com/science/article/pii/S1097276514009216
649 8. Wong-Ho E, Wu T-T, Davis ZH, Zhang B, Huang J, Gong H, et al. Unconventional
650 Sequence Requirement for Viral Late Gene Core Promoters of Murine Gammaherpesvirus
651 68. Frueh K, editor. J Virol [Internet]. 2014 Mar 15;88(6):3411. Available from:
652 http://jvi.asm.org/content/88/6/3411.abstract
653 9. Serio TR, Cahill N, Prout ME, Miller G. A Functionally Distinct TATA Box Required for
654 Late Progression through the Epstein-Barr Virus Life Cycle. J Virol [Internet]. 1998 Oct
655 1;72(10):8338. Available from: http://jvi.asm.org/content/72/10/8338.abstract
656 10. Tang S, Yamanegi K, Zheng Z-M. Requirement of a 12-Base-Pair TATT-Containing
657 Sequence and Viral Lytic DNA Replication in Activation of the Kaposi's Sarcoma-
658 Associated Herpesvirus K8.1 Late Promoter. J Virol [Internet]. 2004 Mar 1;78(5):2609.
659 Available from: http://jvi.asm.org/content/78/5/2609.abstract
660 11. Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription
661 initiation. Nat Rev Mol Cell Biol [Internet]. 2018;19(10):621–37. Available from:

662 https://doi.org/10.1038/s41580-018-0028-8
663 12. Aubry V, Mure F, Mariamé B, Deschamps T, Wyrwicz LS, Manet E, et al. Epstein-Barr
664 Virus Late Gene Transcription Depends on the Assembly of a Virus-Specific Preinitiation
665 Complex. Longnecker RM, editor. J Virol [Internet]. 2014 Nov 1;88(21):12825. Available
666 from: http://jvi.asm.org/content/88/21/12825.abstract
667 13. Davis ZH, Hesser CR, Park J, Glaunsinger BA. Interaction between ORF24 and ORF34 in
668 the Kaposi's Sarcoma-Associated Herpesvirus Late Gene Transcription Factor
669 Complex Is Essential for Viral Late Gene Expression. Sandri-Goldin RM, editor. J Virol
670 [Internet]. 2016 Jan 1;90(1):599. Available from:
671 http://jvi.asm.org/content/90/1/599.abstract
672 14. Wyrwicz LS, Rychlewski L. Identification of Herpes TATT-binding protein. Antiviral
673 Res [Internet]. 2007;75(2):167–72. Available from:
674 http://www.sciencedirect.com/science/article/pii/S0166354207002367
675 15. Gruffat H, Kadjouf F, Mariamé B, Manet E. The Epstein-Barr Virus BcRF1 Gene Product
676 Is a TBP-Like Protein with an Essential Role in Late Gene Expression. J Virol [Internet].
677 2012 Jun 1;86(11):6023. Available from: http://jvi.asm.org/content/86/11/6023.abstract
678 16. Nishimura M, Watanabe T, Yagi S, Yamanaka T, Fujimuro M. Kaposi's sarcoma-
679 associated herpesvirus ORF34 is essential for late gene expression and virus production.
680 Sci Rep [Internet]. 2017;7(1):329. Available from: https://doi.org/10.1038/s41598-017-
681 00401-7
682 17. Gong D, Wu NC, Xie Y, Feng J, Tong L, Brulois KF, et al. Kaposi's Sarcoma-
683 Associated Herpesvirus ORF18 and ORF30 Are Essential for Late Gene Expression
684 during Lytic Replication. Frueh K, editor. J Virol [Internet]. 2014 Oct 1;88(19):11369.

685 Available from: http://jvi.asm.org/content/88/19/11369.abstract
686 18. Brulois K, Wong L-Y, Lee H-R, Sivadas P, Ensser A, Feng P, et al. Association of
687 Kaposi's Sarcoma-Associated Herpesvirus ORF31 with ORF34 and ORF24 Is
688 Critical for Late Gene Expression. Longnecker RM, editor. J Virol [Internet]. 2015 Jun
689 1;89(11):6148. Available from: http://jvi.asm.org/content/89/11/6148.abstract
690 19. Perng Y-C, Campbell JA, Lenschow DJ, Yu D. Human Cytomegalovirus pUL79 Is an
691 Elongation Factor of RNA Polymerase II for Viral Gene Transcription. PLOS Pathog
692 [Internet]. Public Library of Science; 2014 Aug 28;10(8):e1004350-. Available from:
693 https://doi.org/10.1371/journal.ppat.1004350
694 20. Castañeda AF, Glaunsinger BA. The Interaction between ORF18 and ORF30 Is Required
695 for Late Gene Expression in Kaposi's Sarcoma-Associated Herpesvirus. Sandri-
696 Goldin RM, editor. J Virol [Internet]. 2019 Jan 1;93(1):e01488-18. Available from:
697 http://jvi.asm.org/content/93/1/e01488-18.abstract
698 21. Pan D, Han T, Tang S, Xu W, Bao Q, Sun Y, et al. Murine Cytomegalovirus Protein
699 pM91 Interacts with pM79 and Is Critical for Viral Late Gene Expression. Sandri-Goldin
700 RM, editor. J Virol [Internet]. 2018 Sep 15;92(18):e00675-18. Available from:
701 http://jvi.asm.org/content/92/18/e00675-18.abstract
702 22. Djavadian R, Hayes M, Johannsen E. CAGE-seq analysis of Epstein-Barr virus lytic gene
703 transcription: 3 kinetic classes from 2 mechanisms. PLOS Pathog [Internet]. Public
704 Library of Science; 2018 Jun 4;14(6):e1007114-. Available from:
705 https://doi.org/10.1371/journal.ppat.1007114
706 23. Arias C, Weisburd B, Stern-Ginossar N, Mercier A, Madrid AS, Bellare P, et al. KSHV
707 2.0: A Comprehensive Annotation of the Kaposi's Sarcoma-Associated Herpesvirus

708 Genome Using Next-Generation Sequencing Reveals Novel Genomic and Functional
709 Features. PLOS Pathog [Internet]. Public Library of Science; 2014 Jan
710 16;10(1):e1003847-. Available from: https://doi.org/10.1371/journal.ppat.1003847
711 24. Nakamura H, Lu M, Gwack Y, Souvlis J, Zeichner SL, Jung JU. Global Changes in
712 Kaposi's Sarcoma-Associated Virus Gene Expression Patterns following
713 Expression of a Tetracycline-Inducible Rta Transactivator. J Virol [Internet]. 2003 Apr
714 1;77(7):4205. Available from: http://jvi.asm.org/content/77/7/4205.abstract
715 25. Jenner RG, Albà MM, Boshoff C, Kellam P. Kaposi's Sarcoma-Associated
716 Herpesvirus Latent and Lytic Gene Expression as Revealed by DNA Arrays. J Virol
717 [Internet]. 2001 Jan 15;75(2):891. Available from:
718 http://jvi.asm.org/content/75/2/891.abstract
719 26. Paulose-Murphy M, Ha N-K, Xiang C, Chen Y, Gillim L, Yarchoan R, et al. Transcription
720 Program of Human Herpesvirus 8 (Kaposi's Sarcoma-Associated Herpesvirus). J
721 Virol [Internet]. 2001 May 15;75(10):4843. Available from:
722 http://jvi.asm.org/content/75/10/4843.abstract
723 27. Arumugaswami V, Wu T-T, Martinez-Guzman D, Jia Q, Deng H, Reyes N, et al. ORF18
724 Is a Transfactor That Is Essential for Late Gene Transcription of a Gammaherpesvirus. J
725 Virol [Internet]. 2006 Oct 1;80(19):9730. Available from:
726 http://jvi.asm.org/content/80/19/9730.abstract
727 28. McKenzie J, Lopez-Giraldez F, Delecluse H-J, Walsh A, El-Guindy A. The Epstein-Barr
728 Virus Immunoevasins BCRF1 and BPLF1 Are Expressed by a Mechanism Independent of
729 the Canonical Late Pre-initiation Complex. PLOS Pathog [Internet]. Public Library of
730 Science; 2016 Nov 17;12(11):e1006008-. Available from:

731 https://doi.org/10.1371/journal.ppat.1006008
732 29. Brulois KF, Chang H, Lee AS-Y, Ensser A, Wong L-Y, Toth Z, et al. Construction and
733 Manipulation of a New Kaposi's Sarcoma-Associated Herpesvirus Bacterial
734 Artificial Chromosome Clone. J Virol [Internet]. 2012 Sep 15;86(18):9708. Available
735 from: http://jvi.asm.org/content/86/18/9708.abstract
736 30. Bruce AG, Barcy S, DiMaio T, Gan E, Garrigues HJ, Lagunoff M, et al. Quantitative
737 Analysis of the KSHV Transcriptome Following Primary Infection of Blood and
738 Lymphatic Endothelial Cells. Pathog (Basel, Switzerland) [Internet]. MDPI; 2017 Mar
739 19;6(1):11. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28335496
740 31. Bailey L. T, Elkan C. Fitting a mixture model by expectation maximization to discover
741 motifs in biopolymers. In: Proceedings of the Second International Conference on
742 Intelligent Systems for Molecular Biology [Internet]. Menlo Park; 1994. p. 28–36.
743 Available from:
744 https://www.ncbi.nlm.nih.gov/pubmed/?term=Fitting+a+mixture+model+by+expectation+
745 maximization+to+discover+motifs+in+biopolymers
746 32. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif.
747 Bioinformatics [Internet]. 2011 Feb 16;27(7):1017–8. Available from:
748 https://dx.doi.org/10.1093/bioinformatics/btr064
749 33. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of
750 Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for
751 Macrophage and B Cell Identities. Mol Cell [Internet]. 2010;38(4):576–89. Available
752 from: http://www.sciencedirect.com/science/article/pii/S1097276510003667
753 34. Djavadian R, Chiu Y-F, Johannsen E. An Epstein-Barr Virus-Encoded Protein Complex

754 Requires an Origin of Lytic Replication In Cis to Mediate Late Gene Transcription. PLOS
755 Pathog [Internet]. Public Library of Science; 2016 Jun 27;12(6):e1005718-. Available
756 from: https://doi.org/10.1371/journal.ppat.1005718
757 35. Lin CL, Li H, Wang Y, Zhu FX, Kudchodkar S, Yuan Y. Kaposi's sarcoma-associated
758 herpesvirus lytic origin (ori-Lyt)-dependent DNA replication: identification of the ori-Lyt
759 and association of K8 bZip protein with the origin. J Virol [Internet]. American Society
760 for Microbiology; 2003 May;77(10):5578–88. Available from:
761 https://www.ncbi.nlm.nih.gov/pubmed/12719550
762 36. AuCoin DP, Colletti KS, Xu Y, Cei SA, Pari GS. Kaposi's Sarcoma-Associated
763 Herpesvirus (Human Herpesvirus 8) Contains Two Functional Lytic Origins of DNA
764 Replication. J Virol [Internet]. 2002 Aug 1;76(15):7890. Available from:
765 http://jvi.asm.org/content/76/15/7890.abstract
766 37. Yella VR, Bhimsaria D, Ghoshdastidar D, Rodríguez-Martínez JA, Ansari AZ, Bansal M.
767 Flexibility and structure of flanking DNA impact transcription factor affinity for its core
768 motif. Nucleic Acids Res [Internet]. 2018/11/05. Oxford University Press; 2018 Dec
769 14;46(22):11883–97. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30395339
770 38. Deng W, Roberts SGE. A core promoter element downstream of the TATA box that is
771 recognized by TFIIB. Genes Dev [Internet]. Cold Spring Harbor Laboratory Press; 2005
772 Oct 15;19(20):2418–23. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16230532
773 39. Wong E, Wu T-T, Reyes N, Deng H, Sun R. Murine Gammaherpesvirus 68 Open Reading
774 Frame 24 Is Required for Late Gene Expression after DNA Replication. J Virol [Internet].
775 2007 Jun 15;81(12):6761. Available from: http://jvi.asm.org/content/81/12/6761.abstract
776 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

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

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/. Available from: 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

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://dx.doi.org/10.1093/bioinformatics/btu638

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

47. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al.

800 Integrative genomics viewer. Nat Biotechnol [Internet]. Nature Publishing Group, a
801 division of Macmillan Publishers Limited. All Rights Reserved.; 2011 Jan 10;29:24.
802 Available from: https://doi.org/10.1038/nbt.1754
803 48. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-
804 performance genomics data visualization and exploration. Brief Bioinform [Internet].
805 2012 Apr 19;14(2):178–92. Available from: https://dx.doi.org/10.1093/bib/bbs017
806 49. Zhou T, Yang L, Lu Y, Dror I, Dantas Machado AC, Ghane T, et al. DNAshape: a method
807 for the high-throughput prediction of DNA structural features on a genomic scale. Nucleic
808 Acids Res [Internet]. 2013/05/22. Oxford University Press; 2013 Jul;41(Web Server
809 issue):W56–62. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23703209
810

811 **Figure Legends**

812 **Fig 1: The ORF24-RNA pol II interaction is essential for late gene expression and contributes**
813 **to viral DNA replication**

814 iSLK cells harboring WT BAC16, the ORF24$_{RAAAG}$ mutant or the ORF24$_{RAAAG}$ mutant rescue
815 (MR) were reactivated with doxycycline and sodium butyrate for 72 hours. (A) Western blot
816 showing expression of an early protein (ORF59) and a late protein (K8.1) in WT, ORF24$_{RAAAG}$
817 and ORF24$_{RAAAG}$-MR. Vinculin was used as a loading control. (B) Progeny virion production was
818 measured by supernatant transfer on to 293T cells and quantified by flow cytometry. (C) DNA
819 replication of different mutants and MRs was measured by qPCR and normalized to the respective
820 uninduced sample. The fold induction of all mutants was normalized to the WT to get the relative
821 change in DNA replication compared to WT.
822

823 **Fig 2: Differential expression analysis of ORF24-Pol II interaction mutant and mutant**
824 **rescue**
825 iSLK cells harboring $ORF24_{RAAAG}$ and $ORF24_{RAAAG}$ -MR were reactivated with doxycycline and
826 sodium butyrate for 24 and 48 h, whereupon total RNA was extracted and subjected to library
827 preparation and RNA sequencing. (A) Bar plot showing log_2 fold change of the viral genes in the
828 $ORF24_{RAAAG}$ -MR at 48 h relative to 24 h. Genes were color coded by kinetic class as defined in
829 (30). (B) Bar plot showing log_2 fold change of the viral genes in the mutant relative to the MR.
830 Genes were color coded by kinetic class as defined in (30). (C) Heatmap of the scaled log_2 fold
831 change of the genes induced in the MR between 24 to 48 h and genes repressed in the mutant
832 compared to MR at 48 h. Gene clusters with different kinetics and dependence on ORF24-Pol II
833 interaction are indicated. (D) Consensus motif identified by MEME analysis of promoters from 14
834 genes that showed greater than one standard deviation repression in the mutant relative to the MR.
835
836 **Fig 3: ChIP-Seq with HA-ORF34 shows multiple binding sites across the viral genome**
837 (A) ChIP-Seq was performed using an anti-HA antibody to immunoprecipitate DNA prepared
838 from iSLK HA-ORF34 cells reactivated for 48 h with doxycycline and sodium butyrate. iSLK-
839 WT cells lacking the HA tag and input DNA were used as negative controls. The alignment files
840 were converted to the tiled data file (tdf) format and the ChIP background from the untagged cells
841 was subtracted and visualized on the Integrated Genome Viewer (IGV) (B) ChIP-qPCR was
842 performed using an anti-HA antibody to immunoprecipitate DNA prepared from iSLK HA-ORF24
843 cells reactivated for 48 h with doxycycline and sodium butyrate. The DNA was quantified by
844 qPCR with primers specific to the promoter regions of the indicated genes. Data were normalized
845 to the level of input DNA and are presented as percentages of the input. Data shown are from three

846 biological replicates. (C) Consensus motif identified by MEME analysis of promoters from 19
847 genes whose promoters had a ChIP peak.
848
849 **Fig 4: Identifying direct targets of the vPIC**
850 (A) Heatmap showing genes repressed in the ORF24$_{RAAAG}$ mutant compared to the MR adjacent
851 to the list of genes that show ORF34 binding at their promoter by ChIP-Seq analysis. Also
852 indicated are viral genes that were identified to have the TATTWAA motif by FIMO analysis. The
853 direct targets were identified by their combined dependence on the ORF24-Pol II interaction and
854 the presence of a ChIP signal at the promoter. (B) Consensus motif identified by MEME analysis
855 of promoters from 13 direct targets.
856
857 **Fig 5: The motif identified from direct targets is important for transcription from the**
858 **promoter**
859 (A) Schematic of plasmid-based promoter activation assay. iSLK-WT or ORF24$_{RAAAG}$ mutant
860 cells reactivated with doxycycline and sodium butyrate were transfected with reporter plasmid
861 carrying the indicated promoter sequence upstream of the firefly luciferase gene and a control
862 renilla luciferase expressing plasmid. 48 h post transfection and reactivation, the cells were
863 assessed for luciferase activity. (B) Reactivated iSLK-WT cells were transfected with the K8.1
864 promoter reporter plasmid with or without the KSHV left lytic origin of replication cloned in *cis*
865 and assessed for luciferase activity. (C) Reactivated iSLK-WT cells were transfected with the K8.1
866 and ORF57 promoter reporter plasmids in the presence or absence of DNA replication inhibitor
867 (PAA) and assessed for luciferase activity. (D) Reactivated iSLK-WT and iSLK- ORF24$_{RAAAG}$
868 mutant cells were transfected with K8.1 or ORF57 promoter reporter plasmids and assessed for

869 luciferase activity. (E-F) iSLK-WT cells were transfected with plasmids containing the WT K8.1
870 promoter or the indicated single or double point mutant promoters and assessed for luciferase
871 activity. The data for each mutant was normalized to the WT K8.1 promoter. All data shown are
872 average of 3-5 biological replicates.
873
874 **Fig 6: The expanded motif is important for DNA binding by vPIC**
875 (A) Schematic of the transient ChIP assay. iSLK-WT or iSLK-HAORF34 cells reactivated with
876 doxycycline and sodium butyrate were transfected with plasmid carrying the indicated promoter
877 sequence. 48 h post transfection and reactivation, cells were subjected to ChIP with an anti-HA
878 antibody, and the associated DNA was quantified using primers specific to the promoter on the
879 plasmid and the genomic K8.1 promoter locus. (B) iSLK-HAORF34 cells were transfected with
880 K8.1 promoter containing plasmid, mutant plasmids or ORF57 promoter plasmid (as a control).
881 iSLK-WT cells without the HA tag transfected with K8.1 plasmid served as a control for the IP.
882 The ChIP DNA was quantified using plasmid specific primers (left panel) and K8.1 promoter
883 specific primers (right panel), normalized to the level of input DNA and presented as percent input.
884 Data shown are an average of 3-6 biological replicates.
885
886 **Fig 7: The ORF24-34 binding element at the origin is essential for late gene activation**
887 (A) (Top panel) Schematic of the left lytic origin of replication of the KSHV genome. (Bottom
888 panel) IGV image of the ChIP peak for HA-ORF34 at the origin of replication. (B) iSLK-WT cells
889 were transfected with the K8.1 promoter plasmid containing either the full length Ori (WT), the
890 Ori lacking the ORF24-34 binding element (Delta ORF24-34 BE), or the Ori lacking the minimal
891 Ori-lyt and retaining only the ORF24-34 binding element (ORF24-34 BE). 48 h post transfection

892 and reactivation, the cells were assessed for luciferase activity. (C) iSLK-HAORF34 cells were
893 transfected with a K8.1 promoter containing plasmid containing either the full-length Ori (WT),
894 the Ori lacking the ORF24-34 binding element (Delta ORF24-34 BE), or no Ori. Untagged iSLK-
895 WT cells transfected with K8.1 plasmid served as a control for the IP. The ChIP DNA was
896 quantified using plasmid specific primers (left panel) and genomic K8.1 promoter specific primers
897 (right panel), normalized to the level of input DNA and presented as percent input. Data shown
898 are an average of 3-4 biological replicates.
899
900 **Fig S1: Validation of mutant BAC DNA integrity**
901 The integrity of the various BAC mutants used in this study was verified by *Rsr*II digestion.
902
903 **Fig S2: Kinetics of early and late gene transcripts in ORF24$_{RAAAG}$ mutant and MR**
904 iSLK-ORF24$_{RAAAG}$ and MR were reactivated with doxycycline and sodium butyrate for the
905 indicated time points, whereupon total RNA was extracted, converted to cDNA, and quantified by
906 qPCR using primers specific to the CDS of K8.1 (late gene) or ORF57 (early gene). The fold
907 induction was measured relative to uninduced sample.
908
909 **Fig S3: ChIP-Seq data with HA-ORF34**
910 (A) Western blot showing expression of HA-ORF24 and HA-ORF34. iSLK cells harboring WT
911 BAC16, HA-ORF34 BAC or HA-ORF24 BAC were reactivated with doxycycline and sodium
912 butyrate for 72 hours. Protein lysate was immunoprecipitated with HA beads and visualized by
913 SDS-PAGE-Western blot. (B) iSLK cells harboring WT BAC16, HA-ORF34 BAC or HA-ORF24
914 BAC were reactivated with doxycycline and sodium butyrate for 72 hours. Progeny virion

915 production was measured by supernatant transfer on to 293T cells and quantified by flow
916 cytometry. (B) IGV image showing ChIP-Seq data from iSLK-HA-ORF34 cells reactivated with
917 doxycycline and sodium butyrate for 48 h and immunoprecipitated with anti-HA antibody. Also
918 shown are plots for control data from untagged iSLK cells and input DNA for two biological
919 replicates
920
921 **Fig S4: Effect of mutations on DNA structural parameters**
922 Four structural parameters for 40 bp of the WT and expanded motif mutants were calculated using
923 the DNAShape program in R (49). Average of each structural parameter for the mutated base pairs
924 and the two adjacent base pairs at either end (a total of 7 bp) was calculated for each DNA sequence
925 analyzed. The two adjacent base pairs were included in the averaging as some of the structural
926 parameters refer to a base step rather than a specific base. The average for each mutant was
927 normalized to the WT K8.1 promoter.
928
929
930
931
932
933
934

Figure 1

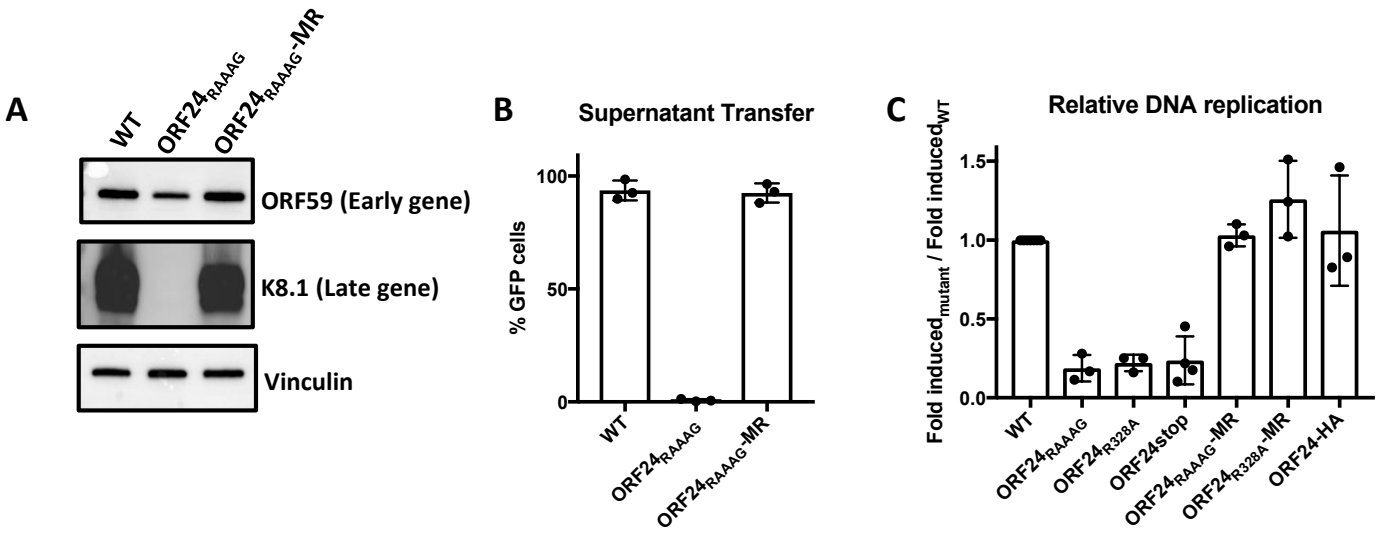

A
WT
ORF24RAAAG
ORF24RAAAG-MR
ORF59 (Early gene)
K8.1 (Late gene)
Vinculin
B
Supernatant Transfer
% GFP cells
100
50
0
WT
ORF24RAAAG
ORF24RAAAG-MR
C
Relative DNA replication
Fold inducedmutant / Fold inducedWT
1.5
1.0
0.5
0.0
WT
ORF24RAAAG
ORF24R328A
ORF24stop
ORF24RAAAG-MR
ORF24R328A-MR
ORF24-HA

Figure S1

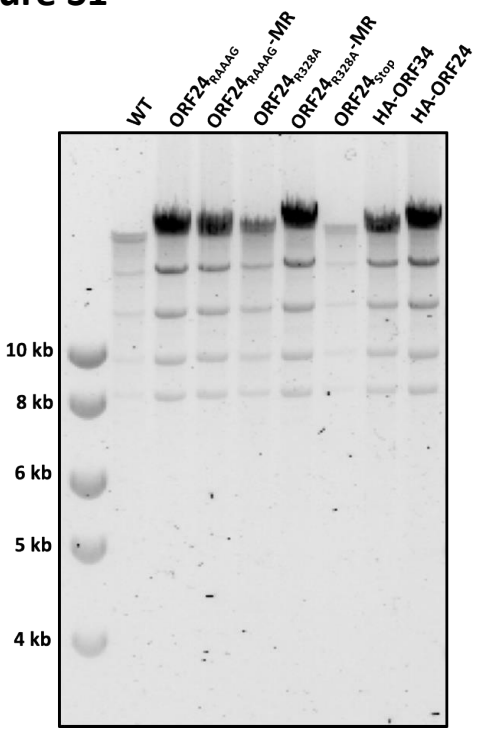

WT
ORF24RAAAG-MR
ORF24RAAAG
ORF24R328A-MR
ORF24R328A
ORF24Stop
HA-ORF34
HA-ORF24
10 kb
8 kb
6 kb
5 kb
4 kb

Figure S2

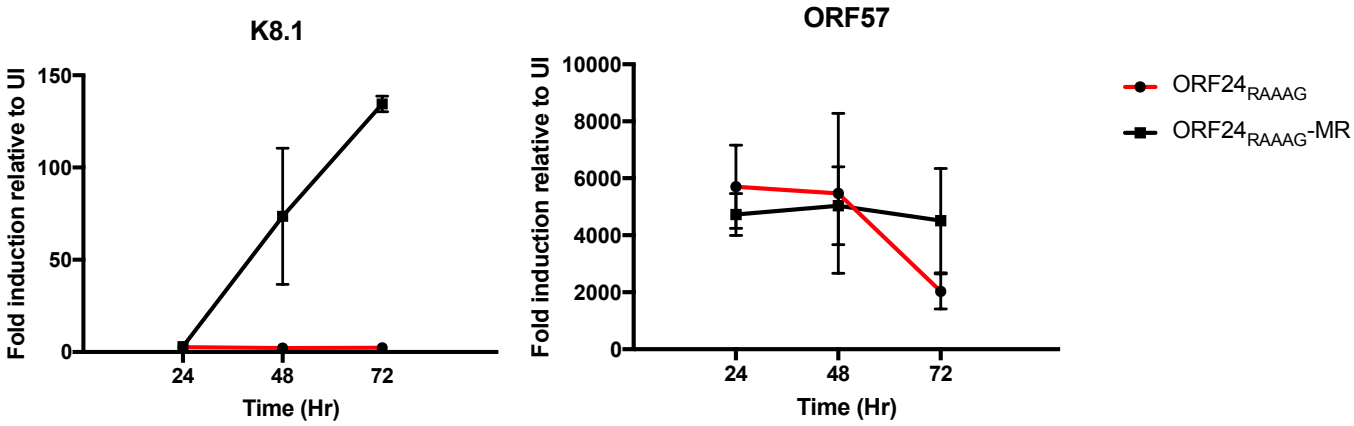
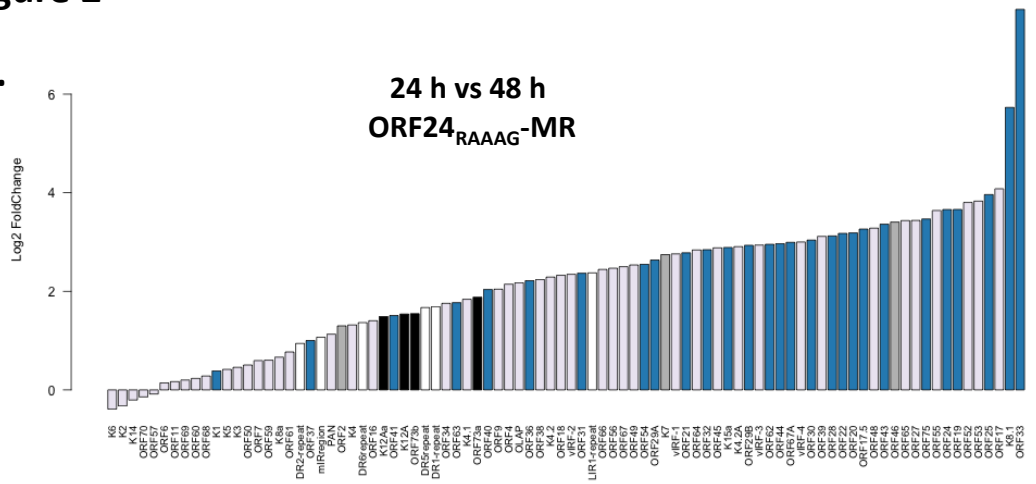
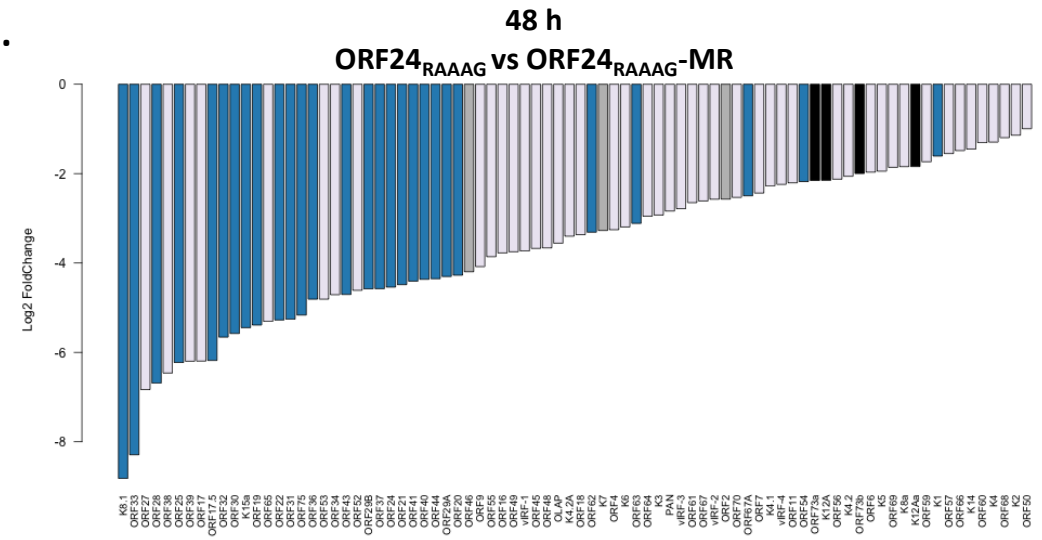

K8.1
Fold induction relative to UI
150
100
50
0
24
48
72
Time (Hr)
ORF57
Fold induction relative to UI
10000
8000
6000
4000
2000
0
24
48
72
Time (Hr)
ORF24RAAAG
ORF24RAAAG-MR

Figure 2

A.

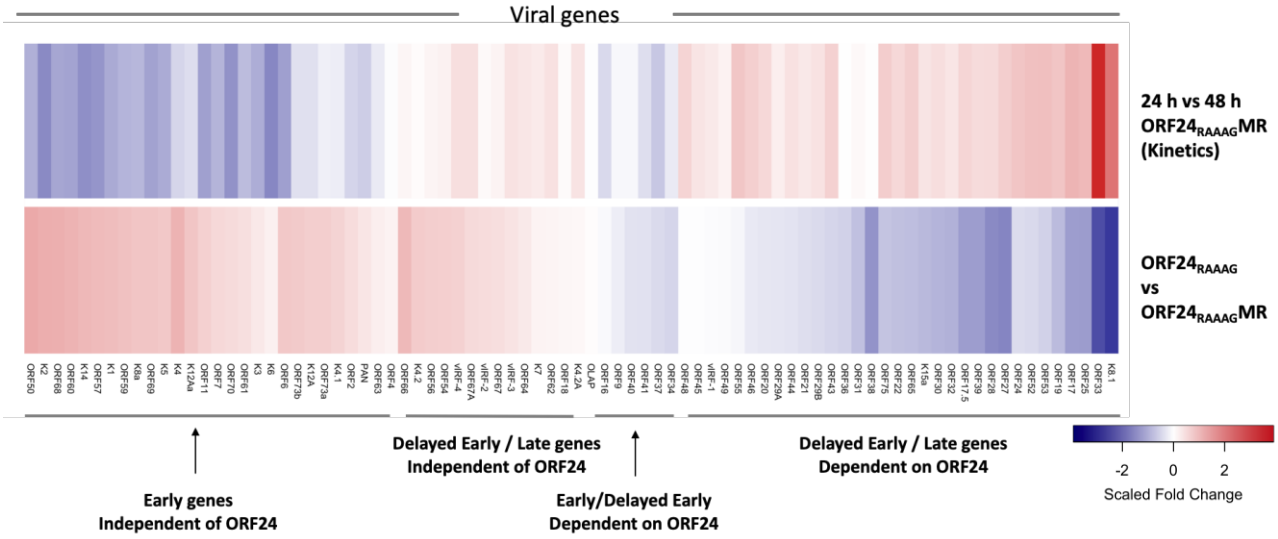


B.



Latent
Early/Delayed Early
Late
Unknown

C.

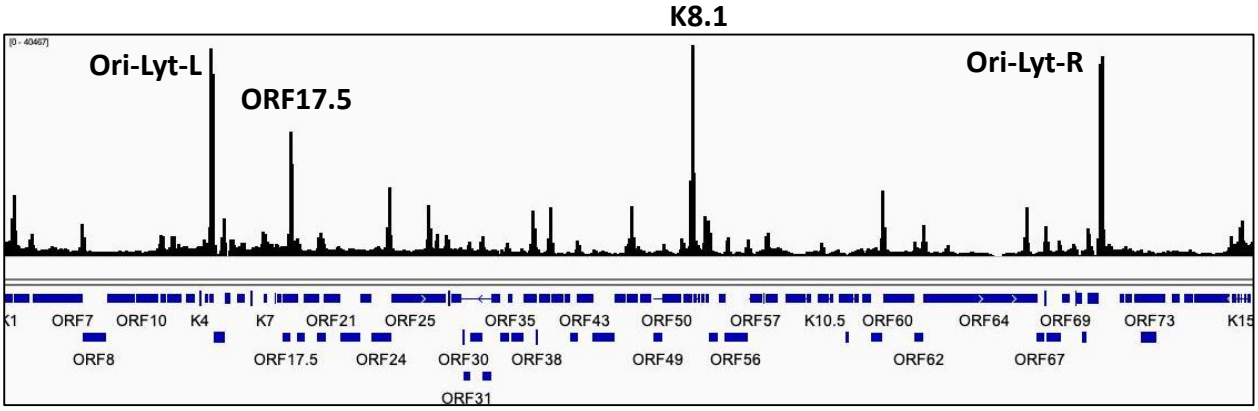


D.

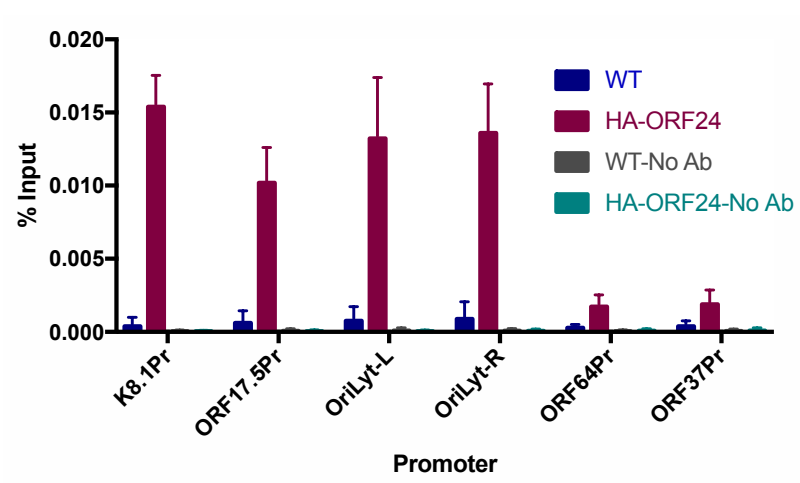
***E*-value:** 3.4e-006

Figure 3

A


K8.1
Ori-Lyt-L
ORF17.5
Ori-Lyt-R
K1
ORF7
ORF10
K4
K7
ORF21
ORF25
ORF35
ORF43
ORF50
ORF57
K10.5
ORF60
ORF64
ORF69
ORF73
K15
ORF8
ORF17.5
ORF24
ORF30
ORF38
ORF49
ORF56
ORF62
ORF67
ORF31

B


0.020
0.015
0.010
0.005
0.000
% Input
WT
HA-ORF24
WT-No Ab
HA-ORF24-No Ab
K8.1Pr
ORF17.5Pr
OriLyt-L
OriLyt-R
ORF64Pr
ORF37Pr
Promoter

C

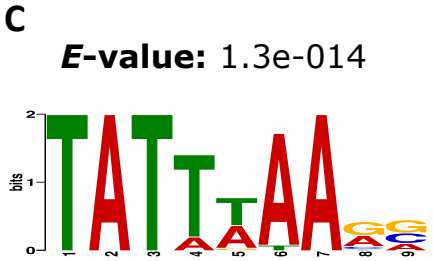
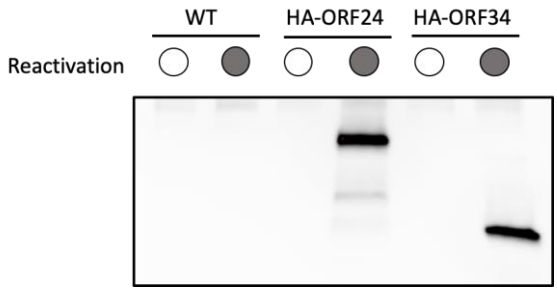
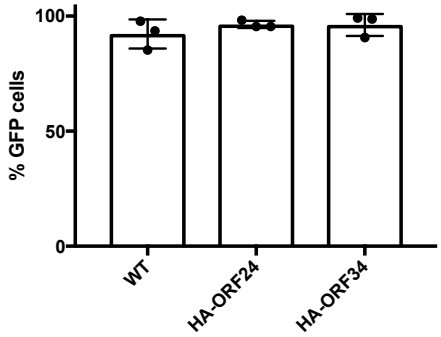

E-value: 1.3e-014
bits
2
1
0
1
2
3
4
5
6
7
8
9

Figure S3

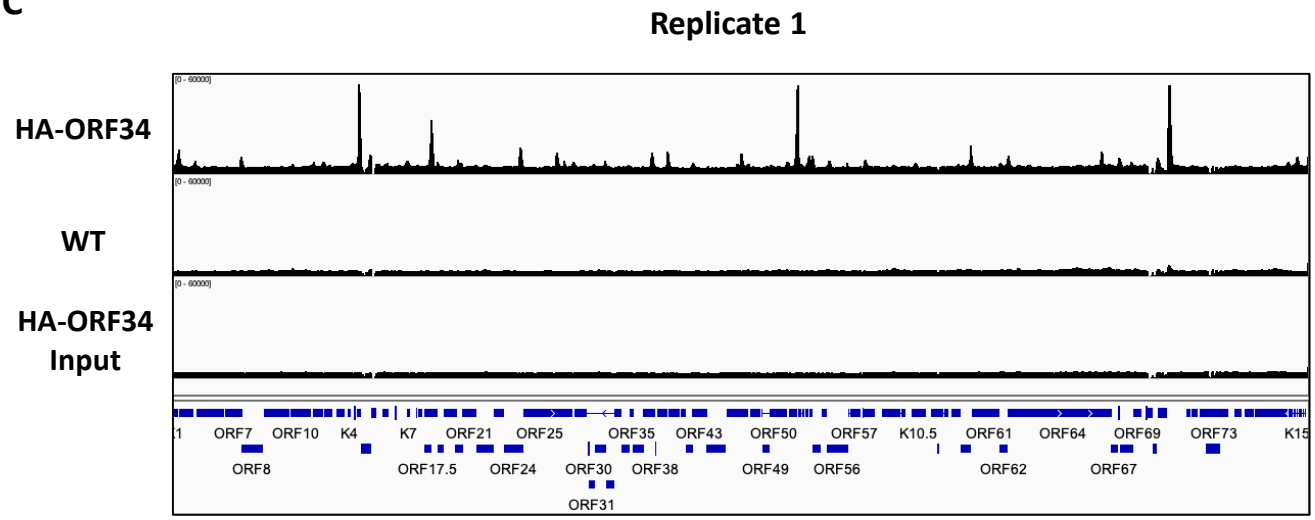
A



B



C



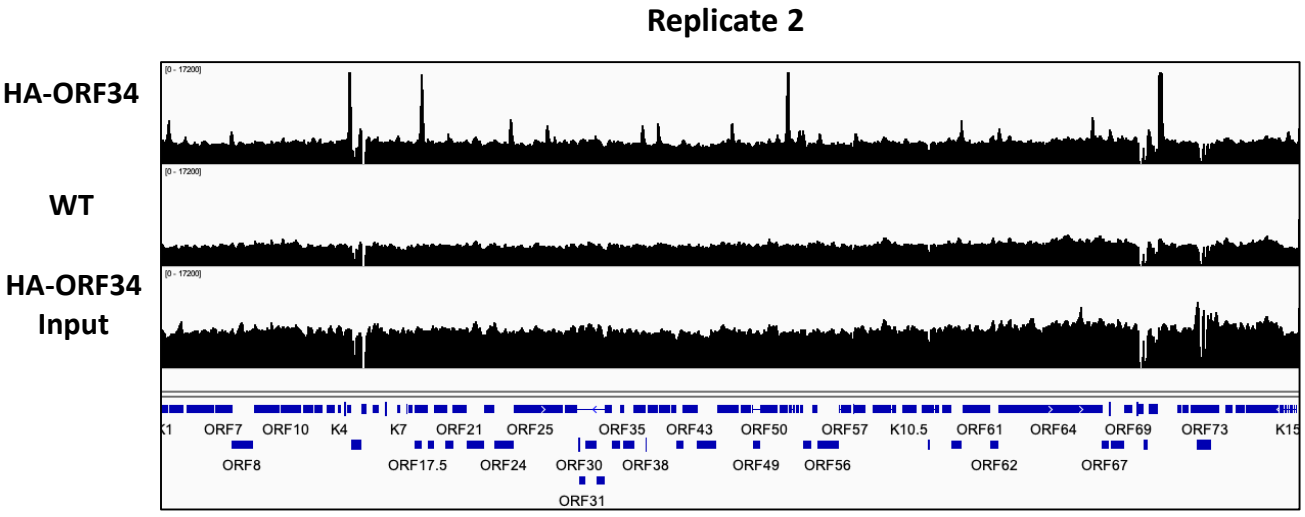
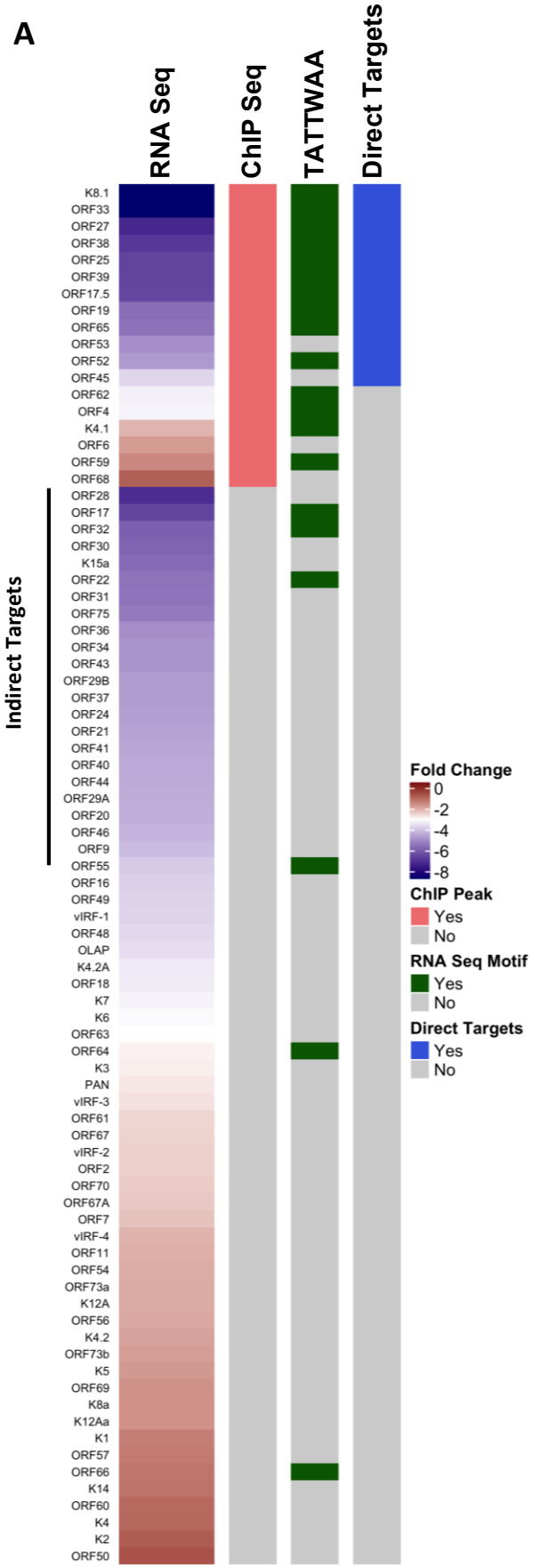


Figure 4

A



B

TATTWAARVNYS

***E*-value:** 5.8e-009

Figure 5

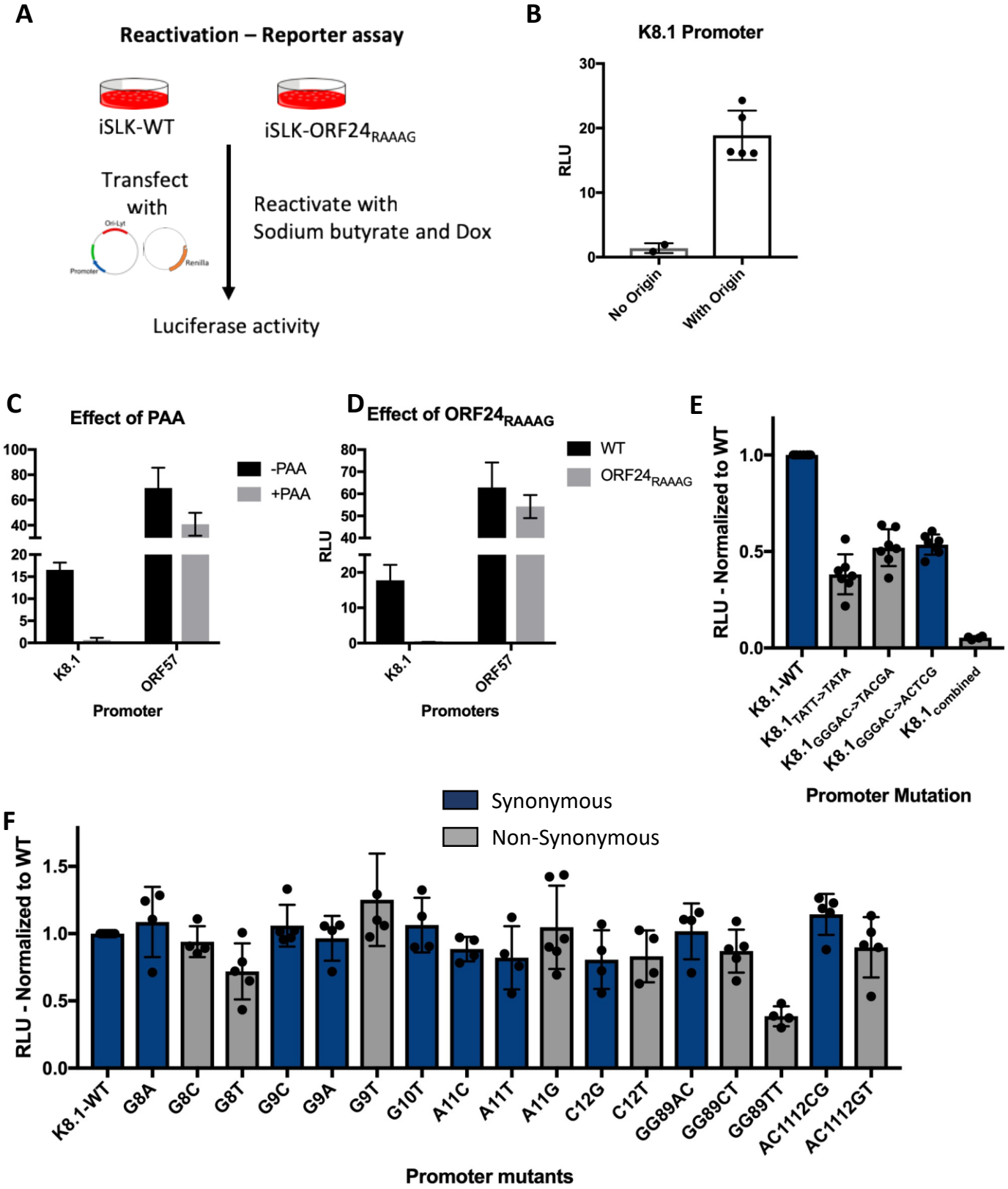
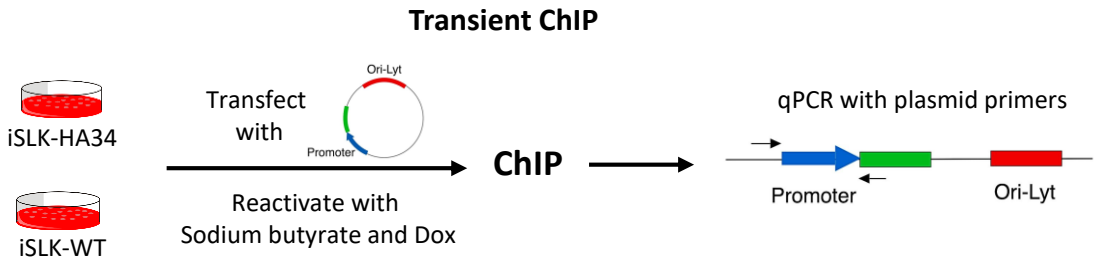

A
Reactivation – Reporter assay
iSLK-WT
iSLK-ORF24RAAAG
Transfect with
Ori-Lyt
Promoter
Renilla
Reactivate with
Sodium butyrate and Dox
Luciferase activity
B
K8.1 Promoter
RLU
30
20
10
0
No Origin
With Origin
C
Effect of PAA
RLU
100
80
60
40
20
15
10
5
0
-PAA
+PAA
K8.1
ORF57
Promoter
D
Effect of ORF24RAAAG
RLU
80
70
60
50
40
20
10
0
WT
ORF24RAAAG
K8.1
ORF57
Promoters
E
RLU - Normalized to WT
1.0
0.5
0.0
K8.1-WT
K8.1TATT->TATA
K8.1GGGAC->TACGA
K8.1GGGAC->ACTCG
K8.1combined
Promoter Mutation
F
Synonymous
Non-Synonymous
RLU - Normalized to WT
1.5
1.0
0.5
0.0
K8.1-WT
G8A
G8C
G8T
G9C
G9A
G9T
G10T
A11C
A11T
A11G
C12G
C12T
GG89AC
GG89CT
GG89TT
AC1112CG
AC1112GT
Promoter mutants

Figure 6

A


Transient ChIP
iSLK-HA34
iSLK-WT
Transfect with
Ori-Lyt
Promoter
Reactivate with
Sodium butyrate and Dox
ChIP
qPCR with plasmid primers
Promoter
Ori-Lyt

B

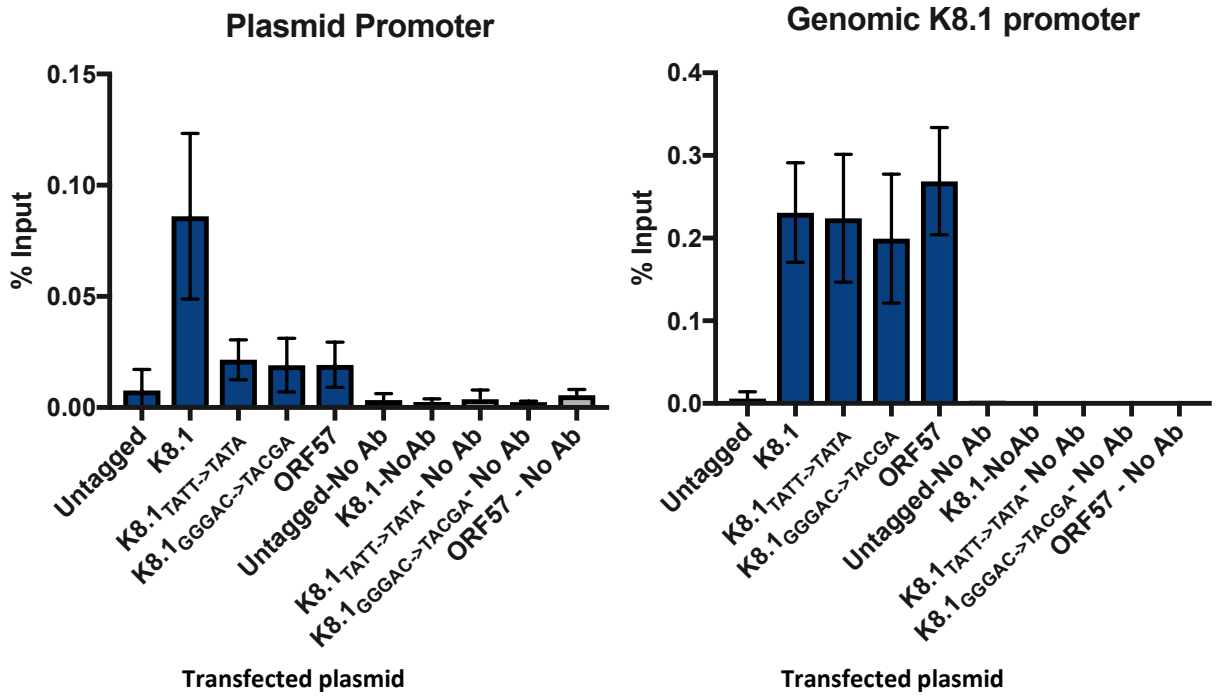
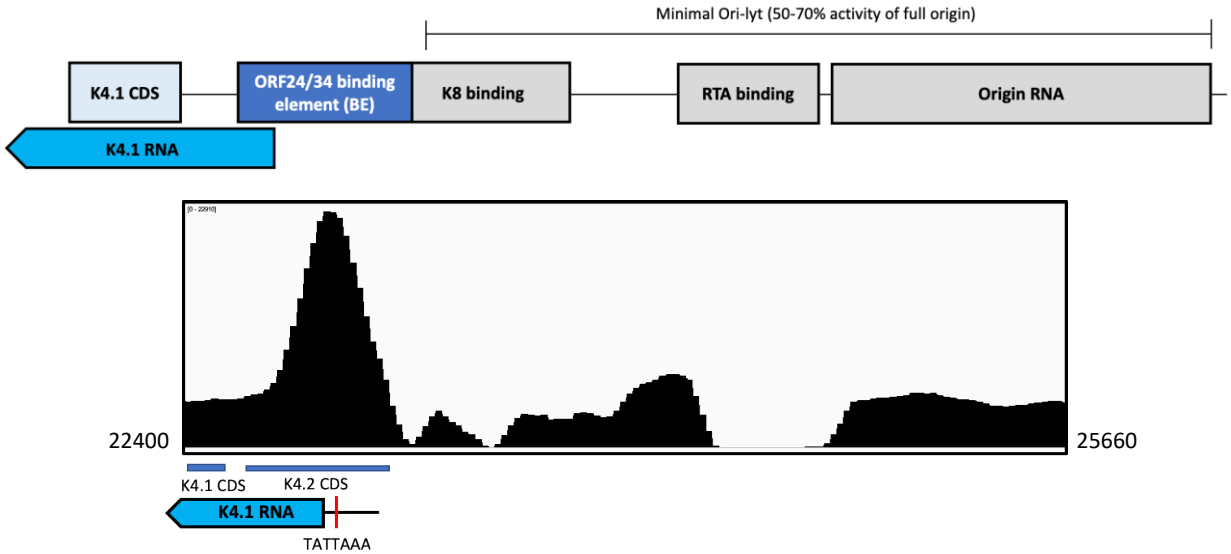
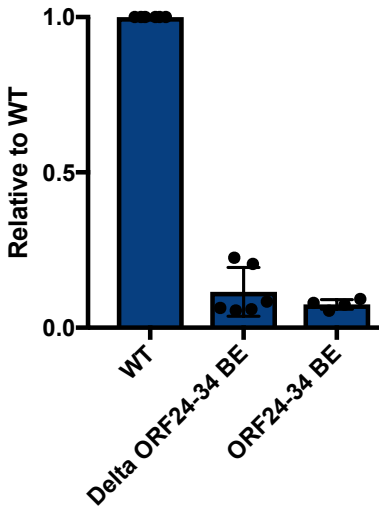

Plasmid Promoter
% Input
0.15
0.10
0.05
0.00
Untagged
K8.1
K8.1TATT->TATA
K8.1GGGAC->TACGA
ORF57
Untagged-No Ab
K8.1-NoAb
K8.1TATT->TATA - No Ab
K8.1GGGAC->TACGA - No Ab
ORF57 - No Ab
Transfected plasmid
Genomic K8.1 promoter
% Input
0.4
0.3
0.2
0.1
0.0
Untagged
K8.1
K8.1TATT->TATA
K8.1GGGAC->TACGA
ORF57
Untagged-No Ab
K8.1-NoAb
K8.1TATT->TATA - No Ab
K8.1GGGAC->TACGA - No Ab
ORF57 - No Ab
Transfected plasmid

Figure 7

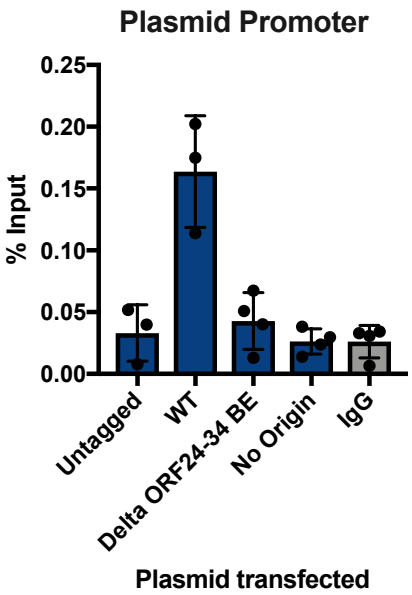
A



B **Luciferase assay**



C **Transient ChIP**



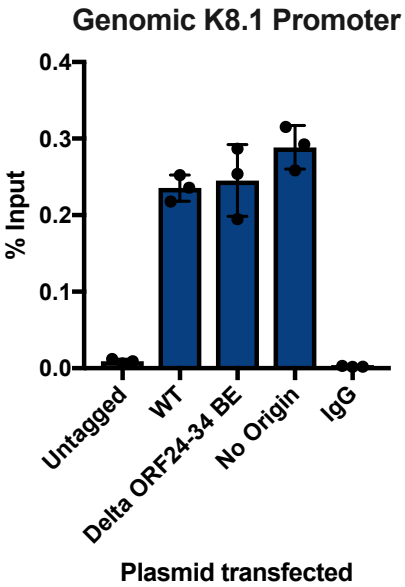


Figure S4

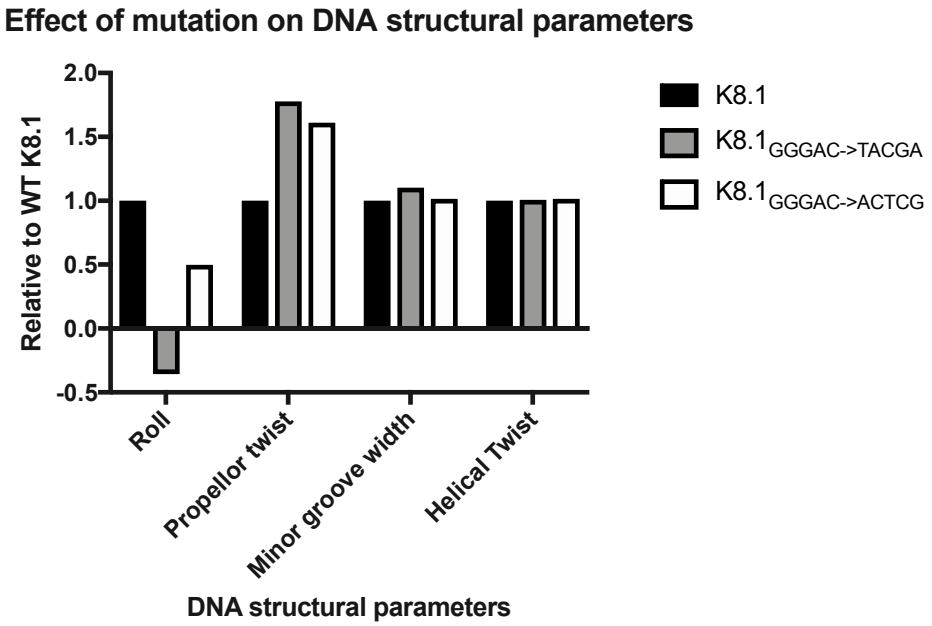

Effect of mutation on DNA structural parameters
Relative to WT K8.1
2.0
1.5
1.0
0.5
0.0
-0.5
Roll
Propellor twist
Minor groove width
Helical Twist
DNA structural parameters
K8.1
K8.1GGGAC->TACGA
K8.1GGGAC->ACTCG

Table 1. Occurrence of TATTWAA motif in KSHV

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

ORFs shown in red do not have a known Transcription Start Site (TSS)

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#)	*TATAAAAGG TATTAAACC
ORF62	Assembly/DNA maturation	L	-	32	TATTTAAAA
ORF33	Capsid binding	L	+	30	TATTAAAGG
ORF8	Glycoprotein	L	+	29	TATTTAAAG
ORF17.5	Viral capsid assembly	L	-	29 (73#)	TATTTAAAG #TATTGAAAA
ORF19	Capsid vertex		-	113	TATTTAAGA
ORF26	Minor capsid protein	L	+	29	TATTAAAGC
ORF45	Immune evasion (Tegument)	E	-	No motif	No motif
ORF6	Single strand DNA binding protein	E	+	89 (31#)	*TATATAAAC #TATTTTAGG
K4.1	Homologs of cellular chemokines	E	-	32	TATTAAACC
ORF59	DNA polymerase processivity factor	E	-	29	TATTTAAGG
ORF68	Packaging protein	E	+	No motif	No motif
ORF4	Immune evasion	E	+	29	TATTAAAAG
ORF65	Minor capsid protein	L	-	31	TATTAAAGC
ORF38	Viral maturation and egress (Tegument protein)	L	+	29	TATTAAAGC

Second site with lower consensus to TATTWAARV motif
Direct targets are shown in gray
* Promoters with TATA instead of the TATT sequence

Supplementary Table 1:
List of all DNA sequences used in the study

Name	Sequence	Used for
K8.1Pr_pGL4.16_F	TGGCCTAACTGGCCGGTACCTGTGCCGTCGTCCGGGA	Cloning
K8.1Pr_pGL4.16_R	CCGGATTGCCAAGCTTAGGATTAAGGGATTAACTTCGGTCCC	Cloning
ORF57Pr_pGL4.16_F	TGGCCTAACTGGCCGGTACCAGTGTAACAATAATGTTCCCACGGC	Cloning
ORF57Pr_pGL4.16_R	CCGGATTGCCAAGCTTTTTTGGGAACCTGGCAGCCAGGTTA	Cloning
Orilyt_KSHV_F	GGTCTGACAGCGGCCGCCAGCTTATCCGGTGGCAATCT	Cloning
Orilyt_KSHV_R	AGCACTGGTAACCACTGCAGTGGGCACAGAGATATGTGACGTGACGCCCT	Cloning
K8.1_G8A_1	GGCAGCAATATTAAAAGGACCGAAGTTAATC	Mutagenesis
K8.1_G8A_2	GATTAACTTCGGTCCTTTTAATATTGCTGCC	Mutagenesis
K8.1_G8C_1	GGCAGCAATATTAAACGGACCGAAGTTAATC	Mutagenesis
K8.1_G8C_2	GATTAACTTCGGTCCGTTTAATATTGCTGCC	Mutagenesis
K8.1_G8T_1	GGCAGCAATATTAAATGGACCGAAGTTAATC	Mutagenesis
K8.1_G8T_2	GATTAACTTCGGTCCATTTAATATTGCTGCC	Mutagenesis
K8.1_G9C_1	GCAGCAATATTAAAGCGACCGAAGTTAATCC	Mutagenesis
K8.1_G9C_2	GGATTAACTTCGGTCGCTTTAATATTGCTGC	Mutagenesis
K8.1_G9A_1	GCAGCAATATTAAAGAGACCGAAGTTAATCC	Mutagenesis
K8.1_G9A_2	GGATTAACTTCGGTCTCTTTAATATTGCTGC	Mutagenesis
K8.1_G9T_1	GCAGCAATATTAAAGTGACCGAAGTTAATCC	Mutagenesis
K8.1_G9T_2	GGATTAACTTCGGTCACTTTAATATTGCTGC	Mutagenesis
K8.1_G10T_1	CAGCAATATTAAAGGTACCGAAGTTAATCCC	Mutagenesis
K8.1_G10T_2	GGGATTAACTTCGGTACCTTTAATATTGCTG	Mutagenesis
K8.1_A11C_1	AGCAATATTAAAGGGCCCGAAGTTAATCCCT	Mutagenesis
K8.1_A11C_2	AGGGATTAACTTCGGGCCCTTTAATATTGCT	Mutagenesis
K8.1_A11T_1	AGCAATATTAAAGGGTCCGAAGTTAATCCCT	Mutagenesis
K8.1_A11T_2	AGGGATTAACTTCGGACCCTTTAATATTGCT	Mutagenesis
K8.1_A11G_1	AGCAATATTAAAGGGGCCGAAGTTAATCCCT	Mutagenesis
K8.1_A11G_2	AGGGATTAACTTCGGCCCCTTTAATATTGCT	Mutagenesis
K8.1_C12G_1	GCAATATTAAAGGGAGCGAAGTTAATCCCTT	Mutagenesis
K8.1_C12G_2	AAGGGATTAACTTCGCTCCCTTTAATATTGC	Mutagenesis
K8.1_C12T_1	GCAATATTAAAGGGATCGAAGTTAATCCCTT	Mutagenesis
K8.1_C12T_2	AAGGGATTAACTTCGATCCCTTTAATATTGC	Mutagenesis
K8.1_GG89AC_1	GGCAGCAATATTAAAACGACCGAAGTTAATCCCTT	Mutagenesis
K8.1_GG89AC_2	AAGGGATTAACTTCGGTCGTTTTAATATTGCTGCC	Mutagenesis
K8.1_GG89CT_1	GGCAGCAATATTAAACTGACCGAAGTTAATCCCTT	Mutagenesis
K8.1_GG89CT_2	AAGGGATTAACTTCGGTCAGTTTAATATTGCTGCC	Mutagenesis
K8.1_GG89TT_1	GGCAGCAATATTAAATTGACCGAAGTTAATCCCTT	Mutagenesis
K8.1_GG89TT_2	AAGGGATTAACTTCGGTCAATTTAATATTGCTGCC	Mutagenesis

K8.1_AC1112CG_1	GGCAGCAATATTAAAGGGCGCGAAGTTAATCCCTTAATC	Mutagenesis
K8.1_AC1112CG_2	GATTAAGGGATTAACTTCGCGCCCTTTAATATTGCTGCC	Mutagenesis
K8.1_AC1112GT_1	GGCAGCAATATTAAAGGGGTCGAAGTTAATCCCTTAATC	Mutagenesis
K8.1_AC1112GT_2	GATTAAGGGATTAACTTCGACCCCTTTAATATTGCTGCC	Mutagenesis
K8.1_TATT->TATA_1	GTTATCCGGCAGCAATATAAAAGGGACCGAAGTTAATC	Mutagenesis
K8.1_TATT->TATA_2	GATTAACTTCGGTCCCTTTTATATTGCTGCCGGATAAC	Mutagenesis
K8.1_GGGAC->TACGA_1	GGCAGCAATATTAAATACGACGAAGTTAATCCCTT	Mutagenesis
K8.1_GGGAC->TACGA_2	AAGGGATTAACTTCGTCGTATTTAATATTGCTGCC	Mutagenesis
K8.1_GGGAC->ACTCG_1	GGCAGCAATATTAAAACTCGCGAAGTTAATCCCTT	Mutagenesis
K8.1_GGGAC->ACTCG_2	AAGGGATTAACTTCGCGAGTTTTAATATTGCTGCC	Mutagenesis
K8.1Pr_qPCR_F	GGGAGAACCATGCCAGACTTTG	ChIP-qPCR
K8.1Pr_qPCR_R	GCATAGGATTAGGAGCGCCAC	ChIP-qPCR
ORF59Pr_qPCR_F	AATCCACAGGCATGATTGC	ChIP-qPCR/ DNA replication
ORF59Pr_qPCR_R	CACACTTCCACCTCCCCTAA	ChIP-qPCR DNA replication
CPSF6Pr_qPCR_F	CGAGGAATGTCCCTGTTTGT	DNA replication
CPSF6Pr_qPCR_R	ACTGGCTGTCTCCTCTCAGC	DNA replication
K8.1CDS_qPCR_F	CCGTCGGTGTGTAGGGATAAAG	RT-qPCR
K8.1CDS_qPCR_R	GTCGTTGTAGTGGTGGCAGAAA	RT-qPCR
ORF57CDS_qPCR_F	TTTGACGAATCGAGGGACGACG	RT-qPCR
ORF57CDS_qPCR_R	GCAGTTGAGAACGACCTTGAGAT	RT-qPCR
ORF17.5Pr_qPCR_F	GCTGGCCTTTAAATACGCCG	ChIP-qPCR
ORF17.5Pr_qPCR_R	TTTCGTCAGTCCCCTGGAGA	ChIP-qPCR
OriLyt_L_qPCR_F	CGGCGGAGGTTGTCAATAGT	ChIP-qPCR
OriLyt_L_qPCR_R	GGGTCCTGTTCTGGGATAGC	ChIP-qPCR
OriLyt_R_qPCR_F	CCCTGGGCACGATAACCATA	ChIP-qPCR
OriLyt_R_qPCR_R	AGGGGCGCAAAACAATTGTG	ChIP-qPCR
ORF64Pr_qPCR_F	GCAGAAAGAACTTGAGCTCT	ChIP-qPCR
ORF64Pr_qPCR_R	AGGCAGACCGGGAGTACGTT	ChIP-qPCR
ORF37Pr_qPCR_F	GCTGGCGTTTCAGAAGCAGTG	ChIP-qPCR
ORF37Pr_qPCR_R	CCATCCAGGGTGTCAACCA	ChIP-qPCR
K8.1Pr_DoubleMutant_1	GTTATCCGGCAGCAATATAAAATACGACGAAGTTAATCCCTTAATCC	Mutagenesis
K8.1Pr_DoubleMutant_2	GGATTAAGGGATTAACTTCGTCGTATTTTATATTGCTGCCGGATAAC	Mutagenesis