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. Together, these results identify an elusive element that contributes to vPIC specificity and suggest novel links between KSHV DNA replication and late transcription.

Author summary

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

Introduction

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

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

53 Several studies suggest that the expression of late genes is also uniquely regulated in β- and
54 γ-herpesviruses, which include human cytomegalovirus (HCMV), Epstein-Barr virus (EBV), and
55 Kaposi's sarcoma-associated herpesvirus (KSHV) (6). Unlike the eukaryotic viruses described
56 above, β and γ herpesviruses do not engage cellular TATA binding protein (TBP) during late
57 gene pre-initiation complex formation. Their core late gene promoters are unusually small (12-15
58 base pairs (bp)), with the only recognizable conserved element being a modified TATA box
59 (TATT) ~30 bp upstream of the transcription start site (TSS) (7–9). This is in contrast to
60 eukaryotic and most early herpesviral promoters, which have multiple promoter elements that are
61 important for promoter recognition and activation by RNA polymerase II (RNA pol II) (10).
62 Another notable feature is that activation of these minimal promoters requires a virally-encoded
63 six-protein complex called the viral Pre-Initiation Complex (vPIC) that coordinates mammalian
64 RNA pol II (11,12). The vPIC is presumed to be uniquely adapted to activate minimal promoters,
65 yet the 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 (13–15). 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
71 18, 30, 31, 34, and 66, but little is known about how each of these factors contributes to vPIC
72 activity (14,16–18). ORF34 interacts with multiple components of the vPIC, and is thus thought
73 to function as a scaffolding factor (12). The ORF18 homolog in HCMV (pUL79) has been
74 shown to function as an elongation factor, although whether it plays a similar role in KSHV and
75 EBV is unclear (19). Despite the paucity of functional information on the individual vPIC
76 components, protein-protein interaction mapping has provided insight into the overall complex
77 organization (12,18). Furthermore, point mutants that disrupt individual inter-subunit vPIC
78 interactions abrogate late gene transcription, highlighting the fact that vPIC complex integrity is
79 critical for function (12,20,21).

80 The mechanistic basis for promoter selectivity has also been challenging to grasp in the
81 context of the minimal, largely nondescript nature of late gene promoters. While a modified
82 TATA box with a TATT sequence is enriched in several late promoters of KSHV, MHV68, and
83 EBV (7–9,22), the ORF24 TBP mimic shows partial binding to a mutant promoter in which the
84 TATT sequence is modified to resemble a canonical TATA box (7). However, ORF24 does not
85 bind promoters that inherently have a canonical TATA box (7), 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
89 have 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 three key
101 findings that provide insight into late gene promoter activation in KSHV. First, we defined the
102 set of direct vPIC promoter targets within the KSHV genome and showed that other promoters
103 are indeed nonproductively bound or only indirectly regulated by the vPIC. Second, by
104 restricting our analysis to these direct targets, we revealed a novel promoter motif that
105 contributes significantly to vPIC binding and thus provides insight into features governing vPIC
106 specificity. Notably, this motif did not emerge when the indirect vPIC targets were included in
107 the analysis. Finally, we found that the KSHV TBP mimic ORF24 is required for efficient viral
108 DNA replication. This was unexpected given that the other studied vPIC components have been
109 shown to be dispensable for this process, and may have implications for the connection between
110 viral DNA replication and late gene transcription.

111

112 **Results**

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

114 Three leucine residues in the N-terminal domain of KSHV ORF24 (amino acids 73-75) are
115 essential for its interaction with RNA Pol II and its ability to promote late gene transcription

116 (14). Given that most viral proteins are multifunctional, we engineered the ORF24$_{RLLLG->RAAAG}$
117 mutation (ORF24$_{RAAAG}$) into KSHV BAC16 to precisely evaluate the impact of the
118 transcriptional function of ORF24 during lytic infection. We also generated the corresponding
119 mutant rescue (ORF24$_{RAAAG}$-MR) to ensure that any observed phenotypes were not the effect of
120 secondary mutations in the BAC. Identical banding patterns from *Rsr*II digestion of WT and
121 mutant BAC16 DNAs verified that there were no gross rearrangements of the viral genome
122 during mutagenesis (Figure S1-lanes1-3). We established stable cell lines latently infected with
123 WT, ORF24$_{RAAAG}$, and ORF24$_{RAAAG}$-MR KSHV in the iSLK background, which contains a
124 doxycycline-inducible version of the KSHV lytic transactivator protein ORF50 (RTA) to enable
125 lytic reactivation (29). During lytic reactivation, these three viruses expressed similar levels of
126 the early protein ORF59, while the ORF24$_{RAAAG}$ mutant was selectively unable to express the
127 K8.1 late protein or produce infectious virions (Fig 1A, 1B). These findings are consistent with
128 previous complementation data (14), and confirm the importance of the ORF24-Pol II interaction
129 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 (12,14,16–18). This defect was rescued in the
133 ORF24$_{RAAAG}$-MR cell line, indicating that it was not due to secondary mutations in the BAC
134 (Fig. 1C). However, to independently confirm this observation, we engineered two additional
135 ORF24 viral mutants: the point mutant ORF24$_{R328A}$, which abrogates its essential interaction
136 with the ORF34 component of the vPIC (12), and a stop mutant (ORF24$_{Stop}$) that has a premature
137 stop codon in the ORF24 coding sequence. Inclusion of the ORF24$_{Stop}$ virus allowed us to
138 evaluate whether the point mutants were functioning in a dominant negative manner, for example

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

162 Combining the above ORF24$_{RAAAG}$ mutant and MR induction data into a heatmap confirmed
163 that in general, genes with later expression kinetics tended to be more severely impacted by the
164 ORF24$_{RAAAG}$ mutation compared to genes with early expression kinetics (Figure 2C). (The fold
165 change data was scaled to allow for easy visualization). However, not all genes fit this model.
166 Some genes categorized as early based on their induction pattern showed a moderate dependence
167 on the ORF24-Pol II interaction (e.g. ORFs 9, 40, 34, 41, 16 and 37). Similarly, a subset of genes
168 that show late kinetics were less dependent on the ORF24-Pol II interaction (e.g. ORFs 67A, 62,
169 K4.2A, 64, K7, 54, 67, 56, 66, K4.2 and vIRFs-2, 3 and 4). Classification as early or late and
170 dependence on the ORF24-Pol II interaction was determined using mean log_2 fold change.
171 Together, these data suggest that not all kinetic late genes are regulated by the vPIC and there
172 may be subset of early genes that are direct or indirect targets of the vPIC.
173

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

176 We then looked for motifs enriched in the promoters of genes most strongly
177 downregulated in the ORF24$_{RAAAG}$ mutant using Multiple EM for Motif Elicitation (MEME)
178 (31), a program that discovers un-gapped motifs that are enriched in a given set of sequences.
179 We defined the promoter sequence as 100 bp upstream from the transcription start site (TSS) or
180 300 bp upstream from the translation start site when the TSS was unknown. Analysis of
181 promoter sequences from 14 of the most strongly down-regulated genes (Log_2 fold change
182 greater than one standard deviation from the mean) showed that TATTWAA is the only motif
183 that is enriched in this set of sequences (Figure 2D).

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

194 Notably, a subset of the genes containing the TATTWAA motif showed less than average
195 repression by the ORF24$_{RAAAG}$ mutation (ORF55, ORF64, ORF66, ORF59, K4.1, ORF4, ORF8,
196 ORF58, ORF26 and ORF62), suggesting that the motif may not be sufficient to mark regulation
197 by the vPIC, which is consistent with a similar observation in MHV68 (7). Furthermore, several
198 genes (ORF28, ORF30, K15, ORF31 and ORF75) were strongly affected by the mutation (Log_2
199 fold change greater than one standard deviation from the mean) but lack the TATTWAA motif.
200 The number of genes in this set increases to 26 if we broaden the criteria to include all the genes
201 that show greater than average repression in the presence of the mutation. These genes could
202 either be down regulated by an indirect mechanism in the absence of the ORF24-Pol II
203 interaction or could be bound by the vPIC in a manner independent of the TATT motif.

204

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

206 Our RNA Seq experiments identified KSHV genes whose expression is influenced by the
207 transcription regulatory function of ORF24. However, RNA Seq does not distinguish genes that
208 are directly bound by the vPIC from those that are indirect transcriptional targets (e.g. controlled
209 by other factors whose expression is induced by ORF24). We therefore sought to identify
210 promoters bound by the KSHV vPIC complex using ChIP Seq. Previous studies suggest that
211 ORF34 acts as a scaffolding factor that brings together all components of the vPIC (12)
212 including ORF24, which binds both RNA Pol II and TATT-containing DNA. We reasoned that
213 ORF34 promoter occupancy would therefore be a good proxy for vPIC assembly, and
214 additionally would avoid potential pitfalls associated with any non-productive or independent
215 interactions between ORF24 and DNA.

216 To facilitate ChIP Seq, we added an N-terminal HA tag to endogenous ORF34 in KSHV
217 BAC16 (HA-ORF34), and generated a stable, doxycycline-inducible iSLK cell line containing
218 this tagged virus as previously described. HA-ORF34 produced WT levels of infectious virions
219 in a supernatant transfer assay (Figure S3A), indicating that the tag does not disrupt ORF34
220 activity. The integrity of the BAC DNA was verified by *Rsr*II digestion (Figure S1-Lane7). We
221 performed the ChIP Seq experiment in duplicate at 48 h post reactivation to ensure active late
222 gene expression. Reactivated iSLK cells with untagged BAC16 served as a negative control for
223 the immunoprecipitation. The individual ChIP peaks for the two replicates as well as the
224 respective inputs and controls are shown in Figure S3B and supplementary file 1. The ChIP Seq
225 data showed that HA-ORF34 bound several locations across the viral genome (Figure 3A) but
226 was not detectably associated with the host genome. A peak calling algorithm (HOMER) (33)
227 identified 25 common peaks in the viral genome across two replicates that were significantly
228 enriched in HA-ORF34 relative to untagged WT or input DNA (Supplementary File 1 and Figure

229 S3B). Notably, 20 of the 25 peaks were found upstream of ORFs, with ~80% of the peaks
230 centered less than 200 bp from the closest TSS (Supplementary File 1). Given the link between
231 DNA replication and late transcription in herpesviruses, it was particularly interesting that two of
232 the most prominent peaks are upstream of the two minimal lytic origins of replication in the
233 KSHV genome (Figure 3A). One of the peaks could not be uniquely assigned to a single ORF
234 and was not considered further. Of the remaining 19 ORFs associated with peaks, 13 are
235 involved in packaging, capsid assembly (or form part of capsid itself), 4 are involved in immune
236 evasion and 2 in DNA replication (Table 2). While several of them have also been previously
237 classified as late genes (12/20), it is notable that 6 of them have been classified as early genes,
238 again suggesting that not all targets of the vPIC are late genes.

239 To confirm that the observed binding pattern was representative of vPIC assembly, we
240 independently validated several prominent ChIP Seq peaks by ChIP-qPCR from cells infected
241 with KSHV containing HA-tagged ORF24 (HA-ORF24). Similar to HA-ORF34, the HA-ORF24
242 virus produced WT levels of infectious virions (Figure S3A). Indeed, the HA-ORF24 ChIP-
243 qPCR data were consistent with the HA-ORF34 ChIP-Seq data (Figure 3B), with peaks showing
244 higher percent input values compared to control regions (ORF64 and ORF37 promoters).

245 A MEME analysis of the promoters associated with the peaks again showed that the only
246 motif enriched in these promoters is TATTWAARV (Figure 3C). 17 of the 19 promoters (except
247 ORF45 and ORF68) have this motif and in a majority of the sequences, it is present ~30 bp from
248 the TSS (Table 2). Two key points can be drawn from the ChIP seq data. First, several genes
249 whose promoters contain a TATTWAA motif do not show ORF34 binding (e.g. ORFs 17, 32,
250 22, 55, 58, 64 and 66). Second, while most promoters bound by ORF34 have a TATTWAA
251 motif, there are some exceptions (ORF45, ORF68 and ORF53). These observations indicate that

252 while the TATTWAA motif is enriched in promoters regulated by the vPIC, it is not sufficient to
253 ensure binding of this specialized transcription complex.
254
255 **Thirteen KSHV genes are direct targets of the vPIC**
256 Combining the RNA Seq and ChIP Seq data, we defined the direct targets of the vPIC as
257 genes that are both downregulated in the absence of the ORF24-Pol II interaction and show
258 ORF34 binding in the promoter region. By this measure, a total of 13 genes are direct targets of
259 the vPIC (Figure 4 and Table 2 – gray rows). Of the 13 direct targets, 11 code for structural
260 proteins or proteins involved in viral egress while two are tegument proteins involved in immune
261 evasion. Genes whose expression was impaired in the KSHV ORF24$_{RAAAG}$ mutant but lacked
262 ORF34 binding at their promoters are likely to be indirect targets of the vPIC (Figure 4A).
263 Finally, it is interesting to note that not all genes that have a ChIP peak are down regulated in the
264 RNA Seq data (e.g. ORFs 4, 6, K4.1, 59, 62 and 68). These could be false positive ChIP peaks
265 and/or have alternate mechanisms of regulation that ensure their transcription even in the
266 absence of the vPIC.
267
268 **The direct vPIC targets reveal an expanded promoter motif**
269 A defining feature of characterized late gene promoters is that they are minimalistic, with
270 no conserved motifs beyond the TATTWAA signature recognized by ORF24. Yet, our data
271 indicate that the TATTWAA motif is insufficient to confer activation by the late gene
272 transcription complex. We hypothesized that the inclusion of potentially indirect vPIC targets
273 into prior motif searches may have hindered identification of other regulatory sequences within
274 these promoters. Indeed, repeating the MEME analysis on the 13 direct targets revealed that an

275 additional 5 bp motif (RVNYS) downstream of the TATTWAA motif was enriched in these
276 promoter sequences (Figure 4B).

277 To explore the contribution of the 5 bp region downstream of the TATTWAA sequence
278 for regulation by the vPIC, we used a plasmid-based promoter activation assay in which
279 expression of a luciferase reporter was driven by either the K8.1 late or ORF57 early promoter
280 (Figure 5A). The KSHV left lytic origin of replication (ori-lyt) was included on the plasmids, as
281 data from MHV68, KSHV and EBV indicate that the presence of ori-lyt in *cis* significantly
282 boosts late gene promoter activation during infection (7,9,34). Each plasmid was transfected
283 together with a plasmid constitutively expressing renilla luciferase (to normalize for transfection
284 efficiency) into lytically reactivated KSHV-positive iSLK cells and the luciferase signal was
285 measured 48 h post transfection and reactivation. As expected, the presence of ori-lyt enhanced
286 K8.1 promoter-driven luciferase expression by 15 – 20-fold (Figure 5B). Furthermore, unlike the
287 consistently robust ORF57 promoter activity, K8.1 promoter activity on the plasmid was
288 completely repressed in the presence of DNA replication inhibitor PAA (Figure 5C) or in cells
289 lytically infected with the ORF24$_{RAAAG}$ mutant (Figure 5D). Taken together, these data confirm
290 that the plasmid assay accurately recapitulates late gene promoter activity.

291 Notably, mutating the 5 bp RVNYS motif in the K8.1 promoter decreased transcription
292 by > 50% (Figure 5E). This decrease was comparable to mutating the T at position four to an A
293 in the TATTWAA sequence, which has been shown to be important (although not sufficient) for
294 recognition and binding by ORF24 (9). Replacement of the 5 bp RVNYS motif with either
295 synonymous (-S) or nonsynonymous (-NS) bases yielded a similar decrease in promoter activity
296 (Figure 5E-K8.1$_{GGGAC\text{->}TACGA\text{-}S}$, K8.1$_{GGGAC\text{->}ACTCG\text{-}NS}$), suggesting a role for the overall
297 architecture of this motif rather than the primary sequence. In support of this hypothesis, a

298 preliminary analysis of four structural features of the expanded motif in the WT and mutant
299 predict that the average roll of the DNA in the expanded motif is lower in the mutant promoter
300 compared to the WT promoter (Figure S4). Furthermore, with the exception of GG89TT-NS,
301 systematic single and double base substitutions in the RVNYS motif were well tolerated, arguing
302 that this region may not be contributing to base specific contacts with the vPIC (Figure 5F). In
303 summary, while the expanded RVNYS motif appears important for transcriptional activation, we
304 hypothesize it may contribute to vPIC specificity by impacting the local topology of the
305 promoter DNA rather than through base specific contacts.
306
307 **The expanded promoter motif is important for DNA binding by vPIC**
308 The RVNYS motif could either contribute to vPIC binding to its target promoters, as is
309 the case for the TATTWAA sequence, or it could modulate downstream events involved in
310 transcription initiation at vPIC bound promoters. To distinguish these possibilities, we used ChIP
311 to measure HA-ORF34 occupancy on the luciferase plasmids containing WT or mutant K8.1
312 promoter in lytically reactivated HA-ORF34 iSLK cells. We engineered qPCR primers that
313 enabled distinction of the K8.1 promoter on the plasmid DNA from the endogenous viral locus
314 (Figure 6A).
315 Notably, both the K8.1$_{TATT\text{->}TATA}$ and K8.1$_{GGGAC\text{->}ACTCG}$ mutations similarly impaired
316 HA-ORF34 binding to the plasmid promoter DNA (Figure 6B- left panel). The ChIP signal was
317 specific for HA-ORF34, as we observed only a background signal when the assay was performed
318 in the WT iSLK cells containing an untagged ORF34 locus. Furthermore, there was comparable
319 vPIC assembly at the genomic K8.1 locus in cells transfected with WT or the mutant plasmids,
320 confirming similar immunoprecipitation efficiencies across samples (Figure 6B - right panel).

321 Together, these data indicate that the expanded promoter motif is important for vPIC assembly
322 on the promoter DNA.
323
324 **Discussion**
325
326 Late gene expression is uniquely regulated in KSHV by a set of viral proteins collectively
327 known as the vPIC. In this study, we generated a comprehensive list of direct targets of the vPIC
328 by combining transcriptome and genome-wide binding data of the complex. Among the direct
329 targets, we identified an expanded promoter recognition motif that contributes to stable binding
330 of the vPIC on the promoter. This is particularly useful to understanding alternate mechanisms of
331 transcription initiation by RNA pol II because of the unique ability of the vPIC to directly recruit
332 RNA Pol II to highly compact promoters.
333 Several key points emerge from independent analysis of the RNA Seq and ChIP Seq data.
334 First, the previously identified TATTWAA motif is insufficient to mark binding and regulation
335 by the vPIC, as demonstrated by the multiple ORFs that have the motif in their promoter (23/83
336 ORFs) but are not repressed in the $ORF24_{RAAAG}$ mutant (10/23 ORFs) or bound by ORF34 (7/23
337 ORFs). Second, the TATTWAA motif is not a pre-requisite for regulation or binding by the
338 vPIC. For example, although ORF53 expression is dependent on the ORF24-Pol II interaction
339 and its promoter is bound by ORF34, its promoter does not have the conserved TATTWAA
340 sequence but a TATATAA sequence in its place. Third, several genes that exhibit late kinetics
341 escape repression in the mutant, indicating that not all late genes are direct or indirect targets of
342 vPIC. It is notable that several of the kinetic late genes that escape repression are involved in
343 immune regulation, consistent with a similar observation in EBV (28). In particular, EBV BPLF1
344 is a late gene whose expression is independent of the vPIC, and we observe that its homolog

345 ORF64 also shows a similar pattern in KSHV. Finally, while there are several early genes that
346 are indirect targets of the vPIC, only 1 of 13 direct targets exhibited early kinetics (ORF45). It is
347 notable that ORF45 is a tegument protein, which might explain the requirement for its continued
348 regulation at later times of infection.

349 Our integrative analysis revealed that a distinct 5 bp motif is integral to vPIC promoter
350 recognition. Rather than engaging in base specific contacts (as is the case for the TATTWAA
351 sequence) (14), we instead hypothesize that the importance of this motif lies in its shape or
352 structure. In this regard, a preliminary analysis of basic structural properties of the expanded
353 motif suggest that the average roll (representing the rotational flexibility) of the DNA in the
354 expanded motif is lower in the 5 bp mutant promoter compared to WT K8.1 promoter. Structure-
355 based recognition and a contribution of base pairs flanking the core binding site to binding
356 affinity is a common feature among transcription factors, although the specific structural
357 properties that influence binding differ for various classes of factors (35). A previous study of the
358 bZIP family of transcription factors identified a correlation between the roll of DNA sequences
359 that flank the core recognition element and binding affinity (35). Consistent with that, we
360 observed a defect in the assembly of the vPIC when the expanded motif was mutated, suggesting
361 a stabilizing role for the flanking sequence.

362 The location of the expanded motif is reminiscent of the downstream recognition element
363 of the general transcription factor TFIIB (BRE^d) (36). Interestingly, the BRE^d element also
364 shows low sequence conservation and does not require an active promoter sequence to conform
365 with the consensus sequence at all positions (36), as seen with the expanded motif. Interaction of
366 TFIIB with the BRE^d element is thought to stabilize the TFIIB-TBP interaction on the promoter
367 DNA, leading to activation or repression of the promoter in a context-dependent manner. While

368 there is some evidence to suggest that TFIIB is present at late gene promoters (14), whether the
369 interaction between the vPIC and the expanded motif is mediated by TFIIB or any other factors
370 remains to be elucidated.

371 Notably, we found that KSHV DNA replication was reduced in the absence of functional
372 ORF24, including with an ORF24 null mutant or ORF24 point mutants that fail to bind either
373 RNA Pol II or the vPIC component ORF34. This was surprising, given that deletion or mutation
374 of several other vPIC components has been shown not to affect DNA replication (12,14,17,18).
375 Furthermore, we previously failed to observe a replication defect in an ORF24 null virus (14).
376 We hypothesize that the discrepancy for the ORF24 null mutant relates to the methods used to
377 establish the infected cell lines. In our earlier study, the viral BAC DNA was directly transfected
378 into iSLK cells, resulting in only a ~5-7-fold increase in genome copy number upon reactivation.
379 It is likely that the majority of viral DNA was present in a non-reactivatable state, providing a
380 level of background that significantly reduced the dynamic range for DNA replication
381 measurements. Here, the iSLK cells were established through co-culture with reactivated 293T
382 cells harboring the BAC DNA, i.e. through infection rather than transfection. This reproducibly
383 yields ~50-70-fold increases in genome copy number upon lytic reactivation, more closely
384 representing actual infection. That said, the role of ORF24 in viral DNA replication may be
385 unique to KSHV, as this link has not been observed upon depletion of its homologs in EBV and
386 MHV68 (28,37).

387 One possibility for how KSHV ORF24 might contribute to DNA replication is through its
388 indirect transcriptional targets. We observed that components of the helicase/primase complex
389 (ORFs 40, 41), which are involved in DNA replication, show more than average repression in the
390 absence of functional ORF24 (Figure 2). Importantly, these genes were not repressed in previous

391 studies using stop mutants of the vPIC components ORF18, ORF30 and ORF31 (17,18), raising
392 the possibility that ORF24 has additional gene regulatory functions outside the context of the
393 vPIC. For example, our observed binding of ORF24 and ORF34 at the lytic origins of replication
394 could facilitate DNA replication. Consistent with this hypothesis, two previous studies that
395 defined the boundaries of the lytic origin of replication observed a decrease in replicated DNA
396 upon deletion of the ORF24-ORF34 binding site identified here (38,39). In theory, this
397 interaction could also contribute to the link between DNA replication and late gene activation,
398 and thus exploring the functional consequences of origin binding will be an exciting area for
399 future study.
400
401 **Methods**
402 **Cell lines**
403 HEK293T cells were maintained in DMEM (Invitrogen) + 10% Fetal Bovine Serum
404 (FBS). iSLK BAC16 cells (29) were maintained in DMEM + 10% FBS and 0.5 mg/mL
405 hygromycin B. HEK293T cells stably expressing ORF24-3xFlag were maintained in DMEM
406 (Invitrogen) + 10% Fetal Bovine Serum (FBS) and 325 ug/ml of zeocin. iSLK-puro cells were
407 maintained in DMEM + 10% FBS.
408
409 **Plasmids**
410 A 100 bp fragment of the K8.1 promoter or ORF57 promoter was generated by PCR from
411 the BAC16 genome and cloned into the KpnI and HindIII sites of the pGL4.16 reporter plasmid
412 using Infusion cloning (Clontech). The left origin of replication was similarly generated by PCR
413 from the BAC16 genome and cloned into the NotI and BstXI sites of the K8.1Pr-pGL4.16

414 plasmid. The promoter mutants were generated by site-directed mutagenesis using KAPA HiFi
415 DNA polymerase using the primers shown in Supplementary table 1. To make the ORF24-
416 3xFlag pJLM1 plasmid for lentiviral transduction, the ORF24-3xFlag fragment was generated by
417 PCR using ORF24-3xFlag-pCDNA4 as the template and cloned in to pLJM1 plasmid digested
418 with AgeI and EcoRI. All primers used for cloning and mutagenesis are listed in Supplementary
419 table 1.
420

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

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

427 To establish iSLK cell lines latently infected with the modified virus, HEK293T cells
428 were transfected with 5-10 ug of the modified BAC16 using linear polyethylenimine (PEI, MW
429 ~25,000) at a 1:3 DNA:PEI ratio. The following day, the 293T cells were mixed 1:1 with iSLK-
430 puro cells (which contain inducible RTA but lack KSHV) and treated with 30 nM 12-O-
431 Tetradecanoylphorbol-13-acetate (TPA) and 300 nM sodium butyrate for 4 days to induce lytic
432 replication. Cells were then grown in selection media containing 300 ug/ml hygromycin B, 1
433 ug/ml puromycin and 250 ug/ml G418. The hygromycin concentration was gradually increased
434 to 500 ug/ml and then 1 mg/ ml until all the HEK293Ts cells died. To generate the $ORF24_{RAAAG}$,
435 $ORF24_{R328A}$, and $ORF24_{Stop}$ (which are defective in infectious virion production), HEK293T
436 cells that stably express ORF24-3xFlag were used for the initial transfection and co-culture.

437

438 **Supernatant transfer assay**

439 For supernatant transfer assays to assess virion production, BAC16-containing iSLK cells
440 were plated at a density of $1x10^6$ cells in 10 cm dishes and treated with 1 ug/mL doxycycline and
441 1 mM sodium butyrate the following day. 72 h post induction, the supernatant was filtered
442 through a 0.45 uM syringe filter and 2 ml of the supernatant was mixed with $1x10^6$ freshly
443 trypsinized HEK293T cells in a 6 well plate and centrifuged at 1,500x*g* for 2 h at 37˚C. The
444 following day, cells were trypsinized, fixed in 4% paraformaldehyde and the percentage of cells
445 expressing GFP was determined by flow cytometry (BD Csampler, BD Biosciences).

446

447 **Viral genome replication**

448 To measure viral genome replication, iSLK-BAC16 cells were reactivated for 72 h as
449 described above. The cells were then scraped into the media and the media + cells were digested
450 with proteinase K (80µg/mL) (Promega) in 5x proteinase K digestion buffer (50mM Tris-HCl
451 pH 7.4, 500mM NaCl, 5mM EDTA, 2.5% SDS) overnight at 55˚C. The gDNA was isolated
452 using Zymo Quick gDNA Miniprep Kit according to the manufacturer's instructions.
453 Quantitative PCR (qPCR) was performed on the isolated DNA using iTaq Universal SYBR
454 Green Supermix on a QuantStudio3 Real-Time PCR machine. DNA levels were quantified using
455 relative standard curves with primers specific for KSHV ORF59 promoter and human CPSF6
456 promoter (Supplementary table 1). The relative genome numbers were normalized to CPSF6 to
457 account for loading differences and to uninduced samples to account for differences in starting
458 genome copy number.

459

460 **RT-qPCR**

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

467

468 **Western blotting**

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

476

477 **RNA Seq sample preparation and analysis**

478 Total RNA was isolated from reactivated iSLK cells at indicated time points using Zymo
479 Direct-Zol RNA kit following the manufacturer's protocol. The samples were ribo-depleted and
480 stranded libraries were prepped by the QB3 Berkeley core sequencing facility. Multiplex
481 sequencing was performed using the Illumina HiSeq 4000 to generate 100 bp paired end reads.
482 The quality of the raw reads was checked using FastQC (42) and adapters were trimmed using

483 TrimGalore (43) using default parameters. The trimmed reads were aligned to both the viral
484 (NC_009333.1) and host genomes (hg19) using STAR aligner (ver2.5.3a) (44). The aligned
485 reads were counted using htseq-counts (45) and differential expression was assessed using
486 DESeq2 (46).

487 To compensate for the problem of quantitative assessment of transcripts due to the
488 overlapping transcripts in the KSHV genome, the primary counts of the transcripts were
489 calculated as recommended in Bruce et al (30) with the following small modifications. Length
490 correction was not applied to allow for differential expression analysis by DESeq2. Transcripts
491 for which the counts became negative in all 3 replicates when calculating primary counts were
492 removed before running DESeq2. Bar plots showing Log_2 Fold Change were generated using R.
493 The heatmap was generated using the heatmap.2 package in R. The fold change values were
494 scaled by column before plotting to ease visualization of the data.

495

496 **Chromatin Immunoprecipitation (ChIP)**

497 ChIP was performed on 3*15-cm plates of iSLK-cells (WT KSHV, $KSHV_{HA\text{-}ORF34}$ and
498 $KSHV_{ORF24\text{-}HA}$) reactivated for 48 h with 1 ug/ml dox and 1 mM sodium butyrate. Cells were
499 crosslinked in 2% formaldehyde for 15 min at room temperature, quenched in 0.125 M glycine,
500 and washed twice with ice-cold PBS. Crosslinked cell pellets were mixed with 1 ml ice-cold
501 ChIP lysis buffer (5mM PIPES pH 8.0, 85mM KCl, 0.5% NP-40) and incubated on ice for 10
502 min, whereupon the lysate was dounce homogenized to release nuclei and spun at 1700 g for 5
503 min at 4°C. Nuclei were then resuspended in 800 µl of nuclei lysis buffer (50mM Tris-HCl pH
504 8.0, 0.3% SDS, 10mM EDTA) and rotated for 10 min at 4°C followed by sonication using a
505 QSonica Ultrasonicator with a cup horn set to 25 amps for 20 min total (5 min on, 5 min off).

506 Chromatin was spun at 16,000 x *g* for 10 min at 4°C and the pellet was discarded. The chromatin
507 was precleared with protein A + protein G beads blocked with 5 ug/ml glycogen, 50 ug/ml BSA,
508 100 ug/ml Ecoli tRNA in 1ml dilution buffer (1.1% Triton-X-100, 1.2mM EDTA, 16.7mM Tris-
509 HCl pH 8, 167mM NaCl) for 2 hours.

510 350 µl of chromatin was diluted in ChIP dilution buffer (16.7 mM Tris-HCl pH 8.0, 1.1%
511 Triton X-100, 1.2 mM EDTA, 167 mM NaCl) to 500 µl and incubated with 10 µg anti-HA
512 antibody (Cell Signaling C29F4) overnight, whereupon samples were rotated with 20 µl pre-
513 blocked protein A + G beads (Thermofisher) for 2 h at 4°C. Beads were washed with low salt
514 immune complex (20 mM Tris pH 8.0, 1% Triton-x-100, 2 mM EDTA, 150 mM NaCl, 0.1%
515 SDS), high salt immune complex (20 mM Tris pH 8.0, 1% Triton-x-100, 2 mM EDTA, 500 mM
516 NaCl, 0.1% SDS), lithium chloride immune complex (10mM Tris pH 8.0, 0.25 M LiCl, 1% NP-
517 40, 1% Deoxycholic acid, 1 mM EDTA), and Tris-EDTA for 10 min each at 4°C with rotation.
518 DNA was eluted from the beads using 100 µl of elution buffer (150 mM NaCl, 50 µg/ml
519 Proteinase K) and incubated at 50 °C for 2 h, then 65 °C overnight. DNA was purified using the
520 Zymo PCR purification kit. For ChIP-qPCR, purified DNA was quantified by qPCR using iTaq
521 Universal SYBR Mastermix (BioRad) and the indicated primers (Supplementary table 1). Each
522 sample was normalized to its own input.

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

527

528 **ChIP Seq Analysis**

529 The quality of the raw reads was checked using FastQC (42) and adapters were trimmed
530 using TrimGalore (43) using default parameters. The trimmed reads were aligned to both the
531 viral and host genomes (hg19) using Bowtie2 aligner using default parameters. The aligned reads
532 were filtered for uniquely mapped reads using SAMTools. The filtered reads were converted to
533 tdf format using IGV Tools and visualized on Integrated Genome Viewer (47,48). We used
534 HOMER (33) with the following minor adjustments to default parameters to call peaks for the
535 immunoprecipitated and input samples from cells that have HA tagged ORF34 and untagged
536 ORF34 (WT) as a control. The fold over input and local fold change were set to 1.3 as
537 recommended for smaller and denser viral genomes (personal communication with Chris Benner,
538 UCSD). The peak size and fragment length were also set at 200 and 150 respectively. From the
539 resulting peaks, only the ones that were present in both conditions (HA-ORF34 vs HA-Input,
540 HA-ORF34 vs WT) and both replicates were retained for further analysis. The common peaks
541 identified were verified visually on the Integrated Genome Viewer and were manually annotated.
542 To identify the direct targets, transcripts that had a Log_2 fold change lower than the mean
543 Log_2 fold change in the $ORF24_{RAAAG}$ mutant vs MR were considered to be down regulated in the
544 mutant. Although ORF8 and ORF26 were not included in the differential expression analysis,
545 they were considered to be downregulated because of negative values for their levels in the
546 mutant and not in the mutant rescue on calculation of primary transcript levels.
547
548 **Promoter Analyses**
549 We defined the promoter sequence as 100 base pairs (bp) upstream from the transcription
550 start site (TSS) or 300 bp upstream from the translation start site when the TSS was unknown.
551 The TSS data was obtained from Table 1 in (23) and a custom python script was used to extract

552 the sequences. These were then run through MEME (31) allowing for multiple occurrences of a
553 motif in a sequence and only searching the given strand to identify enriched motifs. The
554 identified motif was then used to search for similar motifs across all 83 promoter sequences from
555 the viral genome, again only allowing for searching the given strand.
556
557 **Reporter assays**
558 $5*10^5$ iSLK-BAC16 (WT or ORF24$_{RAAAG}$) cells were plated in a 6 well plate 24 h before
559 transfection and reactivation. The cells were reactivated with 1 μg/ml doxycycline and 1 mM
560 Sodium Butyrate and simultaneously transfected with 0.95 μg luciferase reporter plasmid and
561 0.05 μg renilla plasmid using polyjet transfection reagent. 48 h post transfection and reactivation,
562 the cells were rinsed twice with room temperature PBS, lysed by rocking for 15 min at room
563 temperature in 300 μl of Passive Lysis Buffer (Promega), and clarified by centrifuging at 14,000
564 g for 1 min. 20 μl of the clarified lysate was added in triplicate to a white chimney well
565 microplate (Greiner bio-one) to measure luminescence on a Tecan M1000 using a Dual
566 Luciferase Assay Kit (Promega). The firefly luminescence was normalized to the internal Renilla
567 luciferase control. The induced sample was then normalized to the uninduced (but transfected)
568 control to look at the fold induction during reactivation.
569
570 **Transient ChIP-qPCR**
571 $3.5x10^6$ iSLK-BAC16 were plated in a 15 cm dish 24 h before transfection and
572 reactivation. The cells were reactivated with 1 μg/ml doxycycline and 1 mM Sodium Butyrate
573 and simultaneously transfected with 8 μg of reporter plasmid using polyjet transfection reagent.
574 48 h post reactivation and transfection, cells were crosslinked and lysed as mentioned above.

575 Nuclei were sonicated using a Covaris S220 (Covaris) for 5 min with the recommended
576 sonication settings (Power: 140 W, Bursts per cycle: 200, Duty cycle: 5%). The chromatin was
577 pre-cleared as mentioned above and the concentration measured using a qubit. 25 μg of pre-
578 cleared chromatin was used in the IP and the following steps were same as mentioned in the
579 ChIP protocol above.
580

581 **Acknowledgements**

582 We thank all members of the Glaunsinger and Coscoy labs for their helpful suggestions.
583 In particular, we thank Angelica Casteñeda for providing the HEK293T cells stably expressing
584 ORF24-3xFlag and Chris Benner (UCSD) for suggestions on proper usage of HOMER for
585 organisms with small genomes. This work was supported by NIH R01 AI122528 to B.G., who is
586 also a Howard Hughes Medical Institute Investigator.
587

588 **References**

589 1. Payne S. Chapter 28 - Introduction to DNA Viruses. In: Payne S, editor. Viruses
590 [Internet]. Academic Press; 2017. p. 231–6. Available from:
591 http://www.sciencedirect.com/science/article/pii/B9780128031094000283
592 2. Grula MA, Buller PL, Weaver RF. α-Amanitin-Resistant Viral RNA Synthesis in Nuclei
593 Isolated from Nuclear Polyhedrosis Virus-Infected Heliothis zea
594 Larvae and Spodoptera frugiperda Cells. J Virol [Internet]. 1981
595 Jun 1;38(3):916. Available from: http://jvi.asm.org/content/38/3/916.abstract
596 3. Kochetkov SN, Rusakova EE, Tunitskaya VL. Recent studies of T7 RNA polymerase
597 mechanism. FEBS Lett [Internet]. 1998;440(3):264–7. Available from:

598 http://www.sciencedirect.com/science/article/pii/S0014579398014847
599 4. Rong M, He B, McAllister WT, Durbin RK. Promoter specificity determinants of T7
600 RNA polymerase. Proc Natl Acad Sci U S A [Internet]. The National Academy of
601 Sciences; 1998 Jan 20;95(2):515–9. Available from:
602 https://www.ncbi.nlm.nih.gov/pubmed/9435223
603 5. Young CSH. The Structure and Function of the Adenovirus Major Late Promoter. In:
604 Doerfler W, Böhm P, editors. Adenoviruses: Model and Vectors in Virus-Host
605 Interactions: Virion-Structure, Viral Replication and Host-Cell Interactions [Internet].
606 Berlin, Heidelberg: Springer Berlin Heidelberg; 2003. p. 213–49. Available from:
607 https://doi.org/10.1007/978-3-662-05597-7_8
608 6. Gruffat H, Marchione R, Manet E. Herpesvirus Late Gene Expression: A Viral-Specific
609 Pre-initiation Complex Is Key. Front Microbiol [Internet]. Frontiers Media S.A.; 2016 Jun
610 6;7:869. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27375590
611 7. Wong-Ho E, Wu T-T, Davis ZH, Zhang B, Huang J, Gong H, et al. Unconventional
612 Sequence Requirement for Viral Late Gene Core Promoters of Murine Gammaherpesvirus
613 68. Frueh K, editor. J Virol [Internet]. 2014 Mar 15;88(6):3411. Available from:
614 http://jvi.asm.org/content/88/6/3411.abstract
615 8. Serio TR, Cahill N, Prout ME, Miller G. A Functionally Distinct TATA Box Required for
616 Late Progression through the Epstein-Barr Virus Life Cycle. J Virol [Internet]. 1998 Oct
617 1;72(10):8338. Available from: http://jvi.asm.org/content/72/10/8338.abstract
618 9. Tang S, Yamanegi K, Zheng Z-M. Requirement of a 12-Base-Pair TATT-Containing
619 Sequence and Viral Lytic DNA Replication in Activation of the Kaposi's Sarcoma-
620 Associated Herpesvirus K8.1 Late Promoter. J Virol [Internet]. 2004 Mar 1;78(5):2609.

621 Available from: http://jvi.asm.org/content/78/5/2609.abstract
622 10. Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription
623 initiation. Nat Rev Mol Cell Biol [Internet]. 2018;19(10):621–37. Available from:
624 https://doi.org/10.1038/s41580-018-0028-8
625 11. Aubry V, Mure F, Mariamé B, Deschamps T, Wyrwicz LS, Manet E, et al. Epstein-Barr
626 Virus Late Gene Transcription Depends on the Assembly of a Virus-Specific Preinitiation
627 Complex. Longnecker RM, editor. J Virol [Internet]. 2014 Nov 1;88(21):12825. Available
628 from: http://jvi.asm.org/content/88/21/12825.abstract
629 12. Davis ZH, Hesser CR, Park J, Glaunsinger BA. Interaction between ORF24 and ORF34 in
630 the Kaposi's Sarcoma-Associated Herpesvirus Late Gene Transcription Factor
631 Complex Is Essential for Viral Late Gene Expression. Sandri-Goldin RM, editor. J Virol
632 [Internet]. 2016 Jan 1;90(1):599. Available from:
633 http://jvi.asm.org/content/90/1/599.abstract
634 13. Wyrwicz LS, Rychlewski L. Identification of Herpes TATT-binding protein. Antiviral
635 Res [Internet]. 2007;75(2):167–72. Available from:
636 http://www.sciencedirect.com/science/article/pii/S0166354207002367
637 14. Davis ZH, Verschueren E, Jang GM, Kleffman K, Johnson JR, Park J, et al. Global
638 Mapping of Herpesvirus-Host Protein Complexes Reveals a Transcription Strategy for
639 Late Genes. Mol Cell [Internet]. 2015;57(2):349–60. Available from:
640 http://www.sciencedirect.com/science/article/pii/S1097276514009216
641 15. Gruffat H, Kadjouf F, Mariamé B, Manet E. The Epstein-Barr Virus BcRF1 Gene Product
642 Is a TBP-Like Protein with an Essential Role in Late Gene Expression. J Virol [Internet].
643 2012 Jun 1;86(11):6023. Available from: http://jvi.asm.org/content/86/11/6023.abstract

644 16. Nishimura M, Watanabe T, Yagi S, Yamanaka T, Fujimuro M. Kaposi's sarcoma-
645 associated herpesvirus ORF34 is essential for late gene expression and virus production.
646 Sci Rep [Internet]. 2017;7(1):329. Available from: https://doi.org/10.1038/s41598-017-
647 00401-7
648 17. Gong D, Wu NC, Xie Y, Feng J, Tong L, Brulois KF, et al. Kaposi's Sarcoma-
649 Associated Herpesvirus ORF18 and ORF30 Are Essential for Late Gene Expression
650 during Lytic Replication. Frueh K, editor. J Virol [Internet]. 2014 Oct 1;88(19):11369.
651 Available from: http://jvi.asm.org/content/88/19/11369.abstract
652 18. Brulois K, Wong L-Y, Lee H-R, Sivadas P, Ensser A, Feng P, et al. Association of
653 Kaposi's Sarcoma-Associated Herpesvirus ORF31 with ORF34 and ORF24 Is
654 Critical for Late Gene Expression. Longnecker RM, editor. J Virol [Internet]. 2015 Jun
655 1;89(11):6148. Available from: http://jvi.asm.org/content/89/11/6148.abstract
656 19. Perng Y-C, Campbell JA, Lenschow DJ, Yu D. Human Cytomegalovirus pUL79 Is an
657 Elongation Factor of RNA Polymerase II for Viral Gene Transcription. PLOS Pathog
658 [Internet]. Public Library of Science; 2014 Aug 28;10(8):e1004350-. Available from:
659 https://doi.org/10.1371/journal.ppat.1004350
660 20. Castañeda AF, Glaunsinger BA. The Interaction between ORF18 and ORF30 Is Required
661 for Late Gene Expression in Kaposi's Sarcoma-Associated Herpesvirus. Sandri-
662 Goldin RM, editor. J Virol [Internet]. 2019 Jan 1;93(1):e01488-18. Available from:
663 http://jvi.asm.org/content/93/1/e01488-18.abstract
664 21. Pan D, Han T, Tang S, Xu W, Bao Q, Sun Y, et al. Murine Cytomegalovirus Protein
665 pM91 Interacts with pM79 and Is Critical for Viral Late Gene Expression. Sandri-Goldin
666 RM, editor. J Virol [Internet]. 2018 Sep 15;92(18):e00675-18. Available from:

667 http://jvi.asm.org/content/92/18/e00675-18.abstract
668 22. Djavadian R, Hayes M, Johannsen E. CAGE-seq analysis of Epstein-Barr virus lytic gene
669 transcription: 3 kinetic classes from 2 mechanisms. PLOS Pathog [Internet]. Public
670 Library of Science; 2018 Jun 4;14(6):e1007114-. Available from:
671 https://doi.org/10.1371/journal.ppat.1007114
672 23. Arias C, Weisburd B, Stern-Ginossar N, Mercier A, Madrid AS, Bellare P, et al. KSHV
673 2.0: A Comprehensive Annotation of the Kaposi's Sarcoma-Associated Herpesvirus
674 Genome Using Next-Generation Sequencing Reveals Novel Genomic and Functional
675 Features. PLOS Pathog [Internet]. Public Library of Science; 2014 Jan
676 16;10(1):e1003847-. Available from: https://doi.org/10.1371/journal.ppat.1003847
677 24. Nakamura H, Lu M, Gwack Y, Souvlis J, Zeichner SL, Jung JU. Global Changes in
678 Kaposi's Sarcoma-Associated Virus Gene Expression Patterns following
679 Expression of a Tetracycline-Inducible Rta Transactivator. J Virol [Internet]. 2003 Apr
680 1;77(7):4205. Available from: http://jvi.asm.org/content/77/7/4205.abstract
681 25. Jenner RG, Albà MM, Boshoff C, Kellam P. Kaposi's Sarcoma-Associated
682 Herpesvirus Latent and Lytic Gene Expression as Revealed by DNA Arrays. J Virol
683 [Internet]. 2001 Jan 15;75(2):891. Available from:
684 http://jvi.asm.org/content/75/2/891.abstract
685 26. Paulose-Murphy M, Ha N-K, Xiang C, Chen Y, Gillim L, Yarchoan R, et al. Transcription
686 Program of Human Herpesvirus 8 (Kaposi's Sarcoma-Associated Herpesvirus). J
687 Virol [Internet]. 2001 May 15;75(10):4843. Available from:
688 http://jvi.asm.org/content/75/10/4843.abstract
689 27. Arumugaswami V, Wu T-T, Martinez-Guzman D, Jia Q, Deng H, Reyes N, et al. ORF18

690 Is a Transfactor That Is Essential for Late Gene Transcription of a Gammaherpesvirus. J
691 Virol [Internet]. 2006 Oct 1;80(19):9730. Available from:
692 http://jvi.asm.org/content/80/19/9730.abstract
693 28. McKenzie J, Lopez-Giraldez F, Delecluse H-J, Walsh A, El-Guindy A. The Epstein-Barr
694 Virus Immunoevasins BCRF1 and BPLF1 Are Expressed by a Mechanism Independent of
695 the Canonical Late Pre-initiation Complex. PLOS Pathog [Internet]. Public Library of
696 Science; 2016 Nov 17;12(11):e1006008-. Available from:
697 https://doi.org/10.1371/journal.ppat.1006008
698 29. Brulois KF, Chang H, Lee AS-Y, Ensser A, Wong L-Y, Toth Z, et al. Construction and
699 Manipulation of a New Kaposi's Sarcoma-Associated Herpesvirus Bacterial
700 Artificial Chromosome Clone. J Virol [Internet]. 2012 Sep 15;86(18):9708. Available
701 from: http://jvi.asm.org/content/86/18/9708.abstract
702 30. Bruce AG, Barcy S, DiMaio T, Gan E, Garrigues HJ, Lagunoff M, et al. Quantitative
703 Analysis of the KSHV Transcriptome Following Primary Infection of Blood and
704 Lymphatic Endothelial Cells. Pathog (Basel, Switzerland) [Internet]. MDPI; 2017 Mar
705 19;6(1):11. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28335496
706 31. Bailey L. T, Elkan C. Fitting a mixture model by expectation maximization to discover
707 motifs in biopolymers. In: Proceedings of the Second International Conference on
708 Intelligent Systems for Molecular Biology [Internet]. Menlo Park; 1994. p. 28–36.
709 Available from:
710 https://www.ncbi.nlm.nih.gov/pubmed/?term=Fitting+a+mixture+model+by+expectation+
711 maximization+to+discover+motifs+in+biopolymers
712 32. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif.

713 Bioinformatics [Internet]. 2011 Feb 16;27(7):1017–8. Available from:
714 https://dx.doi.org/10.1093/bioinformatics/btr064
715 33. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of
716 Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for
717 Macrophage and B Cell Identities. Mol Cell [Internet]. 2010;38(4):576–89. Available
718 from: http://www.sciencedirect.com/science/article/pii/S1097276510003667
719 34. Djavadian R, Chiu Y-F, Johannsen E. An Epstein-Barr Virus-Encoded Protein Complex
720 Requires an Origin of Lytic Replication In Cis to Mediate Late Gene Transcription. PLOS
721 Pathog [Internet]. Public Library of Science; 2016 Jun 27;12(6):e1005718-. Available
722 from: https://doi.org/10.1371/journal.ppat.1005718
723 35. Yella VR, Bhimsaria D, Ghoshdastidar D, Rodríguez-Martínez JA, Ansari AZ, Bansal M.
724 Flexibility and structure of flanking DNA impact transcription factor affinity for its core
725 motif. Nucleic Acids Res [Internet]. 2018/11/05. Oxford University Press; 2018 Dec
726 14;46(22):11883–97. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30395339
727 36. Deng W, Roberts SGE. A core promoter element downstream of the TATA box that is
728 recognized by TFIIB. Genes Dev [Internet]. Cold Spring Harbor Laboratory Press; 2005
729 Oct 15;19(20):2418–23. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16230532
730 37. Wong E, Wu T-T, Reyes N, Deng H, Sun R. Murine Gammaherpesvirus 68 Open Reading
731 Frame 24 Is Required for Late Gene Expression after DNA Replication. J Virol [Internet].
732 2007 Jun 15;81(12):6761. Available from: http://jvi.asm.org/content/81/12/6761.abstract
733 38. Lin CL, Li H, Wang Y, Zhu FX, Kudchodkar S, Yuan Y. Kaposi's sarcoma-associated
734 herpesvirus lytic origin (ori-Lyt)-dependent DNA replication: identification of the ori-Lyt
735 and association of K8 bZip protein with the origin. J Virol [Internet]. American Society

736 for Microbiology; 2003 May;77(10):5578–88. Available from:
737 https://www.ncbi.nlm.nih.gov/pubmed/12719550
738 39. AuCoin DP, Colletti KS, Xu Y, Cei SA, Pari GS. Kaposi's Sarcoma-Associated
739 Herpesvirus (Human Herpesvirus 8) Contains Two Functional Lytic Origins of DNA
740 Replication. J Virol [Internet]. 2002 Aug 1;76(15):7890. Available from:
741 http://jvi.asm.org/content/76/15/7890.abstract
742 40. Gardner MR, Glaunsinger BA. Kaposi's Sarcoma-Associated Herpesvirus ORF68
743 Is a DNA Binding Protein Required for Viral Genome Cleavage and Packaging. Sandri-
744 Goldin RM, editor. J Virol [Internet]. 2018 Aug 15;92(16):e00840-18. Available from:
745 http://jvi.asm.org/content/92/16/e00840-18.abstract
746 41. Labo N, Miley W, Marshall V, Gillette W, Esposito D, Bess M, et al. Heterogeneity and
747 Breadth of Host Antibody Response to KSHV Infection Demonstrated by Systematic
748 Analysis of the KSHV Proteome. PLOS Pathog [Internet]. Public Library of Science;
749 2014 Mar 27;10(3):e1004046. Available from:
750 https://doi.org/10.1371/journal.ppat.1004046
751 42. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data
752 [Internet]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
753 43. Trim Galore - http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. Available
754 from: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
755 44. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
756 universal RNA-seq aligner. Bioinformatics [Internet]. 2012/10/25. Oxford University
757 Press; 2013 Jan;29(1):15–21. Available from:
758 https://www.ncbi.nlm.nih.gov/pubmed/23104886

759 45. Pyl PT, Anders S, Huber W. HTSeq—a Python framework to work with high-throughput
760 sequencing data. Bioinformatics [Internet]. 2014 Sep 25;31(2):166–9. Available from:
761 https://dx.doi.org/10.1093/bioinformatics/btu638
762 46. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
763 RNA-seq data with DESeq2. Genome Biol [Internet]. 2014;15(12):550. Available from:
764 https://doi.org/10.1186/s13059-014-0550-8
765 47. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al.
766 Integrative genomics viewer. Nat Biotechnol [Internet]. Nature Publishing Group, a
767 division of Macmillan Publishers Limited. All Rights Reserved.; 2011 Jan 10;29:24.
768 Available from: https://doi.org/10.1038/nbt.1754
769 48. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-
770 performance genomics data visualization and exploration. Brief Bioinform [Internet].
771 2012 Apr 19;14(2):178–92. Available from: https://dx.doi.org/10.1093/bib/bbs017
772 49. Zhou T, Yang L, Lu Y, Dror I, Dantas Machado AC, Ghane T, et al. DNAshape: a method
773 for the high-throughput prediction of DNA structural features on a genomic scale. Nucleic
774 Acids Res [Internet]. 2013/05/22. Oxford University Press; 2013 Jul;41(Web Server
775 issue):W56–62. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23703209
776

777 **Figure Legends**

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

780 iSLK cells harboring WT BAC16, the ORF24$_{RAAAG}$ mutant or the ORF24$_{RAAAG}$ mutant rescue
781 (MR) were reactivated with doxycycline and sodium butyrate for 72 hours. (A) Western blot

782 showing expression of an early protein (ORF59) and a late protein (K8.1) in WT, ORF24$_{RAAAG}$
783 and ORF24$_{RAAAG}$-MR. Vinculin was used as a loading control. (B) Progeny virion production
784 was measured by supernatant transfer on to 293T cells and quantified by flow cytometry. (C)
785 DNA replication of different mutants and MRs was measured by qPCR and normalized to the
786 respective uninduced sample. The fold induction of all mutants was normalized to the WT to get
787 the relative change in DNA replication compared to WT.
788
789 **Fig 2: Differential expression analysis of ORF24-Pol II interaction mutant and mutant**
790 **rescue**
791 iSLK cells harboring ORF24$_{RAAAG}$ and ORF24$_{RAAAG}$ -MR were reactivated with doxycycline and
792 sodium butyrate for 24 and 48 h, whereupon total RNA was extracted and subjected to library
793 preparation and RNA sequencing. (A) Bar plot showing $\log_2$ fold change of the viral genes in the
794 ORF24$_{RAAAG}$ -MR at 48 h relative to 24 h. Genes were color coded by kinetic class as defined in
795 (30). (B) Bar plot showing $\log_2$ fold change of the viral genes in the mutant relative to the MR.
796 Genes were color coded by kinetic class as defined in (30). (C) Heatmap of the scaled $\log_2$ fold
797 change of the genes induced in the MR between 24 to 48 h and genes repressed in the mutant
798 compared to MR at 48 h. Gene clusters with different kinetics and dependence on ORF24-Pol II
799 interaction are indicated. (D) Consensus motif identified by MEME analysis of promoters from
800 14 genes that showed greater than one standard deviation repression in the mutant relative to the
801 MR.
802
803 **Fig 3: ChIP-Seq with HA-ORF34 shows multiple binding sites across the viral genome**

804 (A) ChIP-Seq was performed using an anti-HA antibody to immunoprecipitate DNA prepared
805 from iSLK HA-ORF34 cells reactivated for 48 h with doxycycline and sodium butyrate. iSLK-
806 WT cells lacking the HA tag and input DNA were used as negative controls. The alignment files
807 were converted to the tiled data file (tdf) format and the ChIP background from the untagged
808 cells was subtracted and visualized on the Integrated Genome Viewer (IGV) (B) ChIP-qPCR was
809 performed using an anti-HA antibody to immunoprecipitate DNA prepared from iSLK HA-
810 ORF24 cells reactivated for 48 h with doxycycline and sodium butyrate. The DNA was
811 quantified by qPCR with primers specific to the promoter regions of the indicated genes. Data
812 were normalized to the level of input DNA and are presented as percentages of the input. Data
813 shown are from three biological replicates. (C) Consensus motif identified by MEME analysis of
814 promoters from 19 genes whose promoters had a ChIP peak.
815
816 **Fig 4: Identifying direct targets of the vPIC**
817 (A) Heatmap showing genes repressed in the ORF24$_{RAAAG}$ mutant compared to the MR adjacent
818 to the list of genes that show ORF34 binding at their promoter by ChIP-Seq analysis. Also
819 indicated are viral genes that were identified to have the TATTWAA motif by FIMO analysis.
820 The direct targets were identified by their combined dependence on the ORF24-Pol II interaction
821 and the presence of a ChIP signal at the promoter. (B) Consensus motif identified by MEME
822 analysis of promoters from 13 direct targets.
823
824 **Fig 5: Motif identified from direct targets is important for transcription from the promoter**
825 (A) Schematic of plasmid-based promoter activation assay. iSLK-WT or ORF24$_{RAAAG}$ mutant
826 cells reactivated with doxycycline and sodium butyrate were transfected with reporter plasmid

827 carrying the indicated promoter sequence upstream of the firefly luciferase gene and a control
828 renilla luciferase expressing plasmid. 48 h post transfection and reactivation, the cells were
829 assessed for luciferase activity. (B) Reactivated iSLK-WT cells were transfected with the K8.1
830 promoter reporter plasmid with or without the KSHV left lytic origin of replication cloned in *cis*
831 and assessed for luciferase activity. (C) Reactivated iSLK-WT cells were transfected with the
832 K8.1 and ORF57 promoter reporter plasmids in the presence or absence of DNA replication
833 inhibitor (PAA) and assessed for luciferase activity. (D) Reactivated iSLK-WT and iSLK-
834 ORF24$_{RAAAG}$ mutant cells were transfected with K8.1 or ORF57 promoter reporter plasmids and
835 assessed for luciferase activity. (E-F) iSLK-WT cells were transfected with plasmids containing
836 the WT K8.1 promoter or the indicated single or double point mutant promoters and assessed for
837 luciferase activity. The data for each mutant was normalized to the WT K8.1 promoter. All data
838 shown are average of 3-5 biological replicates.
839
840 **Fig 6: Expanded motif is important for DNA binding by vPIC**
841 (A) Schematic of the transient ChIP assay. iSLK-WT or iSLK-HAORF34 cells reactivated with
842 doxycycline and sodium butyrate were transfected with plasmid carrying the indicated promoter
843 sequence. 48 h post transfection and reactivation, cells were subjected to ChIP with an anti-HA
844 antibody, and the associated DNA was quantified using primers specific to the promoter on the
845 plasmid and the genomic K8.1 promoter locus. (B) iSLK-HAORF34 cells were transfected with
846 K8.1 promoter containing plasmid, mutant plasmids or ORF57 promoter plasmid (as a control).
847 iSLK-WT cells without the HA tag transfected with K8.1 plasmid served as a control for the IP.
848 The ChIP DNA was quantified using plasmid specific primers (left panel) and K8.1 promoter

849 specific primers (right panel), normalized to the level of input DNA and presented as percent
850 input. Data shown are an average of 3-6 biological replicates.
851
852 **Fig S1: Validation of mutant BAC DNA integrity**
853 The integrity of the various BAC mutants used in this study was verified by *Rsr*II digestion.
854
855 **Fig S2: Kinetics of early and late gene transcripts in ORF24$_{RAAAG}$ mutant and MR**
856 iSLK-ORF24$_{RAAAG}$ and MR were reactivated with doxycycline and sodium butyrate for the
857 indicated time points, whereupon total RNA was extracted, converted to cDNA, and quantified
858 by qPCR using primers specific to the CDS of K8.1 (late gene) or ORF57 (early gene). The fold
859 induction was measured relative to uninduced sample.
860
861 **Fig S3: ChIP-Seq data with HA-ORF34**
862 (A) iSLK cells harboring WT BAC16, HA-ORF34 BAC or HA-ORF24 BAC were reactivated
863 with doxycycline and sodium butyrate for 72 hours. Progeny virion production was measured by
864 supernatant transfer on to 293T cells and quantified by flow cytometry. (B) IGV image showing
865 ChIP-Seq data from iSLK-HA-ORF34 cells reactivated with doxycycline and sodium butyrate
866 for 48 h and immunoprecipitated with anti-HA antibody. Also shown are plots for control data
867 from untagged iSLK cells and input DNA for two biological replicates
868
869 **Fig S4: Effect of mutations on DNA structural parameters**
870 Four structural parameters for 40 bp of the WT and expanded motif mutants were calculated
871 using the DNAShape program in R (49). Average of each structural parameter for the mutated

872 base pairs and the two adjacent base pairs at either end (a total of 7 bp) was calculated for each
873 DNA sequence analyzed. The two adjacent base pairs were included in the averaging as some of
874 the structural parameters refer to a base step rather than a specific base. The average for each
875 mutant was normalized to the WT K8.1 promoter.
876
877
878
879
880
881
882

Figure 1

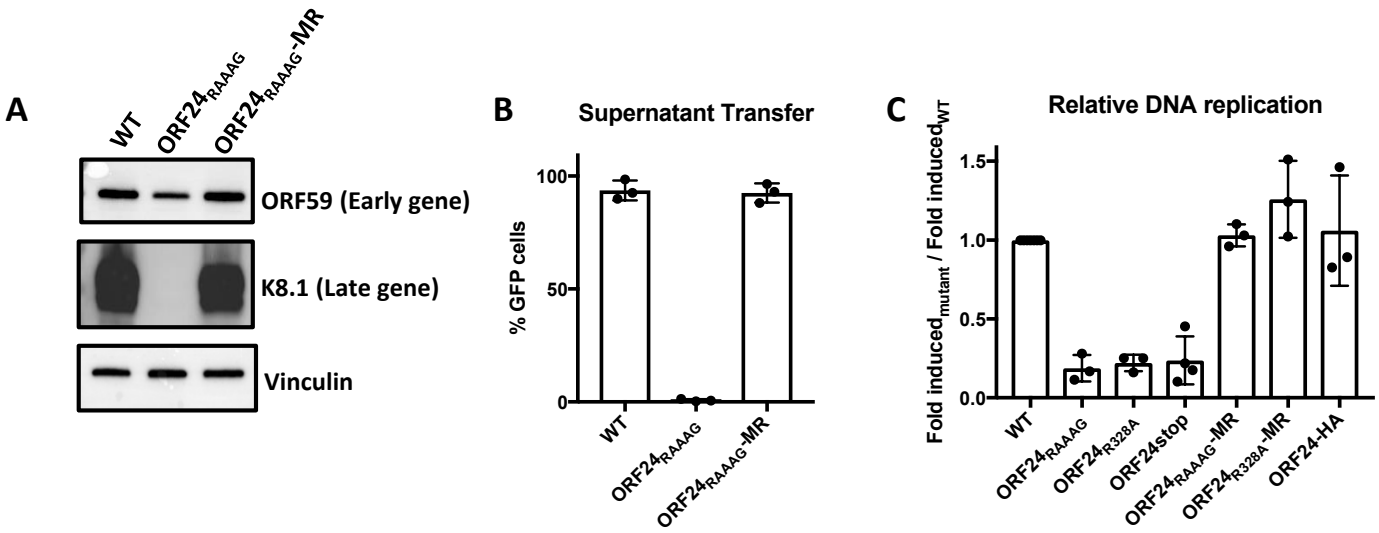

A
WT
ORF24RAAAG
ORF24RAAAG-MR
ORF59 (Early gene)
K8.1 (Late gene)
Vinculin
B
Supernatant Transfer
% GFP cells
100
50
0
WT
ORF24RAAAG
ORF24RAAAG-MR
C
Relative DNA replication
Fold inducedmutant / Fold inducedWT
1.5
1.0
0.5
0.0
WT
ORF24RAAAG
ORF24R328A
ORF24stop
ORF24RAAAG-MR
ORF24R328A-MR
ORF24-HA

Figure S1

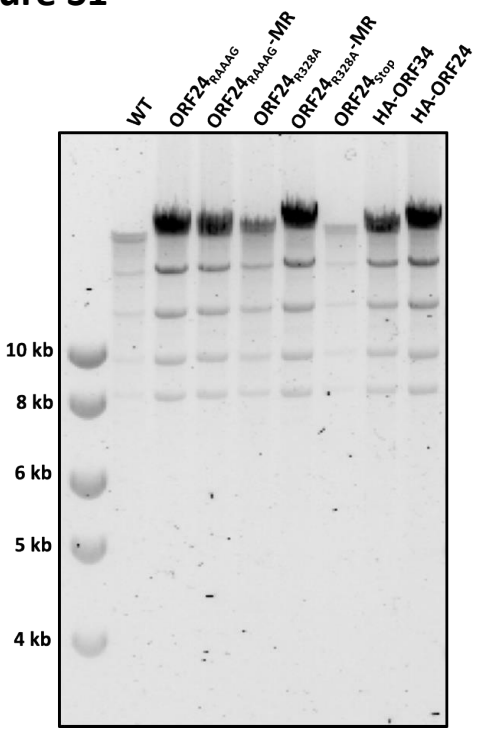

WT
ORF24RAAAG-MR
ORF24RAAAG
ORF24R328A-MR
ORF24R328A
ORF24Stop
HA-ORF34
HA-ORF24
10 kb
8 kb
6 kb
5 kb
4 kb

Figure S2

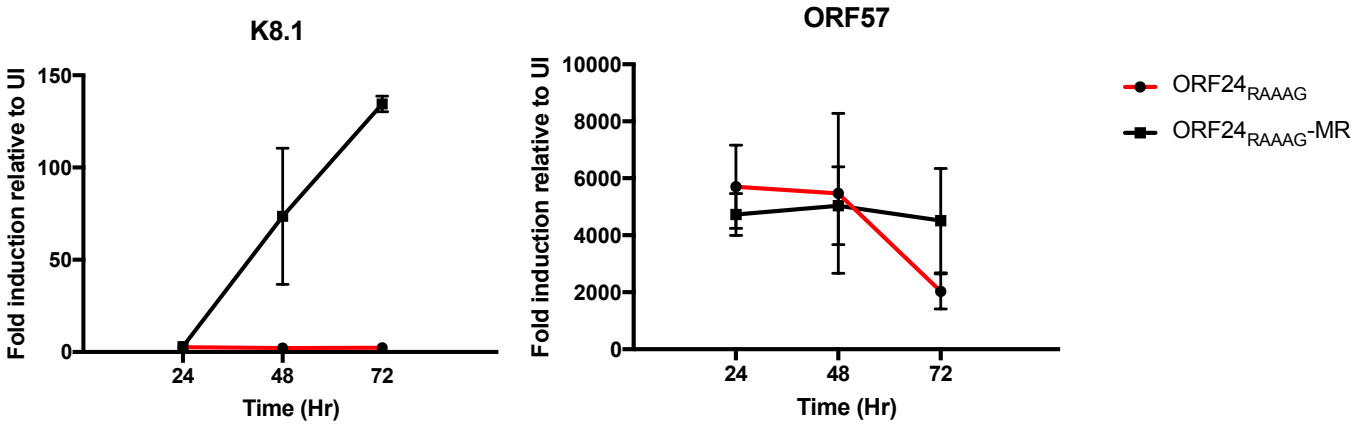
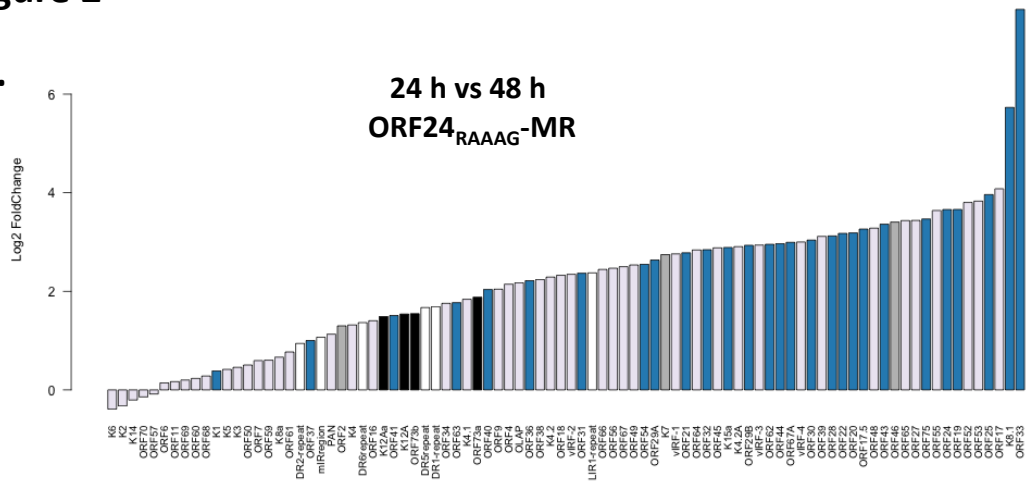
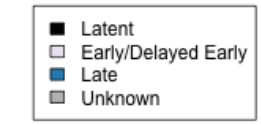

K8.1
Fold induction relative to UI
150
100
50
0
24
48
72
Time (Hr)
ORF57
Fold induction relative to UI
10000
8000
6000
4000
2000
0
24
48
72
Time (Hr)
ORF24RAAAG
ORF24RAAAG-MR

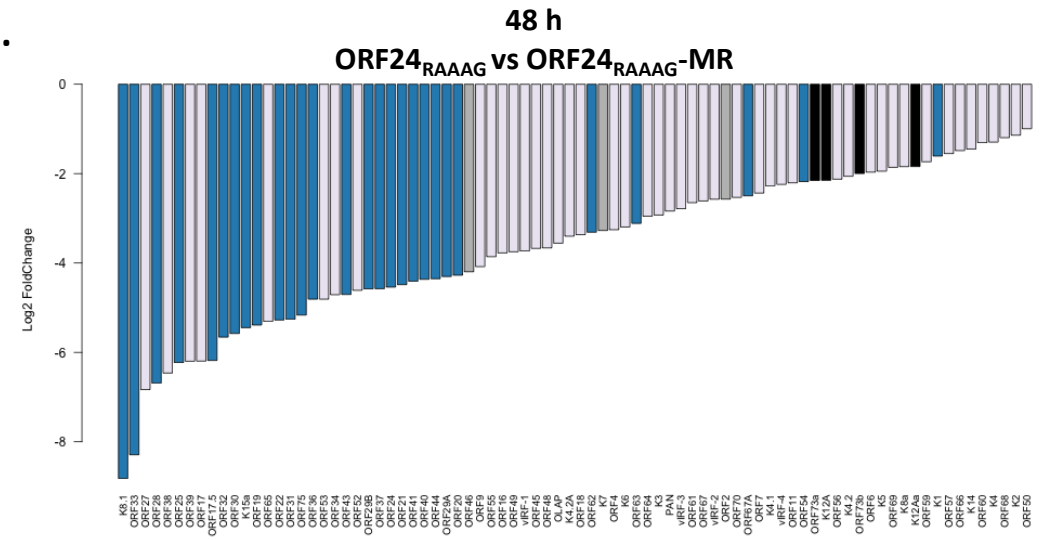
Figure 2

A.

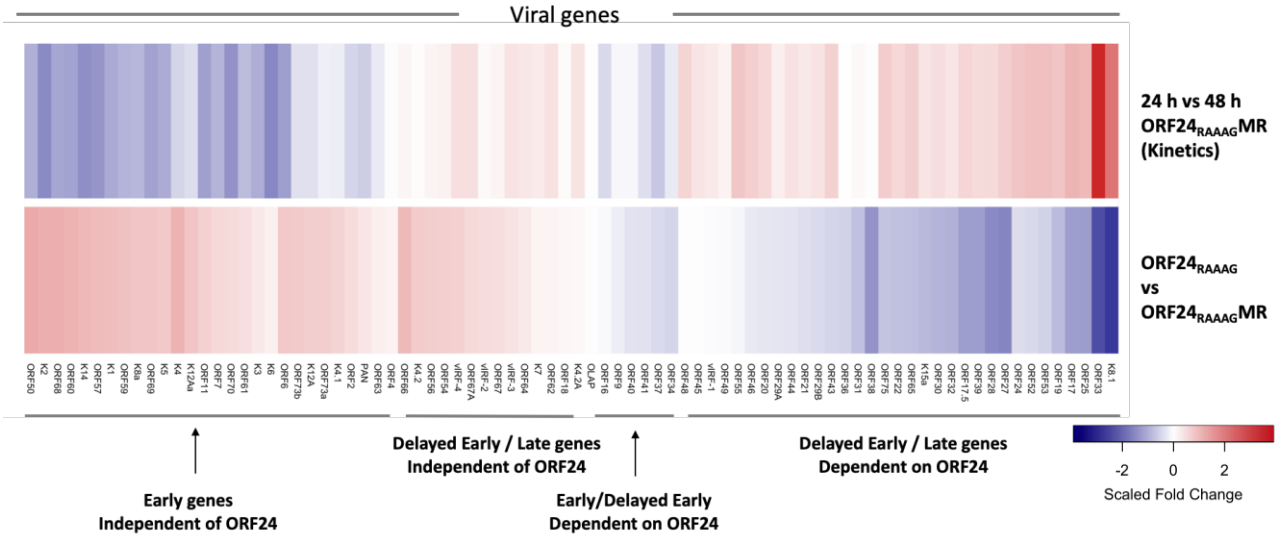




B.



C.

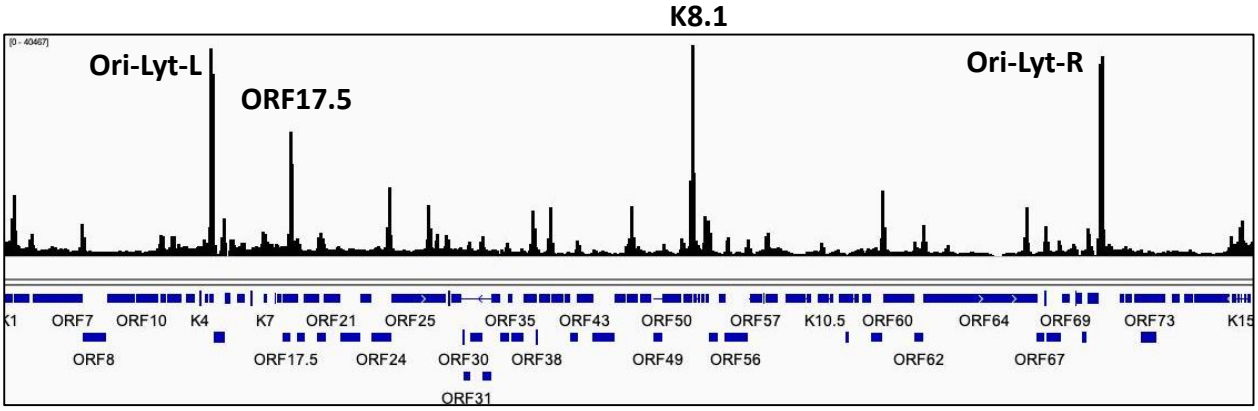


D.

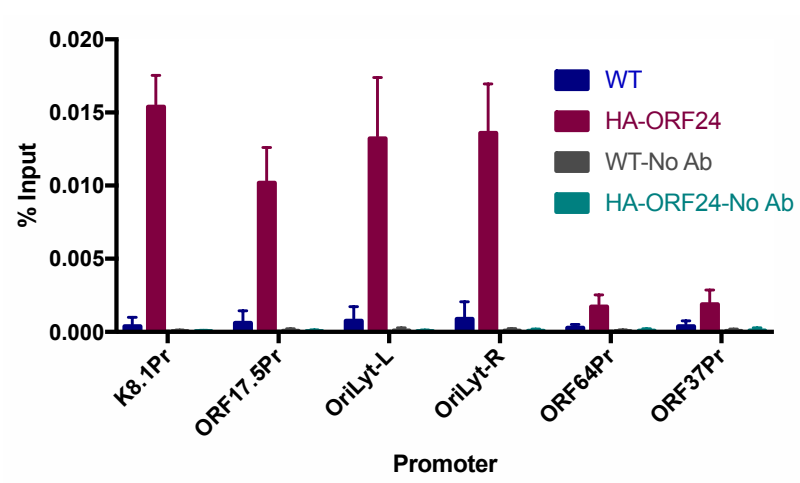
***E*-value:** 3.4e-006

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

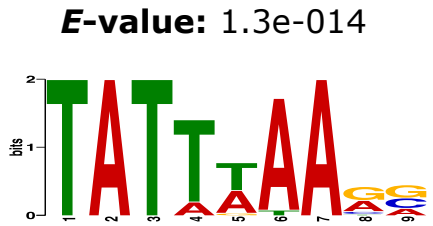
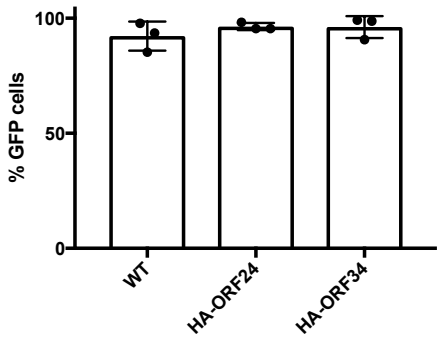

E-value: 1.3e-014
bits
2
1
0
1
2
3
4
5
6
7
8
9

Figure S3

A


Supernatant transfer
% GFP cells
100
50
0
WT
HA-ORF24
HA-ORF34

B

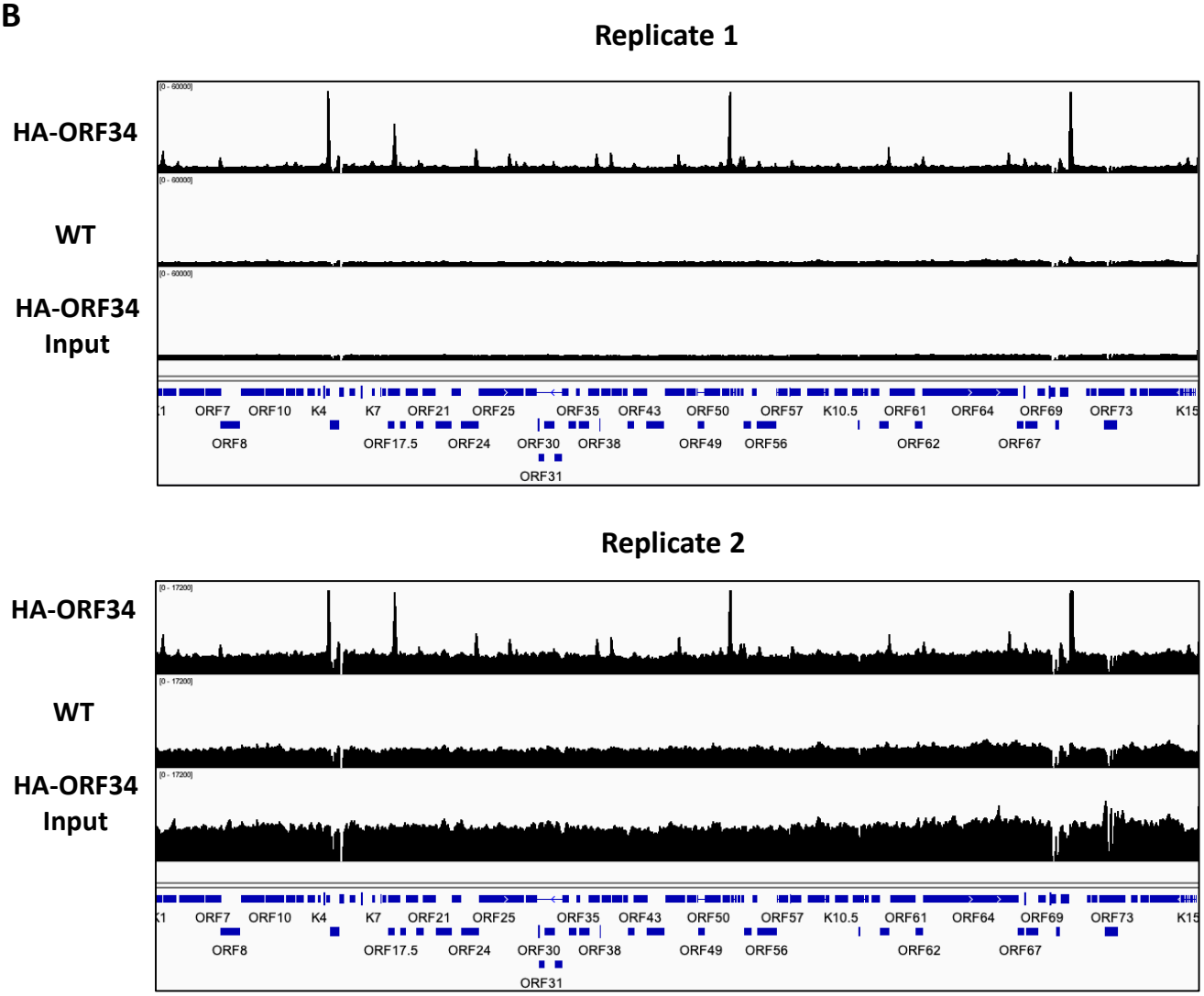
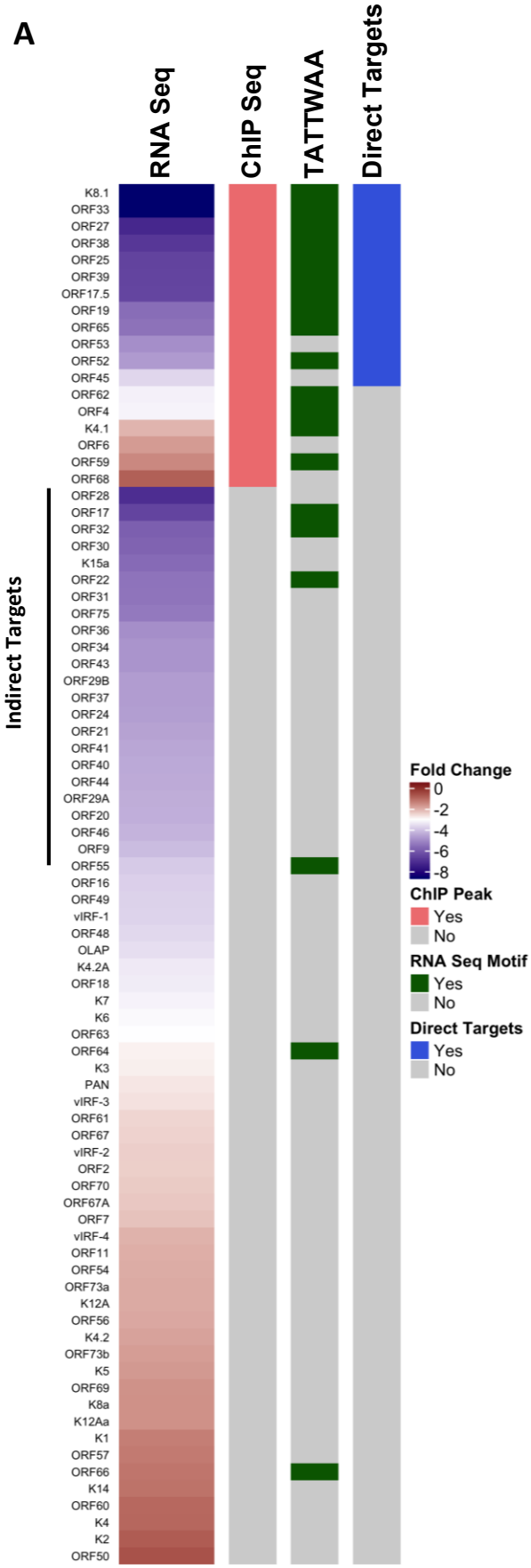

Replicate 1
HA-ORF34
WT
HA-ORF34 Input
ORF7
ORF10
K4
K7
ORF21
ORF25
ORF35
ORF43
ORF50
ORF57
K10.5
ORF61
ORF64
ORF69
ORF73
ORF8
ORF17.5
ORF24
ORF30
ORF38
ORF49
ORF56
ORF62
ORF67
ORF31
Replicate 2
HA-ORF34
WT
HA-ORF34 Input
ORF7
ORF10
K4
K7
ORF21
ORF25
ORF35
ORF43
ORF50
ORF57
K10.5
ORF61
ORF64
ORF69
ORF73
ORF8
ORF17.5
ORF24
ORF30
ORF38
ORF49
ORF56
ORF62
ORF67
ORF31

Figure 4



B

TATTWAARVNYS

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

Figure 5

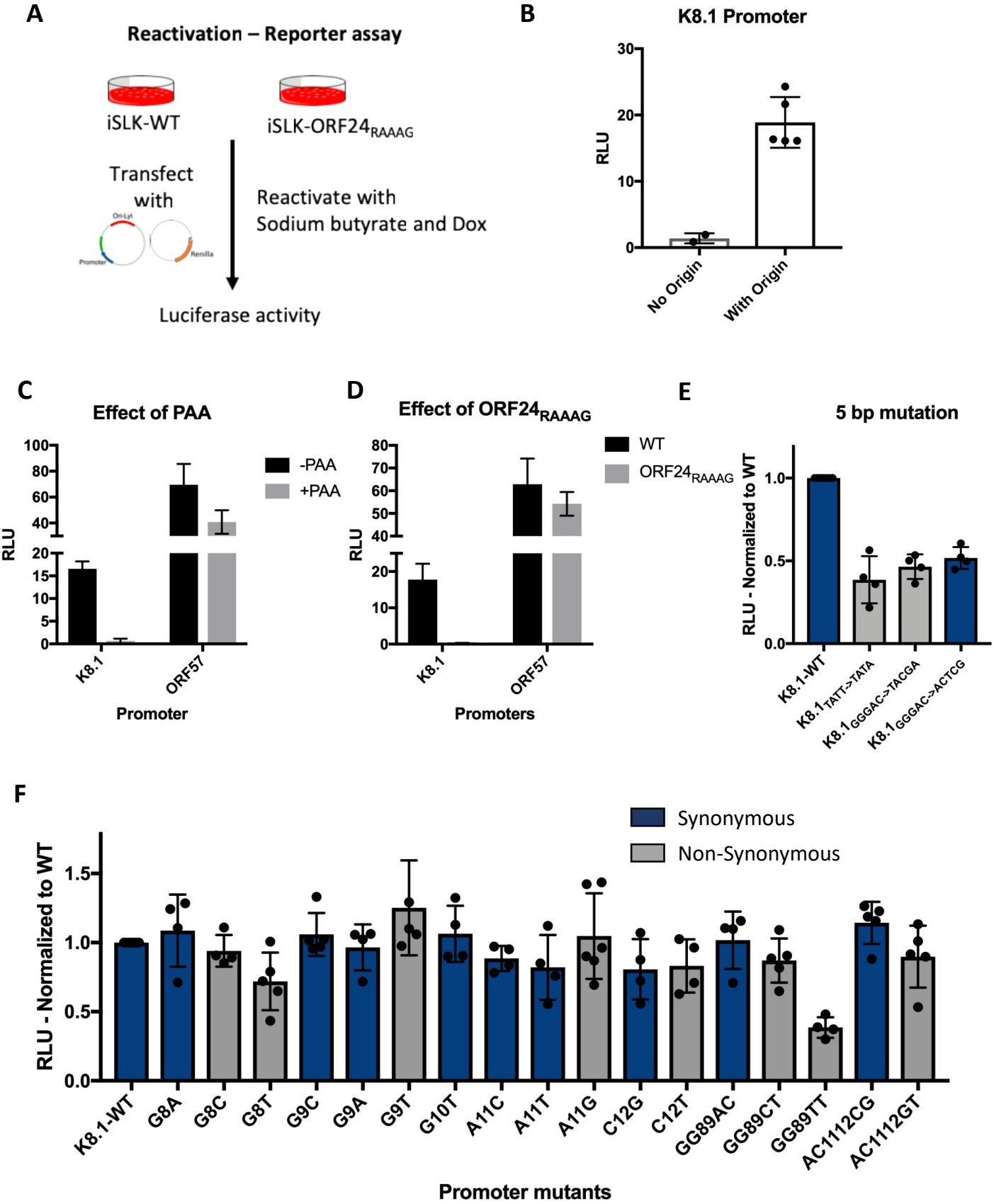
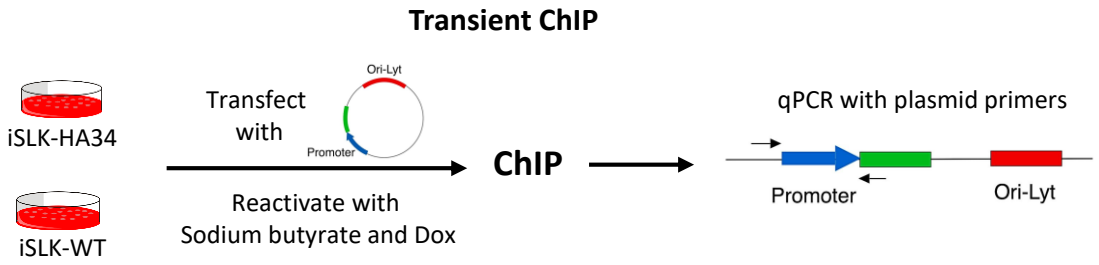

A
Reactivation – Reporter assay
iSLK-WT
iSLK-ORF24RAAAG
Transfect with
Ori-Lyt
Promoter
Renilla
Reactivate with
Sodium butyrate and Dox
Luciferase activity
B
K8.1 Promoter
RLU
30
20
10
0
No Origin
With Origin
C
Effect of PAA
RLU
100
80
60
40
20
15
10
5
0
-PAA
+PAA
K8.1
ORF57
Promoter
D
Effect of ORF24RAAAG
RLU
80
70
60
50
40
20
10
0
WT
ORF24RAAAG
K8.1
ORF57
Promoters
E
5 bp mutation
RLU - Normalized to WT
1.0
0.5
0.0
K8.1-WT
K8.1TATT->TATA
K8.1GGGAC->TACGA
K8.1GGGAC->ACTCG
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

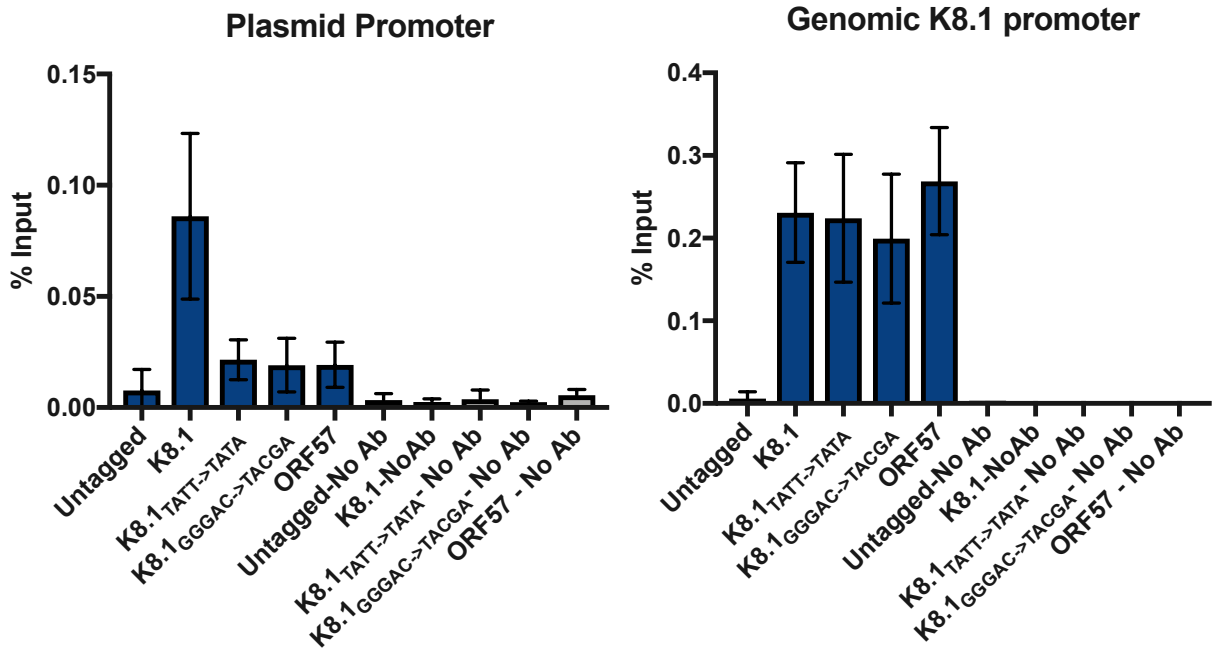

Plasmid Promoter
% Input
0.15
0.10
0.05
0.00
Untagged
K8.1
K8.1TATT->TATA
K8.1GGGAC->TACGA
ORF57
Untagged-No Ab
K8.1-NoAb
K8.1TATT->TATA - No Ab
K8.1GGGAC->TACGA - No Ab
ORF57 - No Ab
Genomic K8.1 promoter
% Input
0.4
0.3
0.2
0.1
0.0
Untagged
K8.1
K8.1TATT->TATA
K8.1GGGAC->TACGA
ORF57
Untagged-No Ab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

Figure S4

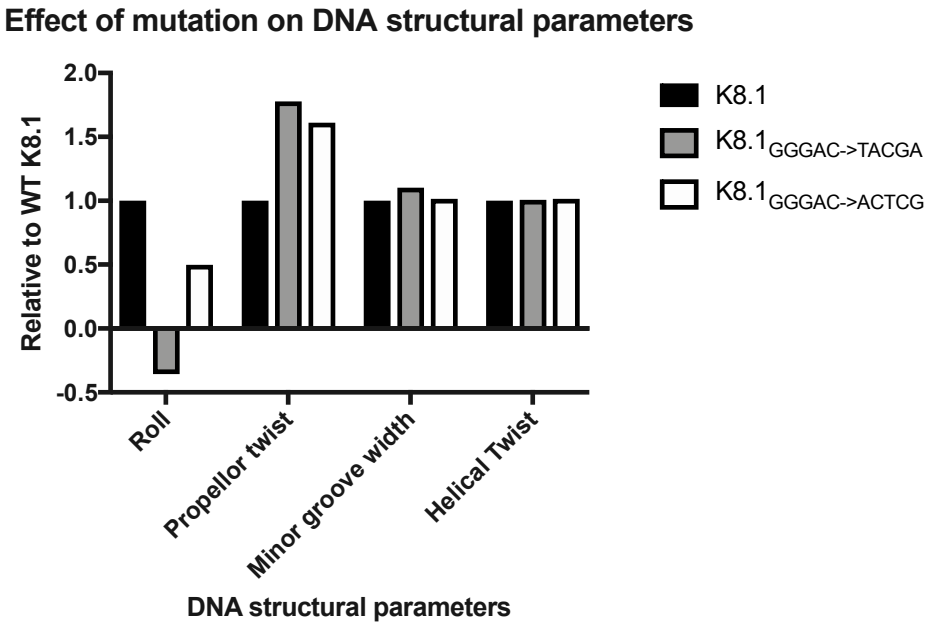

Effect of mutation on DNA structural parameters
Relative to WT K8.1
2.0
1.5
1.0
0.5
0.0
-0.5
Roll
Propellor twist
Minor groove width
Helical Twist
DNA structural parameters
K8.1
K8.1GGGAC->TACGA
K8.1GGGAC->ACTCG

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
K8.1	Glycoprotein	L	+	32
ORF52	Immune Evasion	L	-	29
ORF27	Glycoprotein	L	+	103
ORF53	Glycoprotein	L	-	29
ORF25	Major capsid protein	L	+	89 (29#)
ORF62	Assembly/DNA maturation	L	-	32
ORF33	Capsid binding	L	+	30
ORF8	Glycoprotein	L	+	29
ORF17.5	Viral capsid assembly	L	-	29 (73#)
ORF19	Capsid vertex		-	113
ORF26	Minor capsid protein	L	+	29
ORF45	Immune evasion (Tegument)	E	-	No motif
ORF6	Single strand DNA binding protein	E	+	89 (31#)
K4.1	Homologs of cellular chemokines	E	-	32
ORF59	DNA polymerase processivity factor	E	-	29
ORF68	Packaging protein	E	+	No motif
ORF4	Immune evasion	E	+	29
ORF65	Minor capsid protein	L	-	31
ORF38	Viral maturation and egress (Tegument protein)	L	+	29

Second site with lower consensus to TATTWAARV motif
Direct targets are shown in gray

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