The γ-herpesviral TATA box binding protein directly interacts with RNA polymerase II to direct late gene transcription

Short title: KSHV ORF24 directly interacts with the C-terminal domain of host RNA polymerase II

Angelica F. Castañeda[1#], Allison L. Didychuk[1#], Robert K. Louder[2,3*], Chloe O. McCollum[4], Zoe H. Davis[5], Eva Nogales[2,4,6,7], Britt A. Glaunsinger[1,4,6,7]

[1]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

[2]Molecular Biophysics and Integrative Bio-Imaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

[3]Biophysics Graduate Group, University of California, Berkeley, CA 94720, USA

[4]Department of Molecular and Cell Biology, University of California Berkeley, CA 94720, USA

[5]Division of Infectious Diseases and Immunity, School of Public Health, University of California, Berkeley, CA 94720, USA

[6]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720, USA.

[7]Howard Hughes Medical Institute, Berkeley, CA 94720, USA

Correspondence should be addressed to Britt Glaunsinger (glaunsinger@berkeley.edu)

#These authors contributed equally to this work

*Present address: Department of Biology, Johns Hopkins University, Baltimore, MD 21218, USA

ABSTRACT

β- and γ-herpesviruses are important human pathogens, and include the oncogenic viruses Kaposi's sarcoma-associated virus (KSHV) and Epstein-Barr virus (EBV), and human cytomegalovirus (HCMV), which is a significant cause of congenital disease. Near the end of their replication cycle, these viruses transcribe their late genes in a manner distinct from host transcription. Late gene transcription requires six virally-encoded proteins, one of which is a functional mimic of host TATA-box-binding protein (TBP). We found that this protein, encoded by ORF24 in KSHV, makes a direct protein-protein contact with the C-terminal domain of host RNA polymerase II (Pol II). This interaction is mediated by the N-terminal domain (NTD) of ORF24 through a conserved motif that is shared in its β- and γ-herpesvirus homologs. Thus, β- and γ-herpesviruses employ an unprecedented strategy in eukaryotic transcription, wherein promoter recognition and polymerase recruitment are facilitated by a single transcriptional activator with functionally distinct domains.

INTRODUCTION

Eukaryotic transcription begins with the formation of a pre-initiation complex (PIC) at the core promoter. The assembly pathway for the canonical PIC starts with binding of TFIID to the core promoter and deployment of TATA-box-binding protein (TBP) onto the TATA box or pseudo-TATA box region upstream of the transcription start site (TSS), followed by recruitment of the other general transcription factors (GTFs). In this process, the GTFs position the 12-subunit RNA polymerase II (Pol II) at the core promoter (Thomas and Chiang, 2006). The largest Pol II subunit, Rpb1, has a low-complexity carboxyl terminal domain (CTD) that in humans is composed of 52 heptapeptide repeats with a consensus sequence of YSPTSPS. The CTD is a regulatory hub responsible for coordinating signals throughout the different stages of transcription and RNA processing (Harlen and Churchman, 2017). The phosphorylation state of the CTD controls progression through different states of transcription, as well as interactions

53 with other cellular machinery (Buratowski, 2009, Harlen and Churchman, 2017). Pol II with a
54 hypophosphorylated CTD is recruited into the PIC (Lu et al., 1991), and phosphorylation signals
55 release from the PIC into an elongating complex.

56 DNA viruses hijack the host transcriptional machinery to direct their own gene
57 expression. Given that the mechanisms governing transcription from viral promoters are often
58 similar to those at host promoters, viruses have been invaluable models for understanding
59 transcription complex assembly and regulation (Bauer et al., 2018, Chi and Carey, 1996, Haigh
60 et al., 1990). A conserved feature of double-stranded DNA (dsDNA) viruses is the temporal
61 cascade of gene expression that begins with the expression of two classes of early genes,
62 followed by viral DNA replication, and ending with the expression of late genes. In the β- and γ-
63 herpesviruses, immediate early and early genes are transcribed in a manner similar to host
64 genes. In contrast, the mechanism underlying the regulation of late gene transcription remains
65 poorly understood, yet is known to be distinct from host and early viral gene transcription.

66 Late gene transcription in the β- and γ-herpesviruses is divergent from that of the α-
67 herpesviruses and has a number of unique features. First, β/γ late gene transcription is
68 regulated in part by a core promoter sequence 12-15 base pairs in length that has a TATT motif
69 followed by a RVNYS motif in lieu of the canonical TATA box found in early viral promoters and
70 in some cellular promoters (Nandakumar and Glaunsinger, 2019, Serio et al., 1998, Tang et al.,
71 2004, Wong-Ho et al., 2014). Additionally, late gene expression requires at least six viral
72 proteins, called viral transcriptional activators (vTAs), which form a complex at late gene
73 promoters (Aubry et al., 2014, Castaneda and Glaunsinger, 2019, Davis et al., 2016, Didychuk
74 et al., 2019, Pan et al., 2018, Gruffat et al., 2012). Little is known about the functional role these
75 vTAs play in late gene transcription.

76 The best studied protein in the vTA complex is a virally-encoded TBP mimic (vTBP),
77 encoded by ORF24 in the γ-herpesvirus Kaposi's sarcoma-associated herpesvirus (KSHV).
78 ORF24 is predicted to have a TBP-like domain in the central portion of the protein, which was

79 identified through an *in silico* protein fold threading analysis performed with BcRF1, the homolog
80 of ORF24 from Epstein-Barr virus (Wyrwicz and Rychlewski, 2007). Indeed, ORF24 replaces
81 TBP at late gene promoters during infection, and a virus with mutations in the predicted DNA-
82 binding residues in the TBP-like domain of ORF24 is unable to transcribe late genes (Davis et
83 al., 2015). Thus, β- and γ-herpesviruses encode their own vTBP, which promotes efficient
84 transcription from a distinct set of late gene promoters.

85 While both cellular TBP and ORF24 (vTBP) bind DNA, a unique feature of vTBP is that it
86 additionally interacts with Pol II (Aubry et al., 2014, Davis et al., 2015). In host transcription,
87 recruitment of Pol II to promoters is mediated by TFIIB instead of a direct protein-protein
88 interaction between TBP and Pol II (Chen and Hahn, 2003, Tsai and Sigler, 2000). In contrast,
89 prokaryotic transcription is dependent on sigma factors that facilitate both promoter selection
90 and polymerase recruitment (Feklistov et al., 2014). We were therefore intrigued by the
91 possibility that vTBP is a unique bifunctional eukaryotic transcriptional activator and sought to
92 understand the basis of its ability to recruit Pol II.

93 Here, we demonstrate that ORF24 makes a direct protein-protein contact with the CTD
94 of Pol II. Using recombinantly expressed proteins, we identified a minimal N-terminal domain
95 (ORF24-NTD) of ORF24 sufficient for interaction with Pol II. Using negative stain electron
96 microscopy and pulldown assays, we determined that ORF24-NTD directly interacts with the Pol
97 II CTD, and demonstrate that at least four heptapeptide CTD repeats are required for this
98 interaction. We also demonstrate that interaction with Pol II is conserved across the ORF24 β-
99 and γ-herpesvirus homologs, and that the primary function of the N-terminal domain is the
100 recruitment of Pol II. These findings suggest that vTBP is a fundamentally unique protein when
101 compared to other eukaryotic Pol II-interacting proteins, as it both directly interacts with Pol II
102 and binds promotor DNA to coordinate late gene expression.

103

104

105 **RESULTS**

106 **The N-terminal domain of ORF24 is sufficient for interaction with Pol II**

107 We previously revealed that KSHV ORF24 co-immunoprecipitates with Pol II in cells, in
108 a manner dependent on three conserved leucine residues (L73-75) in the N-terminus of ORF24
109 (Davis et al., 2015). While this interaction requires no other viral proteins, whether ORF24 binds
110 Pol II directly or indirectly via bridging cellular factors was unknown. Additionally, it had not been
111 determined whether Pol II binding occurred through a domain that was fully separable from the
112 regions required for binding the ORF34 vTA (which links ORF24 to the rest of the late gene
113 transcription complex (Davis et al., 2016)) and binding to promoter DNA. To address these
114 unknowns, we first constructed a series of N-terminal ORF24 fragments to identify a minimal
115 region of the protein that retained the Pol II interaction (**Figure 1A**). Plasmids encoding full-
116 length or truncated FLAG-tagged ORF24 were transfected into HEK293T cells and
117 immunoprecipitated using FLAG antibodies to determine which ORF24 segments retained

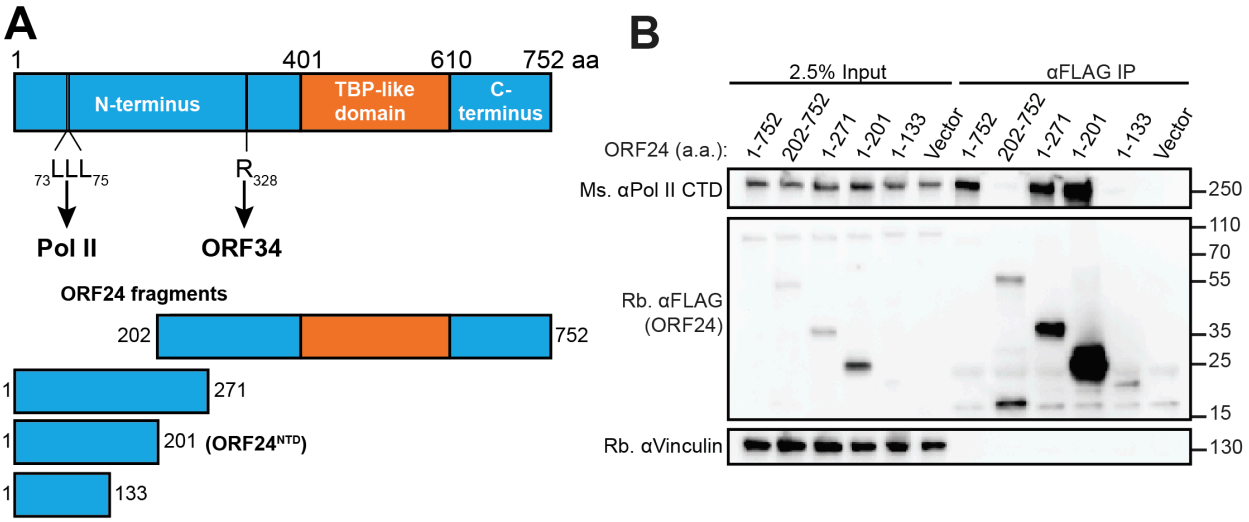


Figure 1. The N-terminal domain of ORF24 (ORF24-NTD) binds Pol II and can be recombinantly purified.
(A) Schematic of constructs used to identify a minimal N-terminal domain of ORF24 showing the predicted boundaries for the N-terminal domain, the TBP-like domain, and the C-terminal domain, including residues known to be required for Pol II binding (amino acids 73-75) and interaction with ORF34 (amino acid 328).
(B) HEK293T cells were transiently transfected with full-length or truncated FLAG-tagged ORF24 and co-immunoprecipitated with anti-FLAG beads followed by western blotting with the indicated antibodies to detect ORF24 and Pol II.

118 binding to endogenous Pol II. The smallest fragment of ORF24 that retained the Pol II
119 interaction consisted of amino acids (a.a.) 1-201, which we termed the ORF24 N-terminal
120 domain (ORF24-NTD) (**Figure 1B**). Notably, the remainder of the protein (a.a. 202-752), which
121 includes the vTBP domain and a region known to interact with the vTA ORF34 (Davis et al.,
122 2016), failed to co-immunoprecipitate Pol II. Thus, ORF24 contains an NTD that is both
123 necessary and sufficient for Pol II binding, which is separable from the other two known
124 functions of the protein.
125
126 **Negative stain electron microscopy of PICs with GST-ORF24-NTD suggests an**
127 **interaction with the Pol II stalk**
128 We next sought to determine how ORF24-NTD interacts with Pol II within a minimal PIC.
129 To generate purified ORF24-NTD for this purpose, we appended it to an N-terminal glutathione-
130 S-transferase (GST) tag, and achieved robust expression of the protein in *E. coli*, similar to that
131 of GST alone (**Figure 2A**). We confirmed that GST-tagged ORF24-NTD retained the ability to
132 interact with Pol II using a GST pulldown with GST-ORF24-NTD and whole cell lysate from
133 HEK293T cells, suggesting that recombinantly expressed GST-ORF24-NTD is well-folded
134 (**Figure 2B**).
135 To assemble a minimal PIC that includes ORF24-NTD, we used a streptavidin-
136 immobilized DNA scaffold based on the super core promoter element (He et al., 2013), which
137 contains the TATA, BRE, and INR core promoter elements (**Figure 2C**). Human TBP, TFIIA,
138 TFIIB, and TFIIF were purified from *E. coli*, and Pol II was immunopurified from HeLa cell
139 extracts (Revyakin et al., 2012). We expected that TBP binding to the TATA motif would initiate
140 formation of the PIC with binding of TFIIA and TFIIB, followed by Pol II and TFIIF recruitment,
141 and then binding of GST-ORF24-NTD. Therefore, any additional density beyond that seen in the
142 well-characterized minimal PIC (DNA/TBP/TFIIA/TFIIB/Pol II/TFIIF) (He et al., 2013) could be
143 attributed to ORF24-NTD.

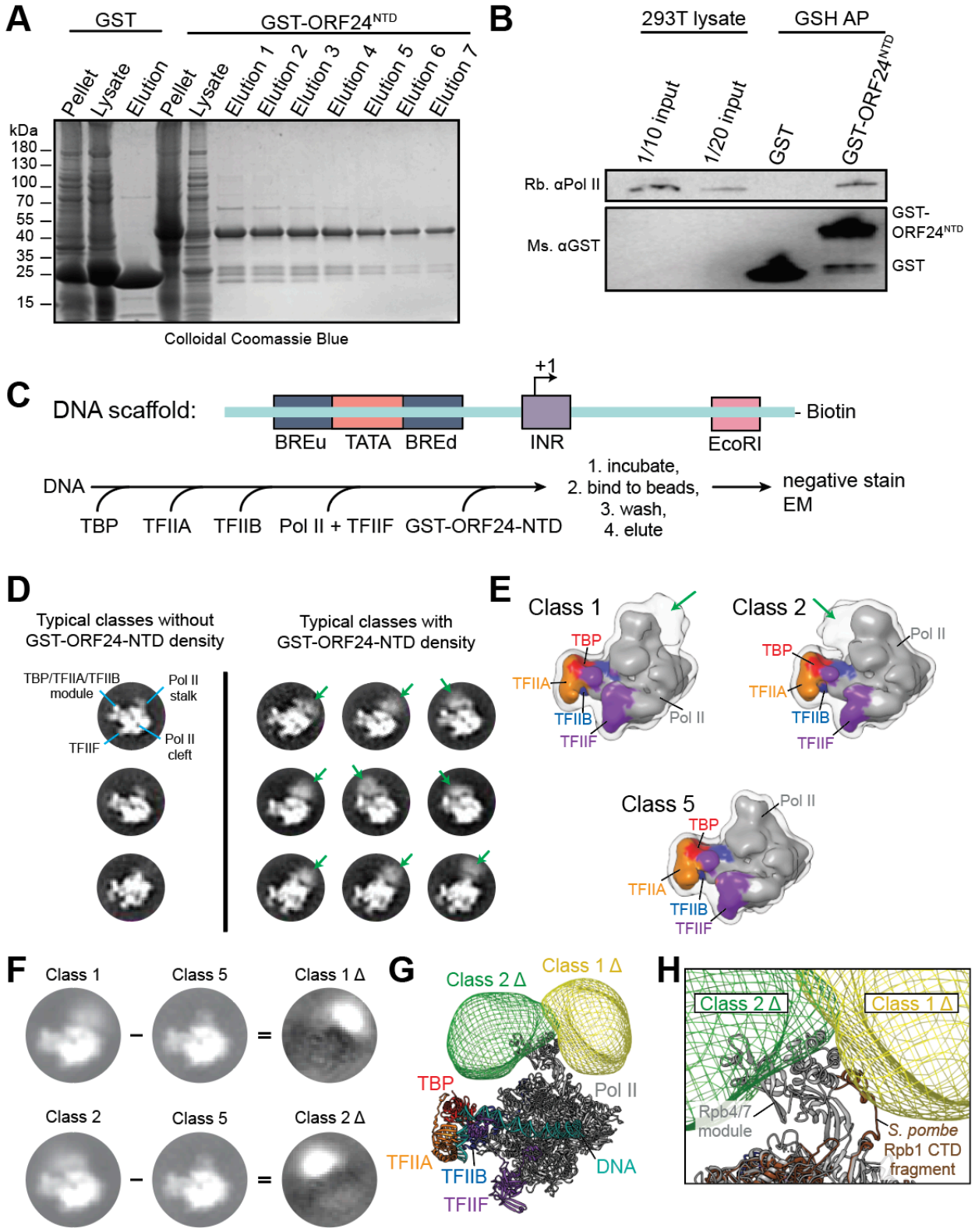

A
GST
GST-ORF24NTD
Pellet
Lysate
Elution
Pellet
Lysate
Elution 1
Elution 2
Elution 3
Elution 4
Elution 5
Elution 6
Elution 7
kDa
180
130
100
70
55
40
35
25
15
Colloidal Coomassie Blue
B
293T lysate
GSH AP
1/10 input
1/20 input
GST
GST-ORF24NTD
Rb. αPol II
Ms. αGST
GST-ORF24NTD
GST
C
DNA scaffold:
+1
BREu
TATA
BREd
INR
EcoRI
- Biotin
DNA
TBP
TFIIA
TFIIB
Pol II + TFIIF
GST-ORF24-NTD
1. incubate,
2. bind to beads,
3. wash,
4. elute
negative stain EM
D
Typical classes without GST-ORF24-NTD density
Typical classes with GST-ORF24-NTD density
TBP/TFIIA/TFIIB module
Pol II stalk
TFIIF
Pol II cleft
E
Class 1
Class 2
Class 5
TBP
TFIIA
TFIIB
TFIIF
Pol II
F
Class 1
Class 5
Class 1 Δ
Class 2
Class 5
Class 2 Δ
G
Class 2 Δ
Class 1 Δ
TBP
Pol II
TFIIA
TFIIB
TFIIF
DNA
H
Class 2 Δ
Class 1 Δ
Rpb4/7 module
S. pombe Rpb1 CTD fragment

Figure 2. ORF24-NTD binds Pol II in minimal PICs.
(A) Colloidal Coomassie gel demonstrating that GST and GST-ORF24-NTD can be recombinantly expressed in *E. coli* and purified by glutathione sepharose purification.
(B) GST or GST-ORF24-NTD was incubated in HEK293T whole cell lysate, then subjected to affinity purification using glutathione magnetic beads followed by western blotting.
(C) Sequential reconstitution strategy for a minimal PIC containing GST-ORF24-NTD.
(D) Representative reference-free two-dimensional class averages of negatively stained minimal PICs (TBP/TFIIA/TFIIB/TFIIF/Pol II/DNA) assembled in the presence of GST-ORF24-NTD. Three classes on the left show different views of the minimal PIC alone, with the class average in the upper-left annotated with the main features of a minimal PIC particle. The nine class averages on the right show diffuse density in various positions around the Pol II stalk attributed to bound GST-ORF24-NTD (green arrows).
(E) Representative three-dimensional class averages of negatively stained minimal PICs assembled in the presence of GST-ORF24-NTD. Classes 1 and 2 exhibit two major areas occupied by bound GST-ORF24-NTD proximal to the Pol II stalk, while class 5 does not exhibit any such density near the Pol II stalk. Solid surfaces are colored by subunit, while a lower intensity iso-surface is shown in transparency to reveal the weaker density attributed to bound GST-ORF24-NTD (green arrows).
(F) Difference mapping of the densities attributed to bound GST-ORF24-NTD. Shown on the left are two-dimensional projections of class 1 (top) and 2 (bottom) from (E), and on the right are the difference maps, called "Class 1 Δ" and "Class 2 Δ", calculated by subtracting Class 5 from each of the respective classes.
(G) Three-dimensional difference maps corresponding to the extra density attributed to bound GST-ORF24-NTD, mapped onto the structure of the minimal PIC (PDB 5IYA).
(H) Zoomed in view of (G) with the structure of *Schizosaccharomyces pombe* Rpb1 (PDB 3H0G) superposed onto the human structure to show the location of the beginning portion of the Rpb1 CTD within the Pol II stalk. Note that only the very N-terminal portion of the Rpb1 CTD is visible in this structure, with >450 amino acids following this sequence in the CTD of human Rpb1.

145

146 The minimal PIC was assembled by sequentially incubating the DNA scaffold with TBP,
147 TFIIA and TFIIB, Pol II and TFIIF, and GST-ORF24-NTD, followed by immobilization on
148 streptavidin-coated beads, and finally washing and elution from the beads (**Figure 2C**). We
149 performed single particle negative stain EM of this assembled complex containing GST-ORF24-
150 NTD, and searched for particles that displayed extra density beyond the well-defined minimal
151 PIC structure. A total of 79,381 PIC particle images were analyzed by two-dimensional and
152 three-dimensional classification, and approximately 25% showed clear extra density in the
153 resulting class averages (**Figure 2D** and **Supplemental Figure 1**). Two of the three-
154 dimensional classes (called Class 1 and 2) exhibit extra density on opposing faces of the Pol II
155 stalk (**Figure 2D**), while a third class (Class 5) is comprised of the majority of particles without
156 extra density near the Pol II stalk. Thus, Class 5 was used to subtract the density corresponding

157 to the minimal PIC from Classes 1 and 2, resulting in difference maps that clearly show the
158 region occupied by bound GST-ORF24-NTD (**Figure 2F**).

159 By superimposing the three-dimensional difference maps representing the extra density
160 attributed to GST-ORF24-NTD onto the cryo-EM structure of the human minimal PIC (**Figure**
161 **2G**) (He et al., 2016), it became evident that the ORF24-NTD is flexibly bound to a region of Pol
162 II near the Rpb4/7 stalk module. An inherently flexible domain of Pol II is the CTD of the largest
163 subunit, Rpb1, which contains a linker followed by 52 heptapeptide repeats (Harlen and
164 Churchman, 2017) that is not visualized by EM due to its disordered structure. Docking of the
165 crystal structure of Rpb1 from *Schizosaccharomyces pombe* (PDB 3H0G) (Spahr et al., 2009),
166 in which the structure of the CTD linker region is partially resolved and can be seen to extend
167 along the surface of the stalk module, further indicates that the ORF24-NTD may be interacting
168 with the flexible CTD of Pol II (**Figure 2H**).

169

170 **ORF24-NTD binds the CTD repeats of Rpb1**

171 Based on the EM results, we hypothesized that ORF24 could be interacting with either
172 Rpb4/7, the Rpb1 linker, or the Rpb1 CTD heptapeptide repeats. We therefore assessed
173 whether recombinantly purified versions of each of these factors bound purified ORF24-NTD.
174 We first performed GST-pulldown assays with GST-tagged Rpb1 CTD or Rpb1 linker and
175 maltose-binding protein (MBP)-tagged ORF24-NTD. As a control, we also purified an MBP-
176 tagged mutant version of ORF24-NTD, in which the three conserved leucines at positions L73-
177 75 were mutated to alanines (ORF24-NTD-3L_A). This mutation renders the protein unable to
178 interact with Pol II in cells (Davis et al., 2015). WT ORF24-NTD bound to the Rpb1 CTD
179 repeats, but not to the Rpb1 linker region (**Figure 3A**). This interaction was specific and
180 mimicked the results in mammalian cell lysate, as no binding to either CTD fusion was observed
181 with ORF24-NTD-3L_A (**Figure 3A**). We also performed Strep pulldown assays using ORF24-
182 NTD-Strep and the heterodimeric GST-Rpb4/Rpb7-His complex. Again, ORF24-NTD interacted

183 with GST-CTD, but not with GST-Rpb4/Rpb7-His (**Figure 3B**). Thus, recombinant ORF24-NTD
184 directly interacts with recombinant Pol II CTD repeats *in vitro*, and enrichment using tags on
185 either the Pol II CTD or on ORF24 allows for the isolation of the other. The lack of interaction
186 with the CTD linker domain or the Rpb4/7 stalk suggests that the CTD repeats are likely to be

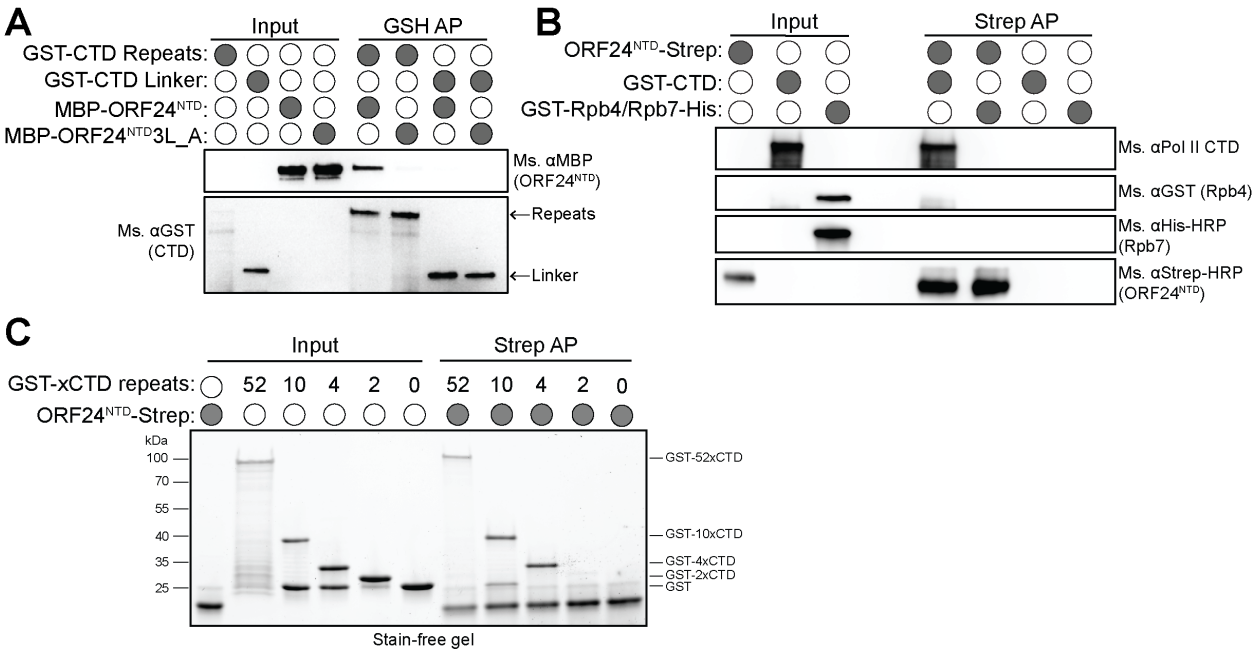


Figure 3. ORF24-NTD directly interacts with four heptapeptide repeats of the Pol II CTD.
(A) Recombinantly purified GST-CTD repeats or the GST-CTD linker were incubated with either purified MBP-ORF24-NTD or MBP-ORF24-NTD-3L_A protein, then subjected to a glutathione pulldown. Samples were resolved by SDS-PAGE and analyzed by western blot.
(B) Recombinantly purified GST-CTD repeats or GST-Rpb4/His-Rpb7 heterodimer were incubated with recombinantly purified Strep-tagged ORF24-NTD, then subjected to a StrepTactinXT pulldown. Samples were resolved by SDS-PAGE and analyzed by western blot.
(C) Recombinantly purified full-length GST-CTD repeats (52 repeats) shorter GST-CTD constructs (10, 4, or 2 repeats), or GST itself were incubated with recombinantly purified Strep-tagged ORF24-NTD, then subjected to a StrepTactinXT pulldown. Samples were resolved by SDS-PAGE and stained with colloidal Coomassie.

187 the primary point of Pol II contact with ORF24.

188

189 The CTD is a key regulatory component of Pol II, responsible for coordinating signals
190 through interactions with multiple transcriptional modulators (Harlen and Churchman, 2017).
191 CTD binding proteins engage with the heptapeptide repeats in a variety of ways, from
192 recognition of a few residues of a repeat (in the case of kinases that promote phosphorylation at

193 conserved serines, threonines, or tyrosines within a given repeat), to multiple consecutive
194 repeats (in the case of Mediator (Robinson et al., 2012)), to recognition of the intrinsically
195 disordered entire domain (in the case of newly-appreciated phase-separated interactions (Lu et
196 al., 2018, Boehning et al., 2018)). To determine which of these types of interactions occur
197 between the CTD and ORF24, we performed Strep pulldown assays with purified ORF24-NTD-
198 Strep and recombinant GST-CTD constructs containing either 2, 4, 10, or 52 (full length)
199 heptapeptide repeats. The recombinant GST-CTD used in these assays was unphosphorylated,
200 as our previous results demonstrated that ORF24 co-immunoprecipitates only
201 hypophosphorylated Pol II from mammalian cells (Davis et al., 2015). Notably, ORF24
202 interacted with 4x, 10x, and 52xCTD, but failed to interact with either 2xCTD repeats or with
203 control GST (**Figure 3C**). Therefore, similar to the Mediator complex (Robinson et al., 2012),
204 ORF24 likely engages four tandem repeats to interact with the Pol II CTD.
205
206 **Conserved leucine residues are essential for interaction with Pol II**
207 Given the importance of the ORF24-Pol II interaction for late gene transcription, we
208 sought to more precisely define which ORF24 residues were required for Pol II binding in cells.
209 We first evaluated whether all three of the leucine residues previously studied (3L_A; (Davis et
210 al., 2015)) contribute equally to the interaction. Each of the leucine residues was mutated to
211 alanine individually and in pairs to generate all possible combinations of leucine to alanine
212 mutations. The individual leucine mutants had distinct phenotypes: L75A almost completely
213 disrupted the ORF24-Pol II interaction, whereas the L74A and L73A mutations caused
214 moderate or mild binding defects, respectively (**Supplemental Figure 2A**). Combinations of any
215 two leucine mutations completely abolished the interaction between ORF24 and Pol II
216 (**Supplemental Figure 2A**). These results suggest that the leucine residues have an additive
217 effect on the Pol II interaction and do not contribute to the interaction equally. Mutation of other
218 residues within the RLLLG motif (ORF24 R72A, which is conserved) and close to the RLLLG

219 motif (ORF24 R68A, which is not conserved) similarly ablated the ORF24-Pol II interaction
220 (**Supplemental Figure 2B, 3**). Thus, although the RLLLG motif is necessary for Pol II
221 recruitment, additional contacts elsewhere in the domain are required. These additional contacts
222 are not perfectly conserved between the β- and γ-herpesviruses and are suggestive of plasticity
223 in the mechanism of Pol II binding.
224

225 **Identification of a minimal Pol II-interaction domain in the β- and γ-herpesviruses**

226 β- and γ-herpesviral homologs of ORF24 from murine gammaherpesvirus 68 (MHV68;
227 mu24), Epstein-Barr virus (EBV; BcRF1), and human cytomegalovirus (HCMV; UL87) also
228 interact with Pol II in cells (Aubry et al., 2014, Davis et al., 2015, Nobre et al., 2019). The RLLLG
229 motif is well-conserved in all β- and γ-herpesvirus homologs, despite overall poor sequence
230 identity (**Supplemental Figure 3**), suggesting that the NTD of ORF24 homologs may also be
231 responsible for Pol II recruitment. To test if the three conserved leucine residues are involved in
232 the homolog-Pol II interactions, we generated triple leucine mutants (3L_A) in UL87, mu24, and
233 BcRF1. Similarly to ORF24, mutation of the RLLLG motif ablated their interactions with Pol II
234 (**Figure 4**), suggesting that vTBPs interact with Pol II through their respective N-terminal
235 domains in a manner dependent upon this highly conserved patch of residues, although other
236 contacts are likely required (**Supplemental Figure 2**).

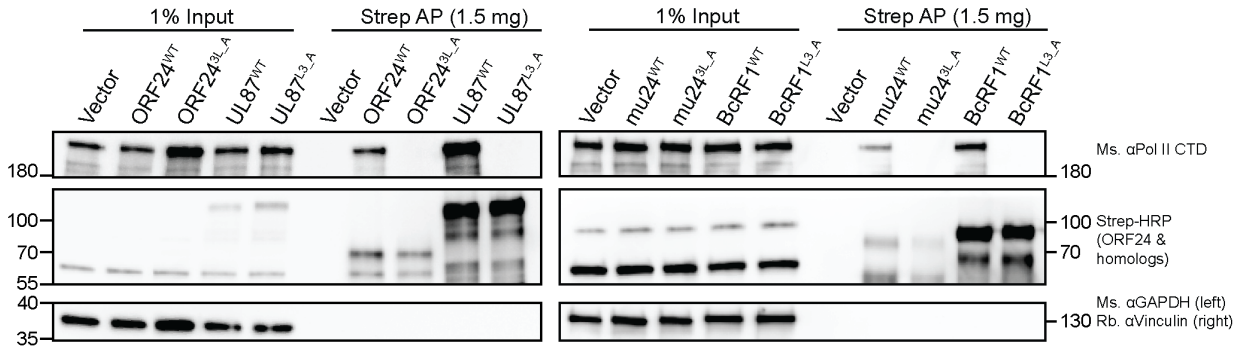


Figure 4. The RLLLG motif in ORF24 homologs from other β- and γ-herpesviruses is required for interaction with Pol II.
Full-length WT or 3L_A mutants, which mutates the conserved RLLLG motif, of Strep-tagged ORF24 or homologs from MHV68 (mu24), EBV (BcRF1), and HCMV (UL87) were transiently transfected into HEK293T cells then coimmunoprecipitated with StrepTactinXT beads followed by Western blotting.

237 We next sought to determine if the NTD of homologs of ORF24 was sufficient for Pol II
238 interaction, with the goal of identifying whether the β- and γ-herpesviruses share a common
239 domain for polymerase recruitment. Based on the boundary of the ORF24-NTD identified in
240 Figure 1 and sequence alignments (**Supplemental Figure 3**), we designed constructs for mu24,
241 BcRF1, and UL87 that encompassed either an analogous NTD, or versions of all the homologs
242 wherein the domain is truncated by ten a.a. or extended by 25 a.a. (**Figure 5A**). Based on
243 isoelectric point (pI) calculations (Kozlowski, 2016), the full-length proteins are predicted to be
244 electropositive, but the N-terminal protein-protein interaction domains are predicted to be
245 electronegative (**Figure 5B**). The BcRF1 NTD appears to be an outlier in terms of predicted pI,
246 which may change its physical properties when taken out of context of the full-length protein.
247 As expected, Strep-tagged ORF24 a.a. 1-201 (ORF24-NTD) or a.a. 1-226 both
248 interacted with Pol II in Strep affinity purifications from whole cell lysate (**Figure 5C**). Notably,
249 reducing the ORF24-NTD by even ten a.a. eliminated its ability to interact with Pol II, suggesting
250 that ORF24-NTD (a.a. 1-201) is, or is nearly, the minimal domain for the Pol II interaction
251 (**Figure 5C**). The mu24 and UL87 proteins showed a similar pattern, in which the predicted
252 domain equivalent to ORF24-NTD was sufficient for Pol II interaction, but further truncation by
253 10 a.a. eliminated the interaction (**Figure 5D, F**). In contrast, all truncations of BcRF1 failed to
254 interact with Pol II (**Figure 5E**).
255 The minimal Pol II interaction domain identified here varies greatly in length (191 a.a. in
256 mu24 vs. 248 a.a. in UL87) and in sequence, as no residues are conserved in both β- and γ-
257 herpesviruses other than the RLLLG motif (**Supplemental Figure 3**). Despite this significant
258 variation, ORF24 and its homologs have converged on a mechanism for Pol II recruitment that
259 is primarily mediated by specific N-terminal domains.

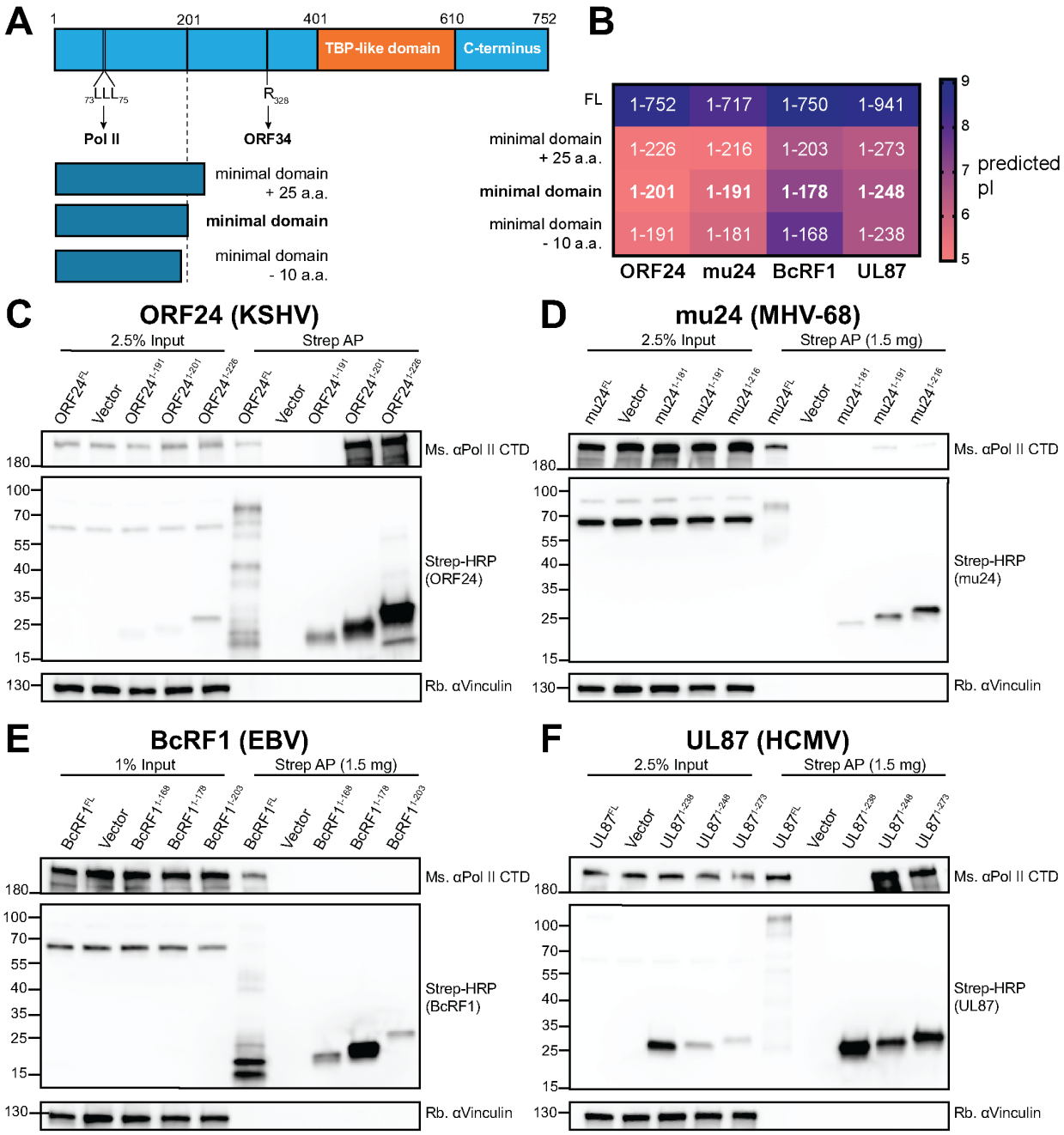


Figure 5. The N-terminal domain of ORF24 homologs from other β- and γ-herpesviruses is sufficient for interaction with Pol II.

(A) Schematic of truncated constructs. The minimal domain of each homolog, corresponding to ORF24 1-201, was predicted based on a sequence alignment (**Supplementary Figure 1**). Additional constructs included ones in which the minimal domain was extended by 25 amino acids or shortened by 10 amino acids.

(B) Predicted isoelectric point for the constructs described in (A).

(C-F) Full-length or truncated Strep-tagged constructs of ORF24 (C) or homologs from MHV68 (mu24; D), EBV (BcRF1; E), and HCMV (UL87; F) were transiently transfected into HEK293T cells then coimmunoprecipitated with StrepTactinXT beads followed by Western blotting.

261 **Chimeric ORF24 retains its ability to interact with Pol II and ORF34**

262 We next wondered if the Pol II recruitment function of the N-terminal domain behaves
263 independently of the other ORF24 functions in vTA and DNA binding. To test this, we generated
264 chimeras of ORF24 wherein the ORF24-NTD (a.a. 1-201) was replaced by the minimal NTD of
265 its homologs (**Figure 6A**). These chimeras retain the region of ORF24 that we previously
266 identified as important for the ORF24-ORF34 interaction (Davis et al., 2016) as well as the
267 ORF24 vTBP domain (Wyrwicz and Rychlewski, 2007) and C-terminal domain. We noted that
268 the full-length homologs of ORF24 do not interact with KSHV ORF34, despite conservation of
269 an arginine (ORF24 R328) essential for the ORF24-ORF34 interaction (Davis et al., 2016)
270 (**Figure 6B**). However, each of the NTD chimeras interacted with both Pol II and KSHV ORF34,
271 suggesting that the minimal NTD is sufficiently well-folded when fused to ORF24 a.a. 202-752
272 (**Figure 6C**). Interestingly, although the minimal domain of BcRF1 alone failed to interact with
273 Pol II by co-IP (**Figure 5E**), it is capable of interaction with Pol II when fused to ORF24 (**Figure**
274 **6C**), suggesting that the isolated fragment may have properties (i.e. isoelectric point- see
275 **Figure 5B**) not compatible with Pol II interaction *in vitro*.

276 Finally, we assessed the ability of the NTD chimeras to functionally complement ORF24
277 using an established transfection-based late gene transcription assay (Castaneda and
278 Glaunsinger, 2019, Davis et al., 2016, Didychuk et al., 2019). The six vTAs (ORFs 18, 30, 31,
279 34, 66 and either WT ORF24, its homologs, or chimeras) were co-transfected into HEK293T
280 cells, along with a firefly luciferase reporter controlled by the K8.1 late gene promoter or, as a
281 control, the early ORF57 promoter. A Renilla luciferase reporter was also included to control for
282 transfection efficiency. Consistent with previous observations (Aubry et al., 2014), none of the
283 full-length homologs could functionally complement ORF24 to activate the K8.1 promoter
284 (**Figure 6D**). However, the mu24-ORF24 chimera promoted transcription to levels ~40% that of
285 wild-type ORF24. Interestingly, neither the BcRF1-ORF24 or UL87-ORF24 chimeras were
286 functional for late gene transcription, despite the fact that they could interact with both ORF34

287 and Pol II **(Figure 6C**). This may suggest that the N-terminal domain has additional functions or
288 interactions beyond polymerase recruitment that mu24 is able to maintain due to sequence
289 similarity, or it may simply be that the mu24 fusion (but not the BcRF1 or UL87 fusions) is folded
290 or oriented in a manner similar enough to ORF24 to be functional. Thus, although the NTD of all
291 vTBPs interacts with Pol II, other contacts or functions may be necessary to successfully
292 promote late gene transcription.
293

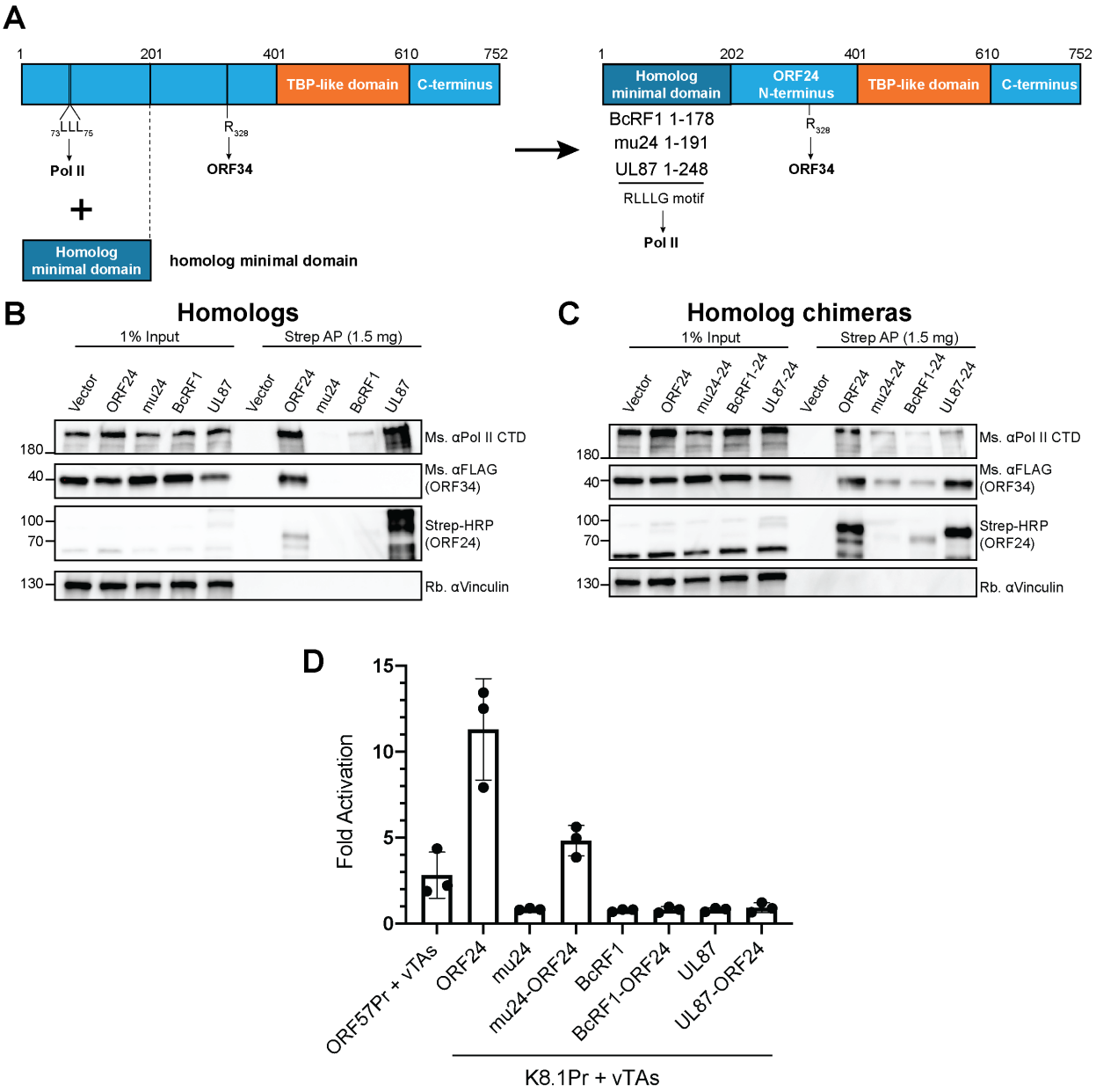


Figure 6. Homolog chimeras of ORF24, but not homologs of ORF24, can interact with ORF34 and transactivate late gene promoters.

(A) Schematic of construct design for ORF24 chimeras. The ORF24-NTD (a.a. 1-201) was replaced with the experimentally identified minimal domain of mu24, BcRF1, and UL87. These chimeric constructs retain the N-terminal ORF24-ORF34 interaction region, the ORF24 vTBP domain, and ORF24 C-terminal tail.

(B) Full-length Strep-tagged homologs of ORF24 were transiently transfected into HEK293T cells along with FLAG-tagged ORF34, then coimmunoprecipitated with StrepTactinXT beads followed by Western blotting.

(C) Full-length Strep-tagged chimeras of ORF24 were transiently transfected into HEK293T cells along with FLAG-tagged ORF34, then coimmunoprecipitated with StrepTactinXT beads followed by Western blotting.

(D) HEK293T cells were transiently transfected with a pGL4.16 firefly luciferase plasmid driven by either the ORF57 (early gene) or K8.1 (late gene) promoter. Plasmids encoding either ORF24, its homologs, or the chimeras, along with the five remaining KSHV vTAs (ORFs 18, 30, 31, 34, and 66) and a pRL-TK renilla luciferase plasmid (as a transfection control) were also transfected. After 24 h, cell lysates were harvested and luciferase activity was measured. Fold activation was calculated by normalizing to the firefly/renilla signal in the absence of vTAs.

295 **DISCUSSION**

296 Late gene transcription in the β- and γ-herpesviruses depends on a set of viral
297 transcriptional activators, including a virally-encoded mimic of host TBP. Here, we demonstrate
298 that the N-terminal domain (NTD) of the vTBP from KSHV, ORF24, recruits Pol II through a
299 direct protein-protein interaction with four heptapeptide repeats of the Pol II C-terminal domain
300 (CTD). Conserved residues in the ORF24-NTD are required for this interaction, suggesting that
301 vTBPs in β- and γ-herpesviruses have converged upon a shared strategy for recruitment of Pol
302 II to late gene promoters. Domain swapping experiments between homologs suggest that while
303 the primary function of the NTD of vTBP in β- and γ-herpesviruses is to recruit Pol II to the late
304 gene promoter, the domain may mediate other interactions or functions. A chimeric vTBP in
305 which the minimal Pol II-interacting domain from ORF24 is replaced with that of mu24 is able to
306 recruit Pol II and maintain interactions with other KSHV vTAs in order to form the vPIC and
307 facilitate transcription. This is the first example of functional interchangeability of any
308 subcomponent of the vTA complex. However, the role of vTBP in late gene transcription
309 extends beyond polymerase recruitment, as vTBP also recruits the remainder of the vTAs to the
310 promoter through protein-protein interactions while also directly binding the late gene promoter
311 DNA. Our work conclusively demonstrates that vTBP is unique among eukaryotic transcriptional
312 activators in its ability to simultaneously bind promoter DNA and Pol II, and that this strategy is
313 conserved across the β- and γ-herpesviruses.

314 Eukaryotic PIC assembly is nucleated by the deployment of TBP onto the promoter DNA
315 by TFIID (Patel et al., 2018) and is followed by recruitment of other GTFs and Pol II. Unlike the
316 direct interaction that occurs between the ORF24 vTBP and Pol II, the interaction between
317 cellular TBP and Pol II is bridged through GTFs. The β- and γ-herpesviruses have thus adopted
318 a solution reminiscent of bacterial sigma factors, wherein vTBP binds promoter DNA while also
319 recruiting the polymerase directly to the promoter (Decker and Hinton, 2013). This raises the
320 question of whether vTBP-mediated Pol II recruitment alters the requirement or roles for other

321 GTFs that contact Pol II. Of particular interest is the role or presence of TFIIB, which both
322 interacts with TBP and is necessary for Pol II recruitment early in PIC formation, functions that
323 may not be necessary in the vPIC given the interactions mediated by vTBP (Chen and Hahn,
324 2003, Liu et al., 2010). The role or presence of other GTFs in vPIC assembly, including TFIIE
325 and TFIIH, which are important for promoter opening and phosphorylation of the Pol II CTD, are
326 also unknown (Nogales et al., 2017). An intriguing possibility is that the other vTAs fulfill some
327 subset of GTF-like functions during formation of the late gene vPIC. In this regard, we have
328 recently demonstrated that the KSHV vTAs ORF30 and ORF66 are essential for stable binding
329 of ORF24 to the K8.1 late gene promoter (Didychuk et al., 2019). Determining which of the
330 canonical GTFs are engaged in the vPIC will be key to understanding how this unique hybrid
331 virus-host complex activates late gene transcription.

332 Our work reveals that ORF24 directly binds hypophosphorylated CTD repeats,
333 consistent with a role facilitating viral PIC formation on late gene promoters. CTD-interacting
334 proteins exhibit exceptional diversity in their strategy for recognition of the CTD and in their
335 preference for phosphorylation (Jeronimo et al., 2013). Relatively few proteins interact with
336 hypophosphorylated CTD; these proteins (TFIID, TFIIE, TFIIF, and Mediator) are all involved in
337 PIC formation, as phosphorylation of the CTD results in release from the PIC into elongating
338 complexes (Eick and Geyer, 2013). The best characterized hypophosphorylated CTD
339 interacting protein is the Mediator coactivator complex (Soutourina, 2018). Mediator is a multi-
340 subunit complex, and its interactions with the Pol II CTD are extensive, involving numerous
341 Mediator subunits (Harper and Taatjes, 2018, Plaschka et al., 2015, Robinson et al., 2012,
342 Robinson et al., 2016). A crystal structure of the Mediator head module with the CTD revealed
343 coordination of nearly four CTD repeats (Robinson et al., 2012). Given the requirement for
344 ORF24-NTD to bind at least four CTD repeats, we are intrigued by the possibility that vTBP
345 functionally or structurally mimics the Mediator head module. One striking difference is the small
346 size of the ORF24-NTD domain that nevertheless efficiently binds the CTD. Mediator is thought

347 to transmit signals from transcription factors bound at regulatory elements to the basal
348 transcriptional machinery (Allen and Taatjes, 2015). Late gene promoters have exceedingly
349 minimal promoters and lack identified enhancer elements (Nandakumar and Glaunsinger,
350 2019); thus, it is unclear if a Mediator-like function is required for transcription, or if recruitment
351 of Pol II to the promoter is sufficient for transcription. The other vTAs may play a role in bridging
352 currently unidentified enhancers or other elements, and may communicate this information to
353 Pol II through the ORF24-ORF34 interaction.

354 In summary, ORF24 is a viral transcriptional activator that replaces TBP at late gene
355 promoters and directly recruits Pol II to transcribe viral late genes. That ORF24 binds both the
356 unphosphorylated CTD of Pol II and promoter DNA makes it unique among known eukaryotic
357 CTD-interacting proteins. Since these two functions of ORF24 are genetically separable, they
358 can be characterized independently of one another. Future work to gain atomic-level insight into
359 how ORF24-NTD coordinates the CTD repeats will advance our understanding of its remarkable
360 role in the regulation of late gene transcription in the β- and γ-herpesviruses.

361

362 **MATERIALS AND METHODS**

363 **Plasmids**

364 All primer sequences are listed in **Table 1**. All plasmids used in this study have been
365 deposited to Addgene. The following ORF24 fragments: residues 1-201 (ORF24-NTD)
366 (Addgene #138420), residues 1-271 (Addgene #138421), and residues 1-133 (Addgene
367 #138422) were PCR amplified from pcDNA4/TO-ORF24-3xFLAG (Davis et al., 2015) (Addgene
368 #138423) with primers to introduce BamHI and NotI sites and cloned into pcDNA4/TO-3xFLAG
369 (C-terminal tag) using T4 DNA ligase (New England Biolabs). pcDNA4/TO-ORF24 202-752-
370 3xFLAG (Addgene #138424) was generated by inverse site-directed mutagenesis with Phusion
371 DNA polymerase (New England Biolabs) using pcDNA4/TO-ORF24-3xFLAG as a PCR

372 template. PCR products from inverse PCR were DpnI treated, then ligated using T4 PNK and
373 T4 DNA ligase (New England Biolabs).

374 To generate the plasmid for GST-ORF24-NTD expression (Addgene #138464), ORF24-
375 NTD (residues 1-201) was PCR amplified from pcDNA4/TO-ORF24-3xFLAG with primers to
376 introduce BamHI and NotI sites and cloned into pGEX4T1 using T4 DNA ligase. pGEX4T1
377 encodes an N-terminal GST tag followed by a thrombin cleavage site. To generate the plasmid
378 for MBP-ORF24-NTD WT (Addgene #138465) and 3L_A (Addgene #138466) expression,
379 ORF24-NTD was PCR amplified from pcDNA4/TO-ORF24-3xFLAG WT (Addgene #138423) or
380 3L_A (Addgene #138425) (Davis et al., 2015) with primers to introduce SacI and BamHI sites
381 and cloned into pMAL-c2X using T4 DNA ligase. pMAL-c2X encodes an N-terminal MBP tag
382 and the plasmids were cloned to express ORF24-NTD with a noncleavable MBP tag. ORF24-
383 NTD was PCR amplified from pcDNA4/TO-ORF24-3xFLAG WT and cloned into the KpnI and
384 EcoRI sites of plasmid p6H-SUMO3 using InFusion. A C-terminal Strep tag on ORF24-NTD was
385 added by inverse PCR to generate p6H-SUMO3-ORF24-NTD-Strep (Addgene #138467).

386 To make the plasmid for GST-Rpb1-linker expression (Addgene #138468), the linker
387 region of Rpb1 (a.a. 1460-1585) was PCR amplified from HEK293T cDNA with primers to
388 introduce BamHI and NotI sites and cloned into pGEX4T1 using T4 DNA ligase. Rpb4 was
389 PCR-amplified from HEK293T cDNA with primers to introduce a BamHI site and cloned into
390 pGEX4T1 using InFusion. The N-terminal BamHI site was regenerated to keep Rpb4 in the
391 same reading frame as the GST fusion tag. Rpb7 was PCR amplified from HEK293T cDNA with
392 primers to introduce a Shine-Delgarno sequence and a C-terminal 6x-His tag and cloned into
393 the EcoRI and NotI sites of pGEX4T1-Rpb4 using T4 DNA ligase to generate pGEX4T1-Rpb4/7
394 (Addgene #138484).

395 The 2x and 4x CTD repeat inserts were ordered as a pair of oligonucleotides from IDT,
396 and the primers were annealed by cooling from 90°C to room temperature in a water bath. The
397 10x CTD repeat insert was ordered as a synthesized gene block from IDT (**Table 2**). All CTD

398 inserts were cloned into the BamHI and NotI sites of pQLink-GST using InFusion cloning
399 (Addgene #138470-138472) (Clontech). pQLink-GST encodes an N-terminal GST tag followed
400 by a TEV protease cleavage site.

401 The following mutations were introduced into pcDNA4/TO-ORF24-3xFLAG using two
402 primer site-directed mutagenesis with KAPA HiFi polymerase (Roche): L73A (Addgene
403 #138426), L74A (Addgene #138427), L75A (Addgene #138428), L73A_L75A (Addgene
404 #138429), L74A_L75A (Addgene #138430), R68A (Addgene #138431), and R72A (Addgene
405 #138432). The L73A_L74A mutation (Addgene #138433) was introduced using InFusion
406 (Clontech) site-directed mutagenesis.

407 Full-length ORF24 with a C-terminal Strep tag (pcDNA4/TO-ORF24-2xStrep) (Addgene
408 plasmid #129742) was previously described (Castaneda and Glaunsinger, 2019). Full-length
409 ORF24 with the 3L_A mutation was subcloned from pcDNA4/TO-ORF24 3L_A-CFLAG
410 (Addgene #138425) into the BamHI and XhoI sites of pcDNA4/TO-2xStrep (C-terminal tag)
411 using InFusion cloning (Addgene #138440). Full-length UL87 was PCR amplified from HCMV
412 Towne BAC DNA with primers to introduced EcoRI and XhoI sites and cloned into pcDNA4/TO-
413 2xStrep (C-terminal tag) using T4 DNA ligase (Addgene #138434). Full-length mu24 was PCR
414 amplified from MHV68-infected 3T3 cell cDNA with primers to introduce BamHI and NotI sites
415 and cloned into pcDNA4/TO-2xStrep (C-terminal tag) using T4 DNA ligase (Addgene #138435).
416 Full-length BcRF1 was PCR amplified from pcDNA4/TO-BcRF1-3xFLAG (Davis et al., 2015)
417 and cloned into the BamHI and XhoI sites of pcDNA4/TO-2xStrep (C-terminal tag) using
418 InFusion cloning (Addgene #138436). Mutations of the RLLLG motif in UL87, mu24, and BcRF1
419 (3L_A mutations) (Addgene #138437-138439) were generated using inverse PCR site-directed
420 mutagenesis.

421 The minimal domains, minimal domain - 10 a.a., and minimal domain + 25 a.a. for
422 ORF24, mu24, BcRF1, and UL87 were PCR amplified from these plasmids and cloned into
423 BamHI/XhoI-cut pcDNA4/TO-2xStrep (C-terminal tag) using InFusion cloning (Addgene

424 #138441-138452). Chimeras of the minimal domain (NTD) ORF24 homologs with ORF24 202-
425 752 were generated using two-insert InFusion cloning (Addgene #138453-138455) into
426 BamHI/XhoI-cut pcDNA4/TO-2xStrep (C-terminal tag).

427 Plasmid K8.1 Pr pGL4.16 (Addgene plasmid #120377) contains the minimal K8.1
428 promoter and ORF57 Pr pGL4.16 (Addgene plasmid #120378) contains a minimal ORF57 early
429 gene promoter and have been described previously (Castaneda and Glaunsinger, 2019).
430 Plasmids pcDNA4/TO-ORF18-2xStrep (Addgene plasmid #120372), pcDNA4/TO-ORF24-
431 2xStrep (Addgene plasmid #129742), pcDNA4/TO-ORF30-2xStrep (Addgene plasmid
432 #129743), pcDNA4/TO-ORF31-2xStrep (Addgene plasmid #129744), pcDNA4/TO-2xStrep-
433 ORF34 (Addgene plasmid #120376) have been previously described (Castaneda and
434 Glaunsinger, 2019). Plasmid pRL-TK (Promega) was kindly provided by Dr. Russell Vance.

435

436 **Tissue Culture and Transfections**

437 HEK293T cells (ATCC CRL-3216) were maintained in DMEM supplemented with 10%
438 FBS (Seradigm). For DNA transfections, HEK293T cells were plated and transfected after 24
439 hours at 70% confluency with PolyJet (SignaGen).

440

441 **Immunoprecipitation and Western Blotting**

442 Cell lysates were prepared 24 hours after transfection by washing and pelleting cells in
443 cold PBS, then resuspending the pellets in IP lysis buffer [50 mM Tris-HCl pH 7.4, 150 mM
444 NaCl, 1 mM EDTA, 0.5% NP-40, and protease inhibitor (Roche)] and rotating for 30 minutes at
445 4°C. Lysates were cleared by centrifugation at 21,000 x *g* for 10 min, then 1-2 mg (as indicated)
446 of lysate was incubated with pre-washed M2 α-FLAG magnetic beads (Sigma) or MagStrep
447 "type3" XT beads (IBA) overnight. Beads were washed 3x for 5 min each with IP wash buffer
448 (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.05% NP-40) and eluted by boiling in 2x
449 Laemmli sample buffer (BioRad).

450 Lysates and elutions were resolved by SDS-PAGE and analyzed by western blot in
451 TBST (Tris-buffered saline, 0.2% Tween 20) using the following primary antibodies: rabbit α-
452 FLAG (Sigma, 1:2500); Strep-HRP (Millipore, 1:2500); rabbit α-Pol II clone N20 (Santa Cruz,
453 1:2500); mouse α-GST clone 8-326 (Pierce, 1:2000); mouse α-MBP (NEB, 1:10000); mouse α-
454 Pol II CTD clone 8WG16 (Abcam, 1:1000), rabbit α-Vinculin (Abcam, 1:1000), mouse anti-
455 GAPDH (1:1,000; Abcam). Following incubation with primary antibodies, the membranes were
456 washed with TBST and incubated with the appropriate secondary antibody. The secondary
457 antibodies used were the following: goat α-mouse-HRP (Southern Biotech, 1:5000) and goat α-
458 rabbit-HRP (Southern Biotech, 1:5000).
459
460 **Protein Expression and Purification**
461 ***GST-ORF24-NTD, GST-CTD linker, and GST.*** Proteins were expressed in Rosetta 2 cells
462 (EMD Millipore) grown in LB at 37°C and induced at an OD_{600} of 0.7 with 0.5 mM IPTG for 16
463 hours at 18°C. The cells were harvested by centrifugation at 6500 x *g* for 10 minutes. The cell
464 pellets were either frozen or immediately resuspended in lysis buffer [50 mM HEPES, pH 7.4,
465 200 mM NaCl, 1 mM EDTA, 1 mM DTT, protease inhibitors (Roche)] and lysed by sonication.
466 The insoluble fraction was removed by centrifugation at 21,000 x *g* for 30 minutes. GST-OR24-
467 NTD, GST-CTD linker, and GST were purified on Glutathione Sepharose (GE Healthcare) by
468 batch purification. The proteins were eluted in wash buffer (50 mM HEPES, pH 7.4, 200 mM
469 NaCl, 1 mM EDTA, 1 mM DTT) containing 10 mM reduced glutathione and dialyzed into storage
470 buffer (50 mM HEPES, pH 7.4, 200 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% glycerol].
471
472 ***GST-xCTD repeats.*** GST-2xCTD repeats, GST-4xCTD repeats, and GST-10xCTD repeats
473 were expressed in BL21 Star (DE3) cells grown in Overnight Express Instant TB Medium (EMD
474 Millipore) at 37°C and induced at an OD_{600} of 1.0 by decreasing the temperature to 18°C and
475 growing for an additional 16 hours. The cells were harvested by centrifugation at 6500 x g for 10

476 minutes. The cell pellets were either frozen or immediately resuspended in lysis buffer [50 mM
477 HEPES, pH 7.4, 300 mM NaCl, 5 mM DTT, 5% glycerol, protease inhibitors (Roche)] and lysed
478 by sonication. The insoluble fraction was removed by centrifugation at 50,000 x *g* for 30
479 minutes. The proteins were purified as described above and eluted in wash buffer (50 mM
480 HEPES, pH 7.4, 300 mM NaCl, 5 mM DTT, 5% glycerol) containing 10 mM reduced glutathione.
481
482 ***MBP-ORF24-NTD.*** The protein was expressed in Rosetta 2 cells grown in LB at 37°C and
483 induced at an OD_{600} of 0.7 with 0.5 mM IPTG for 16 hours at 18°C. The cells were harvested by
484 centrifugation at 6500 x *g* for 10 minutes. The cell pellets were either frozen or immediately
485 resuspended in lysis buffer [50 mM HEPES, pH 7.4, 200 mM NaCl, 1 mM EDTA, 1 mM DTT,
486 protease inhibitors (Roche)] and lysed by sonication. The insoluble fraction was removed by
487 centrifugation at 21,000 x *g* for 30 minutes. MBP-ORF24-NTD was purified by gravity column
488 chromatography with Amylose Resin (New England Biolabs). The protein was eluted in wash
489 buffer (50 mM HEPES, pH 7.4, 200 mM NaCl, 1 mM EDTA, 1 mM DTT) containing 10 mM
490 maltose and dialyzed into storage buffer (50 mM HEPES pH 7.4, 200 mM NaCl, 1mM DTT, 10%
491 glycerol).
492
493 ***ORF24-NTD-Strep.*** The protein was expressed in NiCo21 (DE3) cells (New England Biolabs)
494 grown in Overnight Express Instant TB Medium (EMD Millipore) at 37°C and induced at an
495 OD_{600} of 1.0 by decreasing the temperature to 18°C and growing for an additional 16 hours. The
496 cells were harvested by centrifugation at 6,500 x *g* for 10 minutes. The cell pellets were either
497 frozen or immediately resuspended in lysis buffer [100 mM HEPES, pH 7.5, 500 mM NaCl,
498 0.1% Triton X-100, 10% glycerol, 20 mM imidazole, 1 mM TCEP, protease inhibitors (Roche)]
499 and lysed by sonication. The lysate was cleared by centrifugation at 50,000 x *g* for 30 minutes.
500 The clarified lysate was filtered through a 0.45 µm PES filter (Foxx Life Sciences). The protein
501 was purified on an equilibrated HisTrap (GE Healthcare) and step-eluted in wash buffer (100

502 mM HEPES, pH 7.5, 500 mM NaCl, 0.1% Triton X-100, 10% glycerol, 1 mM TCEP) containing
503 500 mM imidazole. The fractions containing 6xHis-SUMO-ORF24-NTD-Strep were pooled. The
504 SUMO tag was cleaved overnight at 4°C with 1 mg of SenP2 protease. Following cleavage of
505 the SUMO tag, ORF24-NTD-Strep was purified on an equilibrated StrepTrap (GE Healthcare)
506 and step-eluted in wash buffer (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 10 % glycerol, 1 mM
507 TCEP) containing 2.5 mM desthiobiotin (IBA). The Strep elution fractions containing ORF24-
508 NTD-Strep were pooled and sized on a Superdex 200 (GE Healthcare) size exclusion
509 chromatography (SEC) column in SEC buffer (20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM
510 TCEP, 5% glycerol). The fractions containing ORF24-NTD-Strep were pooled and concentrated
511 on a 10K Amicon Ultra-15 concentrator (EMD Millipore). Protein aliquots were flash frozen and
512 stored at -70°C.
513
514 **Pulldown Assays**
515 ***GST pulldowns***
516 To test the interaction between GST-ORF24-NTD and Pol II from mammalian cells, 10
517 µg GST-ORF24-NTD or 10 µg GST was added to 20 µl of washed Glutathione Magnetic
518 Agarose beads (Pierce) along with 250 µg of 293T whole cell lysate. IP wash buffer (50 mM Tris
519 pH 7.4, 150 mM NaCl, 0.05% NP-40, 1 mM EDTA) was added to a final volume of 300 µl. The
520 samples were rotated at 4°C for 1 hour. Following the pulldown, the samples were washed with
521 IP wash buffer three times for 5 minutes each time. After the last wash, the protein was eluted
522 by boiling in 2x Laemmli sample buffer (BioRad). The pulldown to test the interaction between
523 GST-CTD repeats or GST-CTD linker and MBP-ORF24-NTD or MBP-ORF24-NTD 3L_A were
524 performed as described above. The elutions were resolved by SDS-PAGE followed by western
525 blot.
526 ***Strep pulldowns***

527 To test the interaction between ORF24-NTD and Rpb4/7 or xCTD repeats, 5 µg of
528 ORF24-NTD-Strep was added to 10 µl of washed MagStrep "type 3" XT Beads (IBA) along with
529 10 µg of GST-Rpb4/Rpb7-6xHis or GST-xCTD repeats. SEC buffer (20 mM HEPES pH 7.4, 100
530 mM NaCl, 1 mM TCEP, 5% glycerol) including 0.05% NP-40 and 5 mM DTT was added to a
531 final volume of 300 µl. The pulldowns were rotated at 4°C for 1 hour followed by three 5 minute
532 washes. The protein was eluted by boiling in 2x Laemmli sample buffer. Elutions were resolved
533 by SDS-PAGE on a Stain-free gel (BioRad) followed by western blot.
534
535 **Late Gene Reporter Assay**
536 HEK293T cells (1 x 10^6 cells) were plated in 6-well plates and after 24 h each well was
537 transfected with 900 ng of DNA containing 125 ng each of pcDNA4/TO-ORF18-2xStrep, wild-
538 type pcDNA4/TO-ORF24-2xStrep or a homolog of ORF24, pcDNA4/TO-ORF30-2xStrep,
539 pcDNA4/TO-ORF31-2xStrep, pcDNA4/TO-2xStrep-ORF34, pcDNA4/TO-ORF66-2xStrep (or as
540 a control, 750 ng of empty pcDNA4/TO-2xStrep plasmid in lieu of vTA plasmids), with either
541 K8.1 Pr pGL4.16 or ORF57 Pr pGL4.16, along with 25 ng of pRL-TK as an internal transfection
542 control. After 24 h, cells were rinsed twice with PBS, lysed by rocking for 15 min at room
543 temperature in 500 µL of Passive Lysis Buffer (Promega), and clarified by centrifugation at
544 21,000 x g for 2 min. 20 µL of the clarified lysate was added in triplicate to a white chimney well
545 microplate (Greiner Bio-One) to measure luminescence on a Tecan M1000 microplate reader
546 using a Dual Luciferase Assay Kit (Promega). The firefly luminescence was normalized to the
547 internal Renilla luciferase control for each transfection. All samples were normalized to the
548 corresponding control containing empty plasmid.
549
550 **Negative Stain Electron Microscopy**
551 ***PIC assembly and purification.*** TBP, TFIIA, and TFIIB were recombinantly expressed and
552 purified from *E. coli*. Pol II was immunopurified from HeLa cell nuclear extracts following

553 previously published protocols (Knuesel et al., 2009, Revyakin et al., 2012). The DNA construct
554 was previously described (He et al., 2013) and is a SCP (Juven-Gershon et al., 2006)
555 containing a BREu element upstream of the TATA box (Tsai and Sigler, 2000) and a EcoRI
556 restriction enzyme site downstream of the INR element for purification purposes. A biotin tag
557 was coupled to the 5' end of the template strand (Integrated DNA Technologies). The duplex
558 DNA was generated by annealing the single-stranded template strand DNA with equimolar
559 amounts of non-template strand DNA at a final concentration of 50 μM in water. The annealing
560 reaction was carried out by incubating at 98°C for 2 minutes followed by cooling to room
561 temperature at a rate of 1°C per second.

562 PICs were assembled in assembly buffer (20 mM HEPES, pH 7.9, 0.2 mM EDTA, 10%
563 glycerol, 6 mM $MgCl_2$, 80 mM KCl, 1 mM DTT, 0.05% NP-40). DNA was used as a scaffold and
564 purified TBP, TFIIA, TFIIB, Pol II, TFIIF, and GST-ORF24-NTD were sequentially added into the
565 assembly buffer. Following assembly of the PICs, the reaction was incubated at 28°C for 15
566 minutes using a 1:10 dilution of magnetic streptavidin T1 beads (Invitrogen), which had been
567 previously equilibrated in assembly buffer. The beads were washed three times with wash buffer
568 (10 mM HEPES, 3% trehalose, 8 mM $MgCl_2$, 100 mM KCl, 1 mM DTT, 0.025% NP-40). The
569 complex was eluted by incubation at 28°C for 1 hour in digestion buffer (10 mM HEPES, pH 7.9,
570 3% trehalose, 10 mM $MgCl_2$, 50 mM KCl, 1 mM DTT, 0.01% NP-40, 1 unit μL^{-1} EcoRI-HF (New
571 England Biolabs)). After elution, purified PIC was crosslinked on ice in 0.05% glutaraldehyde for
572 5 minutes then immediately used for EM sample preparation.

573

574 ***Electron microscopy.*** Negative stain samples of PIC were prepared on a 400 mesh copper
575 grid containing a continuous carbon supporting layer. The grid was plasma-cleaned for 10
576 seconds using a Solarus plasma cleaner (Gatan). An aliquot (3.5 μL) of the purified sample was
577 placed onto the grid and allowed to absorb for 5 minutes at 100% humidity. The grid was then
578 placed sample-side-down on five successive 75 μL drops of 2% (w/v) uranyl formate solution for

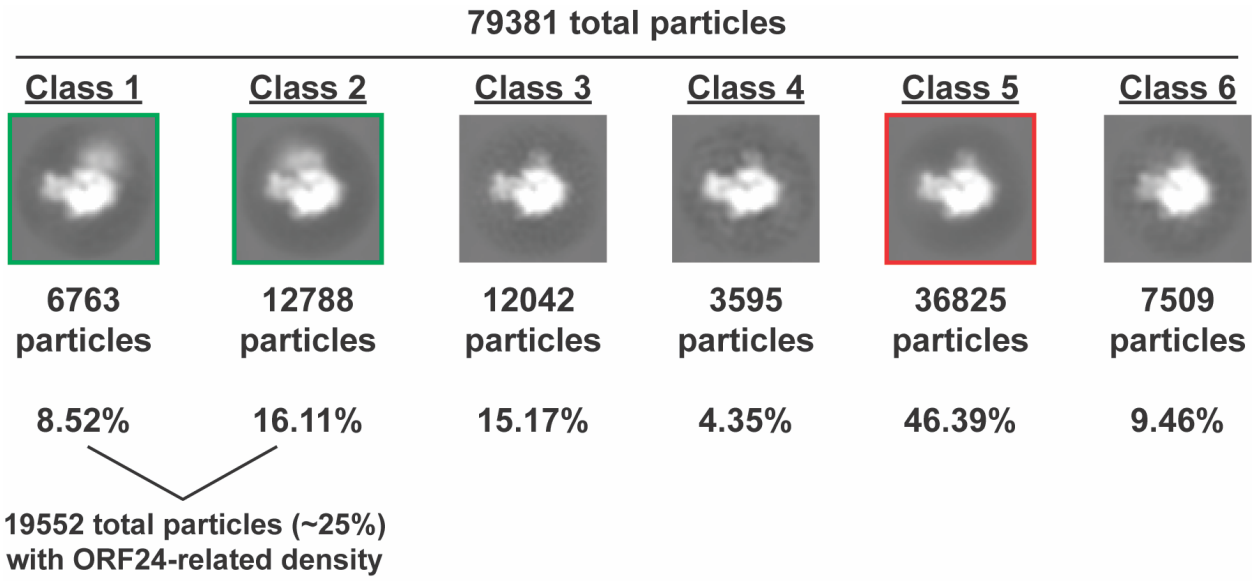
579 10 seconds on each drop followed by blotting to dryness. Data collection was performed on a
580 Tecnai F20 TWIN transmission electron microscope operating at 120 keV at a nominal
581 magnification of 80,000X (1.5 Å/pixel). The data were collected on a 4k X 4k CCD (Gatan) using
582 low-dose procedures (20 e^{-} $Å^{-2}$ total dose per exposure), using Leginon software to
583 automatically focus and collect exposure images.
584
585 ***Image processing.*** Data pre-processing was performed using the Appion processing
586 environment (Lander et al., 2009). Particles were automatically selected from the micrographs
587 using a difference of Gaussians (DoG) particle picker (Voss et al., 2009). The contrast transfer
588 function (CTF) of each micrograph was estimated using both ACE2 and CTFFind (Mallick et al.,
589 2005, Mindell and Grigorieff, 2003). Boxed particle images were extracted using a box size of
590 256 X 256 pixels from images whose ACE2 confidence value was greater than 0.8, phases
591 were flipped, and images were normalized using the XMIPP to remove pixels which were above
592 or below 4.5σ of the mean value (Sorzano et al., 2004). The particle stack was binned by a
593 factor of two and two-dimensional classification was conducted using iterative multireference
594 alignment analysis (MSA-MRA) within the IMAGIC software (van Heel et al., 1996).
595
596 ***Three-dimensional reconstruction*****.** Particles belonging to bad two-dimensional classes were
597 thrown out, resulting in a stack of 79381 single particle images that were used for three-
598 dimensional analysis. Three-dimensional classification was performed within RELION (Scheres,
599 2012), using the cryo-EM structural of a minimal PIC (EMD-2305, (He et al., 2013)), low-pass
600 filtered to 50 Å resolution, as an initial reference, and sorted into 6 classes. The resolution of the
601 reconstructions containing GST-ORF24-NTD were estimated to be ~20 Å.
602
603 **ACKNOWLEDGEMENTS**

We thank Jie Fang for purification of human GTFs and Pol II, and Patricia Grob and Abhiram Chintangal for microscope and computational support, respectively. We are thankful to all members of the Glaunsinger lab, especially Divya Nandakumar, for helpful discussions and suggestions. This research was supported by NIH R01AI122528 to B.G. and R01GM63072 and R35GM127018 to E.N. A.C. was supported by a National Science Foundation Graduate Research Fellowship (DGE 1752814) and a UC Berkeley Chancellor's Fellowship. A.L.D. is The Rhee Family Fellow of the Damon Runyon Cancer Research Foundation (DRG-2349-18). B.G. and E.N. are investigators of the Howard Hughes Medical Institute.

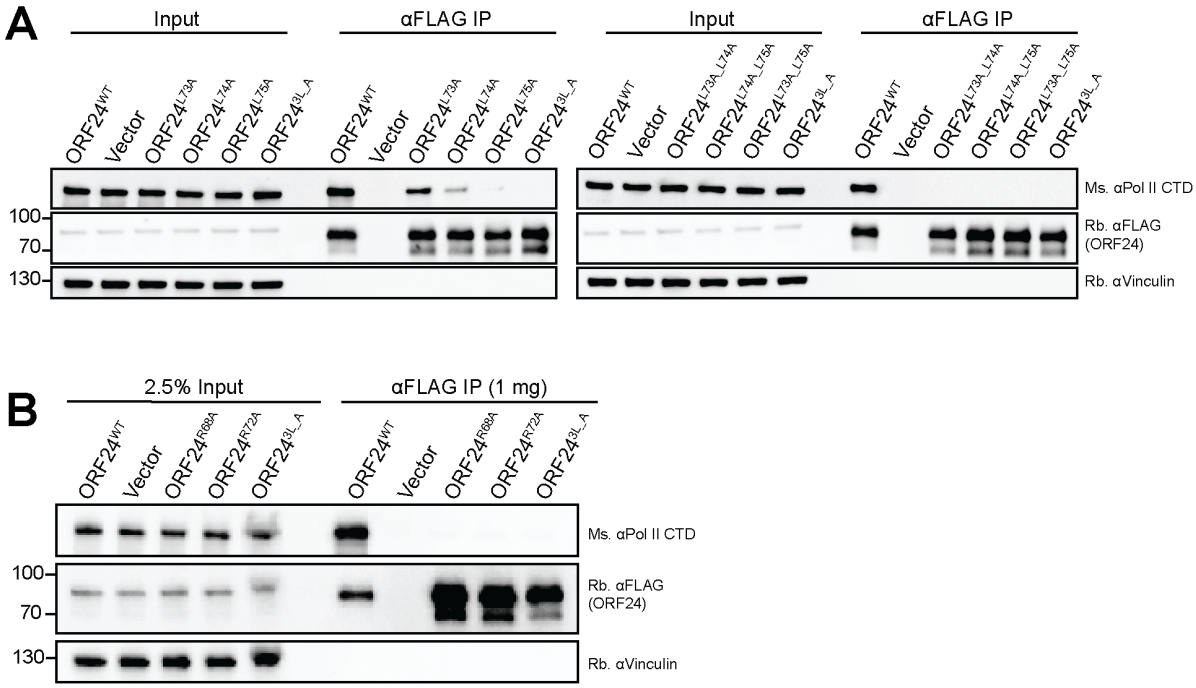
COMPETING INTERESTS

None

SUPPLEMENTARY FIGURES



Supplemental Figure 1. (associated with Figure 2)
Two-dimensional projections of the three-dimensional classes resulting from sorting the single particle EM images of negatively stained minimal PICs (TBP/TFIIA/TFIIB/TFIIF/Pol II/DNA) assembled in the presence of GST-ORF24-NTD. The number of particles assigned to each class and the percent of total particles are indicated. Classes 1, 2, and 5 were used for the difference mapping shown in Figure 2.

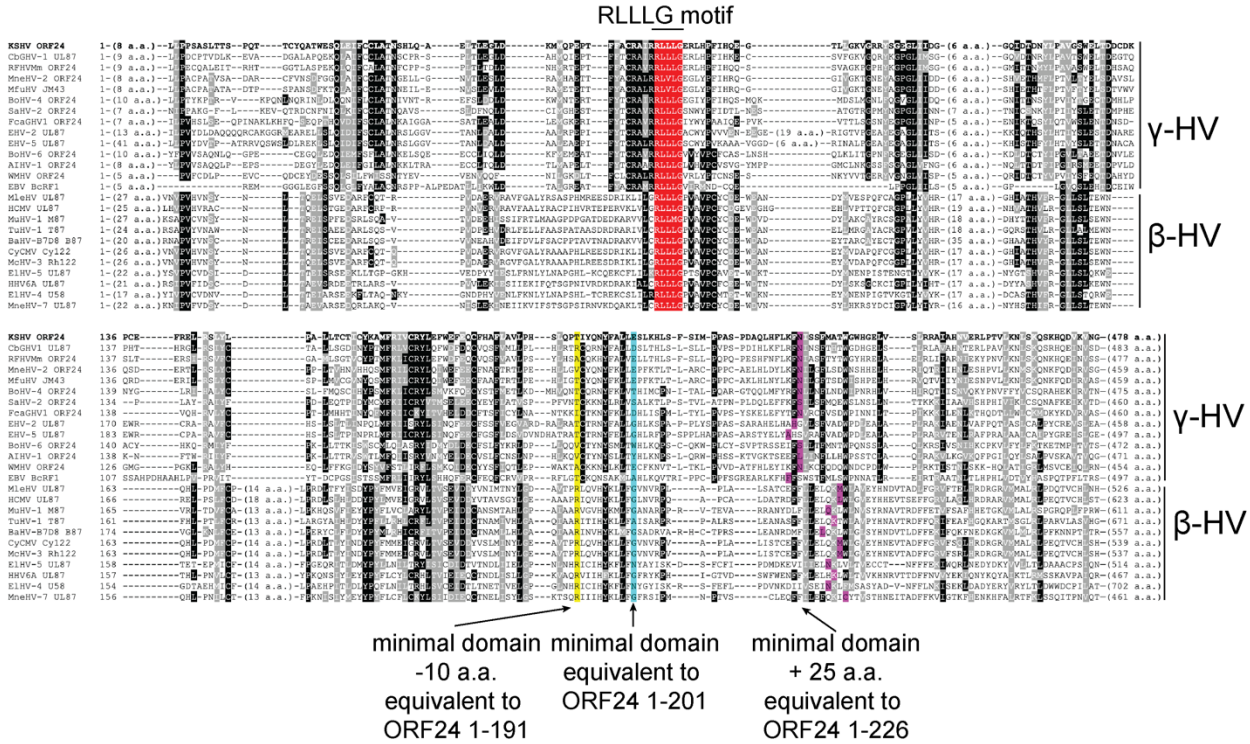


Supplemental Figure 2.

HEK293T cells were transiently transfected with full-length wild-type or mutant FLAG-tagged ORF24 and co-immunoprecipitated with anti-FLAG beads followed by western blotting with the indicated antibodies to detect ORF24 and Pol II.

(A) Mutations in the triple leucine motif are additive in their effect in disrupting the interaction between ORF24 and Pol II.

(B) Mutations outside of the triple leucine motif similarly disrupt the interaction between ORF24 and Pol II.



γ-herpesviruses:

YP_001129377.1 ORF24 [Human gammaherpesvirus 8] (KSHV ORF24)
QDQ69232.1 virion UL87 [Colobine gammaherpesvirus 1] (CbGHV-1 UL87)
AGY30707.1 ORF24 [Retroperitoneal fibromatosis-associated herpesvirus] (RFHVMm ORF24)
AJE29665.1 ORF24 [Macaca nemestrina rhadinovirus 2] (MneHV-2 ORF24)
AAT00020.1 JM43 [Macaca fuscata rhadinovirus] (MfuHV JM43)
AIA82769.1 hypothetical protein [Bovine gammaherpesvirus 4] (BoHV-4 ORF24)
CAA73638.1 hypothetical protein [Saimiriine gammaherpesvirus 2] (SaHV-2 ORF24)
YP_009173899.1 ORF24 [Felis catus gammaherpesvirus 1] (FcaGHV1 ORF24)
NP_042620.2 protein UL87 [Equid gammaherpesvirus 2] (EHV-2 UL87)
YP_009118414.1 protein UL87 [Equid gammaherpesvirus 5] (EHV-5 UL87)
YP_009042003.1 ORF24 [Bovine gammaherpesvirus 6] (BoHV-6 ORF24)
ATQ19112.1 ORF24 [Alcelaphine gammaherpesvirus 1] (AHV-1 ORF24)
ABO48404.1 virion protein [Wood mouse herpesvirus] (WMHV ORF24)
QCG99689.1 BoRF1 [Human gammaherpesvirus 4] (EBV BoRF1)

β-herpesviruses:

AKI29760.1 protein UL87 [Mandrillus leucophaeus cytomegalovirus] (MleHV UL87)
AQN73910.1 protein UL87 [Human betaherpesvirus 5] (HCMV UL87)
CCE56592.1 M87 protein [Murid betaherpesvirus 1] (MuHV-1 M87)
NP_116436.1 T87 [Tupaiid betaherpesvirus 1] (TuHV-1 T87)
AFK83924.1 B87 [Bat betaherpesvirus B7D8] (BaHV-B7D8 B87)
YP_009337527.1 Cy122 [Cynomolgus cytomegalovirus] (CyCMV Cy122)
APT40175.1 Rh122 [Macacine betaherpesvirus 3] (McHV-3 Rh122)
YP_009051991.1 protein UL87 [Elephant endotheliotropic herpesvirus 5] (EIHV-5 UL87)
APO38815.1 protein UL87 [Human betaherpesvirus 6A] (HHV6A UL87)
YP_009179325.1 protein U58 [Elephant endotheliotropic herpesvirus 4] (EIHV-4 U58)
YP_009253965.1 protein UL87 [Macaca nemestrina herpesvirus 7] (MneHV-7 UL87)

Supplemental Figure 3. (associated with Figure 5)

Multiple sequence alignment of homologs of ORF24 from other β- and γ-herpesviruses. The conserved triple leucine motif is highlighted in red. The location of truncations for the constructs used in Figure 5 are highlighted in yellow, teal, and purple.

704
705
706
707
708
709
710

Table 1. List of oligonucleotides used in this study

#	Name	Sequence 5'-3'	Figure	Construct
1	ORF24 1-133_pCDNA_F	GCTCGGATCCATGGCAGCGCTCGAGGG	1	pCDNA4.TO-ORF24 1-133-CFLAG
2	ORF24 1-133_pCDNA_R	TCGAGCGGCCGCCCCACAATCATCGGTAAGTTCCCATGATC	1	pCDNA4.TO-ORF24 1-133-CFLAG
3	ORF24 1-201_pCDNA_F	GCTCGGATCCATGGCAGCGCTCGAGGG	1	pCDNA4.TO-ORF24 1-201-CFLAG
4	ORF24 1-201_pCDNA_R	TCGAGCGGCCGCCCCTCCAGGAGTGCAAAATAATTTTGATAGATTG	1	pCDNA4.TO-ORF24 1-201-CFLAG
5	ORF24 1-271_pCDNA_F	GCTCGGATCCATGGCAGCGCTCGAGGG	1	pCDNA4.TO-ORF24 1-271-CFLAG
6	ORF24 1-271_pCDNA_R	TCGAGCGGCCGCTTCTTGACGTCCTGGTGCTTACTCT	1	pCDNA4.TO-ORF24 1-271-CFLAG
7	ORF24 202-752_pCDNA_F	ATGAGCCTGAAGCATCTCTCG	1	pCDNA4.TO-ORF24 202-752-CFLAG
8	ORF24 202-752_pCDNA_R	GGATCCGAGCTCGGTACCAAG	1	pCDNA4.TO-ORF24 202-752-CFLAG
9	ORF24 1-201_pGEX_F	GCGTGGATCCATGGCAGCGCTCGAGG	1, 2	pGEX4T1-ORF24-NTD
10	ORF24 1-201_pGEX_R	CGATGCGGCCGCTTACTCCAGGAGTGCAAAATAATTTTGATAGATTGTG	1, 2	pGEX4T1-ORF24-NTD
11	ORF24 1-201_pMAL_F	ATTCGAGCTCAATGGCAGCGCTCGAGGG	3	pMAL-c2x-ORF24-NTD, WT and ΔLLL
12	ORF24 1-201_pMAL_R	TAGAGGATCCTTACTCCAGGAGTGCAAAATAATTTTGATAGATTGT	3	pMAL-c2x-ORF24-NTD, WT and ΔLLL
13	Rpb1_Linker aa 1460-1585_F	GCGTGGATCCCTGGGCCAGCTGGCTC	3	pGEX4T1-Rpb1-linker
14	Rpb1_Linker aa 1460-1585_R	CGATGCGGCCGCTGGTGAAGGGATGTAGGGGCT	3	pGEX4T1-Rpb1-linker
15	Rpb4 pGEX4T1_F	GGTTCCGCGTGGATCCATGGCGGCGGGTGG	3	pGEX4T1-hsRpb4
16	Rpb4 pGEX4T1_R	GGAATTCCGGGGATCTTAATACTGAAAGCTGCGCT	3	pGEX4T1-hsRpb4
17	Rpb7 pGEX4T1_F	CCCGGAATTCaggaggTAATTAAATatgttctaccatatctccctagagcac	3	pGEx4T1-hsRpb4/7
18	Rpb7 pGEX4T1_R	CGATGCGGCCGCttagtgatggtggtgatggtggcttacaagccccaagtaatcgt	3	pGEX4T1-hsRpb4/7
19	ORF24_2-201_p6H-SUMO3_F	cagcagacgggagggGCAGCGCTCGAGGGC	3	p6H-SUMO3-ORF24-NTD-Strep
20	ORF24_2-201_p6H-SUMO3_R	gacggagctcgaattTTATTTTTCAAACTGCGGATGGCTCCACTCCAGGAGTGCA AAATAATTTTGATAGATTG	3	p6H-SUMO3-ORF24-NTD-Strep
21	2xCTD_repeat_pQLink_F	TTATTTTCAGGGATCCTATTCTCCGACCTCTCCATCATACAGCCCTACCTC CCCGTCCTAGGCGGCCGCCTAGGACCC	3	pQLink-GST-2xCTD
22	2xCTD_repeat_pQLink_R	GGGTCCTAGGCGGCCGCCTAGGACGGGGAGGTAGGGCTGTATGATGGA GAGGTCGGAGAATAGGATCCCTGAAAATAA	3	pQLink-GST-2xCTD
23	4xCTD_repeat_pQLink_F	TTATTTTCAGGGATCCTATAGCCCCACGAGCCCTTCTTACTCTCCGACCTC ACCATCCTATTCGCCTACTAGCCCGAGTTACAGTCCCACATCTCCGTCCT AGGCGGCCGCCTAGGACCC	3	pQLink-GST-4xCTD
24	4xCTD_repeat_pQLink_R	GGGTCCTAGGCGGCCGCCTAGGACGGAGATGTGGGACTGTAACTCGGG CTAGTAGGCGAATAGGATGGTGAGGTCGGAGAGTAAGAAGGGCTCGTGG GGCTATAGGATCCCTGAAAATAA	3	pQLink-GST-4xCTD
25	ORF24 (L73A)_F	CCCCAGGAGTGCTCTGCGTATCGCTCTGCAG	S2	pCDNA4.TO-ORF24 L73A-CFLAG
26	ORF24 (L73A)_R	CTGCAGAGCGATACGCAGAGCACTCCTGGGG	S2	pCDNA4.TO-ORF24 L73A-CFLAG
27	ORF24 (L74A)_F	GCGTTCCCCCAGGGCTAGTCTGCGTATCGC	S2	pCDNA4.TO-ORF24 L74A-CFLAG
28	ORF24 (L74A)_R	GCGATACGCAGACTAGCCCTGGGGGAACGC	S2	pCDNA4.TO-ORF24 L74A-CFLAG
29	ORF24 (L75A)_F	GGAGGCGTTCCCCCGCGAGTAGTCTGCGTA	S2	pCDNA4.TO-ORF24 L75A-CFLAG
30	ORF24 (L75A)_ R	TACGCAGACTACTCGCGGGGGAACGCCTCC	S2	pCDNA4.TO-ORF24 L75A-CFLAG
31	ORF24 (L73A_L75A)_F	GGCGTTCCCCCGCGAGTGCTCTGCGTATCGCTCTGCAGGC	S2	pCDNA4.TO-ORF24 L73A/L75A-CFLAG
32	ORF24 (L73A_L75A)_R	GCCTGCAGAGCGATACGCAGAGCACTCGCGGGGGAACGCC	S2	pCDNA4.TO-ORF24 L73A/L75A-CFLAG
33	ORF24 (L74A_L75A)_F	GGAGGCGTTCCCCCGCGGCTAGTCTGCGTATCGCTCTGC	S2	pCDNA4.TO-ORF24 L74A/L75A-CFLAG
34	ORF24 (L74A_L75A)_R	GCAGAGCGATACGCAGACTAGCCGCGGGGGAACGCCTCC	S2	pCDNA4.TO-ORF24 L74A/L75A-CFLAG
35	ORF24 (L73A_L74A)_F	GCAGAGCAGCCCTGGGGGAACGCCTCCACC	S2	pCDNA4.TO-ORF24 L73A/L74A-CFLAG
36	ORF24 (L73A_L74A)_R	CCAGGGCTGCTCTGCGTATCGCTCTGCAGG	S2	pCDNA4.TO-ORF24 L73A/L74A-CFLAG
37	ORF24 (R68A)_F	gtagtctgcgtatcgctgcgcaggcgaaaaaggtgg	S2	pCDNA4.TO-ORF24 R68A-CFLAG
38	ORF24 (R68A)_R	ccacctttttcgcctgcgcagcgatacgcagactac	S2	pCDNA4.TO-ORF24 R68A-CFLAG
39	ORF24 (R72A)_F	cccccaggagtagtgcgcgtatcgctctgc	S2	pCDNA4.TO-ORF24 R72A-CFLAG
40	ORF24 (R72A)_R	gcagagcgatacgcgcactactcctggggg	S2	pCDNA4.TO-ORF24 R72A-CFLAG
41	ORF24_F	TACCGAGCTCGGATCCATGGCAGCGCTCGAGGGC	4	pCDNA4.TO-ORF24 3L_A-CSTREP
42	ORF24_R	caccgcctccctcgaggacCAGCGGACG	4	pCDNA4.TO-ORF24 3L_A-CSTREP
43	BcRF1_F	TACCGAGCTCGGATCCATGACACAAGGTAAGAGGGAGATGG	4, 5, 6	pCDNA4.TO-BcRF1-CSTREP
44	BcRF1_R	CACCGCCTCCCTCGAGCACTTGAGCATCACGGCAGTGG	4, 5, 6	pCDNA4.TO-BcRF1-CSTREP
45	BcRF1 (3L_A)_F	GCCGCGGCAggcgttatccgaataaatgactgccaggag	4	pCDNA4.TO-BcRF1 3L_A-CSTREP
46	BcRF1 (3L_A)_R	acgacgcatagcccgacacgcgtaaaaggtg	4	pCDNA4.TO-BcRF1 3L_A-CSTREP
47	mu24_F	TACCGAGCTCGGATCCATGACAATATTCTTACCAGTATTCTGTGATTTGC	4, 5, 6	pCDNA4.TO-mu24-CSTREP
48	mu24_R	CTCCCTCGAGCGGCCGCCCGGAGTCTGGTTGGCAAGG	4, 5, 6	pCDNA4.TO-mu24-CSTREP
49	mu24 (3L_A)_F	GCCGCGGCAggcataagactgtacccaacctgcaatactg	4	pCDNA4.TO-mu24 3L_A-CSTREP
50	mu24 (3L_A)_R	tcttcttaaagctctacataaacagaatgt	4	pCDNA4.TO-mu24 3L_A-CSTREP
51	UL87_F	AAAAGAATTCATGGCCGGCGCTGC	4, 5, 6	pCDNA4.TO-UL87-CSTREP
52	UL87_R	AAAACTCGAGTCGTGATGCAAACCGCAC	4, 5, 6	pCDNA4.TO-UL87-CSTREP
53	UL87 (3L_A)_F	GCCGCGGCAGGACCCGTGGCCGTACCCTGTTTTTGCGAC	4	pCDNA4.TO-UL87 3L_A-CSTREP

54	UL87 (3L_A)_R	ACGACCCAGTACCAGCTTGACACGCTCGGA	4	pCDNA4.TO-UL87 3L_A-CSTREP
55	ORF24_1-191_pCDNA_F	TACCGAGCTCGGATCCATGGCAGCGCTCGAGGGC	5	pCDNA4.TO-ORF24 1-191-CSTREP
56	ORF24_1-191_pCDNA_R	CACCGCCTCCCTCGAGTGTGGGTTGTAGACTATGGGGC	5	pCDNA4.TO-ORF24 1-191-CSTREP
57	ORF24_1-201_pCDNA_F	TACCGAGCTCGGATCCATGGCAGCGCTCGAGGGC	5	pCDNA4.TO-ORF24 1-201-CSTREP
58	ORF24_1-201_pCDNA_R	CACCGCCTCCCTCGAGCTCCAGGAGTGCAAAATAATTTTG	5	pCDNA4.TO-ORF24 1-201-CSTREP
59	ORF24_1-226_pCDNA_F	TACCGAGCTCGGATCCATGGCAGCGCTCGAGGGC	5	pCDNA4.TO-ORF24 1-226-CSTREP
60	ORF24_1-226_pCDNA_R	CACCGCCTCCCTCGAGGTTAAACTTTAAAAAATGTAGC	5	pCDNA4.TO-ORF24 1-226-CSTREP
61	BcRF1_1-168_pCDNA_F	TACCGAGCTCGGATCCATGACACAAGGTAAGAGGGAGATGG	5	pCDNA4.TO-BcRF1 1-168-CSTREP
62	BcRF1_1-168_pCDNA_R	CACCGCCTCCCTCGAGGGTCCCTAGGAAGCGACGC	5	pCDNA4.TO-BcRF1 1-168-CSTREP
63	BcRF1_1-178_pCDNA_F	TACCGAGCTCGGATCCATGACACAAGGTAAGAGGGAGATGG	5	pCDNA4.TO-BcRF1 1-178-CSTREP
64	BcRF1_1-178_pCDNA_R	CACCGCCTCCCTCGAGAGCCAGCATCTTTGCAGAGTTTTGC	5	pCDNA4.TO-BcRF1 1-178-CSTREP
65	BcRF1_1-203_pCDNA_F	TACCGAGCTCGGATCCATGACACAAGGTAAGAGGGAGATGG	5	pCDNA4.TO-BcRF1 1-203-CSTREP
66	BcRF1_1-203_pCDNA_R	CACCGCCTCCCTCGAGGAAGTGGAACTTGAGTCTGGCC	5	pCDNA4.TO-BcRF1 1-203-CSTREP
67	mu24_1-181_pCDNA_F	TACCGAGCTCGGATCCATGACAATATTCTTACCAGTATTC	5	pCDNA4.TO-mu24 1-181-CSTREP
68	mu24_1-181_pCDNA_R	CACCGCCTCCCTCGAGAGCAGACCTGGCCCAGGG	5	pCDNA4.TO-mu24 1-181-CSTREP
69	mu24_1-191_pCDNA_F	TACCGAGCTCGGATCCATGACAATATTCTTACCAGTATTC	5	pCDNA4.TO-mu24 1-191-CSTREP
70	mu24_1-191_pCDNA_R	CACCGCCTCCCTCGAGCTCAGTAAGCTTCAAGTAATTTTTT	5	pCDNA4.TO-mu24 1-191-CSTREP
71	mu24_1-216_pCDNA_F	TACCGAGCTCGGATCCATGACAATATTCTTACCAGTATTC	5	pCDNA4.TO-mu24 1-216-CSTREP
72	mu24_1-216_pCDNA_R	CACCGCCTCCCTCGAGATTGAATTTGATATATTCCAAATGA	5	pCDNA4.TO-mu24 1-216-CSTREP
73	UL87_1-238_pCDNA_F	TACCGAGCTCGGATCCATGGCCGGCGCTGCGCCG	5	pCDNA4.TO-UL87 1-238-CSTREP
74	UL87_1-238_pCDNA_R	CACCGCCTCCCTCGAGGCGCGGCGCCGCGGCCTCG	5	pCDNA4.TO-UL87 1-238-CSTREP
75	UL87_1-248_pCDNA_F	TACCGAGCTCGGATCCATGGCCGGCGCTGCGCCG	5	pCDNA4.TO-UL87 1-248-CSTREP
76	UL87_1-248_pCDNA_R	CACCGCCTCCCTCGAGTCCAAAGAGCAGCTTGTAGTGAACC	5	pCDNA4.TO-UL87 1-248-CSTREP
77	UL87_1-273_pCDNA_F	TACCGAGCTCGGATCCATGGCCGGCGCTGCGCCG	5	pCDNA4.TO-UL87 1-273-CSTREP
78	UL87_1-273_pCDNA_R	CACCGCCTCCCTCGAGAAGCTTTTGCAGCTCCAGC	5	pCDNA4.TO-UL87 1-273-CSTREP
79	BcRF1-ORF24_1_F	TACCGAGCTCGGATCCATGACACAAGGTAAGAGGGAGATGG	6	pCDNA4.TO-BcRF1-ORF24-CSTREP
80	BcRF1-ORF24_1_R	TTCAGGCTagccagcatctttgc	6	pCDNA4.TO-BcRF1-ORF24-CSTREP
81	BcRF1-ORF24_2_F	gctggctAGCCTGAAGCATC	6	pCDNA4.TO-BcRF1-ORF24-CSTREP
82	BcRF1-ORF24_2_R	CACCGCCTCCCTCGAGCAGCGGACGGACGCAACG	6	pCDNA4.TO-BcRF1-ORF24-CSTREP
83	mu24-ORF24_1_F	TACCGAGCTCGGATCCATGACAATATTCTTACCAGTATTC	6	pCDNA4.TO-mu24-ORF24-CSTREP
84	mu24-ORF24_1_R	TTCAGGCTctcagtaagcttcaagtaattttt	6	pCDNA4.TO-mu24-ORF24-CSTREP
85	mu24-ORF24_2_F	tactgagAGCCTGAAGCATCTCT	6	pCDNA4.TO-mu24-ORF24-CSTREP
86	mu24-ORF24_2_R	CACCGCCTCCCTCGAGCAGCGGACGGACGCAACG	6	pCDNA4.TO-mu24-ORF24-CSTREP
87	UL87-ORF24_1_F	TACCGAGCTCGGATCCATGGCCGGCGCTGCGCCG	6	pCDNA4.TO-UL87-ORF24-CSTREP
88	UL87-ORF24_1_R	TTCAGGCTtccaaagagcagct	6	pCDNA4.TO-UL87-ORF24-CSTREP
89	UL87-ORF24_2_F	AGCCTGAAGCATCTCTCGTTTTCAAT	6	pCDNA4.TO-UL87-ORF24-CSTREP
90	UL87-ORF24_2_R	CACCGCCTCCCTCGAGCAGCGGACGGACGCAACG	6	pCDNA4.TO-UL87-ORF24-CSTREP

732
733
734
735
736
737
738
739 **Table 2. Nucleotide sequence of synthetic gene blocks used in this study.**

Geneblock Name	Sequence 5'-3'	Figure	Construct
10xCTD_repeat_pQLink_insert	TTATTTTCAGGGATCCTATAGCCCGACGTCGCCGAGTTATTCACCTACGTCCCC ATCATACTCCCCCACGAGCCCTAGTTATTCGCCAACTTCCCCGAGCTATTCCCC AACATCACCCAGCTATAGTCCCACTTCACCCTCCTATTCACCTACGAGTCCATCT TATTCTCCAACCAGTCCTTCGTACTCACCCACGTCCCCATCGTATTCTCCTACTT CCCCTAGCTAGGCGGCCGCCTAGGACCC	3	pQLink-GST-10xCTD

740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769

770
771
772
773
774
775
776
777 **REFERENCES**

778 ALLEN, B. L. & TAATJES, D. J. 2015. The Mediator complex: a central integrator of
779 transcription. *Nat Rev Mol Cell Biol,* 16**,** 155-66.
780 AUBRY, V., MURE, F., MARIAME, B., DESCHAMPS, T., WYRWICZ, L. S., MANET, E. &
781 GRUFFAT, H. 2014. Epstein-Barr virus late gene transcription depends on the assembly
782 of a virus-specific preinitiation complex. *J Virol,* 88**,** 12825-38.
783 BAUER, D. L. V., TELLIER, M., MARTINEZ-ALONSO, M., NOJIMA, T., PROUDFOOT, N. J.,
784 MURPHY, S. & FODOR, E. 2018. Influenza Virus Mounts a Two-Pronged Attack on
785 Host RNA Polymerase II Transcription. *Cell Rep,* 23**,** 2119-2129 e3.
786 BOEHNING, M., DUGAST-DARZACQ, C., RANKOVIC, M., HANSEN, A. S., YU, T., MARIE-
787 NELLY, H., MCSWIGGEN, D. T., KOKIC, G., DAILEY, G. M., CRAMER, P., DARZACQ,
788 X. & ZWECKSTETTER, M. 2018. RNA polymerase II clustering through carboxy-
789 terminal domain phase separation. *Nat Struct Mol Biol,* 25**,** 833-840.
790 BURATOWSKI, S. 2009. Progression through the RNA polymerase II CTD cycle. *Mol Cell,* 36**,**
791 541-6.
792 CASTANEDA, A. F. & GLAUNSINGER, B. A. 2019. The Interaction between ORF18 and
793 ORF30 Is Required for Late Gene Expression in Kaposi's Sarcoma-Associated
794 Herpesvirus. *J Virol,* 93.
795 CHEN, H. T. & HAHN, S. 2003. Binding of TFIIB to RNA polymerase II: Mapping the binding site
796 for the TFIIB zinc ribbon domain within the preinitiation complex. *Mol Cell,* 12**,** 437-47.
797 CHI, T. & CAREY, M. 1996. Assembly of the isomerized TFIIA--TFIID--TATA ternary complex is
798 necessary and sufficient for gene activation. *Genes Dev,* 10**,** 2540-50.
799 DAVIS, Z. H., HESSER, C. R., PARK, J. & GLAUNSINGER, B. A. 2016. Interaction between
800 ORF24 and ORF34 in the Kaposi's Sarcoma-Associated Herpesvirus Late Gene
801 Transcription Factor Complex Is Essential for Viral Late Gene Expression. *J Virol,* 90**,**
802 599-604.
803 DAVIS, Z. H., VERSCHUEREN, E., JANG, G. M., KLEFFMAN, K., JOHNSON, J. R., PARK, J.,
804 VON DOLLEN, J., MAHER, M. C., JOHNSON, T., NEWTON, W., JAGER, S., SHALES,
805 M., HORNER, J., HERNANDEZ, R. D., KROGAN, N. J. & GLAUNSINGER, B. A. 2015.
806 Global mapping of herpesvirus-host protein complexes reveals a transcription strategy
807 for late genes. *Mol Cell,* 57**,** 349-60.
808 DECKER, K. B. & HINTON, D. M. 2013. Transcription regulation at the core: similarities among
809 bacterial, archaeal, and eukaryotic RNA polymerases. *Annu Rev Microbiol,* 67**,** 113-39.
810 DIDYCHUK, A. L., CASTANEDA, A. F., KUSHNIR, L. O., HUANG, C. J. & GLAUNSINGER, B.
811 A. 2019. Conserved CxnC motifs in Kaposi's sarcoma-associated herpesvirus ORF66
812 are required for viral late gene expression and are essential for its interaction with
813 ORF34. *J Virol*.
814 EICK, D. & GEYER, M. 2013. The RNA polymerase II carboxy-terminal domain (CTD) code.
815 *Chem Rev,* 113**,** 8456-90.
816 FEKLISTOV, A., SHARON, B. D., DARST, S. A. & GROSS, C. A. 2014. Bacterial sigma factors:
817 a historical, structural, and genomic perspective. *Annu Rev Microbiol,* 68**,** 357-76.

818 GRUFFAT, H., KADJOUF, F., MARIAME, B. & MANET, E. 2012. The Epstein-Barr virus BcRF1
819 gene product is a TBP-like protein with an essential role in late gene expression. *J Virol,*
820 86**,** 6023-32.
821 HAIGH, A., GREAVES, R. & O'HARE, P. 1990. Interference with the assembly of a virus-host
822 transcription complex by peptide competition. *Nature,* 344**,** 257-9.
823 HARLEN, K. M. & CHURCHMAN, L. S. 2017. The code and beyond: transcription regulation by
824 the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol,* 18**,** 263-273.
825 HARPER, T. M. & TAATJES, D. J. 2018. The complex structure and function of Mediator. *J Biol*
826 *Chem,* 293**,** 13778-13785.
827 HE, Y., FANG, J., TAATJES, D. J. & NOGALES, E. 2013. Structural visualization of key steps in
828 human transcription initiation. *Nature,* 495**,** 481-6.
829 HE, Y., YAN, C., FANG, J., INOUYE, C., TJIAN, R., IVANOV, I. & NOGALES, E. 2016. Near-
830 atomic resolution visualization of human transcription promoter opening. *Nature,* 533**,**
831 359-65.
832 JERONIMO, C., BATAILLE, A. R. & ROBERT, F. 2013. The writers, readers, and functions of
833 the RNA polymerase II C-terminal domain code. *Chem Rev,* 113**,** 8491-522.
834 JUVEN-GERSHON, T., CHENG, S. & KADONAGA, J. T. 2006. Rational design of a super core
835 promoter that enhances gene expression. *Nat Methods,* 3**,** 917-22.
836 KNUESEL, M. T., MEYER, K. D., BERNECKY, C. & TAATJES, D. J. 2009. The human CDK8
837 subcomplex is a molecular switch that controls Mediator coactivator function. *Genes*
838 *Dev,* 23**,** 439-51.
839 KOZLOWSKI, L. P. 2016. IPC - Isoelectric Point Calculator. *Biol Direct,* 11**,** 55.
840 LANDER, G. C., STAGG, S. M., VOSS, N. R., CHENG, A., FELLMANN, D., PULOKAS, J.,
841 YOSHIOKA, C., IRVING, C., MULDER, A., LAU, P. W., LYUMKIS, D., POTTER, C. S. &
842 CARRAGHER, B. 2009. Appion: an integrated, database-driven pipeline to facilitate EM
843 image processing. *J Struct Biol,* 166**,** 95-102.
844 LIU, X., BUSHNELL, D. A., WANG, D., CALERO, G. & KORNBERG, R. D. 2010. Structure of
845 an RNA polymerase II-TFIIB complex and the transcription initiation mechanism.
846 *Science,* 327**,** 206-9.
847 LU, H., FLORES, O., WEINMANN, R. & REINBERG, D. 1991. The nonphosphorylated form of
848 RNA polymerase II preferentially associates with the preinitiation complex. *Proc Natl*
849 *Acad Sci U S A,* 88**,** 10004-8.
850 LU, H., YU, D., HANSEN, A. S., GANGULY, S., LIU, R., HECKERT, A., DARZACQ, X. & ZHOU,
851 Q. 2018. Phase-separation mechanism for C-terminal hyperphosphorylation of RNA
852 polymerase II. *Nature,* 558**,** 318-323.
853 MALLICK, S. P., CARRAGHER, B., POTTER, C. S. & KRIEGMAN, D. J. 2005. ACE: automated
854 CTF estimation. *Ultramicroscopy,* 104**,** 8-29.
855 MINDELL, J. A. & GRIGORIEFF, N. 2003. Accurate determination of local defocus and
856 specimen tilt in electron microscopy. *J Struct Biol,* 142**,** 334-47.
857 NANDAKUMAR, D. & GLAUNSINGER, B. 2019. An integrative approach identifies direct
858 targets of the late viral transcription complex and an expanded promoter recognition
859 motif in Kaposi's sarcoma-associated herpesvirus. *PLoS Pathog,* 15**,** e1007774.
860 NOBRE, L. V., NIGHTINGALE, K., RAVENHILL, B. J., ANTROBUS, R., SODAY, L., NICHOLS,
861 J., DAVIES, J. A., SEIRAFIAN, S., WANG, E. C., DAVISON, A. J., WILKINSON, G. W.,
862 STANTON, R. J., HUTTLIN, E. L. & WEEKES, M. P. 2019. Human cytomegalovirus
863 interactome analysis identifies degradation hubs, domain associations and viral protein
864 functions. *Elife,* 8.
865 NOGALES, E., LOUDER, R. K. & HE, Y. 2017. Structural Insights into the Eukaryotic
866 Transcription Initiation Machinery. *Annu Rev Biophys,* 46**,** 59-83.

867 PAN, D., HAN, T., TANG, S., XU, W., BAO, Q., SUN, Y., XUAN, B. & QIAN, Z. 2018. Murine
868 Cytomegalovirus Protein pM91 Interacts with pM79 and Is Critical for Viral Late Gene
869 Expression. *J Virol,* 92.
870 PATEL, A. B., LOUDER, R. K., GREBER, B. J., GRUNBERG, S., LUO, J., FANG, J., LIU, Y.,
871 RANISH, J., HAHN, S. & NOGALES, E. 2018. Structure of human TFIID and mechanism
872 of TBP loading onto promoter DNA. *Science,* 362.
873 PLASCHKA, C., LARIVIERE, L., WENZECK, L., SEIZL, M., HEMANN, M., TEGUNOV, D.,
874 PETROTCHENKO, E. V., BORCHERS, C. H., BAUMEISTER, W., HERZOG, F., VILLA,
875 E. & CRAMER, P. 2015. Architecture of the RNA polymerase II-Mediator core initiation
876 complex. *Nature,* 518**,** 376-80.
877 REVYAKIN, A., ZHANG, Z., COLEMAN, R. A., LI, Y., INOUYE, C., LUCAS, J. K., PARK, S. R.,
878 CHU, S. & TJIAN, R. 2012. Transcription initiation by human RNA polymerase II
879 visualized at single-molecule resolution. *Genes Dev,* 26**,** 1691-702.
880 ROBINSON, P. J., BUSHNELL, D. A., TRNKA, M. J., BURLINGAME, A. L. & KORNBERG, R.
881 D. 2012. Structure of the mediator head module bound to the carboxy-terminal domain
882 of RNA polymerase II. *Proc Natl Acad Sci U S A,* 109**,** 17931-5.
883 ROBINSON, P. J., TRNKA, M. J., BUSHNELL, D. A., DAVIS, R. E., MATTEI, P. J.,
884 BURLINGAME, A. L. & KORNBERG, R. D. 2016. Structure of a Complete Mediator-RNA
885 Polymerase II Pre-Initiation Complex. *Cell,* 166**,** 1411-1422 e16.
886 SCHERES, S. H. 2012. RELION: implementation of a Bayesian approach to cryo-EM structure
887 determination. *J Struct Biol,* 180**,** 519-30.
888 SERIO, T. R., CAHILL, N., PROUT, M. E. & MILLER, G. 1998. A functionally distinct TATA box
889 required for late progression through the Epstein-Barr virus life cycle. *J Virol,* 72**,** 8338-
890 43.
891 SORZANO, C. O., MARABINI, R., VELAZQUEZ-MURIEL, J., BILBAO-CASTRO, J. R.,
892 SCHERES, S. H., CARAZO, J. M. & PASCUAL-MONTANO, A. 2004. XMIPP: a new
893 generation of an open-source image processing package for electron microscopy. *J*
894 *Struct Biol,* 148**,** 194-204.
895 SOUTOURINA, J. 2018. Transcription regulation by the Mediator complex. *Nat Rev Mol Cell*
896 *Biol,* 19**,** 262-274.
897 SPAHR, H., CALERO, G., BUSHNELL, D. A. & KORNBERG, R. D. 2009. Schizosacharomyces
898 pombe RNA polymerase II at 3.6-A resolution. *Proc Natl Acad Sci U S A,* 106**,** 9185-90.
899 TANG, S., YAMANEGI, K. & ZHENG, Z. M. 2004. Requirement of a 12-base-pair TATT-
900 containing sequence and viral lytic DNA replication in activation of the Kaposi's sarcoma-
901 associated herpesvirus K8.1 late promoter. *J Virol,* 78**,** 2609-14.
902 THOMAS, M. C. & CHIANG, C. M. 2006. The general transcription machinery and general
903 cofactors. *Crit Rev Biochem Mol Biol,* 41**,** 105-78.
904 TSAI, F. T. & SIGLER, P. B. 2000. Structural basis of preinitiation complex assembly on human
905 pol II promoters. *EMBO J,* 19**,** 25-36.
906 VAN HEEL, M., HARAUZ, G., ORLOVA, E. V., SCHMIDT, R. & SCHATZ, M. 1996. A new
907 generation of the IMAGIC image processing system. *J Struct Biol,* 116**,** 17-24.
908 VOSS, N. R., YOSHIOKA, C. K., RADERMACHER, M., POTTER, C. S. & CARRAGHER, B.
909 2009. DoG Picker and TiltPicker: software tools to facilitate particle selection in single
910 particle electron microscopy. *J Struct Biol,* 166**,** 205-13.
911 WONG-HO, E., WU, T. T., DAVIS, Z. H., ZHANG, B., HUANG, J., GONG, H., DENG, H., LIU,
912 F., GLAUNSINGER, B. & SUN, R. 2014. Unconventional sequence requirement for viral
913 late gene core promoters of murine gammaherpesvirus 68. *J Virol,* 88**,** 3411-22.
914 WYRWICZ, L. S. & RYCHLEWSKI, L. 2007. Identification of Herpes TATT-binding protein.
915 *Antiviral Res,* 75**,** 167-72.
916