Title
Global mapping of herpesvirus-host protein complexes reveals a novel transcription strategy for late genes

Author
Davis, Zoe Hartman

Publication Date
2015

Peer reviewed|Thesis/dissertation

Global mapping of herpesvirus-host protein complexes reveals a novel transcription strategy for late genes

By

Zoe Hartman Davis

A dissertation submitted in partial satisfaction of the

Requirements for the degree of

Doctor of Philosophy

in

Infectious Disease and Immunity

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Britt A. Glaunsinger, Chair
Professor Laurent Coscoy
Professor Qiang Zhou

Spring 2015

Abstract

Global mapping of herpesvirus-host protein complexes reveals a novel transcription strategy for late genes

By Zoe Hartman Davis

Doctor of Philosophy in Infectious Diseases and Immunity

University of California, Berkeley

Professor Britt A. Glaunsinger, Chair

Mapping host-pathogen interactions has proven instrumental for understanding how viruses manipulate host machinery and how numerous cellular processes are regulated. DNA viruses such as herpesviruses have relatively large coding capacity and thus can target an extensive network of cellular proteins. To identify the host proteins hijacked by this pathogen, we systematically affinity tagged and purified all 89 proteins of Kaposi's sarcoma-associated herpesvirus (KSHV) from human cells. Mass spectrometry of this material identified over 500 high-confidence virus-host interactions. KSHV causes AIDS-associated cancers and its interaction network is enriched for proteins linked to cancer and overlaps with proteins that are also targeted by HIV-1. This work revealed many new interactions between viral and host proteins. I have focused on one interaction in particular, that of a previously uncharacterized KSHV protein, ORF24, with cellular RNA polymerase II (RNAP II).

All DNA viruses encode a class of genes that are expressed only late in the infectious cycle, following replication of the viral genome. These 'late' genes are subject to tight regulation and must be robustly induced during a period of significant cell stress, yet in many cases have strikingly minimal promoters. For most viruses, the mechanism underlying the activation of these genes has remained largely unknown. I have shown that late gene expression in KSHV requires ORF24, which nucleates a non-canonical transcription complex. I have found that ORF24 is a modular protein that binds directly and specifically to viral late gene TATA-like boxes through its central domain, which was previously predicted to structurally mimic TATA-box-binding protein (TBP). Chromatin immunoprecipitation analyses show that ORF24 functionally replaces human TBP at these promoters. This is highly unusual, as cellular TBP canonically serves as a critical nucleation factor for the assembly of transcription pre-initiation complexes (PICs) at promoters. Through residues in its N-terminus, ORF24 then binds RNAP II to activate late gene expression. This strategy appears to be conserved in multiple related herpesviruses, and represents a novel, streamlined mechanism for transcriptional activation.

As the late gene PIC appears deviates from the canonical complex, I have worked to define the viral and cellular proteins assembled at these promoters. Foundational work in related herpesviruses has revealed a complex of six viral proteins (including

ORF24) to be essential for late gene expression. We have mapped the pairwise interactions between these viral proteins and found that the four other late gene viral transactivators (vTFs) interact with ORF24 through the viral protein ORF34. The interaction between ORF24 and ORF34 is essential for the expression of late genes. Aside from ORF24 and ORF34, the precise roles of the other vTFs in late gene expression remain unknown and are an ongoing area of research.

TABLE OF CONTENTS

FIGURES

TABLES

Chapter 1 - Introduction

Herpesviridae Family

The *Herpesviridae* family is comprised of large, enveloped, double-stranded DNA viruses, several of which infect and cause disease in humans. Between the icosahedral capsid and the envelope lies the proteinaceous tegument layer containing viral proteins important for infection of new cells. The *Herpesviridae* family is divided into the α-, β- and γ- subfamilies based on genome sequence, length of the replication cycle, and tissue tropism. Human α-herpesviruses are characterized by their short reproductive cycle and establishing latency in sensory ganglia. They include herpes simplex virus 1 and 2 (HSV-1 and HSV-2, HHV-1 and HHV-2, respectively), which cause oral and genital blisters and varicella zoster virus (VZV, HHV-3), which causes chickenpox and with subsequent reactivations, shingles (Wald and Corey, 2007). Human β-herpesviruses have a long replication cycle and maintain latency in lymphoid cells. Human cytomegalovirus (HCMV, HHV-5) is the best known of the human β-herpesviruses. It causes approximately 10% of cases of infectious mononucleosis and can cause severe birth defects when initial exposure and infection occurs during pregnancy (Bravender, 2010). HHV-6 and HHV-7 are both β-herpesviruses that are extremely common and cause rashes. They are typically contracted during childhood (Tanaka et al., 1994; Yamanishi et al., 1988). Both of the γ-herpesviruses known to infect humans, Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus (KSHV) can cause lymphoproliferative disease, including several cancers. Epstein-Barr virus (EBV, HHV-4) is the causative agent of approximately 90% of infectious mononucleosis and can cause Burkitt's lymphoma, nasopharyngeal carcinoma, and Hodgkin's lymphoma (Bravender, 2010). The last human herpesvirus is the most recently identified: Kaposi's sarcoma-associated herpesvirus (KSHV, HHV-8) the causative agent of Kaposi's sarcoma (KS), one of the main AIDS defining illnesses of the 1980's. KSHV remains a leading cause of cancer in patients with untreated AIDS and other immunosuppressed individuals (Ganem, 2010).

Kaposi's Sarcoma-Associated Herpesvirus (KSHV)

KS remains a leading cause of cancer in patients with untreated AIDS and other immunosuppressed individuals. The majority of the incidence and ill effects of this virus occur in resource poor settings where the coincidence of HIV creates a population of immunosuppressed individuals. KSHV seroprevalence is highest in sub-Saharan Africa, where estimates range from between 20% - 70% (Butler et al., 2009). However, KSHV related diseases are also experienced in more developed settings in individuals with untreated HIV as well in patients deliberately immunosuppressed for medical reasons (for example, from organ transplant or autoimmune disease) (Ganem, 2010).

The KSHV genome consists of a long unique region where open reading frames (ORFs) are encoded flanked by terminal repeat regions on either side. With in the

long unique region, this large virus encodes for 90 unique protein coding ORFs, the significant majority of which have no known function, or function is inferred by homology to other herpesviral proteins (Chang et al., 1994). Importantly, many of these ORFs are not necessarily part of the mechanics of entry, replication and exit – the basic viral lifecycle – but they serve to modulate host-virus interactions *in vivo*, for example, through evading immune detection by the host (Song et al., 2005).

The Viral Lifecycle

KSHV virions first attach to cells by binding to the ubiquitous cell surface molecule heparan sulfate. Both attachment and entry receptors can vary depending on cell type (Jarousse et al., 2008). For instance, DC-SIGN can be used for attachment to certain cell types (Rappocciolo et al., 2006). Next, in some cell types, macropinocytosis is triggered through a combination of the virus binding integrins and the EphA2 receptor (Boshoff, 2012; Dutta et al., 2013). In contrast, entry into human foreskin fibroblast cells occurs by clathrin-mediated endocytosis. Once entry has occurred and the viral capsid and tegument proteins are deposited into the cytoplasm, the capsid attaches to microtubules and is transported to the nucleus. However, once at the nuclear pore, the capsid does not enter into the nucleus. Rather, the viral linear DNA genome is "injected" into the nucleus, leaving an empty capsid on the cytoplasmic face of the nuclear pore. Once the genome enters the nucleus, it is rapidly circularized.

KSHV has the classical lytic and latent biphasic lifecycle characteristic of the *Herpesviridae* family. The latent transcriptional program is marked by expression of a minimal set of viral genes necessary to maintain the viral episome in dividing cells. During latency, the episome is tethered to the cellular genome by one of the hallmark viral latency proteins, latency-associated nuclear antigen (LANA). While the virus uses its own polymerase and accessory factors for genome replication during the lytic transcriptional program, during latency, the host cell DNA polymerase and replication machinery replicates the viral genome. Currently, no available drugs or vaccines are able to target latent viruses being covertly maintained in cells due in large part to the paucity of expressed antigens to identify a cell as being infected. Reactivation is typically limited to small number of cells at any one time.

In contrast to latency, during the lytic program, the majority of the 130 kilobase genome is expressed in regulated waves of gene expression, subdivided into early, delayed early, and late genes. Each wave of gene expression requires the previous expression stage for its own expression, with the replication and transcriptional activator (RTA, ORF50) protein being required for launching the initiation of the viral lytic cycle. Typically, early genes are involved in shaping the host response to viral infection, while delayed early genes include the highly conserved machinery for replicating the viral genome [e.g. polymerase (ORF9), processivity factor (ORF59), primase (ORF40, ORF41), helicase (ORF56, ORF44), and single-stranded DNA binding proteins (ORF6)].

The KSHV genome is replicated by a rolling circle replication mechanism. As previously discussed, in the virion, the viral genome is linear. Upon entry into the nucleus, the genome is circularized to facilitate replication. After circularizing, one

strand of the genome is nicked. From the 3' end of the nicked strand, continuous DNA synthesis occurs, displacing the 5' portion of the nicked fragment so it can then serve as a template for discontinuous DNA synthesis. DNA genomes are then made as concatemers (Lehman and Boehmer, 1999). Only once the viral genome is replicated does viral late gene synthesis occur.

Late genes are mostly structural components of the virions. After sufficient late gene products have been produced, viral assembly commences. Capsid assembly occurs in the nucleus. Capsid proteins assemble around a protein scaffold (ORF17.5). Once the shell is constructed, the scaffolding protein is digested away by a viral protease (ORF17), leaving an empty capsid that is ready for the DNA genome. Concatemeric viral DNA is then pushed into the capsid through the portal protein (ORF43) in the capsid shell. Once a "headful" of DNA is achieved and termination sequences are recognized by the terminase complex (ORF7, ORF29, ORF67), the genome is cleaved into unit length (Flint and American Society for Microbiology., 2009).

Next, the capsid must exit the nucleus, acquire the tegument and the envelope and exit the cell. First, some tegument proteins are acquired in the nucleus and move to the cytoplasm with the virion as it gets its first envelope by budding through the inner nuclear membrane. The envelope is subsequently lost as the particle fuses into the cytoplasm, still maintaining the nucleocapsid-associated tegument proteins from the nucleus. Next, the virus buds into the Golgi and acquires a new envelope along with membrane associated tegument proteins. The virion then buds out of the trans Golgi and is brought to the cell surface in a secretory vesicle. Lastly, the enveloped virion is released to the extracellular environment upon fusion of the vesicle with the cell membrane (Flint and American Society for Microbiology., 2009).

Late Gene Expression

Late genes are characterized by their dependence on DNA replication – their defining feature is the requirement for genome replication prior to their expression. As a defined expression class, late genes are remarkably conserved, seen in all double-stranded DNA viruses from bacteriophages to giant pandoraviruses, to human cell-infecting viruses. Late genes are typically structural components of the virion and are thus essential for the production of infectious virions. The relationship between DNA replication and late gene expression has made the regulation of late gene expression very challenging to study by traditional, biochemical transcriptional assays, as a robust, defined assay for KSHV in vitro replication does not exist. As a result, until very recently, almost all assays have had to be performed in the poorly defined, highly complex context of full viral infection in the background of a human cell. Thus, while much is known about early and delayed early gene expression, far more basic questions remain unanswered in the field of late gene expression.

Nevertheless, several compelling observations have been collected over the years in the herpesvirus literature. Late gene promoters are surprisingly minimal, consisting of as little as 12-15 base pairs (Tang et al., 2004). Canonical promoter sequences contain TATA box core elements. Late gene promoters, by contrast,

contain TATT boxes. In addition to their dependence on DNA replication, late genes have been found to require six viral proteins (ORF18, ORF24, ORF30, ORF31, ORF34, and ORF66) for their activation. These have been identified in KSHV as well as in related γ- and β-herpesviruses (Aubry et al., 2014; Gong et al., 2014; Isomura et al., 2011; Wong et al., 2007). Mutants of any one of the six viral late gene transactivators (vTFs) have normal early and delayed early gene expression as well as normal DNA replication. However, they are unable to produce late genes. The late gene production defect appears to occur at the level of transcription, as late gene mRNAs are not produced. While our data presented herein has done much to improve our understanding of the role of ORF24 in the vPIC, how the other vTFs contribute to late gene expression still remains unknown.

Molecular Piracy

Many of the genes of KSHV can be found to share sequence or structural homology with proteins (frequently involved in cell survival and proliferation) originating from other organisms (this is a feature shared with other herpesviruses and with many other DNA viruses, as well). Several well-studied examples include: vFLIP (ORF71) (Thome et al., 1997), vCyclin (ORF72) (Chang et al., 1996), vIL-6 (K2) (Moore et al., 1996), and vBCL-2 (ORF16) (Kugel and Goodrich, 2002). How do these genes end up in viral genomes? This is a challenging question to definitively answer as the timescale over which theses genes were acquired and have since evolved is immense. Nevertheless, two main mechanisms for host gene acquisition are generally considered.

First, recombination can occur between the viral and the host genome as an unintended result of viral replication. Second, recombination can occur between a cDNA of a cellular mRNA produced during a retrovirus co-infection (Ouyang et al., 2014). As an artifact of direct gene capture from the host, the intro-exon structure of the gene is maintained (at least initially). Over time, if direct host-viral genome recombination was the method of gene acquisition, shortening or loss of introns may occur.

Transcription initiation

The initiation of transcription in eukaryotic cells is a highly regulated process involving the coordination of several large complexes and gene specific proteins. RNA transcription is carried out by three polymerase complexes in eukaryotic cells: polymerase I (transcribes rRNA with the exception of 5s), polymerase II [transcribes mRNA, some small nuclear RNAs (snRNAs)], and polymerase III (transcribes tRNA, 5s, some snRNAs). Subsequent discussion will concern only polymerase II (RNAP II, Pol II) transcription of mRNA in eukaryotic cells.

First, the polymerase needs to find the core promoter, the binding site for RNA polymerase and the location where transcription begins. The core promoter typically consists of the TATA box sequence (the binding site for the TBP subunit of the general transcription factor TFIID), the TFIIB response element (BRE; located immediately upstream of the TATA-box), the initiator region (INR; a consensus site that defines the +1 transcription start site, where the polymerase actually begins

making the RNA), and the downstream promoter element (DPE). Importantly, most core promoters do not contain all four elements, demonstrating the tremendous flexibility of cells in service of faithful transcription initiation. On its own, the polymerase is unable to recognize and differentiate specific DNA sequences – for that, it requires transcription factors to bind the DNA of promoters in a sequence-specific manner and help to activate and recruit the polymerase to those specific genes. In addition to transcription factors, there is a critical set of general transcription factors (GTFs) needed for initiation. While it is not entirely clear to what degree the transcription pre-initiation complex (PIC) may assemble prior to binding DNA, I will describe the widely held model of step-wise complex assembly (Thomas and Chiang, 2006). First, TATA-box binding protein (TBP) binds the TATA-box, triggering a dramatic bending and opening of the promoter DNA. TBP along with the TBP-associated factors (TAFs) comprise the GTF TFIID. While TBP is almost universally used regardless of promoter, in humans, there are 14 TAFs, and their usage can vary widely depending on factors such as cell type, developmental stage, and which genes are being transcribed (Freiman et al., 2002; Herrera et al., 2014; Maston et al., 2012; Zhou et al., 2013). Next, sequential assembly of the GTFs occurs in the following order: TFIIA, TFIIB (to promote proper start site selection and stabilize promoter-bound TFIID), Pol II along with TFIIF, TFIIE and lastly, TFIIH (acting to aid in the melting of promoter DNA as well as in the escape of the complex from promoter DNA) (Thomas and Chiang, 2006). After ~10 nucleotides are synthesized, the polymerase is committed and begins elongation.

Chapter 2 - Global mapping of herpesvirus-host protein complexes reveals a novel transcription strategy for late genes

Introduction

Viruses reshape the intracellular environment during infection, both to co-opt processes necessary for viral amplification and to subvert antiviral defenses. Studies of virus-host interactions have thus provided a wealth of insight into host biology, including how the manipulation of specific pathways can contribute to disease. Due to genome size constraints, viral proteins are generally multifunctional and have evolved to target diverse cellular machinery. The number of interactions coordinated by individual viral proteomes is therefore anticipated to be substantial, as indicated by recent high throughput proteomics analyses of virus-host protein-protein interactions (PPIs) in mammalian cells (Pichlmair et al., 2012; Rozenblatt-Rosen et al., 2012). Systems-level analyses can also reveal infection-linked patterns within cells, as well as pathways or machinery that serve as hubs for viral perturbation (Hirsch, 2010; Navratil et al., 2011).

The first comprehensive analyses of protein complexes hijacked by viruses in mammalian cells were recently documented for the RNA viruses human immunodeficiency virus 1 (HIV-1) and hepatitis C virus (HCV) using affinity tag/purification mass spectrometry (AP/MS) (Germain et al., 2013; Jager et al., 2011a). Similar systematic mass spectrometry-based approaches have yet to be applied to DNA viruses, although a number of binary interaction screens using yeast-2-hybrid assays have been reported (Calderwood et al., 2007; Lee et al., 2011; Rajagopala et al., 2011; Uetz et al., 2006). DNA viruses can have significantly greater coding capacity relative to their RNA virus counterparts and generally exhibit genome amplification and gene expression strategies that more closely mimic those of the host.

Herpesviruses are among the largest mammalian DNA viruses identified to date, encoding 70 to over 230 proteins (Stern-Ginossar et al., 2012). Divided into three subfamilies (α, β, and γ), herpesvirus infections have diverse pathogenic outcomes that are frequently serious in immunocompromised individuals. For example, the majority of lethal AIDS-associated cancers are caused by human γ-herpesviruses, including Kaposi's sarcoma (KS). The etiologic agent of KS is a γ-herpesvirus termed Kaposi's sarcoma-associated herpesvirus (KSHV), which is also associated with the B cell lymphoproliferative disorders multicentric Castleman's disease and primary effusion lymphoma.

The KSHV life cycle is divided into lytic and latent transcriptional programs. Latency is the stage primarily linked to neoplastic disease, as the restricted subset of viral genes expressed during this phase generally manipulate growth regulatory pathways. All viral proteins are expressed during lytic replication, which is when

progeny virion production occurs. Both lytic and latent KSHV infection result in broad changes in cellular metabolism and gene expression. KSHV encodes an estimated 89 proteins, including immune modulators and signaling proteins that have been 'pirated' from the host, as well as proteins broadly conserved within the herpesvirus family involved in viral replication. However, the majority of KSHV-encoded proteins remain uncharacterized with a relatively small number of PPIs identified.

Here, we sought to gain a global perspective on how a large DNA virus interfaces with its host by assembling a PPI network for KSHV proteins in human cells. This network is the largest host-pathogen interactome constructed to date, as well as the first comprehensive PPI map for a DNA virus in mammalian cells. We use it to study virus-human hybrid transcription pre-initiation complex (PIC) with an essential role in directing viral late gene expression. This PIC incorporates functional mimicry of the human TATA-box-binding protein (TBP) with direct recruitment of cellular RNA polymerase II, suggesting a system that merges principles underlying both eukaryotic and prokaryotic transcriptional regulation.

Results

Assembly of the KSHV-human interactome.

To systematically construct a comprehensive interaction network map for KSHV, we cloned each of the 89 KSHV open reading frames (ORFs) from infected cells and fused them to a strep epitope affinity tag. Protein expression was confirmed by Western blotting with strep antibodies (data not shown). Where antibodies were available, we also compared the expression level of transfected viral ORFs to the endogenous version expressed during lytic KSHV infection (Figure 2.1A). The levels were generally comparable, although the level of some ORFs (e.g. K8.1) was significantly higher in the 293T cells, which could impact the number of interactions detected. To assemble the network, constructs were individually transfected in a minimum of three biological replicates into 293T cells and viral ORFs together with their associated protein complexes were affinity purified (AP) over Strep-Tactin beads and analyzed by mass spectrometry (MS) (Figure 2.2A). The resulting data set, assembled from over 500 AP-MS runs, contained a total of 50,835 interactions**.**

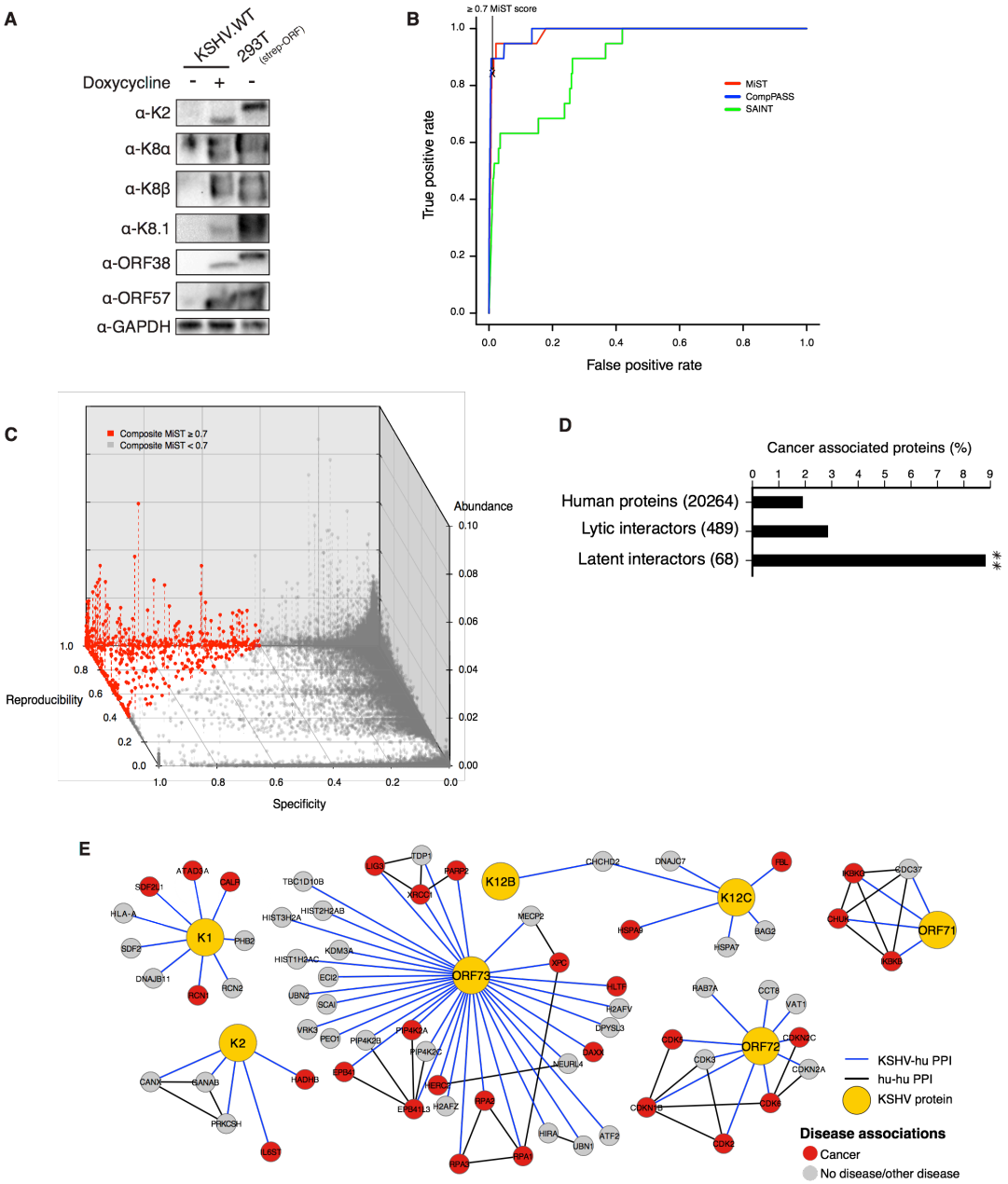

Figure 2.1: Protein expression levels and KSHV latency protein interactors linked to cancer.

(A) Plasmids expressing strep-tagged KSHV ORFs were transfected into 293T cells and their expression level compared to that in unreactivated (no doxycycline) and reactivated (with doxycycline) KSHV.WT iSLK cells using the indicated antibodies against each viral protein. Viral proteins in 293T cells are slightly larger than in infected cells due to the strep tag. **(B)** ROC curves comparing the optimal MiST feature weights to CompPASS (Sowa et al., 2009) and SAINT (Choi et al., 2011) scoring algorithms. The high-confidence data set (MiST ≥ 0.7) is indicated with an **x**. **(C)** Viral and cellular protein pairs were mapped in three-dimensional space

showing the reproducibility, abundance, and specificity features of the MiST score. Interaction pairs with MiST scores ≥ 0.7 are indicated in red. **(D)** Bar graph showing the percentage of all human proteins, versus the subset that interact with KSHV lytic or latent proteins, that are associated with cancer using the cBioPortal (TCGA) cancer mutation dataset (Cerami et al., 2012). ** Indicates a p-value of < 0.005 as determined by the hyper-geometric test. For further details on the data composing this figure, reference (Davis et al., 2015). **(E)** Network of KSHV latency protein interactions showing enrichment for cellular proteins linked to cancer. The KSHV latency protein-human interactome is shown with viral proteins as gold nodes and associated cellular proteins as red (associated with cancer) or grey (not associated with cancer) nodes. Disease terms from the DisGeNET (Bauer-Mehren et al., 2011) database were manually collapsed to create curated terms for the network. Interactions between cellular proteins present in the interaction network for each individual viral protein are displayed using black edges. Host-host edges were manually curated from CORUM (Ruepp et al., 2010) and the literature.

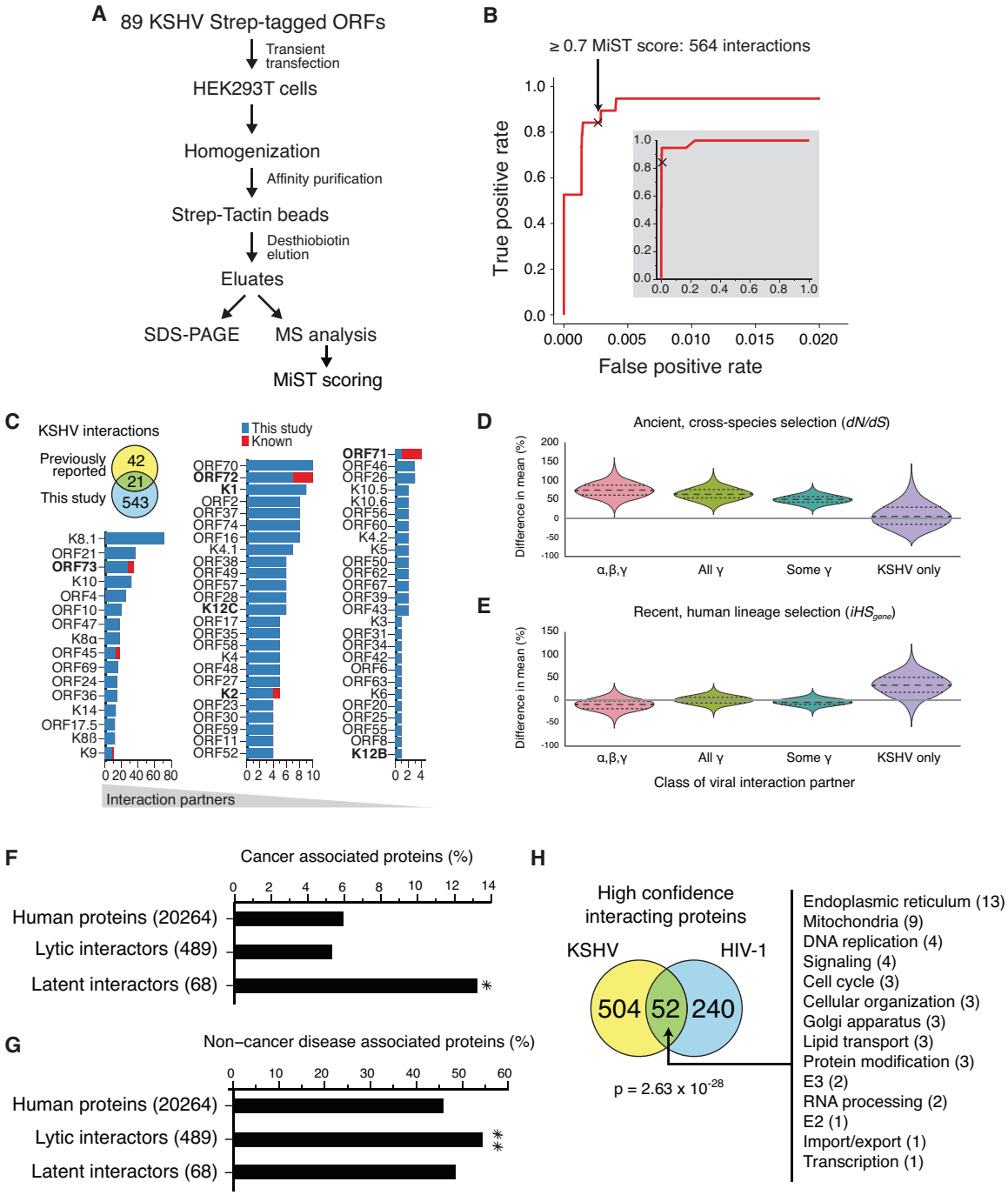

Figure 2.2: Assembly and broad characterization of the KSHV-human interactome.

(A) Summary of the workflow used for the assembly of the PPI network. **(B)** The MiST feature weights and score threshold were subjected to an exhaustive parameter grid search and the optimal values were empirically determined by maximizing the area under the curve defined by the True Positive rate (Y-axis) versus False Positive rate (X-axis). The depicted receiver operating characteristic (ROC) curve illustrates MiST prediction accuracy with optimal feature weights for specificity (0.5) and reproducibility (0.5) with the curve inflection point marked

with an **x**. The ROC curve was used to help define an appropriate MiST threshold of 0.7. This corresponded to 564 high confidence interactions. **(C)** Venn diagram showing the overlap between previously reported KSHV-host interactions and the high confidence PPI network. The bar graph shows the number of host proteins associating with each viral protein, ranked in descending order, with the previously reported interactions noted in red. **(D-E)** Host genes that interact with viral proteins unique to KSHV show significantly elevated signs of recent selection. Shown are the distributions of mean differences from background across 10,000 bootstrap samples. Signatures of selection are measured using a mammalian based dN/dS statistic for ancient selection (D), and the iHS statistic based on the 1000 Genomes Project for recent selection (E). **(F-G)** Bar graphs showing the percentage of all human proteins, versus the subset that interact with KSHV lytic or latent proteins, that are associated with cancer (F) or non-cancer diseases (G) in the DisGeNET database. * Indicates a p-value < 0.05 and ** indicates a p-value < 0.005 as determined by the hyper-geometric test. For further details on the data composing this figure, reference (Davis et al., 2015). **(H)** Venn diagram depicting the overlap between the high confidence HEK293T interaction partners of KSHV and HIV-1. The interaction partners are broken out into the number of cellular proteins in each category. The p-value was determined by the hyper-geometric test.

A set of high confidence PPIs was established by applying and analyzing all interactions using the mass spectrometry interaction statistics (MiST) scoring algorithm, which we designed for scoring AP-MS derived host-pathogen PPIs (Jager et al., 2011a; Verschueren et al., 2014). MiST scores are a weighted combination of the specificity, reproducibility, and abundance features of each detected interaction. We trained the feature weights and validated our prediction accuracy on a set of published interactions to establish a MiST score of ≥ 0.7 for our high confidence interaction set (Figure 2.2B, Figure 2.1B-C). A total of 21 of the 28 published interactions detected (75%, 14 trained and 7 in the validation group, $p = 7.32 \times 10^{-19}$) had MiST scores ≥ 0.7, demonstrating the ability of MiST to enrich for biologically relevant interactions. When applied to all detected PPIs, MiST scoring prioritized 564 KSHV-human interactions, which we refer to as the 'high confidence' data set. Of the 89 KSHV proteins tested, 67 displayed high confidence interactions with at least one human protein (Figure 2.1C). In all likelihood, our stringent cutoff excludes at least some biologically relevant interactions, however, the significant majority of previously known interactions in the data set were maintained and further revealed a multitude of novel interactions with interconnected human complexes. Indeed, comparison of this set with published KSHV-human interactions curated from the literature revealed 44% overlap (28 out of 63 published PPIs) (Table 2.2). We also independently validated 16 out of an attempted 28 newly described, high confidence interactions through co-IP followed by Western blotting (Table 2.1).

Table 2.1 Validated interactions from AP-MS.
The attempted validations by co-IP done for high confidence interaction partners identified by the interactome. Positive interactions are indicated by +; the absence

of a detected interaction by co-IP is indicated by −. Interaction partners were epitope-tagged expression plasmids unless otherwise indicated.

ORF	Interaction Partner	Source	Interaction
K4	SART3	(Song et al., 2010)	+
K8.1	ERLEC1	ORFeome V8.1	−
ORF2	PPP2R2D	ORFeome V8.1	+
ORF10	AKT1S1	(Oshiro et al., 2007)	−
ORF10	NUP62	ORFeome V8.1	+
ORF10	RAE1	ORFeome V8.1	+
ORF17	KPNA2	(Kumar et al., 2011)	+
ORF17.5	GMDS	ORFeome V8.1	−
ORF21	SNX9	(Lin et al., 2002)	+
ORF23	PPP1CA	(Lee et al., 2010)	+
ORF23	PPP1CB	(Lee et al., 2010)	+
ORF24	POLR2A	Santa Cruz Biotechnology (antibody)	+
ORF24	RPAP3	ORFeome V8.1	+
ORF24	URI1	(Mita et al., 2011)	−
ORF31	TPP1	ORFeome V8.1	+
ORF35	CPVL	ORFeome V8.1	−
ORF35	MLLT3	Bethyl (antibody)	−
ORF38	BZW1	ORFeome V8.1	−
ORF38	DCAF10	ORFeome V8.1	−
ORF38	PPT1	ORFeome V8.1	+
ORF39	CRK	ORFeome V8.1	+
ORF45	DNMT1	(Peng et al., 2011)	+
ORF45	MORF4L2	ORFeome V8.1	+
ORF45	USP7	(Song et al., 2008)	+
ORF48	PPP6R1	(Stefansson and Brautigan, 2006)	−
ORF48	PPP6R2	(Stefansson and Brautigan, 2006)	−
ORF49	CNOT6	(Wagner et al., 2007)	−
ORF49	CNOT7	(Wagner et al., 2007)	−

While some of the interactions that we were unable to verify by co-IP/Western blot validation may represent false positives originating from the AP/MS, it is possible that others are non-binary interactions that require a more limiting endogenous bridging factor.

Table 2.2: Previously described interactions for KSHV viral proteins.
Prey proteins found in the high (H) confidence data set correspond to a MiST score of ≥0.7. Proteins found below the cutoff are labeled low (L) while proteins absent from the MS, regardless of score are labeled not detected (ND).

ORF	Interaction Partner	Source	**N**ot **D**etected/**H**igh/**L**ow
K2	IL6ST	(Li et al., 2001)	H
K2	VKORC1	(Chen et al., 2012)	ND
K9	EP300	(Burysek et al., 1999)	H
K9	CREBBP	(Lin et al., 2001)	H
K9	NDUFA13	(Seo et al., 2002)	ND
K9	IRF3	(Lin et al., 2001)	ND
K9	IRF7	(Lin et al., 2001)	ND
K10	RBPJ	(Heinzelmann et al., 2010)	ND
K10	PABPC	(Kanno et al., 2006)	ND
K12A	CYTH1	(Kliche et al., 2001)	ND
K12B	MAPKAPK2	(McCormick and Ganem, 2005)	ND
ORF45	SIAH1	(Abada et al., 2008)	H
ORF45	SIAH2	(Abada et al., 2008)	H
ORF45	RSK1	(Kuang et al., 2008)	H
ORF45	RSK2	(Kuang et al., 2008)	H
ORF45	MAPK1	(Kuang et al., 2009)	H
ORF45	KIF3A	(Sathish et al., 2009)	ND
ORF45	IRF7	(Zhu et al., 2002)	ND
ORF50	PARP1	(Gwack et al., 2003)	L
ORF50	POU2F1	(Carroll et al., 2007)	ND
ORF50	RBPJ	(Liang et al., 2002)	ND
ORF57	ALYREF	(Malik et al., 2004)	L
ORF57	WIBG	(Boyne et al., 2010)	L
ORF57	NUP62	(Malik et al., 2012)	ND
ORF57	FYTTD1	(Jackson et al., 2011)	ND
ORF71	IKBKG	(Liu et al., 2002)	H
ORF71	IKBKB	(Liu et al., 2002)	H
ORF71	CHUK	(Liu et al., 2002)	H
ORF71	TRAF2	(Guasparri et al., 2006)	ND
ORF72	CDK2	(Platt et al., 2000)	H
ORF72	CDK5	(Platt et al., 2000)	H
ORF72	CDK6	(Platt et al., 2000)	H
ORF72	CDK4	(Platt et al., 2000)	ND
ORF73	BRD2	(Platt et al., 1999)	L
ORF73	KDM3A	(Kim et al., 2013)	H
ORF73	H2A	(Barbera et al., 2006)	H
ORF73	H2B	(Barbera et al., 2006)	L

ORF73	RPA1	(Shamay et al., 2012)	H
ORF73	RPA2	(Shamay et al., 2012)	H
ORF73	DAXX	(Murakami et al., 2006)	H
ORF73	MECP2	(Krithivas et al., 2002)	H
ORF73	TP53	(Friborg et al., 1999)	L
ORF73	TRF1	(Shamay et al., 2012)	L
ORF73	XPC	(Shamay et al., 2012)	H
ORF73	ANG	(Paudel et al., 2012)	ND
ORF73	ANXA2	(Paudel et al., 2012)	ND
ORF73	BRD4	(Ottinger et al., 2006)	ND
ORF73	DEK	(Krithivas et al., 2002)	ND
ORF73	CBX5	(Lim et al., 2003)	ND
ORF73	VHL	(Cai et al., 2006)	ND
ORF73	RB1	(Radkov et al., 2000)	ND
ORF73	CENPF	(Shamay et al., 2012)	ND
ORF73	YIPF3	(Shamay et al., 2012)	ND
ORF73	HMGA1	(Shamay et al., 2012)	ND
ORF73	HMGA2	(Shamay et al., 2012)	ND
ORF73	PYGO2	(Shamay et al., 2012)	ND
ORF73	PPP2R2A	(Shamay et al., 2012)	ND
ORF73	TIP60	(Shamay et al., 2012)	ND
ORF73	CUL5	(Cai et al., 2006)	ND
ORF73	RBX1	(Cai et al., 2006)	ND
ORF73	TCEB1	(Cai et al., 2006)	ND
ORF73	CBP	(Lim et al., 2001)	ND

We next looked for patterns in the KSHV interaction partners by grouping the viral proteins according to their conservation within the α-, β-, and γ-herpesviral subfamilies and analyzing the evolutionary selection profile of the interacting cellular proteins. We made the assumption that the degree of conservation of viral proteins between subfamilies reflects in part the evolutionary age of the protein, with proteins conserved amongst the herpesviruses likely being "older", while those unique to γ-herpesviruses or KSHV likely being "younger". We found that interactors of the more conserved viral proteins showed strong signals of ancient selection as measured by dN/dS ratios across mammals (Figure 2.2D). In contrast, interactors of KSHV-specific viral proteins had strong signals of recent human-specific selection as measured by iHS (Voight et al., 2006) using data from the 1000 Genomes Project (Genomes Project et al., 2010), exhibiting an average increase in positive selection of 33% above background across bootstrap samples. (Figure 2.2E). Thus, there is a dichotomy of natural selection acting broadly across mammals for interactors of ancient herpesvirus proteins compared to the human-specific signatures of selection for interactors of viral proteins exclusive to KSHV. This is suggestive of a

longstanding evolutionary interplay between herpesviruses and their hosts that continues to shape mammalian genomes.

A biologically relevant interaction network should be enriched for proteins linked to KSHV-induced diseases. The viral latency factors are the primary drivers of KSHV-induced cancers, although lytic cycle proteins provide paracrine signals to enhance the tumor microenvironment (Mesri et al., 2014). We therefore looked for overlap between the PPI networks of the lytic or latent proteins and disease associated cellular proteins annotated in the DisGeNET (Bauer-Mehren et al., 2011) or cBioPortal (TCGA) databases. Indeed, the interaction partners of latent proteins were significantly enriched for host factors associated with cancer (Figure 2.2D and Figure 2.1D-E), whereas partners of lytic proteins were instead enriched in non-cancer disease associations (Figure 2.2E and Figure 2.3). This correlation suggests that a number of KSHV proteins target host factors whose disruption may contribute to disease.

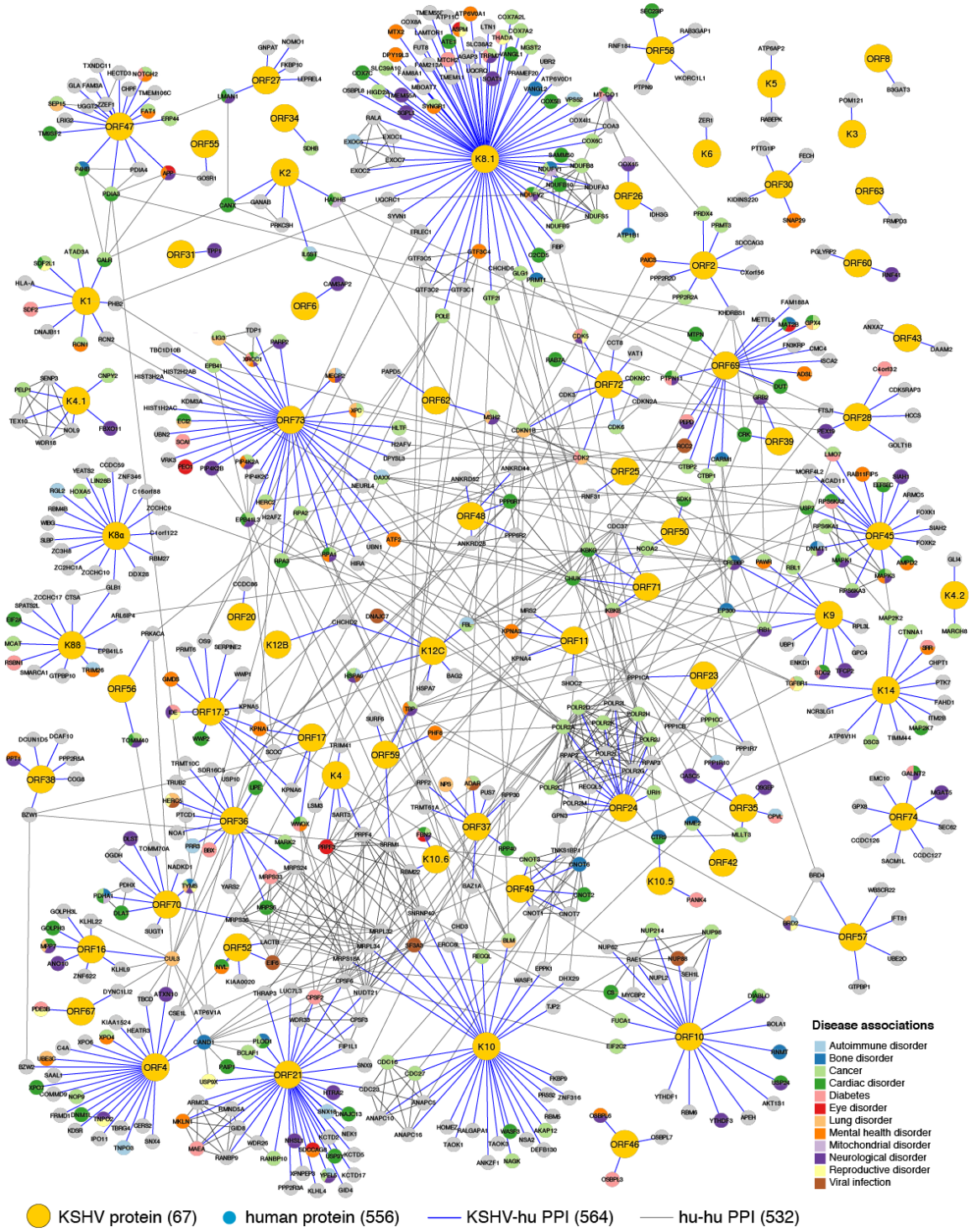

Figure 2.3: Full high-confidence KSHV interactome network overlaid with all host-host connections and cellular disease-linked proteins.

The KSHV-human interactome is shown with viral proteins as gold nodes and associated cellular proteins as pie chart nodes, with the coloring indicating the association of that cellular protein with a particular disease (specified by the colors in the legend). Host proteins without disease association are shown in grey. Disease terms from the DisGeNET (Bauer-Mehren et al., 2011) database were manually

collapsed to create curated terms for the network. All known interactions between cellular proteins present in the interactome are displayed using black and grey edges. Host-host edges were manually curated from CORUM (Ruepp et al., 2010) and the literature. For further details on the data composing this figure, reference (Davis et al., 2015).

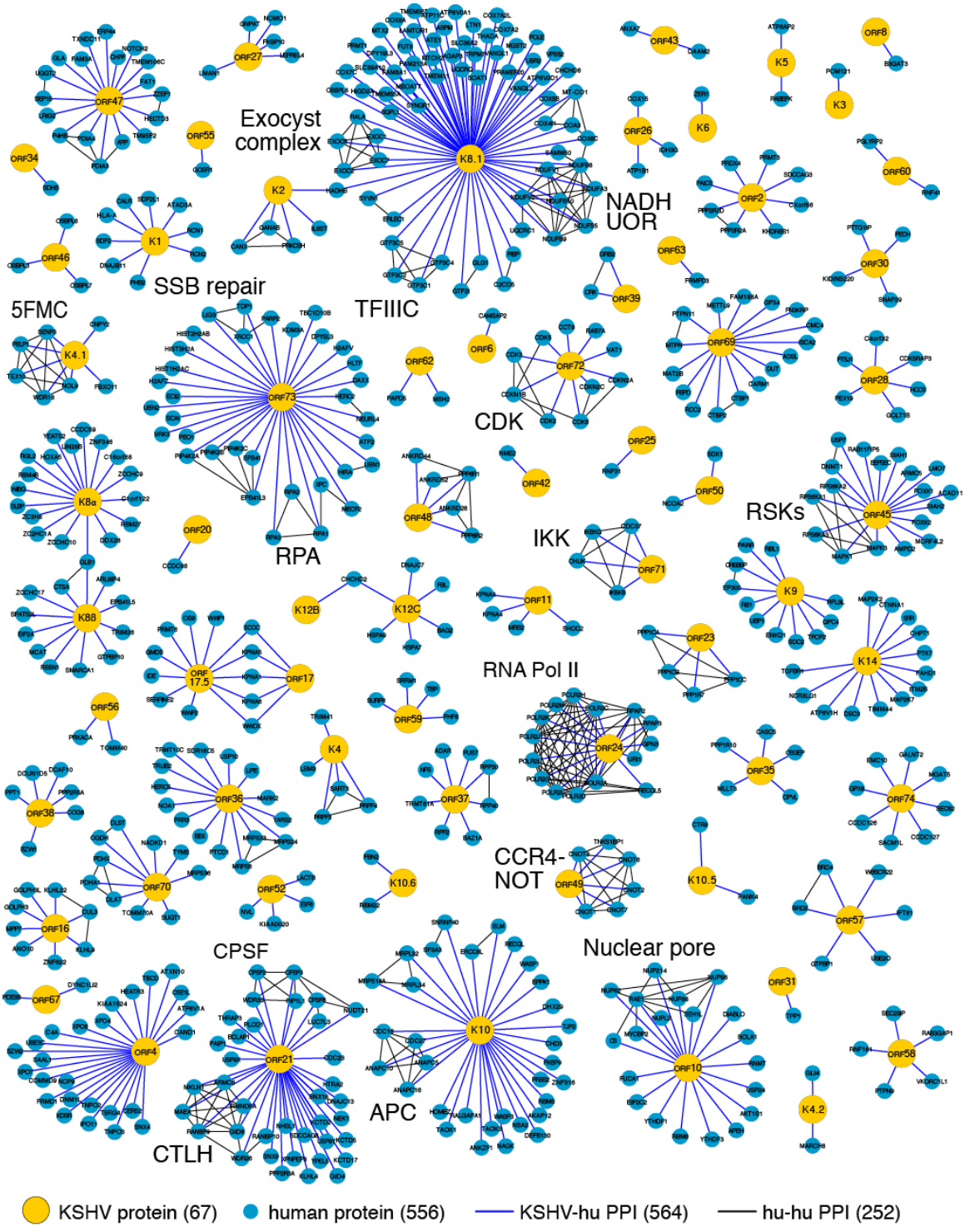

Figure 2.4: Network representation of the KSHV-host interactome.
The full, high confidence interaction network contains 67 KSHV proteins (gold nodes) and 556 cellular proteins (blue nodes). Black edges indicate known interactions between cellular proteins associated with each individual viral protein. A subset of the identified complexes is labeled. NADH UOR: ubiquinone oxidoreductase; TFIIIC: transcription factor for polymerase III C; SSB repair: single-

stranded break repair; 5FMC: 5 friends of methylated Chtop; CDK: cyclin dependent kinase; RSKs: ribosomal s6 kinases; IKK: IκB kinase complex; RPA: replication protein A complex; RNA Pol II: RNA polymerase II; CPSF: cleavage and polyadenylation specificity factor complex; APC: anaphase-promoting complex; CTLH: C-terminal to LisH motif complex.

The KSHV and HIV-1 proteins have significant overlap in their interaction partners.

Epidemic KS is closely associated with HIV/AIDS patients, and we recently assembled an HIV-1-human interactome using the same pipeline described herein (Jager et al., 2011a). Notably there was a significant degree of overlap in high confidence interactors identified in both sets, with nearly all HIV-1 proteins and 22 KSHV proteins interacting with a common group of 52 cellular proteins (Figure 2.2F and Table 2.3). The degree of overlap is particularly striking when the dramatic difference in lifecycle and replication strategies of the viruses are considered. Almost every HIV protein and polyprotein examined is represented in this overlap while almost 25% of the KSHV proteins examined in this study are included. Of the KSHV proteins involved, glycoproteins are overrepresented in the HIV-overlap set, likely due to the glycoproteins of HIV and KSHV interacting with a common set of ER and Golgi resident cellular proteins. These overlapping proteins may be participants in cellular processes broadly required for viral amplification, or proteins selectively contributing to the KSHV-HIV-1 interaction. Also included in the overlapping set are many mitochondrial proteins from the KSHV glycoprotein K8.1 and the HIV accessory proteins Vpr and Vpu. Cellular proteins involved in diverse processes including DNA replication and cell cycle are also covered in the overlap group, demonstrating the significant diversity in this set (Table 2.3).

Table 2.3: KSHV-HIV-1 high confidence overlapping proteins.
Cellular proteins interacting with at least one HIV-1 and KSHV protein are divided by functional category. The viral proteins interacting with the cellular protein are written after in parentheses with the KSHV protein(s) first in black and the HIV-1 protein(s) second in orange. For further details on the data composing this table, reference (Davis et al., 2015).

Term	Cellular protein (KSHV protein, HIV protein)
Endoplasmic reticulum (13)	GANAB (K2, Gp120 and Gp160); ERLEC1 (K8.1, Gp41 and Gp160); RCN1 (K1, Pol); DNAJC7 (K12C, Vif); CALR (K1, Gp160); CANX (K2, Gp160); PDIA4 (ORF47, IN and Pol); PRKCSH (K2, Gp120 and Gp160); PLOD1 (ORF21, IN); PDIA3 (ORF47, Gp41); DNAJB11 (K1, Gp120 and Gp160); P4HB (ORF47, Gp160); SDF2L1 (K1, Gp120 and Gp160)
Mitochondria (9)	ATAD3A (K1, Vpu); TOMM40 (ORF56, Gp41); TMEM55B (K8.1, Vpu); TMEM11 (K8.1, Vpr); MTCH2 (K8.1, Vpr); MRS2 (ORF11, Rev); COX4IL (K8.1, Gp41 and Gp160); HADHB (K2 and K8.1, Vpu); NDUFB10 (K8.1, Gp41)
DNA replication (4)	POLε (K8.1, Pr); RECQL (K10, Pol); PRMT1 (K8.1, Gag and NC); MSH2 (ORF62, Pr)
Signaling (4)	FBN2 (K10.6, Tat); GRB2 (ORF39, RT); ATXN10 (ORF4, Vpu); PRDX4 (ORF2, Gp41)
Cell cycle (3)	CDKN2A (ORF72, Nef); APP (ORF47, Gp160); TBRG4 (ORF4, Vpu)
Cellular organization (3)	DSC3 (K14, Gp160); EPPK1 (K10, Pr); SDCCAG8 (ORF21, MA)
Golgi apparatus (3)	FUT8 (K8.1, Gp160); GOLPH3 (ORF16, Vpr); LMAN1 (ORF27, Gp41 and Gp160)
Lipid transport (3)	OSBPL6 (ORF46, Pr); OSBPL3 (ORF46, Pol); SGPL1 (K8.1, Vpu)
Protein modification (3)	DPY19L3 (K8.1, Vpr); ITM2B (K14, Gp160); SDF2 (K1, Gp120 and Gp160)
E3 (2)	HERC2 (ORF73, Pr); UBR2 (K8.1, Vif)
RNA processing (2)	YTHDF3 (ORF10, NC); THRAP3 (ORF21, Gag)
E2 (1)	UBE2O (ORF57, Tat)
Import/export (1)	XPO4 (ORF4, Vpu)
Transcription (1)	PHB2 (K1, Gp41)


A
Golgi apparatus
Nuclear import/export
Transcription
Ribosome biogenesis
RNA processing
Cell division
Ubiquitination
Phosphorylation regulation
RNA metabolism
Innate immunity
Endoplasmic reticulum
Mitochondrial translation
ATP biosynthesis
Lipid transport
Cell cycle
DNA repair
MAPK cascade
< 5E−4
5E−3
5E−2
0
q-value
K8.1 K1 K2 ORF27 ORF47 ORF55 ORF74 ORF36 K4.1 ORF37 ORF52 K9 ORF24 ORF72 ORF4 ORF11 ORF17 ORF10 ORF17.5 ORF2 ORF23 K8.6 ORF26 ORF70 ORF46 ORF49 ORF71 K14 ORF57 ORF45 ORF73 K10 ORF21
B
ARM repeat
IBB domain
BTB (POZ) domain
BACK (BTB/Kelch associated) domain
Kelch repeat
PDZ (DHR) domain
BAR domain
CTLH domain
LisH domain
J domain
USP domain
MIF4G domain
Fe2OG dioxygenase domain
PX (phox homology) domain
Ig−like C1−type (immunoglobulin−like) domain
MIR domain
ANK repeat
EF−hand domain
TPR repeat
WD repeat
C2H2−type zinc finger
WH2 domain
PHD−type zinc finger
Matrin−type zinc finger
G−patch domain
Chromo domain
RanBP2−type zinc finger
Bromo domain
ZZ−type zinc finger
KIX domain
TAZ−type zinc finger
HECT (E6AP−type E3 ubiquitin−protein ligase) domain
C2 domain
WW domain
PH domain
RING−type zinc finger
LRR (leucine−rich) repeat
CHCH domain
PRKCSH domain
SAM−dependent MTase PRMT−type domain
Solcar repeat
UBR−type zinc finger
SH2 domain
SH3 domain
NET domain
Lipoyl−binding domain
GED domain
HEAT repeat
Importin N−terminal domain
AGC−kinase C−terminal domain
Fork−head DNA−binding domain
SIAH−type zinc finger
FHA domain
MRG domain
C3H1−type zinc finger
CCHC−type zinc finger
DOC domain
Thioredoxin domain
Ig−like V−type (immunoglobulin−like) domain
Cadherin domain
Ig−like C2−type (immunoglobulin−like) domain
FERM domain
BRCT domain
PIPK domain
JmjC domain
AT hook DNA−binding domain
PARP alpha−helical domain
PPR (pentatricopeptide) repeat
HMG box DNA−binding domain
RCC1 repeat
YTH domain
Brix domain
DRBM (double−stranded RNA−binding) domain
SAM (sterile alpha motif) domain
Tyrosine−protein phosphatase domain
L−type lectin−like domain
PPIase FKBP−type domain
B30.2/SPRY domain
S1 motif domain
CH (calponin−homology) domain
Fibronectin type−III domain
PWI domain
Ricin B−type lectin domain
CSD (cold−shock) domain
Homeobox DNA−binding domain
Protein kinase domain
RRM (RNA recognition motif) domain
Helicase ATP−binding domain
Helicase C−terminal domain
K10 ORF21 ORF17.5 ORF11 ORF17 ORF45 K8.6 K9 ORF4 ORF37 ORF57 ORF16 ORF48 ORF72 ORF36 K10.6 K4 ORF70 ORF39 ORF46 ORF27 K6 ORF69 ORF50 K12C ORF38 ORF59 K12B ORF6 ORF59 ORF74 ORF47 K1 K14 K8.1 ORF73

Figure 2.5: Gene ontology and domain enrichment for the KSHV interactome.
(A) Heat map depicting the enrichment of biological processes of cellular proteins interacting with KSHV baits via a GO term analysis. The color scale corresponds to the significance of the bait-term association q-values (p-values adjusted for FDR). **(B)** Heat map representing the enrichment q-values for the domains found in the interacting proteins of KSHV. For further details on the data composing this figure, reference (Davis et al., 2015).

The KSHV interactome as a tool for functional prediction.

The high confidence KSHV-human PPI network comprised 564 interactions between 67 viral proteins and 556 human proteins (Figure 2.4). When annotated to include known human-human interactions either within the network of individual viral proteins (Figure 2.4) or the network as a whole (Figure 2.3), enrichment for components of many distinct host complexes emerged. Many are new KSHV-host interactions, including the C-terminal to LisH motif (CTLH) and cleavage and polyadenylation specificity factor (CPSF) complexes with ORF21, RNA polymerase II with ORF24, the ribosome biogenesis 5 friends of methylated CHTOP (5FMC) complex with K4.1 (vCCL3), and the anaphase promoting complex (APC) with K10 (vIRF-4). We also detected viral interactions with several complexes that had been previously described, including ORF45 with ribosomal S6 kinases (RSKs) and MAP kinases (Kuang et al., 2008; Kuang et al., 2009), K9 (vIRF-1) with CREBBP and p300 (Burysek et al., 1999; Lin et al., 2001), ORF72 (vCyclin) with several CDKs (Platt et al., 2000), ORF71 (vFLIP) with the NFκB regulating IκB kinase (IKK) complex (Liu et al., 2002), and ORF73 (LANA) with the replication protein A (RPA) complex (Shamay et al., 2012). Collectively, these data reveal the breadth of cellular machinery physically targeted by KSHV proteins, which presumably disrupt or hijack these complexes for viral benefit.

A primary goal was to use these data to predict new protein functions, as the majority of KSHV ORFs remain uncharacterized. We therefore performed enrichment tests for gene ontology (GO) terms associated with the set of PPIs for each of the KSHV proteins (Figure 2.5). We also analyzed the individual viral PPI networks for enrichment of specific protein domains to identify functions not well annotated in the GO database (Figure 2.5). Assimilating the data from these two analyses enabled us to link many of the uncharacterized KSHV proteins to specific pathways or processes. For example, ORF46 (unstudied in KSHV, classified as a putative uracil DNA glycosylase by homology) interactors are enriched for the lipid transport term, suggesting a potential role in metabolism or signaling. Interactors for ORF70 (minimally studied in KSHV, known to encode a functional thymidylate synthase) are enriched for the ATP biogenesis term, due in part to the oxoglutarate dehydrogenase complex (OGDH and DLST) and pyruvate dehydrogenase complex (PDHA1, PDHX, DLAT).

New functional groupings also surfaced for previously characterized KSHV proteins. Some prominent examples include a putative role for K8.1 in ATP biosynthesis through the NADH ubiquinone oxidoreductase complex and the cytochrome oxidase complex, a role for the viral thymidine kinase (ORF21) in RNA

processing, and a role for the viral protease and scaffolding proteins (ORF17 and ORF17.5, respectively) in nuclear import/export. Importantly, our enrichment analyses accurately assigned functions to many of the characterized KSHV proteins, including ORF71 (vFLIP) to innate immunity (Liu et al., 2002), ORF45 to the MAPK cascade and innate immunity (Kuang et al., 2009; Zhu et al., 2002), and ORF72 (vCyclin) to the cell cycle (Verschuren et al., 2004). The enrichment analyses accurately assigned functions to many of the characterized KSHV proteins, suggested new additional functional groupings for several of these proteins, as well as enabled us to link many of the uncharacterized KSHV proteins to specific pathways or processes (Figure 2.6).

We identified many new pathways touched on KSHV ORFs, allowing for new inferences about ORF function. ORF23, for instance, is dispensable for replication of a related γ-herpesvirus and remains uncharacterized in KSHV (Ohno et al., 2012). However, in our high-confidence network, it co-purifies with a complex consisting of 4 phosphatases (PPP1CA, PPP1CB, PPP1CC, and PPP1R7), suggesting a role in phosphorylation control. Another example is the viral chemokine agonist K4.1 vCLL3. It interacts with the 5FMC complex, including the SUMO-specific proteases thought to be involved in the regulation of rDNA transcription, indicting it may impact ribosome biogenesis (Fanis et al., 2012; Finkbeiner et al., 2011).

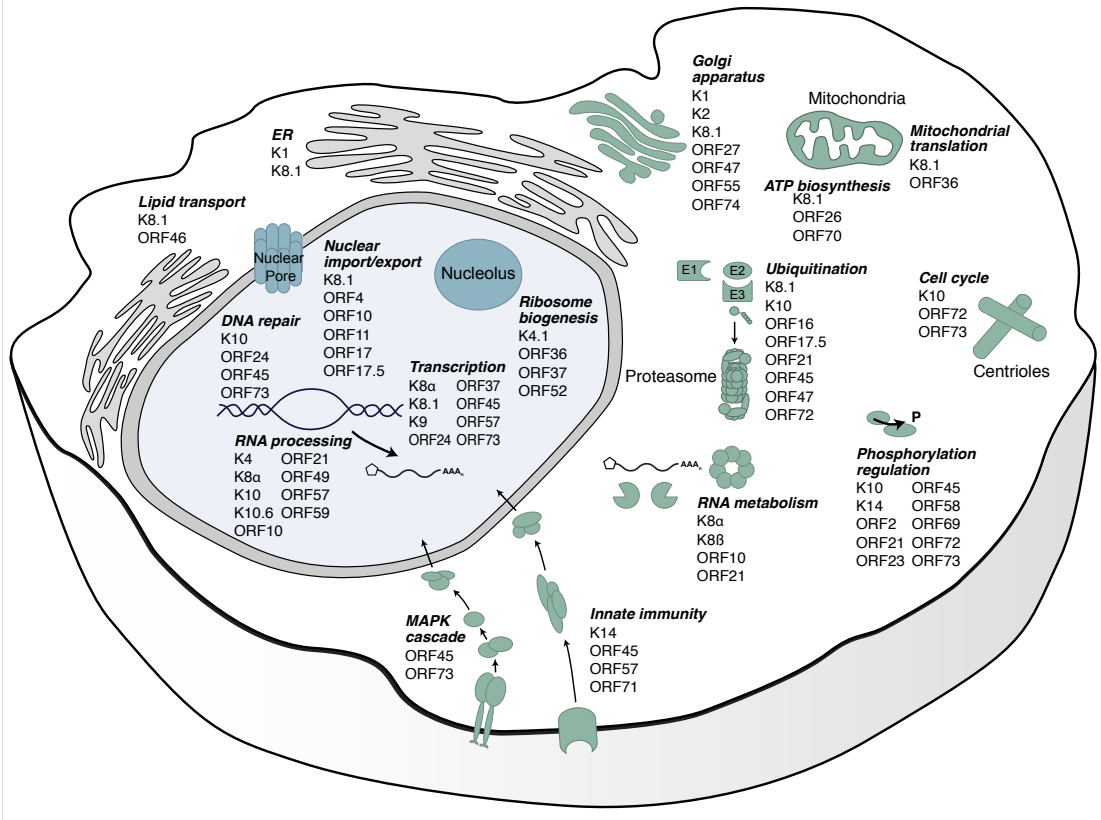

Figure 2.6: Using the KSHV interactome to predict viral protein functions.
Summary of predicted functions of KSHV proteins, derived from GO term and protein domain enrichments within the set of high confidence PPIs associated with

each viral protein. For further details on the data composing this figure, reference (Davis et al., 2015).

ORF24 interaction with RNA Pol II is essential for KSHV late gene expression.

One of the most striking functional predictions was the link between the viral early lytic cycle protein ORF24 and transcription. Though unstudied in KSHV, the ORF24 orthologs in γ- and β-herpesviruses are required for the expression of viral late genes, which are transcribed only after the onset of viral genome replication (Gruffat et al., 2012; Isomura et al., 2011; Wong et al., 2007). In the high confidence data set, exogenously expressed ORF24 co-purified 10 of the 13 human RNA polymerase II (Pol II) subunits (an additional 2 subunits had MiST scores above 0.6), and also with multiple Pol II subunits in lytically reactivated KSHV-positive iSLK.219 cells (Figure 2.7A). We engineered a FLAG tag at the N-terminus of ORF24 within the KSHV genome (KSHV.FLAG.24), and found that Pol II also interacted with this endogenous ORF24 expressed from its native promoter (Figure 2.7B). Pol II subunits were not found in any of the other KSHV protein high confidence networks, including well-characterized KSHV transcriptional factors (ORF50 and K8α) (Figure 2.8A). ORF24 specifically bound the lower mobility, hypophosphorylated form of the RPB1 (POLR2A) subunit of Pol II and not the form phosphorylated at serine 2, 5, or 7 of the RPB1 C-terminal domain (CTD), consistent with a potential role in PIC assembly (Figure 2.7C). The interaction was DNA-independent, suggesting that it may occur prior to Pol II recruitment to promoters (Figure 2.8B). Finally, we observed that Pol II binding, as measured both by AP-MS and IP-Western blotting, was a feature conserved among ORF24 orthologs in human cytomegalovirus (HCMV; UL87), Epstein-Barr virus (EBV; BcRF1), and murine γ-herpesvirus 68 (MHV68; mORF24) (Figure 2.7D, Figure 2.8C-F).

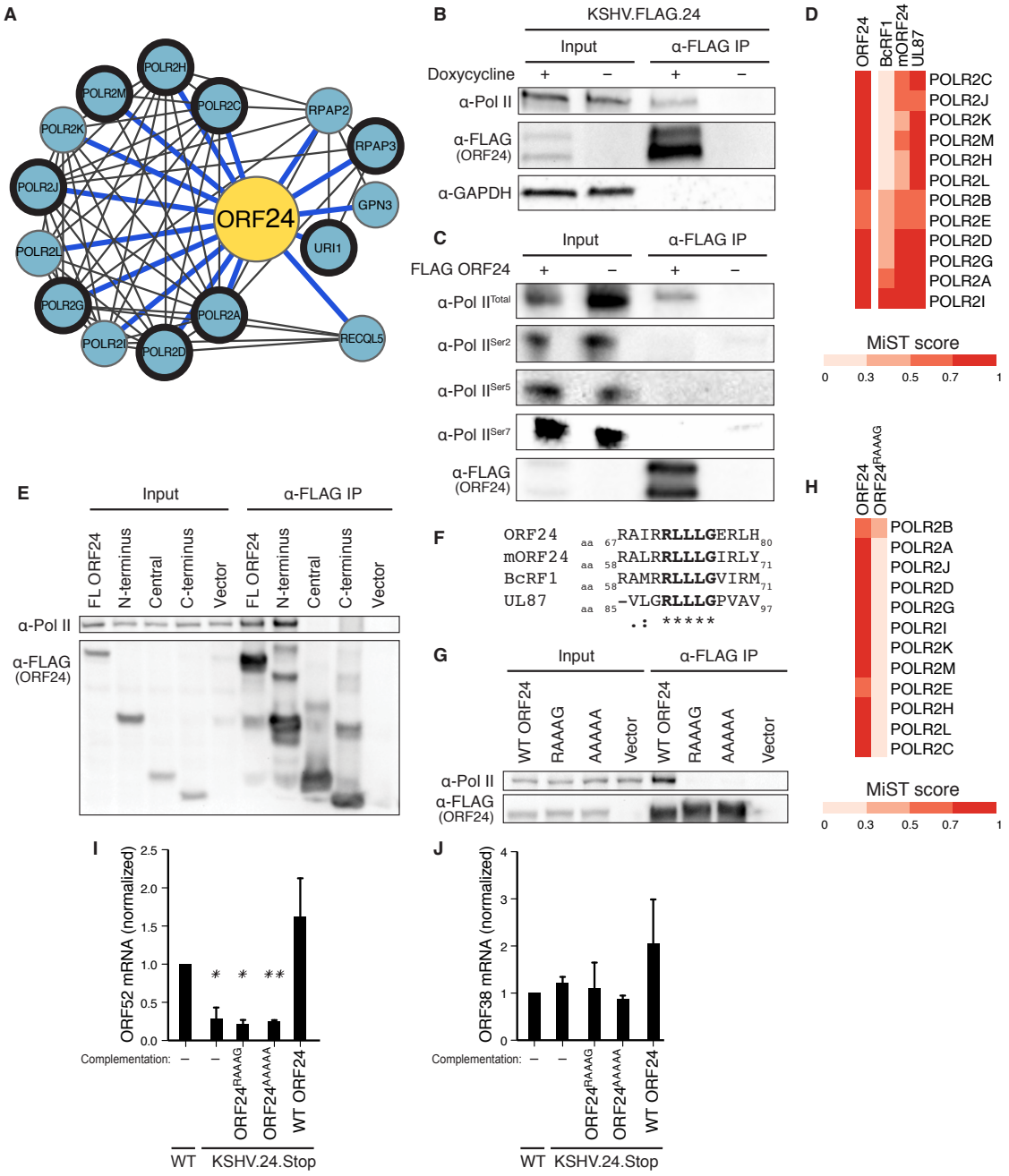

Figure 2.7: KSHV ORF24 binds human RNA polymerase II to drive viral late gene expression.

(A) Network showing the high confidence interaction partners of ORF24 identified in 293T cells (all blue nodes), and those also identified in KSHV infected iSLK.219 cells (blue nodes with heavy black borders). Pol II subunits are depicted to the left of the ORF24 node while Pol II accessory factors are shown to the right. Grey lines indicate known interactions between human proteins. **(B)** Lysates of iSLK cells either latently (- doxycycline) or lytically (+ doxycycline) infected with KSHV.FLAG.24 were subjected to α-FLAG IP and Western blotted for the RPB1 subunit of Pol II. GAPDH was used as a loading and IP specificity control. In this and

all subsequent IPs, input represents 5% of lysate used for IP. **(C)** Lytically induced iSLK cells infected with KSHV.WT were transduced with FLAG-tagged ORF24, and lysates were subjected to α-FLAG IP and Western blotted using antibodies against total RPB1 (Pol II) or the hyperphosphorylated forms of Pol II. **(D)** Heat map representing the MiST scores for Pol II subunits detected in association with the ORF24 orthologs from KSHV (ORF24), MHV68 (mORF24), EBV (BcRF1), and CMV (UL87) via AP-MS from 293T cells. **(E)** 293T cells were transfected with plasmids expressing FLAG-tagged full length (FL) ORF24 or fragments of ORF24 encompassing the N-terminus (amino acids 1-401), central domain (amino acids 402-603), or C-terminus (amino acids 604-752). Cell lysates were subjected to α-FLAG IP and Western blotted for the RPB1 subunit of Pol II. **(F)** A Clustal Omega multiple sequence alignment for the ORF24 orthologs showing the stretch of 5 conserved residues within their N-termini. **(G)** 293T cells were transfected with plasmids expressing FLAG-tagged WT ORF24 or an ORF24 mutant in which the conserved 'RLLLG' motif was mutated to 'RAAAG' or 'AAAAA'. Lysates were subjected to α-FLAG IP and Western blotted with the indicated antibodies. **(H)** Heat map representing the MiST scores for Pol II subunits detected in association with WT ORF24 (ORF24) versus the ORF24^{RAAAG} mutant via AP-MS from 293T cells. **(I-J)** iSLK cells infected with WT KSHV or KSHV lacking ORF24 (KSHV.24.Stop) were transduced with a retroviral vector expressing either WT ORF24 or the indicated ORF24 Pol II binding mutants. Forty-eight hr following lytic reactivation, accumulation of the ORF52 late gene (I) or the ORF38 early gene (J) was measured by RT-qPCR and normalized to levels of 18S, with the viral mRNA levels present during WT infection set to 1. * Indicates a p-value < 0.05 and ** indicates a p-value < 0.005 as determined by Student's t-test.

KSHV ORF24 is a 752 amino acid protein with a central 209 amino acid domain that is highly conserved amongst its orthologs and is flanked by more divergent N- and C-terminal regions. Truncation mutants of KSHV ORF24 encompassing each of these domains demonstrated that the Pol II interaction was mediated exclusively through the N-terminus (amino acids 1-401) (Figure 2.7E). Although the amino acid sequence of the N-terminal domain is poorly conserved, it does contain an invariant 5 amino acid stretch (Figure 2.7F). Mutation of these 5 amino acids in KSHV ORF24 to alanines (ORF24^{AAAAA}), or just the 3 leucines to alanines (ORF24^{RAAAG}) completely abrogated the interaction with Pol II (Figure 2.7G-H).

Similar to previous observations in EBV, CMV, and MHV68 (Gruffat et al., 2012; Isomura et al., 2011; Wong et al., 2007), introduction of a stop codon in the KSHV ORF24 locus (KSHV.24.Stop) impaired late gene expression but did not impact early gene expression or viral DNA replication (Figure 2.8G-H, Figure 2.11A). As expected, repair of the stop mutation in the virus or complementation with wild type (WT) ORF24 restored late gene expression (Figure 2.8G, Figure 2.11A). When WT or Pol II binding mutants of ORF24 were exogenously expressed in cells infected with KSHV.24.Stop (Figure 2.8I), expression of WT ORF24, but not the Pol II binding mutants, restored expression of both the ORF52 (Figure 2.7I) and other late genes (Figure S5A) but did not impact expression of the ORF38 early gene (Figure 2.7J). Furthermore, cells complemented with the ORF24 Pol II binding mutants produced

no detectable infectious virions in supernatant transfer assays (Figure 2.11B). Thus, the ORF24-Pol II interaction is essential for the selective activation of late genes and completion of the viral lifecycle.

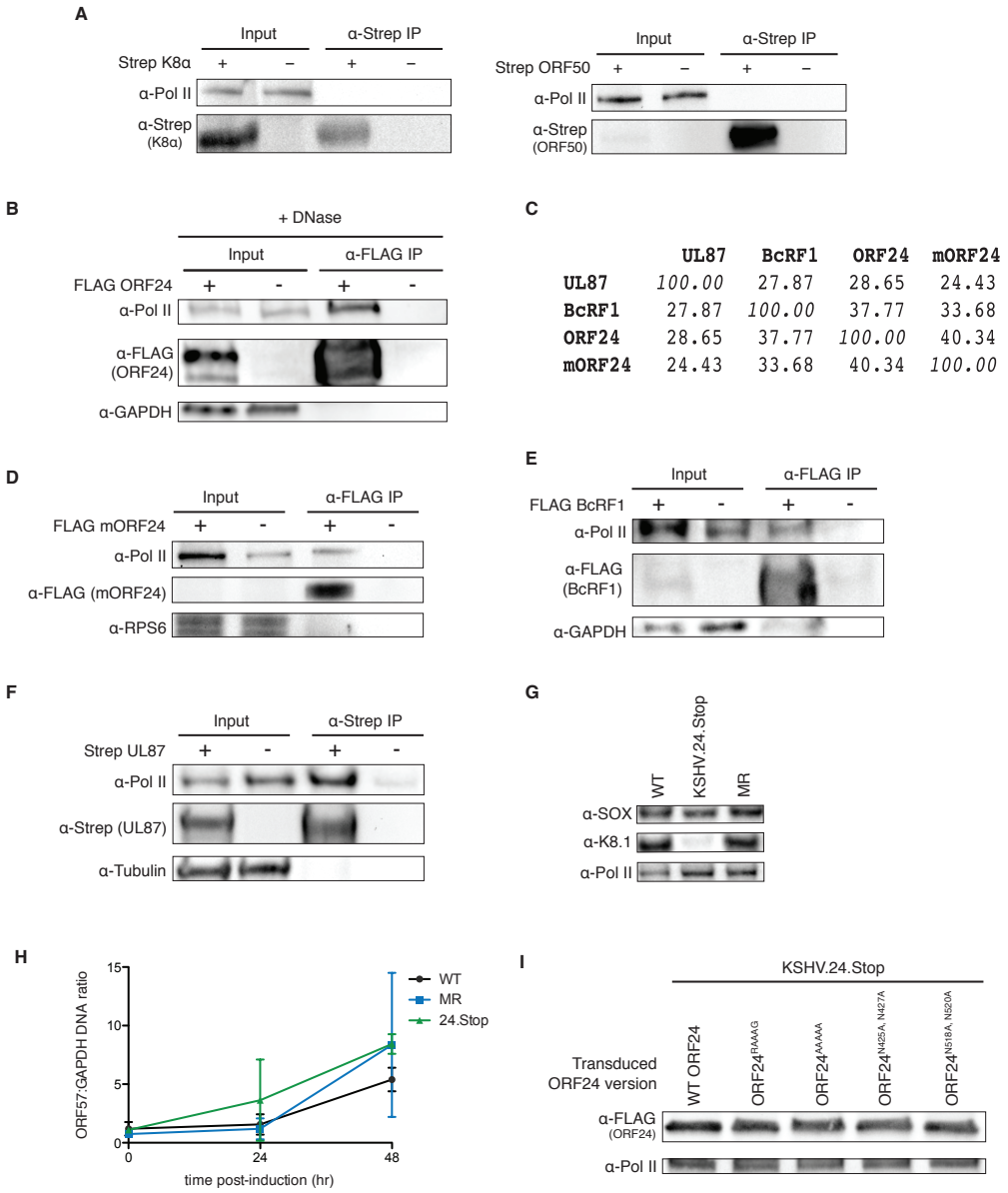

Figure 2.8: ORF24 interacts with Pol II in a DNA-independent manner and the interaction is conserved across ORF24 orthologs.

(A) Pol II interaction is not a conserved feature of viral transcription factors. Lysates of 293T cells transfected with either empty vector or plasmids expressing the strep-tagged KSHV transcription factors K8α or ORF50 (RTA) were subjected to α-strep IP and Western blotted with antibodies against strep and Pol II. Input lanes represent 5% of the lysate used for IP. **(B)** The interaction between ORF24 and Pol II is DNA independent. Lysates were prepared as described above, but were treated with DNase prior to α-FLAG IP. Input lanes represent 5% of the lysate used for IP. GAPDH was used as a loading and IP specificity control. **(C)** Percent identity matrix for

ORF24 orthologs. **(D-F)** ORF24 orthologs from other herpesviruses all interact with Pol II. 293T cells were transfected with plasmids expressing FLAG mORF24 from MHV68 (E), FLAG BcRF1 from EBV (F), or strep UL87 from HCMV (G). The viral proteins were then subjected to IP with the indicated antibody and their interaction with Pol II was monitored by Western blotting with an antibody against RPB1. Input lanes represent 5% of the lysate used for IP. Where indicated, RPS6, GAPDH, and tubulin served as loading and IP specificity controls. **(G)** Deletion of ORF24 from KSHV prevents expression of late but not early genes. Western blots of lysates from iSLK cells lytically infected with WT KSHV, KSHV bearing a stop codon at the 5' end of ORF24 (KSHV.24.Stop), or a mutant rescue KSHV in which the stop mutation was repaired to WT (KSHV.MR). SOX (ORF37) is a KSHV protein expressed with early kinetics, whereas K8.1 is a KSHV late protein. Pol II serves as a loading control. **(H)** ORF24 is not required for viral DNA amplification. qRT-PCR measurements of the DNA levels of the KSHV viral genome in iSLK cells containing WT KSHV, KSHV.24.Stop, or KSHV.MR 24 and 48 hr after induction into the lytic cycle. **(I)** Lysates of the KSHV.24.Stop iSLK cell lines transduced with the indicated WT or mutant ORF24 for the complementation assay were Western blotted with anti-FLAG antibodies to confirm equivalent expression of the ORF24 proteins. Pol II serves as a loading control.

ORF24 functionally resembles cellular TBP.

Mapped late gene promoters are strikingly minimal, as they encompass only 12-15 bp in total and are comprised of little more than a core TATA-like element (TATT) (Homa et al., 1988; Tang et al., 2004; Wong-Ho et al., 2014). The central conserved domain in the ORF24 orthologs was predicted to adopt a TBP-like secondary structure, although the proteins lack significant sequence homology (Wyrwicz and Rychlewski, 2007). Similar to the EBV and MHV68 ORF24 orthologs (Gruffat et al., 2012; Wong-Ho et al., 2014), purified KSHV ORF24 bound to a TATT-containing KSHV K8.1 late gene promoter sequence in DNA gel shift assays, but failed to bind a size-matched segment from the KSHV ORF57 early gene promoter containing a canonical TATA box (Figure 2.9A). K8.1 promoter binding was dependent on its core TATT sequence, as mutating these bases to CCCC abrogated the interaction (Figure 2.9A). However, with MHV68 ORF24, mutating the last T of the TATT sequence had no effect compared to a wildtype late gene probe (ORF52) (Figure 2.10A-B).

TBP has asparagine and valine residues within its inner 'saddle' that contact DNA and contribute to its specificity for the TATA sequence (Kim et al., 1993). Mutation of these positionally conserved residues in KSHV ORF24 ($ORF24^{N425A, N427A}$ and $ORF24^{N518A, N520A}$) did not impair Pol II binding, but abrogated the interaction with the K8.1 late gene promoter as well as the ability to rescue late gene expression and virion production during infection with the KSHV.24.Stop virus (Figure 2.9A-C, Figure 2.11A-B). Thus, KSHV ORF24 is a modular protein containing at least two essential but genetically separable domains required to drive selective transcription from minimalistic viral late promoters.

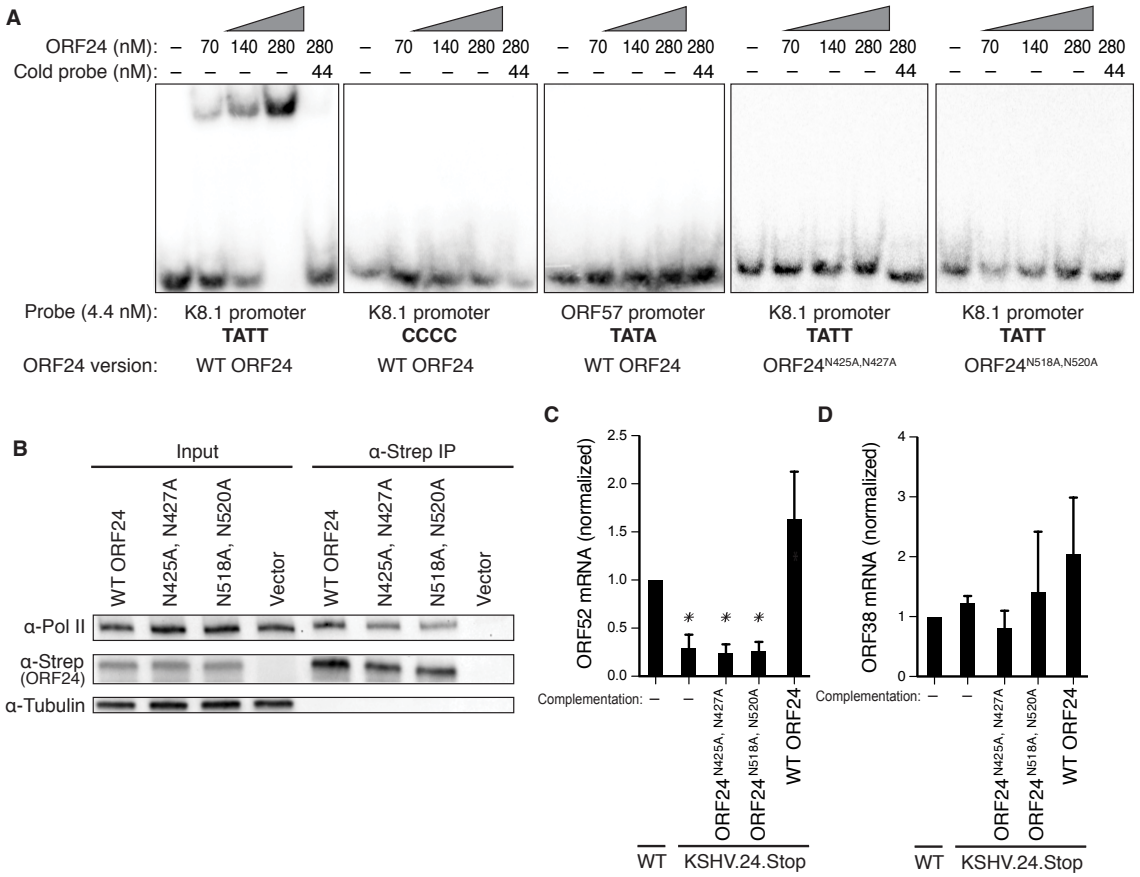

Figure 2.9: ORF24 binds late gene promoter DNA in a manner essential for late gene expression.

(A) Purified WT ORF24 or ORF24 putative DNA-binding mutants (ORF24$^{N425A, N427A}$ and ORF24$^{N518A, N520A}$) (0-280 nM) were subjected to EMSA with 23 nt radiolabeled dsDNA probes encompassing the K8.1 late promoter, which contains a central 'TATT' sequence, or the ORF57 early promoter, which contains a central 'TATA' sequence. Where indicated, the 'TATT' motif in the K8.1 promoter probe was mutated to 'CCCC'. The specificity of the interaction was confirmed by competition with unlabeled (cold) probe. **(B)** 293T cells were transfected with plasmids expressing strep-tagged WT ORF24 or the indicated ORF24 DNA-binding mutants. Lysates were subjected to α-strep IP and Western blotted for the RPB1 subunit of Pol II. Tubulin was used as a loading and IP specificity control. **(C-D)** iSLK cells infected with WT KSHV or KSHV.24.Stop were transduced with a retroviral vector expressing WT ORF24 or the indicated ORF24 DNA-binding mutants. Forty-eight hr following lytic reactivation, accumulation of the ORF52 late gene (C) or the ORF38 early gene (D) was measured by RT-qPCR and normalized to levels of 18S, with the viral mRNA levels present during WT infection set to 1. * Indicates a p-value < 0.05 as determined by Student's t-test.

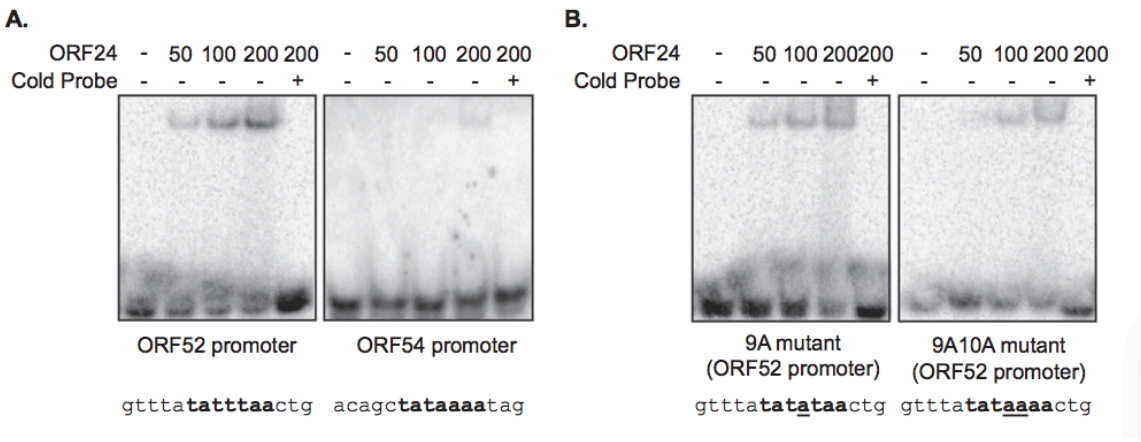

Figure 2.10: ORF24 interacts with TATA-like core promoter sequences.
(A-B) Increasing amounts of FLAG-tagged MHV68 ORF24 were incubated with 1 nM of radiolabeled probe with the listed sequence, with or without competing cold probe. The resulting complexes were resolved on 5% TBE gels. (A) ORF52: late; ORF54: early. (B) 9A mutant: core TAT**T** mutated to TAT**A;** 9A10A mutant: TAT**TT**AA changed to TAT**AA**AA.

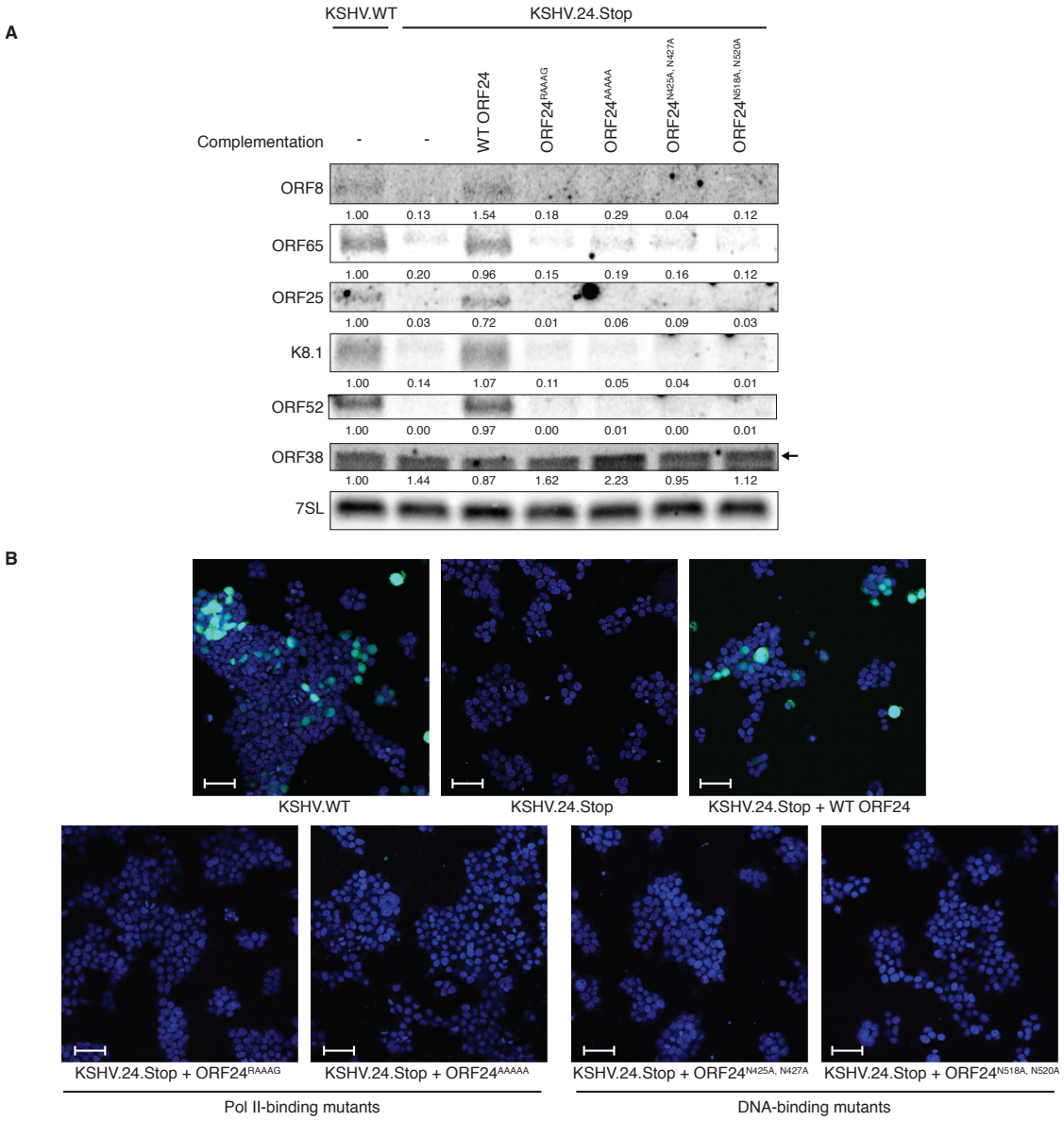

Figure 2.11: ORF24 mutants unable to bind either DNA or Pol II are defective for late gene expression and virion production.

(A) RNA harvested from lytically reactivated KSHV.WT and KSHV.24.Stop iSLK cells complemented with the indicated WT or mutant ORF24 was Northern blotted using strand specific primer probes against the late genes ORF8, ORF65, ORF25, K8.1, and ORF52, the ORF38 early gene, or 7SL (as a loading control). **(B)** Both the DNA binding and Pol II binding activities of ORF24 are required for production of infectious progeny virions. KSHV.WT and KSHV.24.Stop iSLK cells complemented with WT ORF24, ORF24-Pol II binding mutants, or DNA-binding mutants were lytically reactivated for 48 hr and supernatant from these cells was collected and transferred to recipient 293T cells. After 24 hr, the recipient cells were monitored for GFP expression, as infectious virus constitutively expresses GFP. DNA is stained with DAPI (blue). A minimum of 30 fields were monitored for each experiment, with a single representative field shown. No GFP positive cells were detected in any field

of recipient cells containing supernatant from infected cells lacking WT ORF24. Scale bar = 50 μm.

ORF24 replaces TBP in the viral late gene pre-initiation complex.

TBP and TBP associated factors (TAFs) collectively comprise TFIID, which binds the core promoter and serves as a platform for the recruitment of other PIC components. Our data suggested, however, that Pol II might be recruited independently of TFIID to viral late promoters by ORF24.

In agreement with our *in vitro* binding results, chromatin immunoprecipitation (ChIP) assays showed FLAG-ORF24 bound to the K8.1 late gene promoter but not the ORF57 early gene promoter in lytically reactivated iSLK cells infected with KSHV.FLAG.24 virus (Figure 2.12A). In contrast, TBP was present at the ORF57 early promoter but not detectable at the K8.1 late promoter (Figure 2.12A). Unlike Pol II, both TBP and ORF24 remained at their associated promoter (Figure 2.12A). The apparent mutually exclusive binding of ORF24 and TBP suggests that ORF24 acts as a replacement for TBP on late gene promoters.

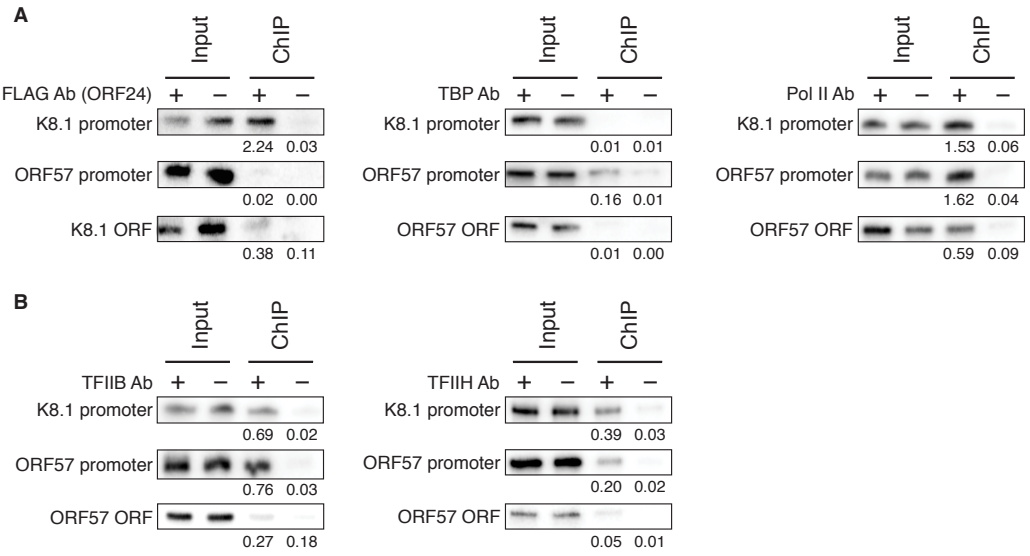

Figure 2.12: ORF24 functionally replaces TBP on the late promoter.
(A-B) Chromatin from iSLK cells lytically infected with KSHV.FLAG.24 was isolated and subjected to ChIP using antibodies against FLAG (to precipitate FLAG ORF24), TBP, and Pol II (A), or TFIIB and TFIIH (ERCC3) (B). Co-precipitating DNA was detected by PCR using radiolabeled nucleotides and primers specific for the indicated promoter or gene body. Experiments were repeated at least three times and a representative image is presented with the ChIPs quantified against input samples.

Cellular TBP does not bind Pol II directly; its interaction with Pol II at the PIC is instead bridged by TFIIB and other general transcription factors (GTFs) (Bushnell et al., 2004). However, these additional PIC components were absent from the KSHV ORF24 MS data, a finding we independently confirmed by Western blotting for TBP,

TFIIB, TFIIE, TFIIF, and TFIIH in KSHV.FLAG.24 immunoprecipitates (Figure 2.13A). Thus, either the recruitment of Pol II to late promoters by ORF24 bypasses the need for additional GTFs or, alternatively, GTF assembly into the late promoter PIC occurs in a TBP-independent manner. We addressed these possibilities by measuring the *in vivo* promoter occupancy of the GTFs TFIIB and TFIIH (ERCC3), two canonical components of PICs found at RNA Pol II transcribed promoters (Figure 2.12B). Despite the absence of TBP, ChIP assays revealed assembly of both TFIIH and TFIIB on the K8.1 late gene promoter (Figure 2.12B). TFIIB occupancy at the K8.1 promoter was lost in the KSHV.24.Stop virus, confirming that recruitment of at least a subset of GTFs to this promoter is ORF24 dependent (Figure 2.13B). Thus, the KSHV ORF24 protein nucleates assembly of a PIC that bypasses the need for TBP but is nonetheless structured to enable subsequent recruitment of other core mammalian transcription factors.

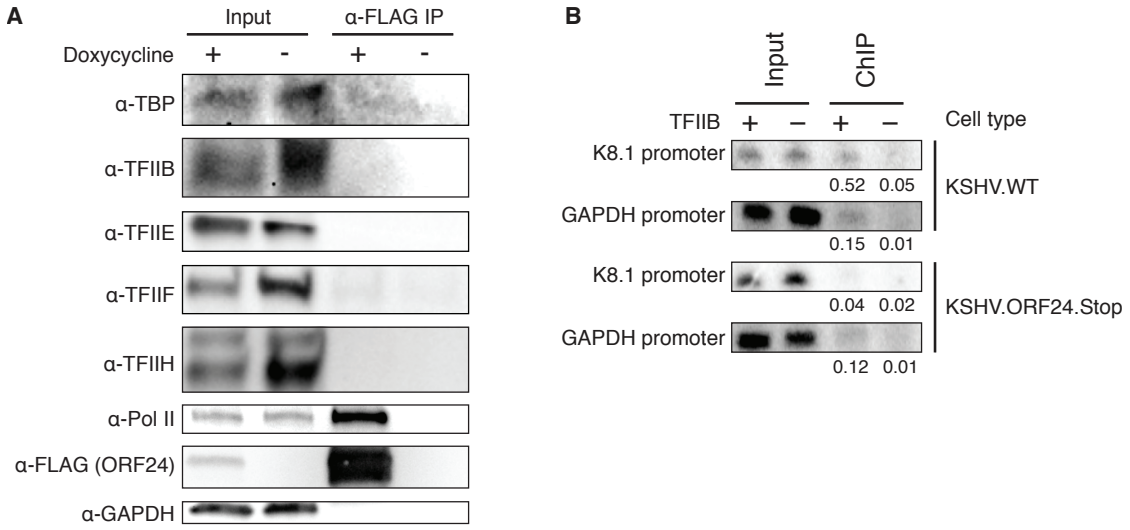

Figure 2.13: ORF24 interacts selectively with Pol II and is needed for late gene PIC assembly.

(A) Lysates of iSLK cells either latently (- doxycycline) or lytically (+ doxycycline) infected with KSHV.FLAG.ORF24 were subjected to α-FLAG IP and Western blotted with antibodies to the indicated general transcription factors or GAPDH as a control. Input lanes represent 5% of the lysate used for IP. **(B)** Chromatin from iSLK cells lytically infected with KSHV.WT or KSHV.24.Stop was isolated and subjected to ChIP using antibodies against TFIIB. Co-precipitating DNA was detected by PCR using radiolabeled nucleotides and primers specific for the indicated promoter. Experiments were repeated at least three times and a representative image is presented with the ChIPs quantified against input samples

Discussion

Here, we present the largest whole virus-host interactome defined to date. Host genes that interact with more highly conserved viral proteins show significantly elevated rates of positive selection across mammals. This effect diminishes for the more recently evolved KSHV-specific viral proteins, consistent with restricted

contact across the expanses of the host-pathogen phylogenetic trees. However, the interactors of these KSHV-specific proteins show a dramatic increase in signatures of recent natural selection within human populations. We therefore speculate that KSHV may have contributed to shaping patterns of human genetic variation over the last 10,000-30,000 years.

There is significant overlap between the virus-host interactomes of KSHV and HIV-1, with high confidence targets of nearly every HIV-1 protein present in the overlapping PPI set. This overlap is notable, because although these two viruses are pathologically linked (particularly in the case of KSHV-induced diseases), they have distinct life cycles and replication strategies. As more systematic interaction data sets emerge, it will be of interest to examine whether any of these overlapping partners are present within additional virus-host interactomes. In this regard, core host pathways that are reiteratively targeted by invading pathogens could generate signals akin to 'patterns of pathogenesis' that are recognized by the innate immune system as markers of viral infection (Vance et al., 2009).

There are several strengths in our systematic approach that lead to a robust PPI network. First is our ability to identify reproducible and specific interactions, which should serve as the basis for future hypothesis-driven studies to probe their role in the KSHV life cycle. We demonstrated that the PPI network could accurately predict known protein functions and, in the case of viral latency factors, was enriched for proteins linked to cancer. Our standardized pipeline also eliminates the inherent difficulties in cross-comparisons of individual PPI experiments. That said, the network likely underestimates the number of interactions coordinated by KSHV. It does not capture PPIs driven by complexes comprised of multiple viral proteins and excludes some biologically relevant interactions that were detected but de-prioritized by MiST due to low specificity.

Directed by the interactome, we gained insight into a poorly understood stage of γ-herpesvirus gene expression (Figure 2.14). Late genes of all DNA viruses become transcriptionally active only after the onset of viral DNA replication and produce proteins necessary for progeny virion assembly. Foundational work in related herpesviruses has revealed a complex of six viral proteins to be essential for late gene expression and recruitment of Pol II (Arumugaswami et al., 2006; Aubry et al., 2014; Jia et al., 2004; Perng et al., 2014; Wong et al., 2007; Wu et al., 2009). Although the precise roles of each of these factors remain unknown, the ORF24 orthologs in related herpesviruses have been shown to exhibit sequence-specific DNA binding activity and display predicted structural homology to TBP (Gruffat et al., 2012; Wyrwicz and Rychlewski, 2007). We find that ORF24 but no other KSHV proteins binds Pol II, indicating that it is the late gene transcription factor that mediates polymerase recruitment to KSHV late promoters. Furthermore, the observation that ORF24 but not TBP occupies the KSHV K8.1 late promoter *in vivo* suggests that ORF24 functionally replaces TBP specifically at late gene PICs. This interaction occurs through the atypical TATA box (TATT) characteristic of late promoters and requires ORF24 residues that are spatially and functionally similar to the DNA binding residues of TBP. These data support the hypothesis that KSHV ORF24 and its orthologs function as the first described viral TBP mimics (Wyrwicz and

Rychlewski, 2007), orchestrating the assembly of a novel virus-host hybrid transcription complex.

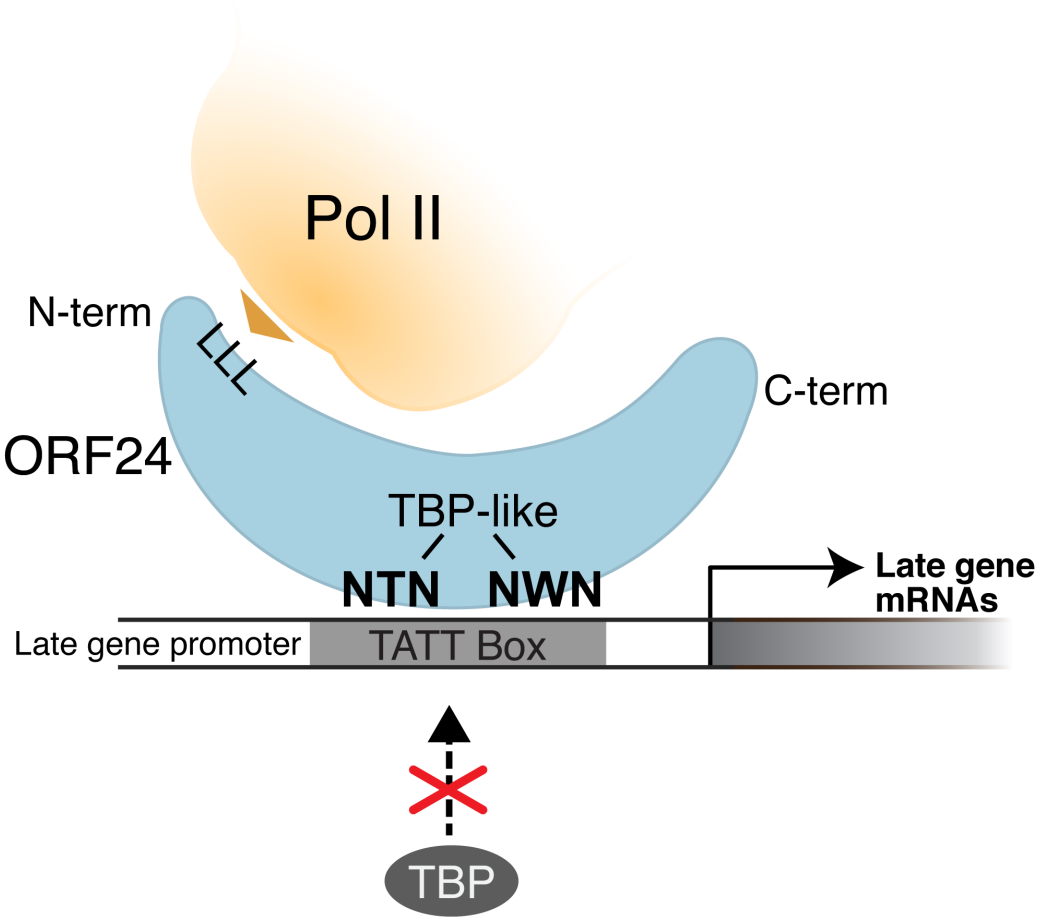

Figure 2.14: Model for ORF24 activation of late gene promoters.
ORF24 interacts with Pol II through conserved leucines in its N-terminal domain. The polymerase is then brought to the late promoter TATT box via the ORF24 TBP-like domain. TBP is excluded from late promoters, and instead ORF24 binding nucleates assembly of the PIC by enabling recruitment of other cellular (GTFs).

TBP serves as a critical nucleation factor for PIC assembly, including at promoters lacking a TATA box (Thomas and Chiang, 2006). However, unlike ORF24, TBP does not directly bind Pol II; the polymerase is instead brought to the TBP-bound promoter through interactions with TFIIB and additional transcription associated factors (Murakami et al., 2013). The low degree of sequence homology between ORF24 and TBP makes it unlikely that ORF24 is able to associate with a similar cohort of TAFs and GTFs, a notion supported by the selectivity of the ORF24-Pol II interaction. Thus, cellular GTF recruitment to late promoters may instead occur through other viral or cellular proteins that perhaps join the ORF24-bound DNA. ORF24-directed recruitment of Pol II could also enable late genes to bypass select regulatory factors required for canonical eukaryotic promoter activation.

While proteins from other viruses have been reported to interact with Pol II, ORF24 is unique in its ability to coordinately bind Pol II and promoter DNA (Dorjsuren et al., 1998; Engelhardt et al., 2005; Mavankal et al., 1996; Takramah et al., 2003; Zhou and Knipe, 2002). This is reminiscent of prokaryotic sigma factors, which compete to bind the RNA polymerase holoenzyme and are required for selective promoter recognition and initiation (Osterberg et al., 2011). It is possible that herpesviruses have evolved an analogous strategy to secure sufficient levels of Pol II for strong late gene expression late in infection, a time when the cell is stressed and perhaps constrained for resources. DNA viruses likely display significant variety in PIC composition, including in the core GTFs as has been shown for some cell cycle-regulated promoters (Guglielmi et al., 2013). Future studies of the regulation of late genes are therefore anticipated to uncover other novel modes of transcriptional regulation with parallels in mammalian cells.

Materials and Methods

Plasmid construction

A library of strep-tagged KSHV ORFs was assembled from RNA from TREx BCBL-1-RTA cells reactivated for 36 hr with TPA (20 ng/mL), ionomycin (500 ng/mL) and doxycycline (1 μg/mL). ORF primers were designed based on sequences from the European Nucleotide Archive and GenBank. Inserts were ligated into pcDNA 4/TO (Invitrogen) modified to include a C-terminal 2x strep tag.

Plasmids expressing ORF63 (Gregory et al., 2011) and ORF64 (Gonzalez et al., 2009) were kindly provided by Blossom Damania and were used for subcloning. Cloning of MHV68 ORF24 and EBV BcRF1 was carried out in the same manner described for KSHV genes using RNA from MHV68 infected 3T3 cells and EBV infected Akata cells reactivated with anti-human IgG (Takada and Ono, 1989). CMV UL87 was cloned using Towne BAC DNA as a template. ORF24 DNA binding and Pol II binding mutants were made by site-directed mutagenesis. ORF24 for expression in BJAB and iSLK.219 cells was PCR amplified and subcloned into the pLPCX retroviral vector (Clontech) using 5' BglII and 3' Hind III restriction sites. WT and mutant versions of ORF24 were cloned into the pQCXIN retroviral vector (Clontech) using 5' AgeI and 3' BamHI restriction sites.

Cell culture

HEK293T (ATCC), Phoenix 293 retroviral packaging cells (Swift et al., 2001), the renal carcinoma cell line iSLK puro (Sturzl et al., 2013) bearing a dox-inducible version of the major KSHV lytic transactivator RTA, KSHV positive iSLK.219 cells (Myoung and Ganem, 2011), and other iSLK puro-derived cell lines were maintained in Dulbecco's modified Eagle medium (DMEM; Invitrogen) with 10% fetal bovine serum (FBS; Invitrogen, HyClone) and penicillin-streptomycin (Invitrogen). The BJAB B cell line was maintained in RPMI (Invitrogen) medium with 10% FBS and penicillin-streptomycin. Stable cell lines expressing tagged viral ORF24 for affinity

purification or ORF24 complementation assays were generated by retroviral transduction.

Transductions

Briefly, pLPCX or pQCXIN retroviral plasmids were co-transfected with VSV G into low passage Phoenix 293 cells. Supernatant containing viral particles was collected 48 hr later, filtered, complemented with 8 μg/mL of polybrene and centrifuged onto target cells. Transfected BJABs transduced with pLPCX constructs were selected for two weeks with 1 μg/mL puromycin (Millipore).

Affinity purification for mass spectrometry

Affinity purification of proteins to be analyzed by MS was performed as previously described (Jager et al., 2011a; Jager et al., 2011e). Briefly, clarified cell lysate was incubated with pre-washed Strep-Tactin beads (IBA) or FLAG beads (Sigma) and allowed to bind for 1-3 hr. Following purification, complexes bound to beads were washed and then eluted with desthiobiotin (IBA) or 3X FLAG peptide (Elim BioPharm) for strep or FLAG tags, respectively.

Affinity purification and Western blotting

Cells were incubated in lysis buffer [0.5% NP-40, 150 mM NaCl, 50 mM Tris pH 7.4, 1 mM EDTA and protease inhibitors (Roche)]. Equivalent amounts of clarified lysate were added to 30 mL pre-washed Strep-Tactin beads (IBA) or FLAG beads (Sigma) and allowed to bind for 1-3 hr, washed 3 times with lysis buffer lacking protease inhibitors and containing 0.05% NP-40, and washed 3 additional times in lysis buffer lacking NP-40. Complexes were eluted using desthiobiotin (IBA) or 3X FLAG peptide (Elim BioPharm) for strep or FLAG tags, respectively. The elutions were divided into thirds and analyzed by Western blotting to confirm protein expression with α-FLAG (ORF24 in iSLK.219 cells) or α-strep (all MS in 293T cells) antibodies, silver stain, and mass spectrometry.

Co-immunoprecipitations for experiments other than MS were performed similarly, except with the omission of the final washes in buffer lacking NP-40. Western blots were performed using the following antibodies: rabbit α-FLAG (1:5000; Sigma), mouse α-FLAG (clone M2; 1:5000; Sigma), mouse α-β-tubulin (clone TUB 2.1; 1:3000; Sigma), mouse α-strep (1:3000; Qiagen), strep direct HRP conjugate antibody (1:10,000; Novagen), rabbit α-RNA Pol II (N20; 1:5000; Santa Cruz Biotechnologies), rabbit α-V5 (1:3000; Invitrogen), rabbit α-TFIIB (clone SI-1; 1:1,000; Santa Cruz Biotechnologies), mouse α-TBP (clone 1TBP18; 1:1000; Abcam), rabbit α-TFIIE (clone C-21; 1:1,000; Santa Cruz Biotechnologies), rabbit α-TFIIF RAP74 (clone N-16; 1:1,000; Santa Cruz Biotechnologies), rabbit α-TFIIH p89 (clone S-19; 1:1,000; Santa Cruz Biotechnologies), rabbit α-GAPDH (1:1,000; Abcam), mouse α-S6 ribosomal protein (clone 54D2; 1:1000; Cell Signaling), rat α-Pol II serine 5 (clone 3E8; 1:10,000; Millipore), rat α-Pol II serine 7 (clone 4E12; 1:10,000;

Millipore), goat α-rat HRP secondary antibody (1:5000; Cell Signaling), and goat α-mouse and goat α-rabbit HRP secondary antibodies (1:5000; Southern Biotech).
For nuclease treatment of whole cell lysate, lysis buffer without EDTA was used [0.5% NP-40, 150 mM NaCl, 50 mM Tris pH 7.4, and protease inhibitors (Roche)]. To 500 μg of cell lysate, 2 μL of DNase I (NEB), 2 μL of micrococcal nuclease, and micrococcal nuclease buffer (NEB) were added and allowed to incubate for 1 hr at 4°C. Following incubation, 10 mM EDTA was added to inactivate the nucleases. Immunoprecipitation was then carried out as described.

Mass spectrometry

Eluates were processed, trypsin digested, and concentrated for LC-MS/MS. Digested peptide mixtures were analyzed by LC-MS/MS on a Thermo Scientific Velos Pro ion trap mass spectrometry system equipped with a Proxeon Easy nLC II high pressure liquid chromatography and autosampler system.

Affinity purified protein eluates were first denatured and reduced in 2 M urea, 10 mM NH_4HCO_3, and 2 mM DTT for 30 min at 60°C, then alkylated with 2 mM iodoacetamide for 45 min at room temperature. Trypsin (Promega) was added at a 1:100 enzyme: substrate ratio and digested overnight at 37°C. Following digestion, samples were concentrated using C18 ZipTips (Millipore) according to the manufacturer's specifications.

Samples were injected onto a pre-column (2 cm x 100 μm I.D. packed with ReproSil Pur C18 AQ 5 μm particles) in 0.1% formic acid and then separated with a 1 hr gradient from 5% to 30% acetonitrile in 0.1% formic acid on an analytical column (10 cm x 75 μm I.D. packed with ReproSil Pur C18 AQ 3 μm particles). The mass spectrometer collected data in a data-dependent fashion, collecting one full scan followed by 20 collision-induced dissociation MS/MS scans of the 20 most intense peaks from the full scan. Dynamic exclusion was enabled for 30 seconds with a repeat count of 1. The resulting raw data was matched to protein sequences by the Protein Prospector algorithm (Clauser et al., 1999). Data were searched against a database containing SwissProt Human protein sequences (downloaded April 13, 2012) and viral sequences, concatenated to a decoy database where each sequence was randomized in order to estimate the false positive rate. The searches considered a precursor mass tolerance of 1 Da and fragment ion tolerances of 0.8 Da, and considered variable modifications for protein N-terminal acetylation, protein N-terminal acetylation and oxidation, glutamine to pyroglutamate conversion for peptide N-terminal glutamine residues, protein N-terminal methionine loss, protein N-terminal acetylation and methionine loss, and methionine oxidation, and constant modification for carbamidomethyl cysteine. Prospector data was filtered using a maximum protein expectation value of 0.01 and a maximum peptide expectation value of 0.05.

Network scoring and benchmarking

Weights of the three MiST features were set by supervised training using two thirds of the known KSHV-host interactions detected in our screen as a positive set

(19/28), 100 times larger random subsets of the interactome as negative sets, and the remaining third of known interactions as a validation set (9/28). An exhaustive grid search was performed on the 4-dimensional vector described by the 3 weight variables and a threshold variable. The domain of these variables was limited to between 0 and 1, with 0.01 increments. The sum of the three variables was constrained to 1. For every complete assignment of the vector, the True Positive Rate (TPR) and False Positive Rate (FPR) in the set of MiST scores higher than the threshold were computed against the high confidence training and validation set or the negative set, respectively, together with the AUC for these rates. The feature weight values that gave rise to a maximal AUC, low FPR, and maximal TPR on the validation set were selected as the final MiST weights for abundance, reproducibility, and specificity. Network representations were plotted using Cytoscape, v. 2.8.3 (Smoot et al., 2011).

Evolutionary analysis

Human proteins interacting with only one KSHV protein were binned according to the classification of their viral binding partner within the herpesviral subfamilies. Within these bins of human proteins, differences in evolutionary pressures compared to a random background of MS-observable human proteins were assessed. Cross species selection (dN/dS) was assessed based on multiple sequence alignments of human, chimp, gorilla, orangutan, and rhesus macaque sequences constructed using MOSAIC (Maher and Hernandez, 2014). iHS scores were calculated by selscan (Szpiech and Hernandez, 2014), then summarized at the gene level by taking a linear combination of the log of the number of nominally significant z-scores and the maximum z-score within a given coding region. Weighting was chosen using principle components analysis.

Disease association analysis

First, all 20,264 proteins in the human reference proteome were mapped to gene-disease associations in DisGeNET (v.1) using Uniprot accession codes. All proteins in the reference proteome were labeled as 'cancer-associated', 'other disease associated' or 'none' based on their DisGeNET disease annotations. Cancer association was called using a regular expression search with tumor and cancer-related keywords. Every protein with at least one cancer-associated term was labeled as a cancer-associated protein. Second, the reference proteome was divided in three sets: all proteins, latent KSHV ORF (K1, K2, K12A, K12B, K12C, ORF71, ORF72 and ORF73) interacting proteins or lytic KSHV ORF interacting proteins (remaining). Using this partitioning, relative fractions of proteins with cancer-associations or other disease associations were computed and the significance of the observed counts was computed with the hyper-geometric test. The nodeCharts plugin for Cytoscape was used to create disease association pie charts (Figure S2).

Comparison to HIV-1 AP-MS data set

The significance of observing an overlap of 52 proteins between the previously published data set of 292 unique HIV host factors characterized by AP-MS in HEK-293T cells and the 556 unique KSHV factors described here given the human reference proteome (n=20,264) was computed by the hyper-geometric test. The 52 overlapping proteins were annotated with GO annotations.

Ontology enrichment analysis of prey proteins

For each analysis, the list of identified preys was joined with the respective ontology mapping using Uniprot accession codes. Then, for each ontology term in this joined list, the average MiST score of all associated prey proteins was compared against a sampled distribution of averages for equally sized random subsets of the data. The probability that the observed term score was higher than the distributed sampled average was computed by the one-tailed cumulative normal distribution test. The set of enrichment p-values for all terms in the ontology mapping was adjusted for multiple hypotheses testing by applying the Benjamini-Hochbach correction (also known as FDR control or q-value). Terms that were significant at an FDR of 0.05 were reported as output.

KSHV Mutagenesis

KSHV mutants were engineered using KSHV BAC16 mutagenesis as previously described using a two-step scarless Red recombination system (Brulois et al., 2012). iSLK derived cells used in this study include a BAC with a premature stop codon in the ORF24 coding region (KSHV.24.Stop) and an N-terminally FLAG tagged ORF24 (KSHV.FLAG.24). Briefly, electrocompetent GS1783 *E. coli* cells containing wildtype KSHV BAC16 and transiently expressing *gam, bet* and *exo* were electroporated with a kanamycin resistance cassette containing an I-SceI restriction site and flanked by homologous arms to the region to be mutated. After sequence verifying the insertion of the resistance gene and the mutation, the second step of recombination was performed to remove the cassette, leaving kanamycin sensitive clones that were screened by replica plating. Pulsed-field gel electrophoresis (PFGE) of NheI digested BAC DNA was used to verify BAC integrity. Purified BAC16 DNA was transfected into low passage iSLK cells using PolyJet (SignaGen) and BAC16-containing cells were selected with 400 μg/mL of hygromycin. Individual clones were amplified and tested for their ability to reactivate by Western blotting and RT-qPCR for early viral gene products (data not shown).

RT-qPCR

RNA isolated from the relevant lytically reactivated iSLK puro derived cells was treated with Turbo DNase (Ambion) and reverse transcribed with AMV RT (Promega) using random 9-mers or gene specific primers in cases of overlapping viral transcripts. cDNA levels were quantified using DyNAmo color flash SYBR green master mix, ROX passive reference dye (Thermo Scientific) and transcript specific

primers. Transcript levels were normalized to 18s. Error bars represent the standard deviation of two to three independent experiments.

Chromatin immunoprecipitation

iSLK cells containing wildtype or mutant KSHV were induced with 1 μg/mL doxycycline for 36 hr before harvesting chromatin. ChIP was performed as described (Listerman et al., 2006) but using a Covaris focused sonicator to shear chromatin. After reversing crosslinks, DNA was purified and eluted in 150 μL of dH_2O. Eluted DNA (0.5-1 μL) was subjected to 20-25 cycles of PCR in reactions spiked with α-^{32}P dCTP. PCR products were resolved on urea PAGE gels, quantified using Image Lab (BioRad) software and normalized to 1% input.

ChIPs were performed with approximately 10 μg of chromatin and 2.5 μg of the following antibodies: rabbit anti-RNA Pol II (N20; Santa Cruz Biotechnologies), rabbit anti-TFIIH p89 (clone S-19; Santa Cruz Biotechnologies), rabbit anti-TFIIB (clone SI-1; Santa Cruz Biotechnologies), mouse anti-TBP (clone 1TBP18; Abcam) and mouse anti-FLAG (clone M2; Sigma).

Electromobility shift assays

Wild type or DNA binding mutants of 3X FLAG ORF24 were isolated from transduced BJABs in lysis buffer (100 mM HEPES, pH 7.9, 500 mM NaCl, 1% NP-40), affinity purified over FLAG beads, and eluted with 3X FLAG peptide. Protein purity was assessed by silver stain. Samples were dialyzed against 20 mM HEPES, pH 7.9, 40% glycerol, 100 mM KCl, 0.5 mM DDT and 0.5 mM PMSF. Shift buffer (1 μg/μL salmon sperm DNA, 0.5 mg/mL BSA, 0.05% triton-X 100, 10% glycerol, 10 mM HEPES pH 7.9, 40 mM KCl, 5 mM $MgCl_2$ and 10 μM BME) was added to binding reactions containing 70-280 nM protein and 4.4 nM radiolabeled probe. Oligonucleotides were end-labeled with γ-^{32}P ATP using T4 PNK (NEB), purified, and reverse complemented sequences were annealed together to make a double-stranded probe. Reactions were incubated at 37°C for 30 min and resolved by native PAGE.

Northern blotting

RNA was collected and purified from lytically reactivated BAC16-containing iSLK cells using Trizol (Invitrogen). Blots were hybridized with oligonucleotide probes end-labeled with γ-^{32}P ATP using T4 PNK (NEB), then scanned using a PharosFX Imager System (BioRad).z

Infectious virion production assay

BAC16-containing iSLK cells were reactivated for three days with sodium butyrate and doxycycline. Supernatant was collected and passed through a 0.45 μm filter. Filtered supernatant containing 8 μg/mL polybrene was added to 293T cells plated on poly-L-lysine-treated glass coverslips. Twenty-four hr later, slides were washed with PBS, fixed in 4% formaldehyde, washed again with PBS, and mounted

on slides with DAPI (Vector Labs). Slides were imaged with an LSM 710 laser scanning confocal microscope (Zeiss). BAC16 constitutively expresses GFP, thus, GFP positive cells indicated the production and transfer of infectious virions.

Table 2.4: List of Primers.
Strep K1, strep K12 (Kaposin A, Kaposin B, Kaposin C), and strep ORF72 (vCyclin) were kind gifts from D. Ganem. Strep ORF73 (LANA) was a kind gift from L. Coscoy.

Primer	Sequence (5'-3')	Orientation F: Forward R: Reverse	Use
K2 (vIL-6) (BamHI)	GCTCGGATCCATGTGCTGGTTCAAGTTGTGGTC	F	Cloning
K2 (vIL-6) (NotI)	TCGAGCGGCCGCCCTGACTTATCGTGGACGTCAGGAG	R	Cloning
K3 (EcoRI)	GGTGGAATTCATGGAAGATGAGGATGTTCCTGTCTG	F	Cloning
K3 (NotI)	TCGAGCGGCCGCCCTGAATGAAACATAAGGGCAGACGAAAC	R	Cloning
K4 (EcoRI)	GGTGGAATTCATGGACACCAAGGGCATCCT	F	Cloning
K4 (NotI)	TCGAGCGGCCGCCCTCCGCGAGCAGTGACTGG	R	Cloning
K4.1 (KpnI)	GCTTGGTACCATGTGGAGCATGTGCTGGG	F	Cloning
K4.1 (NotI)	TCGAGCGGCCGCCCCGAGGGGCATAACCCTTTACC	R	Cloning
K4.2 (EcoRI)	GGTGGAATTCATGCAAATTAGCTTTGCCGAAGTTCT	F	Cloning
K4.2 (NotI)	TCGAGCGGCCGCCCTGATTGAAGCCCAGGCGAC	R	Cloning
K5 (EcoRI)	GGTGGAATTCATGGCGTCTAAGGACGTAGAAGAG	F	Cloning
K5 (NotI)	TCGAGCGGCCGCCCTCCACCGTTGTTTTTTGGATGATTTTTC	R	Cloning
K6 (EcoRI)	GGTGGAATTCATGGCCCCCGTCCACG	F	Cloning
K6 (NotI)	TCGAGCGGCCGCCCCGAAGCTATGGCAGGCAGC	R	Cloning
K7 (EcoRI)	GGTGGAATTCATGGGAACACTGGAGATAAAAGGGG	F	Cloning
K7 (NotI)	TCGAGCGGCCGCCCCGACAACTGGCCTGGAGATTG	R	Cloning
K8α (kBZIP)	AGTGTGGTGGAATTCATGCCCAGAATGAAGGACATACCT	F	Cloning

(EcoRI)			
K8α (kBZIP) (NotI)	CCCTCGAGCGGCCGCCCTGAACATGGTGGGAGTGG C	R	Cloning
K8β (EcoRI)	GGTGGAATTCATGCCCAGAATGAAGGACATACCT	F	Cloning
K8β (NotI)	TCGAGCGGCCGCCCCGATACCTGCTGCAGCTGTCT	R	Cloning
K8.1 (BamHI)	GCTCGGATCCATGAGTTCCACACAGATTCGCAC	F	Cloning
K8.1 (NotI)	TCGAGCGGCCGCCCTGACACTATGTAGGGTTTCTT ACGCC	R	Cloning
K9 (EcoRI)	GGTGGAATTCATGGACCCAGGCCAAAGACC	F	Cloning
K9 (NotI)	TCGAGCGGCCGCCCTGATTGCATGGCATCCCATAA CG	R	Cloning
K10 (EcoRI)	GGTGGAATTCATGGGGTCCTCTGGGACG	F	Cloning
K10 (NotI)	TCGAGCGGCCGCCCTCCTGTAGACTATCCCAAATG GAGC	R	Cloning
K10.5 (EcoRI)	GGTGGAATTCATGTACCACGTGGGACAGGAG	F	Cloning
K10.5 (NotI)	TCGAGCGGCCGCCCTGAGTCATCACATGTAACTGA ACGCA	R	Cloning
K10.6 (EcoRI)	GGTGGAATTCATGGCGGGACGCAGG	F	Cloning
K10.6 (NotI)	TCGAGCGGCCGCCCTGACCTTGGTCTTCTCCGATGC	R	Cloning
K11 (EcoRI)	GGTGGAATTCATGCACAGTTTGTTTTTTGAAGAGC C	F	Cloning
K11 (NotI)	TCGAGCGGCCGCCCTGAGTCTCTGTGGTAAAATGG GGC	R	Cloning
K11.1 (EcoRI)	GGTGGAATTCATGCCTCGCTACACGGAGTC	F	Cloning
K11.1 (NotI)	TCGAGCGGCCGCCCTGAGTCTCTGTGGTAAAATGG GGC	R	Cloning
K13 (vFLIP) (EcoRI)	GGTGGAATTCATGGCCACTTACGAGGTTCTCTG	F	Cloning
K13 (vFLIP) (NotI)	TCGAGCGGCCGCCCCGATGGTGTATGGCGATAGTG TTG	R	Cloning
K14 (vOX2) (EcoRI)	GGTGGAATTCATGATACACACATTTTTTGATTGTC CCGG	F	Cloning

K14 (vOX2) (NotI)	TCGAGCGGCCGCCCTCCCTGGGTGGATAGGGG	R	Cloning
ORF2 (EcoRI)	GGTGGAATTCATGGATCCTACACTTTACTGTGTAGTTGC	F	Cloning
ORF2 (NotI)	TCGAGCGGCCGCCCTGACGAAGTCTCACTGAAGGGC	R	Cloning
ORF4 (EcoRI)	GGTGGAATTCATGGCCTTTTTAAGACAAACACTGTGG	F	Cloning
ORF4 (NotI)	TCGAGCGGCCGCCCCGAACGAAAGAACAGATAGTGAAATAAGGTAATCA	R	Cloning
ORF6 (SSB) (BamHI)	GCTCGGATCCATGGCGCTAAAGGGACCACA	F	Cloning
ORF6 (SSB) (NotI)	TCGAGCGGCCGCCCCGACAAATCCAGGTCAGAGAGCA	R	Cloning
ORF7 (EcoRI)	GGTGGAATTCATGGCAAAGGAACTGGCGG	F	Cloning
ORF7 (NotI)	TCGAGCGGCCGCCCCGAGACCTGGGAGTCATTGTGG	R	Cloning
ORF8 (gB) (EcoRI)	AAGAATTCATGACTCCCAGGTCTAGATTGGC	F	Cloning
ORF8 (gB) (NotI)	TCGAGCGGCCGCGCTCCCTCCCCCGTTTCCG	R	Cloning
ORF9 (Pol) (EcoRI)	GGTGGAATTCATGGATTTTTTCAATCCATTTATCGACCCAAC	F	Cloning
ORF9 (Pol) (NotI)	TCGAGCGGCCGCCCCGAGGGCGTGGGAAAAGTC	R	Cloning
ORF10 (RIF) (EcoRI)	GGTGGAATTCATGCAGACAGAGGCAACGTTC	F	Cloning
ORF10 (RIF) (NotI)	TCGAGCGGCCGCCCTCCCGATTGCATGGGTTCCT	R	Cloning
ORF11 (AflII)	TAAACTTAAGATGGCGCAGGAGTCAGAGC	F	Cloning
ORF11 (NotI)	TGCAGAATTCCGAACTGCGTCCGGTGG	R	Cloning
ORF16 (EcoRI)	GGTGGAATTCATGGACGAGGACGTTTTGCCT	F	Cloning

ORF16 (NotI)	TCGAGCGGCCGCCCTGATCTCCTGCTCATCGCGAC	R	Cloning
ORF17 (Protease) (EcoRI)	GGTGGAATTCATGAGCCTCCTAAGCCCCG	F	Cloning
ORF17 (Protease) (NotI)	TCGAGCGGCCGCCCCGACTGCTTGTTCAGGAGCTC	R	Cloning
ORF17.5 (Scaffold) (EcoRI)	GGTGGAATTCATGAACAGCTCTGGTCAAGAGGAT	F	Cloning
ORF17.5 (Scaffold) (NotI)	TCGAGCGGCCGCCCCGACTGCTTGTTCAGGAGCTC	R	Cloning
ORF18 (EcoRI)	GGTGGAATTCATGCTCGGAAAATACGTGTGTGAGA	F	Cloning
ORF18 (NotI)	TCGAGCGGCCGCCCTGAAACCGCGTTGTTGTTAAACG	R	Cloning
ORF20 (EcoRI)	GGTGGAATTCATGTACGAGGTTTTTACAGACTTTCCC	F	Cloning
ORF20 (NotI)	TCGAGCGGCCGCCCTCCTGGACCTGAACAAGCCG	R	Cloning
ORF21 (TK) (EcoRI)	GGTGGAATTCATGGCAGAAGGCGGTTTTGG	F	Cloning
ORF21 (TK) (NotI)	TCGAGCGGCCGCCCCGAGACCCTGCATGTCTCCT	R	Cloning
ORF22 (gH) (BamHI)	GCTCGGATCCATGCAGGGTCTAGCCTTCTTGG	F	Cloning
ORF22 (gH) (NotI)	TCGAGCGGCCGCCCCGAATAAAGGATGGAAAACAGTCTGTAAAGAAA	R	Cloning
ORF23 (BamHI)	GCTCGGATCCATGTTACGAGTTCCGGACGTGA	F	Cloning
ORF23 (NotI)	TCGAGCGGCCGCCCTGAGACGGTCAATAAAGCGTAGATTTTTAAAAG	R	Cloning
ORF24 (BamHI)	GCTCGGATCCATGGCAGCGCTCGAGGG	F	Cloning
ORF24 (NotI)	TCGAGCGGCCGCCCTGAGACCAGCGGACGGAC	R	Cloning
ORF25 (MCP)	AGTGTGGTGGAATTCATGGAGGCGACCTTGGAGC	F	Cloning

(EcoRI)			
ORF25 (MCP) (NotI)	CTCCCTCGAGCGGCCGCGGCGAATACACCACCTTG TTTCC	R	Cloning
ORF26 (EcoRI)	GGTGGAATTCATGGCACTCGACAAGAGTATAGTGG	F	Cloning
ORF26 (NotI)	TCGAGCGGCCGCCCTGAGCGTGGGGAATACCAACA	R	Cloning
ORF27 (EcoRI)	GGTGGAATTCATGGCGTCATCTGATATTCTGTCGG	F	Cloning
ORF27 (NotI)	TCGAGCGGCCGCCCTGATTTAAAATTTAGAATCAA GGGAGGGGTG	R	Cloning
ORF28 (EcoRI)	GGTGGAATTCATGAGCATGACTTCCCCGTCT	F	Cloning
ORF28 (NotI)	TCGAGCGGCCGCCCCGAATCTGGCATGTATATTGT ACGGTAGG	R	Cloning
ORF29a (EcoRI)	GGTGGAATTCATGCTGCTCAGCCGTCAC	F	Cloning
ORF29a (NotI)	TCGAGCGGCCGCCCTGAAGGCCCTGGGCTTACG	R	Cloning
ORF29b (EcoRI)	GGTGGAATTCATGCTTCAGAAAGACGCCAAGC	F	Cloning
ORF29b (NotI)	TCGAGCGGCCGCCCTGATTGTGGGGATATGGGCTT GT	R	Cloning
ORF30 (EcoRI)	GGTGGAATTCATGGGTGAGCCAGTGGATCC	F	Cloning
ORF30 (NotI)	TCGAGCGGCCGCCCTCCTTTCGCACCGGTGTCTAG	R	Cloning
ORF31 (EcoRI)	GGTGGAATTCATGTCACAAAACAGAAAGACTCTGC C	F	Cloning
ORF31 (NotI)	TCGAGCGGCCGCCCCGACGTATCTTTCGTTGATAG CATGC	R	Cloning
ORF32 (EcoRI)	GGTGGAATTCATGGATGCGCATGCTATCAACG	F	Cloning
ORF32 (NotI)	TCGAGCGGCCGCCCCGAGCCATAGCGGCCTCG	R	Cloning
ORF33 (BamHI)	GCTCGGATCCATGGCTAGCCGGAGGCG	F	Cloning
ORF33 (NotI)	TCGAGCGGCCGCCCTCCATAAGAACGTAAGCCCAG GG	R	Cloning
ORF34 (EcoRI)	GGTGGAATTCATGTTTGCTTTGAGCTCGCTCG	F	Cloning
ORF34 (NotI)	TCGAGCGGCCGCCCTGAGAGTTGGTTGAGTCCATT CTCC	R	Cloning
ORF35	GCTCGGATCCATGGACTCAACCAACTCTAAAAGAG	F	Cloning

(BamHI)	AGTT		
ORF35 (XhoI)	CTCCCTCGAGTGAGGGAGTTTCAGGGCACA	R	Cloning
ORF36 (EcoRI)	GGTGGAATTCATGCGCTGGAAGAGAATGGAGAG	F	Cloning
ORF36 (NotI)	TCGAGCGGCCGCCCTGAGAAAACAAGTCCGCGGG	R	Cloning
ORF37 (SOX) (EcoRI)	GGTGGAATTCATGGAGGCCACCCCCAC	F	Cloning
ORF37 (SOX) (NotI)	TCGAGCGGCCGCCCCGACGGGCTGTGAGGGA	R	Cloning
ORF38 (EcoRI)	GGTGGAATTCATGGGATTTCTCCTATCTATCTGCA AACG	F	Cloning
ORF38 (NotI)	CTCGAGCGGCCGCCCTGAATAAATTGCTTCTTTAT TTTTTTTCTTCTTTTTTAATGCG	R	Cloning
ORF39 (gM) (EcoRI)	GGTGGAATTCATGCGCGCTTCAAAGAGCG	F	Cloning
ORF39 (gM) (NotI)	TCGAGCGGCCGCCCCGAAATGAATATCATTTGCGT TTCGTCGAT	R	Cloning
ORF40 (PAF) (EcoRI)	GGTGGAATTCATGGCAACGAGCGAAGAAACG	F	Cloning
ORF40 (PAF) (NotI)	TCGAGCGGCCGCCCTCCAGCAGGGACAGTAGGTC	R	Cloning
ORF41 (PAF) (EcoRI)	GGTGGAATTCATGGCCGGGTTTACTCTGAAGG	F	Cloning
ORF41 (PAF) (NotI)	TCGAGCGGCCGCCCTCCAAATAAAGATAAAAGCCT GGTCCA	R	Cloning
ORF42 (EcoRI)	GGTGGAATTCATGTCCCTGGAAAGGGCCC	F	Cloning
ORF42 (NotI)	TCGAGCGGCCGCCCTGATTTTGAAAAAAGGGAAAC AATGGGG	R	Cloning
ORF43 (portal) (EcoRI)	GGTGGAATTCATGTTGAGGATGAACCCGGGG	F	Cloning
ORF43 (portal) (NotI)	TCGAGCGGCCGCCCCGATGCACTTCCAGGACAAGG	R	Cloning

ORF44 (HEL) (KpnI)	GCTTGGTACCATGGACAGCTCGGAAGGGTG	F	Cloning
ORF44 (HEL) (NotI)	TCGAGCGGCCGCCCTCCGTAGATCAGAGTAGTCTTGGGG	R	Cloning
ORF45 (EcoRI)	GGTGGAATTCATGGCGATGTTTGTGAGGACCT	F	Cloning
ORF45 (NotI)	TCGAGCGGCCGCCCTCCGTCCAGCCACGGC	R	Cloning
ORF46 (EcoRI)	GGTGGAATTCATGGACGCATGGTTGCAACAG	F	Cloning
ORF46 (NotI)	TCGAGCGGCCGCCCTGACTGCTCCAACAGGCCC	R	Cloning
ORF47 (gL) (EcoRI)	GGTGGAATTCATGGGGATCTTTGCGCTATTTGC	F	Cloning
ORF47 (gL) (NotI)	TCGAGCGGCCGCCCTGATTTTCCCTTTTGACCTGCGTG	R	Cloning
ORF48 (EcoRI)	GGTGGAATTCATGGAGGTGTGTATCCCAATTCCG	F	Cloning
ORF48 (NotI)	TCGAGCGGCCGCCCTCCATCATACTCATCGTCGGAGC	R	Cloning
ORF49 (EcoRI)	GGTGGAATTCATGACATCGAGAAGGCCCCTTAAAG	F	Cloning
ORF49 (NotI)	TCGAGCGGCCGCCCTGATTGTATACTGAACAATGCGTGTTTACAAT	R	Cloning
ORF50 (RTA) (EcoRI)	AGTGTGGTGGAATTCATGAAAGAATGTTCCAAGCTTGGTGC	F	Cloning
ORF50 (RTA) (NotI)	CCCTCGAGCGGCCGCCCTCCGTCTCGGAAGTAATTACGCC	R	Cloning
ORF52 (EcoRI)	GGTGGAATTCATGGCCGCGCCCAGG	F	Cloning
ORF52 (XhoI)	CTCCCTCGAGTCCGTCATCAACCCCCGC	R	Cloning
ORF53 (gN) (EcoRI)	GGTGGAATTCATGACAGCGTCCACGGTGG	F	Cloning
ORF53 (gN) (NotI)	TCGAGCGGCCGCCCCGATGCATGGACCACCTCG	R	Cloning
ORF54	GGTGGAATTCATGAACAACCGCCGAGGC	F	Cloning

(dUTPase) (EcoRI)			
ORF54 (dUTPase) (NotI)	TCGAGCGGCCGCCCCGAAAACCCAGACGACCCC	R	Cloning
ORF55 (EcoRI)	GGTGGAATTCATGTCGTCTCCATGGTACACCTG	F	Cloning
ORF55 (NotI)	TCGAGCGGCCGCCCCGATGTCGAACCTATCGCGC	R	Cloning
ORF56 (PRI) (EcoRI)	GGTGGAATTCATGGAGACGACATACCGCCG	F	Cloning
ORF56 (PRI) (NotI)	TCGAGCGGCCGCCCTGAACTGGCCAGTCCCACT	R	Cloning
ORF57 (MTA) (EcoRI)	AGTTGAATTCATGGTACAAGCAATGATAGACATGGACA	F	Cloning
ORF57 (MTA) (NotI)	ATTGCGGCCGCTGAAGAAAGTGGATAAAAGAATAAACCCTTGTTAAATTTGG	R	Cloning
ORF58 (EcoRI)	GGTGGAATTCATGTGCCGCCTGGACAGT	F	Cloning
ORF58 (NotI)	TCGAGCGGCCGCCCTGAGCCAACAACTTTATTTATTACCGACAG	R	Cloning
ORF59 (PPF) (EcoRI)	GGTGGAATTCATGCCTGTGGATTTTCACTATGGG	F	Cloning
ORF59 (PPF) (NotI)	TCGAGCGGCCGCCCTCCAATCAGGGGGTTAAATGTGGT	R	Cloning
ORF60 (EcoRI)	GGTGGAATTCATGGATTCAGTTGATCGATTTCTGTATACAAG	F	Cloning
ORF60 (NotI)	TCGAGCGGCCGCCCTCCCAAATCGTCAGTCACACAC	R	Cloning
ORF61 (EcoRI)	GGTGGAATTCATGTCTGTCCGGACATTTTGTCAG	F	Cloning
ORF61 (NotI)	TCGAGCGGCCGCCCCGACTGACAGACCAGGCACT	R	Cloning
ORF62 (EcoRI)	GGTGGAATTCATGAAGGTGCAGGCTGAAAATGC	F	Cloning
ORF62 (NotI)	TCGAGCGGCCGCCCTGACAGAAACACAGTCCAGGGG	R	Cloning
ORF63	TACCGAGCTCGGATCCATGGACGGCACAGACGCT	F	Cloning

(BamHI)			
ORF63 (NotI)	CTCCCTCGAGCGGCCGCCCCGATTCGACAAACAGTTTCCG	R	Cloning
ORF64 (EcoRI)	AGTGTGGTGGAATTCATGGCAGCCCAGCCTCT	F	Cloning
ORF64 (NotI)	CTCCCTCGAGCGGCCGCCCTCCCAAGTACCACTTCTTTATTCTGTCA	R	Cloning
ORF65 (KpnI)	GCTTGGTACCATGTCCAACTTTAAGGTGAGAGACCC	F	Cloning
ORF65 (XhoI)	CTCCCTCGAGCGATTTCTTTTTGCCAGAGGGGG	R	Cloning
ORF66 (EcoRI)	GGTGGAATTCATGGCCCTGGATCAGCG	F	Cloning
ORF66 (NotI)	TCGAGCGGCCGCCCTCCGGAGGAACACTTCCCG	R	Cloning
ORF67 (EcoRI)	GGTGGAATTCATGAGTGTCGTTGGTAAGCGTGTAG	F	Cloning
ORF67 (NotI)	TCGAGCGGCCGCCCTCCGCTGGGCCTCATCC	R	Cloning
ORF67.5 (EcoRI)	GGTGGAATTCATGGAGTACGCGTCTGACCAG	F	Cloning
ORF67.5 (NotI)	TCGAGCGGCCGCCCTCCGGGCCGTGCCC	R	Cloning
ORF68 (BamHI)	GCTCGGATCCATGTCACGAGGCAGAAGCTGG	F	Cloning
ORF68 (XhoI)	CTCCCTCGAGTCCAGCGTACAAGTGTGACGTC	R	Cloning
ORF69 (EcoRI)	GGTGGAATTCATGGAGACCCCGATATGCACC	F	Cloning
ORF69 (NotI)	TCGAGCGGCCGCCCTGATAGGGCGTTGACAAGTGC	R	Cloning
ORF70 (EcoRI)	GGTGGAATTCATGTTTCCGTTTGTACCTTTAAGCTTGT	F	Cloning
ORF70 (NotI)	TCGAGCGGCCGCCCCGATACTGCCATTTCCATACGAATGG	R	Cloning
ORF74 (vGPCR) (NotI)	GAAGGGGGCGGCCGCAATGGCGGCCGAGGATTTC	F	Cloning
ORF74 (vGPCR) (NotI)	TAGACTCGAGCGGCCGCCTACGTGGTGGCGCCG	R	Cloning
ORF75 (FGARAT) (BamHI)	TACCGAGCTCGGATCCATGTTTGTCAGCTTTTATACAAGCCGTACG	F	Cloning

ORF75 (FGARAT) (NotI)	CTCCCTCGAGCGGCCGCCCTGAGTGGTGGTCGTTGATCTTCT	R	Cloning
K8.1 WT	CCGGCAGCAATATAAAAGGGACC	F	EMSA
K8.1 WT	GGTCCCTTTTATATTGCTGCCGG	R	EMSA
K8.1 TATT:CCCC	CCGGCAGCAACCCCAAAGGGACC	F	EMSA
K8.1 TATT:CCCC	GGTCCCTTTGGGGTTGCTGCCGG	R	EMSA
ORF57 WT	TTAATCCCACTATATAACCTGGCT	F	EMSA
ORF57 WT	AGCCAGGTTATATAGTGGGATTAA	R	EMSA
ORF24 stop	CCGAGCGCCTCCCTGACGACGAGTCCGCAGACCACGTGTTAACAAGCGACTTGGGAATCAAGGATGACGACGATAAGTAGGG	F	BAC mutagenesis
ORF24 stop	CAGCAGAATATTTCCAGCTGTGATTCCCAAGTCGCTTGTTAACACGTGGTCTGCGGACTCAACCAATTAACCAATTCTGATTAG	R	BAC mutagenesis
ORF24 N FLAG	GGCGAGGTACGGGAAAGGTCGTTGCTCCAAGGTCGCCTCCATGGACTACAAAGACGATGACGACAAGATGGCAGCGCTCGAGGGCCCAGGATGACGACGATAAGTAGGG	F	BAC mutagenesis
ORF24 N FLAG	CGCTCGGTGGCAGTAGTAGGGGGCCCTCGAGCGCTGCCATCTTGTCGTCATCGTCTTTGTAGTCCATGGAGGCGACCTTGGAGCAACAACCAATTAACCAATTCTGATTAG	R	BAC mutagenesis
K8.1 Pr	GGGAGAACCATGCCAGACTTTG	F	qPCR
K8.1 Pr	GCATAGGATTAGGAGCGCCAC	R	qPCR
ORF52 ORF	AAATCGAAGCCAGGGTCAGG	F	qPCR
ORF52 ORF	CTCCTCTTCGTCGCCTGTTATTG	R	qPCR
ORF57 Pr	ATTAGGGTGAGCGAAGTCACG	F	qPCR
ORF57 Pr	GCTTGTACCATGTCCTTTGGTT	R	qPCR
ORF57 ORF	GGTGTGTCTGACGCCGTAAAG	F	qPCR
ORF57 ORF	CCTGTCCGTAAACACCTCCG	R	qPCR
ORF38 ORF	TATCTGCAAACGTCCCTCAC	F	qPCR

ORF38 ORF	CATCCCTCTTCCTCCATCCC	R	qPCR, Northern blot
K8.1 ORF	CCGTCGGTGTGTAGGGATAAAG	F	qPCR
K8.1 ORF	GTCGTTGTAGTGGTGGCAGAAA	R	qPCR
GAPDH Pr	TACTAGCGGTTTTACGGGCG	F	qPCR
GAPDH Pr	TCGAACAGGAGGAGCAGAGAGCGA	R	qPCR
ORF24 N-terminus (BamHI)	GCTCGGATCCATGGCAGCGCTCGAGGG	F	Cloning
ORF24 N-terminus (NotI)	TCGAGCGGCCGCTTGCTGCCCAGAGTCCGCG	R	Cloning
ORF24 central domain (BamHI)	GCTCGGATCCATGTCTACAGTGTCAGACATGCTGGAAC	F	Cloning
ORF24 central domain (NotI)	TCGAGCGGCCGCTTGTTTCTGTGCTCAAGTAGGGAGAATATTCT	R	Cloning
ORF24 C-terminus (BamHI)	GCTCGGATCCATGCGTTCCCAAATACAGACGCTACACA	F	Cloning
ORF24 C-terminus (NotI)	TCGAGCGGCCGCAAGACCAGCGGACGGACG	R	Cloning
N425A, N427A ORF24	AGCTTATCTTAGTGTTAATCACCATGGCGGTGGCGAATATGGTGATCTTGAAGTTTTCCAAACTG	F	Site-directed mutagenesis
N425A, N427A ORF24	CAGTTTGGAAAACTTCAAGATCACCATATTCGCCACCGCCATGGTGATTAACACTAAGATAAGCT	R	Site-directed mutagenesis
N518A, N520A ORF24	TCCGTGCAGGTCCAACGTGGAGGCCCAAGCTGCTATCGATACAGGAAATATG	F	Site-directed mutagenesis

N518A, N520A ORF24	CATATTTCCTGTATCGATAGCAGCTTGGGCCTCCACGTTGGACCTGCACGGA	R	Site-directed mutagenesis
5aa ORF24	CCTTTTTCGCCTGCAGAGCGATACGCGCAGCAGCCGCGGCGGAACGCCTCCACCCTTTTATACATC	F	Site-directed mutagenesis
5aa ORF24	GATGTATAAAAGGGTGGAGGCGTTCCGCCGCGGCTGCTGCGCGTATCGCTCTGCAGGCGAAAAAGG	R	Site-directed mutagenesis
3aa ORF24	TTTTTCGCCTGCAGAGCGATACGCAGAGCAGCCGCGGGGGAACGCCTCC	F	Site-directed mutagenesis
3aa ORF24	GGAGGCGTTCCCCCGCGGCTGCTCTGCGTATCGCTCTGCAGGCGAAAAA	R	Site-directed mutagenesis
UL87 (EcoRI)	AAAAGAATTCATGGCCGGCGCTGC	F	Cloning
UL87 (XhoI)	AAAACTCGAGTCGTGATGCAAACCGCAC	R	Cloning
BcRF1 (EcoRI)	GGTGGAATTCATGACACAAGGTAAGAGGGAGATGG	F	Cloning
BcRF1 (NotI)	TCGAGCGGCCGCGCCACTTGAGCATCACGGCAGTG	R	Cloning
mORF24 (BamHI)	TACCGAGCTCGGATCCATGACAATATTCTTACCAGTATTCTGTGATTTGC	F	Cloning
mORF24 (NotI)	CTCCCTCGAGCGGCCGCCCGGAGTCTGGTTGGCAAGG	R	Cloning
ORF8	CCTTCTTGGTGGTACTTGTCTTTGG	R	Northern probe
ORF65	GTTATACTGCTTCCCGAGACCC	R	Northern probe
ORF25	GAGGTGTGCCTTGAACATGGATA	R	Northern probe
K8.1	GTCGTTGTAGTGGTGGCAGAAA	R	Northern probe
7SL	GCTCCGTTTCCGACCTGGGCC	R	Northern probe

Acknowledgements

We would like to thank the members of the Glaunsinger and Krogan labs and Laurent Coscoy for helpful discussions and critical reading of the manuscript. We would also like to thank Christopher Hann-Soden, Alexandre Mercier, Trever Greene, Andrew Birnberg, Matthew Gardner, Marta Gaglia and Edward Tang for help with cloning, cell culture, and mass spectrometry; Jeffrey Queisser for computational help; Sumit Chanda for ORFeome constructs.

This research was supported by NIH R01 grants CA160556 and CA136367 to BAG, NIH P50 GM082250, P50 GM081879, PO1 AI090935, PO1 AI091575, PO1 CA177322, U54 AI081680 and DARPA-10-93-Prophecy-PA-008 grants to NJK, R01HG007644 and a Sloan Foundation Research Fellowship to RDH, an Irving H. Wiesenfeld Fellowship and a CRCC fellowship to ZHD, a UCSF M. Lloyd Kozloff fellowship and NIH F31 CA180609-01 to MCM. NJK and BG are WM Keck Foundation Young Investigators.

Chapter 3 - KSHV vTFs form a complex essential for late gene expression

Introduction

Kaposi's sarcoma-associated herpesvirus (KSHV) is a γ-herpesvirus and the etiologic agent of primary effusion lymphoma (PEL), multicentric Castleman's disease, and the AIDS-associated neoplasm Kaposi's sarcoma. Like all herpesviruses, KSHV genes are expressed in a temporally regulated, ordered cascade. The first two phases are the immediate early and delayed early genes, which encode gene products that modulate the host and are essential for viral replication. Late genes largely encode proteins required for assembly, maturation and release of infectious virions, and their expression occurs after the onset of viral genome replication.

Evidence suggests that the set of viral factors controlling late gene expression are conserved between the β and γ-herpesviruses. Studies in the related γ-herpesviruses MHV-68 and Epstein-Barr virus (EBV) as well as in the β-herpesvirus human cytomegalovirus (HCMV) have shown that a group of six viral transactivation factors (vTFs) (ORFs 18, 24, 30, 31, 34, 66; MHV68/KSHV nomenclature) are necessary for late gene expression (Arumugaswami et al., 2006; Aubry et al., 2014; Isomura et al., 2011; Jia et al., 2004; Perng et al., 2014; Wong et al., 2007; Wu et al., 2009).

It remains unclear how viral and host factors coordinate to orchestrate the robust induction of late genes at the end of the infectious cycle, a period which coincides with significant cellular stress. All of the mapped late promoters in γ-herpesviruses are extremely minimal (12-15 base pairs), lacking any known response elements to viral or cellular transcriptional transactivators (Tang et al., 2004; Wong-Ho et al., 2014). Late gene promoters contain little more than a TATT variant of the canonical TATA box, flanked by the TFIIB recognition element (BRE). During canonical transcription, the cellular TATA-box binding protein (TBP) serves to nucleate the assembly of the remaining constituents of the transcription pre-initiation complex. However, in β and γ-herpesviruses this function is replaced by the vTF ORF24 in KSHV (BcRF1 in EBV, UL87 in HCMV) (Davis et al., 2015). The central portion of ORF24 displays predicted structural mimicry to TBP, through which it binds the late promoter TATT boxes (Davis et al., 2015). The central portion of ORF24 displays predicted structural mimicry to TBP, through. ORF24 also contains an N-terminal domain that interacts with RNAP II to bring the polymerase to these promoters (Aubry et al., 2014; Davis et al., 2015). At present, how the other vTFs contribute to organization and function of the viral transcription pre-initiation complex is largely unknown, although the HCMV UL79 vTF (KSHV ORF18) has been shown to promote RNAP II elongation (Perng et al., 2014). The six EBV vTFs form a complex, suggesting that interactions between these viral proteins are likely to be important for late gene transcription. Indeed, this has recently been shown to be true for ORF31 and ORF34 of KSHV, which associate with ORF24 at promoters (Brulois et al., 2015).

Here, we defined the architecture of the KSHV late gene vTF complex by identifying all pairwise interactions between the six vTFs. Our resulting model

reveals a highly interconnected set of interactions between several of the vTF components, although the TATT-box binding protein ORF24 exhibits a single vTF-vTF interaction with ORF34. We mapped this ORF24-ORF34 interaction interface to a single amino acid on ORF24, and demonstrated that it plays an essential role in late gene expression and infectious virion production. Finally, we confirmed that all six KSHV vTFs are necessary and sufficient to activate expression of a late gene promoter.

Results

KSHV ORF24 Interacts With RNAP II and ORF34

Aside from ORF24, little is known about how each of the vTFs contributes to the overall architecture of the late gene transcription pre-initiation complex. We first sought to determine whether ORF24 was the only KSHV late gene vTF capable of binding RNAP II or, as has been suggested for the β-herpesvirus HCMV (Perng et al., 2014), other vTFs may also bind the polymerase. Strep-tagged versions of each of the vTFs were transfected into HEK293T cells, affinity purified over Strep-Tactin beads, and the level of co-purifying endogenous RNAP II was evaluated by Western blotting for the RPB1 polymerase subunit (Figure 3.1A). While an interaction was observed for ORF24 and RNAP II, no interaction between the other five vTFs and RNAP II was detected (Figure 3.1A). It should be noted that ORF30 is a very small protein (77 amino acids) and only weakly expressed, making it difficult to detect. Nevertheless, these data are in agreement with the KSHV-human protein-protein interaction network, which similarly found that ORF24 was the only KSHV protein that bound RNAP II (Davis et al., 2015).

The six vTFs have been suggested to form a complex, suggesting that they all participate in the assembly of the β- and γ-herpesviral late promoter transcription pre-initiation complex (vPIC) (Arumugaswami et al., 2006; Aubry et al., 2014; Jia et al., 2004; Wong et al., 2007; Wu et al., 2009). How this complex is organized, as well as the importance of many of the individual vTF protein-protein interactions remain unknown. To begin to dissect the organization of the vTF complex in KSHV, we examined all possible pairwise interactions between each of the six vTFs using an orthogonal interaction detection method called HT-GPCA (high throughput *Gaussia princeps* complementation assay) (Cassonnet et al., 2011). HT-GPCA is based on the reconstitution of a split luciferase, and its power lies in the technically straightforward ability to confirm that a set of interactions occurs within an intact cell, as opposed to post-lysis when all soluble proteins are brought in proximity in solution. Bait and prey proteins are each fused to a half of the *Gaussia princeps* luciferase using Gateway cloning and transfected into mammalian cells. A direct interaction brings the luciferase fragments in close proximity, thereby restoring enzymatic activity and producing a luciferase signal (Figure 3.1B). In previous benchmarking assays using a reference set of protein pairs, this methodology was shown to known to outperform other protein-protein interaction detection assays, in that it exhibited high true positive and low false positive detection rates (Cassonnet et al., 2011).

For this assay, each of the pairs of vTFs being examined for interaction were fused with *Gaussia* luciferase fragment 2 (Gluc2) and fragment 1 (Gluc1). All pairwise combinations of vTFs were transfected into HEK293T cells and individual protein-protein interactions were assessed by measuring luciferase activity. Interactions were detected between ORF18 and ORF30, ORF18 and ORF31, ORF18 and ORF34, ORF18 and ORF66, ORF24 and ORF34, ORF31 and ORF34, ORF31 and ORF66, and ORF34 and ORF66 (Figure 3.1C-D). The pairwise interaction profiles are consistent with the formation a complex comprising all six vTFs. ORF34 interacts with four of the five other vTFs, suggesting that it may act as an important scaffold protein in bringing the vPIC complex to ORF24 and in turn to late gene promoters.

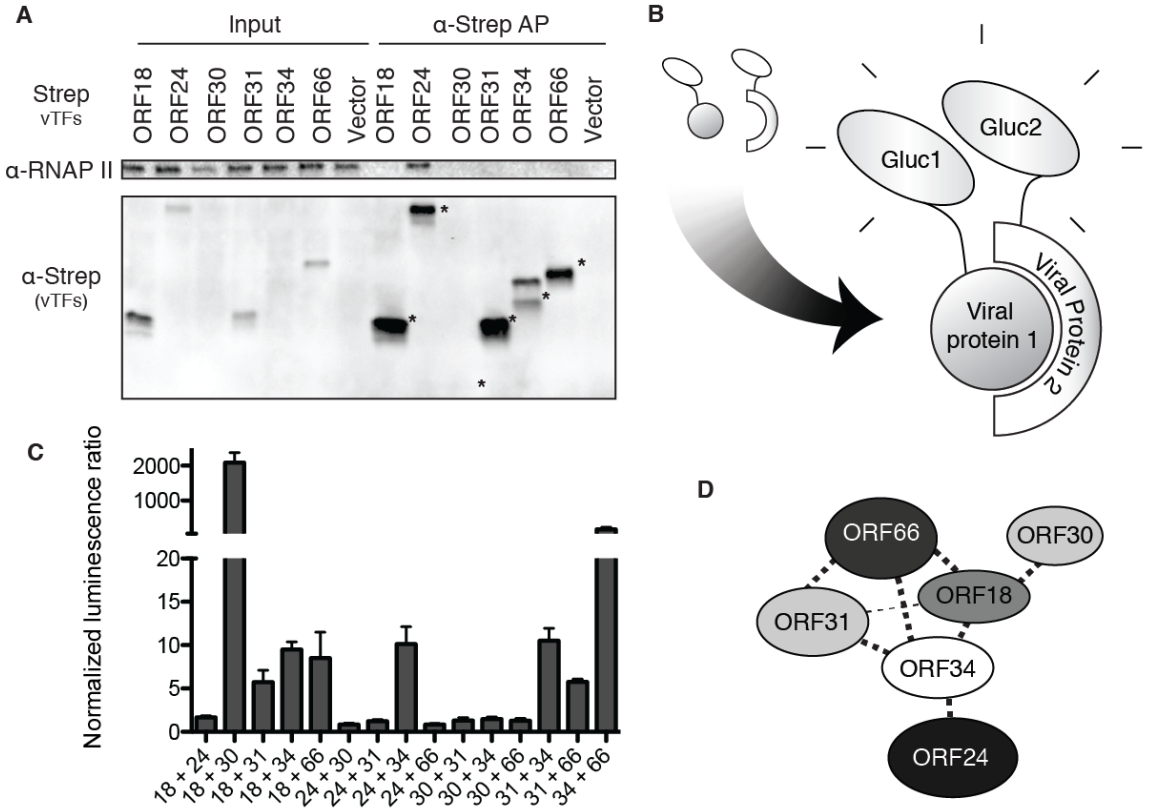

Figure 3.1: vTFs interact to form the vPIC.
(A) ORF24 is the only vTF to interact with RNAP II. Strep-tagged viral ORFs were transfected into 293T cells and lysates were affinity purified over Strep-Tactin beads a d Western blotted for strep and RNAP II. * indicates the strep-tagged viral protein. **(B)** The *Gaussia* split luciferase assay provides a quantitative measurement of protein interactions. The two halves of the *Gaussia* luciferase protein (Gluc1 and Gluc2) are fused to the viral proteins to be assessed for interaction. If the two viral proteins interact, they bring the two inactive *Gaussia* fragments into physical proximity, allowing for the formation of a functional luciferase enzyme that facilitates luminescence on addition of luciferin substrate. **(C)** The results of all pairwise vTF interactions by the split luciferase assay. Luminescence levels were measured and normalized to the signal of the Gluc1 and Gluc2-tagged proteins

transfected on their own. **(D)** Summary of interaction data from (C). Note that only ORF34 interacts with ORF24.

KSHV ORF34 is Essential for Late Gene Expression and Progeny Virion Production

Because ORF34 was the only vTF that interacted with ORF24, we sought to explore this part of the complex in more detail. Although ORF34 is required for late gene expression in MHV68, EBV, and HCMV, its role in KSHV late gene expression had not been confirmed (Aubry et al., 2014; Isomura et al., 2011; Wu et al., 2009). To determine whether ORF34 is similarly necessary for KSHV late gene expression, we engineered a stop codon into the KSHV ORF34 locus (KSHV.ORF34.Stop) to make a mutant virus unable to express ORF34 protein. ORF34 with a premature stop mutation 9 amino acids into the coding region was engineered using the KSHV BAC16 scarless recombination mutagenesis system (Brulois et al., 2012). Gel electrophoresis of the *Eco*RI-digested wildtype and KSHV.ORF34.Stop BAC16 DNA showed identical banding patterns, indicating that no gross rearrangements or deletions occurred during BAC16 mutagenesis (Figure 3.2A). The KSHV.ORF34.Stop BAC16 DNA was transfected into iSLK puro cells and selected with hygromycin to generate stable cell lines expressing the ORF34 mutant.

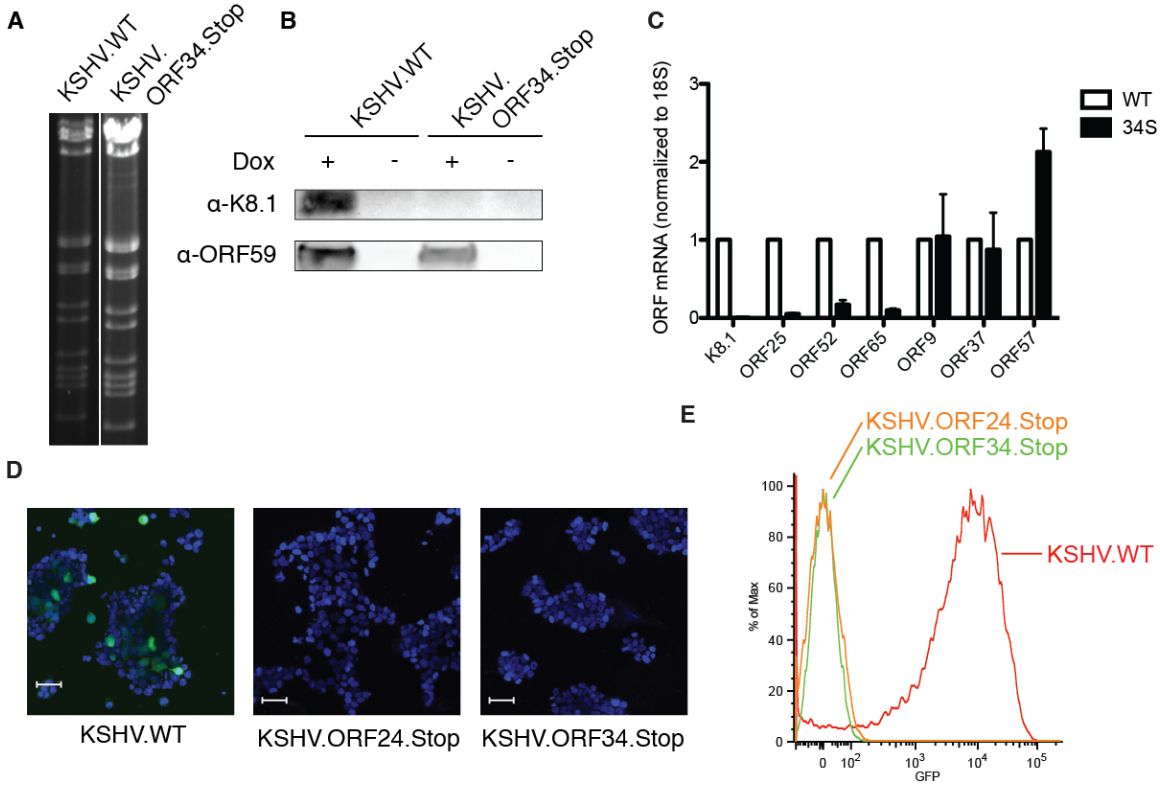

Figure 3.2: ORF34 is required for late gene expression and progeny virion production.

(A) No gross recombination events were detected in the KSHV.ORF34.Stop BAC DNA compared with KSHV.WT. BAC DNA was digested with EcoRI and run out on a 0.7% agarose TAE gel at 30V for 24 hr. **(B-C)** KSHV.WT and KSHV.ORF34.Stop-infected

iSLK cells were lytically reactivated for 72 hr. Protein lysates were collected and analyzed by Western blot for K8.1 (late) and ORF59 (delayed early) (B). RNA was collected and levels were analyzed for early (ORF9, ORF37, and ORF57) and late (K8.1, ORF25, ORF52, and ORF65) genes normalized to 18S. **(D-E)** KSHV.WT, KSHV.ORF24.Stop, and KSHV.ORF34.Stop-infected iSLK cells were lytically reactivated for 72 hr. Supernatant was added to 293T cells and 24 hr later, cells were monitored for GFP expression by confocal microscopy (D) or flow cytometry (E), indicating the transfer of infectious virus. In (D), DNA is stained with DAPI (blue). A minimum of 30 fields were monitored for each experiment, with a single representative field shown. No GFP positive cells were detected in any field of recipient cells containing supernatant from KSHV.ORF24.Stop or KSHV.ORF34.Stop-infected cells. Scale bar = 50 μm. KSHV.ORF24.Stop is shown as a negative control.

Individual clones were amplified and tested for their ability to produce early and late gene products at 72 hr post lytic reactivation using a combination of Western blotting and RT-qPCR. KSHV.ORF34.Stop was defective for production of the late gene products K8.1, ORF25, ORF65, and ORF52 but not the early gene products ORF59, ORF9, and ORF57 (Figure 3.2B-C). The failure of the KSHV.ORF34.Stop to express late genes was not due to a defect in viral genome replication, as viral DNA levels were equivalent between the wildtype and mutant virus as measured by qPCR (data not shown).

Late gene products generally comprise structural components of the virion, so we assessed whether the late gene expression defect of the KSHV.ORF34.Stop virus impacted the production of infectious virions. iSLK puro-derived cell lines were reactivated with sodium butyrate and doxycycline to induce the viral lytic cycle. After 72 hr, viral supernatant was transferred to uninfected HEK293T cells. These cells were then analyzed for the presence of GFP-positive KSHV BAC16 virus by confocal microscopy and flow cytometry. As a negative control, we included the KSHV.ORF24.Stop virus, which has been shown to have a late gene expression defect and consequently fails to generate progeny virions (Davis et al., 2015). Indeed, while cells lytically infected with wildtype KSHV produced virus (as measured by GFP+ cells in supernatant transfer assays), no infectious virion production was detected from cells containing either the KSHV.ORF34.Stop or the KSHV.ORF24.Stop viruses (Figure 3.2D-E). These data are consistent with an essential role of ORF34 in late gene expression, as has been demonstrated for several other γ- and β-herpesviruses (Isomura et al., 2011).

Identification of the ORF34 Binding Site on ORF24

Having established that ORF24 and ORF34 interact, and that both are necessary for late gene expression and subsequent virion production, we next wanted to determine the importance of this interaction for viral late gene expression. To this end, we mapped the region of ORF24 that binds ORF34 with the goal of identifying an ORF24 point mutant selectively defective for interaction with ORF34. FLAG-tagged ORF24 truncation mutants consisting of the N-terminus, TBP-like DNA-binding domain, or C-terminus were tested for their ability to interact with strep-

tagged ORF34 by affinity purification (AP) followed by Western blotting (Figure 3.3A). AP of strep-ORF34 led to co-purification of both full length ORF24 and the N-terminal domain of ORF24, but not the other regions of ORF24 (Figure 3.3B). As would be predicted based on the data in Figure 3.1A, endogenous RNAP II co-purified with ORF34 only in the presence of full length or N-terminal ORF24, both of which contain the RNAP II binding domain (Figure 3.3A-B). This confirms that ORF24 binds both ORF34 and RNAP II, and that the interaction between ORF34 and RNAP II occurs indirectly via ORF24.

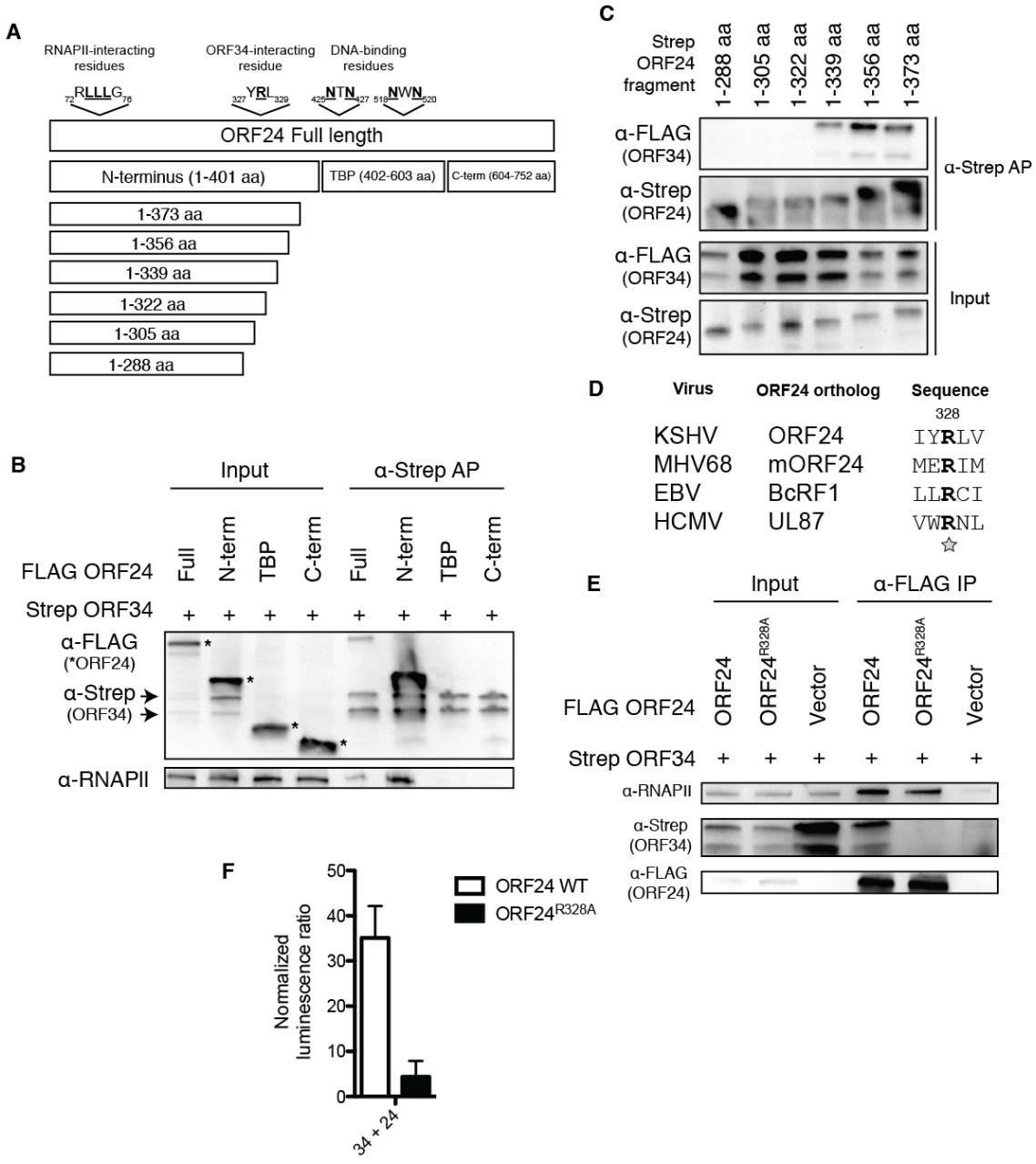

Figure 3.3: ORF24^{R328} mediates the interaction with ORF34.
(A) Map showing the defined functional regions of ORF24 as well as the truncation mutants tested. **(B)** FLAG full-length or ORF24 fragments (N-terminus, TBP-like

domain, or C-terminus) were co-transfected with strep ORF34 in 293T cells. Lysates were affinity purified over Strep-Tactin beads and Western blotted for FLAG (ORF24), strep (ORF34), and RNAP II. Input is 5% of lysate used for purification. **(C)** The N-terminal domain of ORF24 (amino acids 1-402) was further truncated to identify the interaction site with ORF34. Strep ORF24 fragments were co-transfected with FLAG ORF34 into 293T cells. Lysates were affinity purified as described in (B). **(D)** Alignment showing a conserved amino acid (R328 in KSHV ORF24) between ORF24 orthologs from β and γ-herpesviruses. This was the only conserved amino acid found in the 323-339 amino acid region. **(E)** FLAG ORF24 WT or ORF24^{R328A} was co-transfected with strep ORF34 into 293T cells. Lysate was immunoprecipitated with FLAG beads and Western blotted for strep (ORF34) and RNAP II. **(F)** Gaussia split luciferase assay with ORF24 WT and ORF24^{R328A}.

To further refine the interaction site, we created a panel of strep-tagged successive 17 amino acid truncation mutants of the ORF24 N-terminus and tested their ability to co-purify FLAG-ORF34. The minimal interaction site was encompassed by the fragment containing amino acids 1-339 (Figure 3.3C). Given that the ORF24 truncation mutant containing residues 1-322 was unable to interact with ORF34, we hypothesized that the 17 amino acid region of ORF24 containing residues 323-339 might contain critical residues for binding ORF34. We further reasoned that the residue(s) of ORF24 involved in the ORF34 interaction might be conserved across ORF24 orthologs from other γ- and β-herpesviruses that have also been described as essential for late gene expression. Alignment of the N-terminus of KSHV ORF24 with its orthologs UL87 from HCMV, BcRF1 from EBV and mORF24 from MHV68 revealed a single conserved residue (R328) between amino acids 323 and 339 (Figure 3.3D). Mutation of this residue to an alanine (FLAG-ORF24^{R328A}) abrogated the interaction with strep-ORF34, but did not impair the ability of ORF24 to bind RNAP II (Figure 3.3E). This confirms that while ORF24 interacts with both ORF34 and RNAP II via its N-terminus, the interaction interfaces are distinct and that the R328A mutant is selectively defective for ORF34 binding. We also fused ORF24^{R328A} to Gluc2 and analyzed its binding profile to Gluc1 tagged ORF34 using the HT-GPCA assay. This confirmed that its interaction with ORF34 was severely impaired compared to that of wildtype Gluc2-ORF24 (Figure 3.3F).

The ORF24-ORF34 Interaction is Necessary for KSHV Late Gene Expression

To determine the relevance of the interaction between ORF24 and ORF34 in the context of viral late gene expression, we engineered the ORF24^{R328A} mutation in the KSHV BAC and generated KSHV.ORF24^{R328A} stable iSLK cells. To ensure the integrity of the BAC DNA following mutagenesis, the restriction pattern of the KSHV.ORF24^{R328A} BAC was verified by comparison with wildtype BAC16 (Figure 3.4A). iSLK cells containing either wildtype KSHV or KSHV.ORF24^{R328A} were lytically reactivated for 48 hr, whereupon we measured the expression of the K8.1 (late) and ORF59 (delayed early) proteins (Figure 3.4B), the expression of early and late viral mRNAs (Figure 3.4C), as well as viral genome amplification (data not shown) and infectious virion production (Figure 3.4D-E). Similar to the KSHV.ORF34.Stop virus,

KSHV.ORF24^{R328A} was selectively defective for late gene expression, but had no defect in viral genome amplification (Figure 3.4B-C and data not shown). Furthermore, KSHV.ORF24^{R328A} failed to produce infectious progeny virions in supernatant transfer assays, as measured by GFP-positive cells using both confocal microscopy and flow cytometry (Figure 3.4D-E). These data indicate that the ability of ORF24 to bind ORF34 is critical for viral late gene expression, and that a single point mutation that selectively disrupts that interaction prevents completion of the KSHV lifecycle.

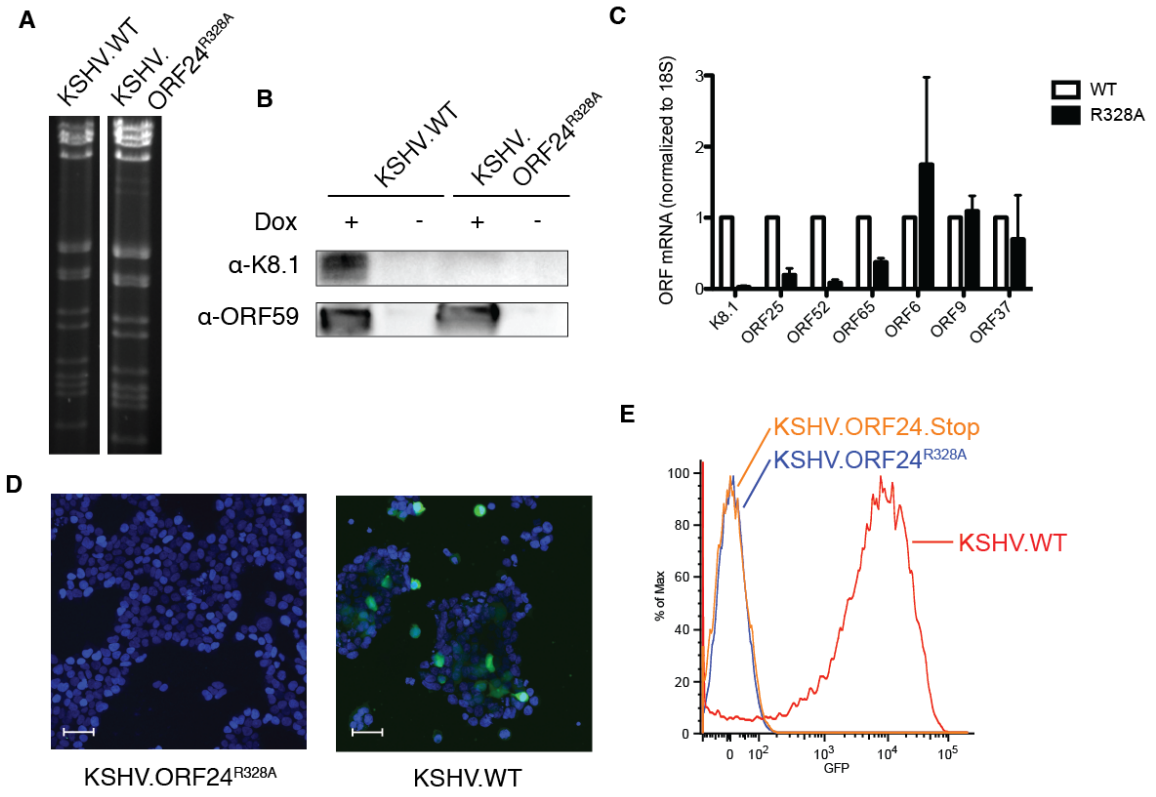

Figure 3.4: The interaction between ORF24 and ORF34 is required for late gene expression and progeny virion production.

(A) No gross recombination events were detected in the KSHV.ORF24^{R328A} BAC DNA compared with KSHV.WT. BAC DNA was digested with EcoRI and run out on a 0.7% agarose TAE gel at 30V for 24 hr. **(B-C)** KSHV.WT and KSHV. ORF24^{R328A}-infected iSLK cells were lytically reactivated for 72 hr. Protein lysates were collected and analyzed by Western blot for K8.1 (late) and ORF59 (delayed early) (B). RNA was collected and levels were analyzed for early (ORF6, ORF9, and ORF37) and late (K8.1, ORF25, ORF52, and ORF65) genes normalized to 18S. **(D-E)** KSHV.WT, KSHV.ORF24.Stop, and KSHV.ORF24^{R328A}-infected iSLK cells were lytically reactivated for 72 hr. Supernatant was added to 293T cells and 24 hr later, cells were monitored for GFP expression by confocal microscopy (D) or flow cytometry (E), indicating the transfer of infectious virus. In (D), DNA is stained with DAPI (blue). A minimum of 30 fields were monitored for each experiment, with a single representative field shown. No GFP positive cells were detected in any field of

recipient cells containing supernatant from KSHV.ORF24^{R328A}-infected cells. Scale bar = 50 μm. KSHV.ORF24.Stop is shown as a negative control.

The Six KSHV vTFs Are Necessary and Sufficient to Drive Transcription From a Late Gene Promoter

A late gene promoter activation assay was recently established for EBV, in which expression of all six vTFs enabled transcription from a co-transfected plasmid containing luciferase driven by an EBV late gene promoter (Aubry et al., 2014; Wu et al., 2009). Removal of any one of the individual vTFs prevented promoter activation in this assay, indicating that the full complement of vTFs is required for transcription from the late promoter (Aubry et al., 2014). This provides the ability to measure late gene transcription in uninfected cells and thus in a manner separable from herpesviral DNA replication, which had not previously been possible.

We first evaluated whether, like in EBV, the six KSHV vTFs were able to drive transcription from a plasmid-based late gene promoter. The K8.1 promoter is the only KSHV late promoter that has been fully mapped (Tang et al., 2004), and thus we cloned this promoter upstream of a *Photinus pyralis* (firefly) luciferase reporter gene. All wells were also transfected with a uniform amount of a constitutive SV-40 promoter driven *Renilla reniformis* luciferase reporter for normalization. Co-transfection of the six KSHV vTFs with the luciferase reporter plasmids led to activation of the K8.1 late promoter (Figure 3.5A). Removal of any one of the vTFs from the transfection eliminated K8.1 promoter activation, confirming the requirement for the complete set of six vTFs (Figure 3.5A). The interaction between ORF34 and ORF24 was essential for late promoter activation, as substitution of ORF24^{R328A} for wildtype ORF24 in the vTF complex resulted in a complete loss of luciferase activity (Figure 3.5A). Similarly, the ORF24^{RAAAG} mutant that is unable to interact with RNAP II (Davis et al., 2015) was unable to activate the K8.1 promoter in this assay. The vTFs were selectively required for late promoter activation, as neither the K14 (vOX2) nor ORF57 (MTA) early gene promoters cloned upstream of the firefly luciferase reporter were activated in the presence of the six vTFs (Figure 3.5B). Thus, ORF24 must interact with ORF34 for the expression of late genes.

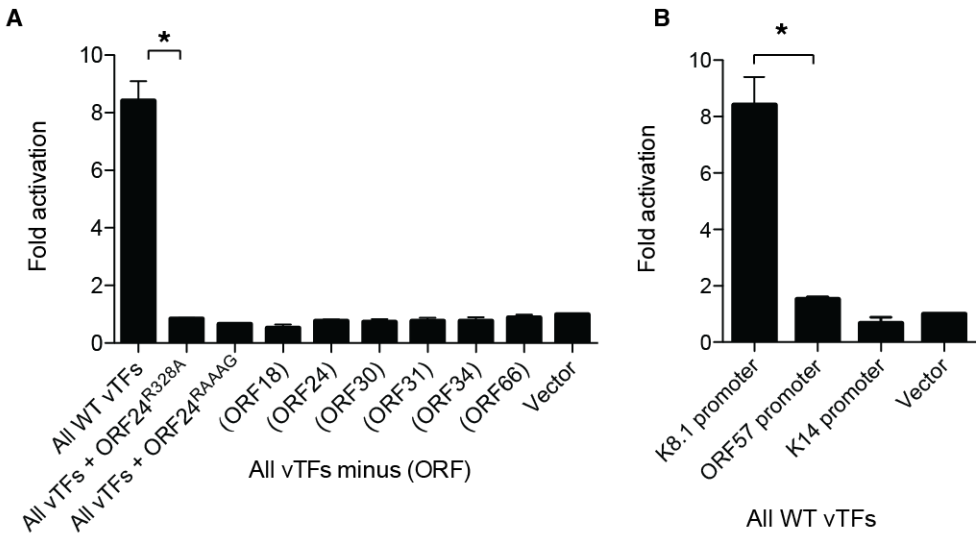

Figure 3.5: All six vTFs and the ORF24-ORF34 interaction are required for activation of a late gene promoter.

(A) 293T cells were transfected with all six vTFs or lacking the vTF in parentheses and assessed for their ability to activate a co-transfected late gene promoter (K8.1)-driven luciferase reporter. **(B)** Cells were transfected as in (A) with all six wildtype vTFs or empty vector and assessed for their ability to activate a luciferase reporter driven by the K8.1 (late), ORF57 (early), or K14 (early) promoter. * indicates a p-value < 0.05 as determined by Student's t test.

Discussion

Late gene transcriptional regulation in β- and γ-herpesviruses is remarkably complex, requiring more viral proteins than any other nuclear replicating viruses characterized to date. Six vTFs, together with a largely undefined set of cellular transcription factors that for KSHV include RNAP II, TFIIB and TFIIH, coordinate to drive specific activation of the minimalistic late gene promoters (Aubry et al., 2014; Davis et al., 2015). Unlike ORF24, the other five vTFs have no described sequence or structural similarity with characterized proteins, and thus aside from a role for HCMV UL79 (ORF18 ortholog) in RNAP II elongation (Perng et al., 2014), little is known about how they contribute to late gene transcription. In EBV, the six vTFs have been shown to form a complex (Aubry et al., 2014). However, the organization of the vTF complex, as well as the importance of many of the individual protein-protein interactions remained unknown. Understanding the regulation and the architecture of the unique β/γ-herpesviral vPIC and, ultimately, how the vTFs coordinate recruitment of cellular machinery may reveal a gene expression strategy that is fundamentally distinct from previous examples of transcription in eukaryotic cells.

Using a high-throughput *in vivo* protein-protein interaction assay, we mapped all of the pairwise interactions between the six KSHV vTFs, enabling us to assemble a model of the general organization of the vTF complex (Figure 3.1D). These data suggest that ORF34 is an essential and highly connected member of the vPIC, as it forms contacts with the majority of the other vTFs and is the only vTF to directly

contact ORF24. Further characterization of this interaction revealed that the ability of ORF24 to bind ORF34 is critical for mediating viral late gene expression, and a single point mutation that selectively disrupts the interaction prevents completion of the KSHV lifecycle. These findings suggest that the interactions we defined between the other vTFs are likely to be similarly important for late gene transcription. Indeed, a recent report demonstrated that the interaction between the KSHV vTFs ORF31 and ORF34 is also essential for late gene expression (Brulois et al., 2015).

The best characterized of the KSHV vTFs is ORF24, which binds both RNAP II and late gene promoter DNA in a sequence-specific manner (Davis et al., 2015). These activities occur through distinct N-terminal and central TBP-like domains of the protein, respectively. Here, we demonstrate that the interaction with ORF34 is a third essential activity, and that it maps to a region of the ORF24 N-terminus distinct from the RNAP II binding site. These findings highlight how different regions of ORF24 coordinate multiple specific protein and DNA binding activities, each of which are essential for KSHV late gene transcription.

A plasmid-based late promoter-luciferase assay was recently described for EBV, which for the first time provides a tractable system to study late promoter activation in a manner separable from viral genome replication (Aubry et al., 2014). During infection, late gene expression must be licensed by the replication of the viral genome by the viral DNA replication machinery (Anders et al., 2007). Furthermore, transfection of the six EBV vTFs is not sufficient to activate late gene expression from the EBV genome, suggesting that the prior steps of replication may remodel late gene promoters to make them permissive to vTF activation (Aubry et al., 2014). The transfected late gene promoter luciferase reporter is not subject to this repression, yet still demonstrates specific activation. These observations suggest that late transcription requires both de-repression of promoters, which is enabled by viral genome replication, and subsequent transcriptional activation, which is driven by the six vTFs. Using the luciferase-based promoter activation assay, we demonstrated that similar to EBV, all six KSHV vTFs are necessary (and together are sufficient) for late promoter activation, and do not activate early promoters. Further, we found that this assay recapitulated our findings from viral mutants, in that ORF24 mutants unable to interact with ORF34 (R328A) or RNAP II (RAAAG) both failed to activate the late promoter (Figure 3.5A). Thus, this reporter system should facilitate efforts to dissect specific features of late promoter vPIC assembly in a manner separable from the complexities of viral DNA replication.

Though the majority of phenotypes described for vTFs are conserved amongst the β- and γ-herpesviruses, we note some differences. For example, ORF24 is the only KSHV vTF for which we detect an interaction with RNAP II, while in HCMV, UL79 (the ORF18 ortholog) was also reported to interact with RNAP II. Further, while we see interactions between many of the vTFs (Figure 3.1D), we do not detect pairwise interactions between all of the vTFs as seen with the EBV orthologs in an *in vitro* GST-pulldown assay (Aubry et al., 2014). It is possible that these distinctions arise because of differences in assay conditions (e.g. *in vitro* versus in cells). Alternatively, they may hint at evolving interaction surfaces that regulate the precise complement of viral and cellular proteins coordinated by the vTFs. The

differences between the vTF interactions help explain the finding that while all six vTFs together of HCMV or EBV origin are able to activate a late promoter, HCMV orthologs are not able to substitute for individual EBV vTFs (Aubry et al., 2014).

While there are some differences in vPIC organization amongst the β- and γ-herpesviruses, late gene activation in these viruses differs substantially from that of the α-herpesviruses. None of the vTFs have orthologs in the α-herpesviruses, which instead require ICP27 (ortholog of KSHV ORF57) (Rice and Knipe, 1990) and ICP8 (ortholog of KSHV ORF6) (Gao and Knipe, 1991), both of which also play critical roles in viral DNA replication. ICP27 associates with the carboxyl terminal domain (CTD) of RNAP II (Dai-Ju et al., 2006), although unlike the γ-herpesvirus ORF24 protein it does not make direct sequence-specific contacts with the late promoters. The viral transcription factor ICP4, which has no orthologs outside of α-herpesviruses, drives transcription of all kinetic classes of genes by interacting with the general transcription factor TFIID (Carrozza and DeLuca, 1996). Thus, the β- and γ-herpesviral vPIC gene expression strategy appears unique to these herpesviruses, although other viruses may incorporate the core functions of the vTF proteins within a smaller subset of proteins. Interestingly, the cytoplasmic-replicating vaccinia virus encodes at least four late gene transcription factors, several of which have been reported to interact with each other (Broyles, 2003; D'Costa et al., 2010; Dellis et al., 2004). The majority of DNA viruses replicate in the nucleus, which allows them to take advantage of cellular transcription and DNA replication components. However, because it replicates in the cytoplasm, vaccinia must supply all of these factors. Thus, the dedication of several open reading frames to the specific regulation of late gene transcription could be a possible mechanism for viruses to reduce their reliance on the host cell.

As demonstrated by the extensive interactions between the vTFs, late gene transcription is tightly regulated in DNA viruses. This may, in part, help delay the formation of immune-stimulatory dsRNA products produced once condensed viral genomes are being fully transcribed and producing mRNAs overlapping on both DNA strands (Arias et al., 2014). Late genes are distributed throughout the KSHV genome, which could allow them to act as intergenic barriers to prevent dsRNA production during early gene expression, in addition to generating virion components needed post genome replication.

Methods

Plasmid construction

For the generation of the Gaussia luciferase (GLuc) fragment expression constructs, vTF ORFs (Davis et al., 2015) were first amplified and cloned into the pDON207 entry vector (Invitrogen) using Gateway recombination cloning (Invitrogen). vTFs were then transferred into the GLuc fragment 1 and 2 (GLuc1 and GLuc2; pSPICA-N1 and pSPICA-N2) destination vectors (Muller et al., 2012). ORF24 domain fragments and N-terminal truncation mutants were generated by restriction enzyme cloning into a pCDNA4/TO (Invitrogen) vector with a C-terminal 3X FLAG or 2X strep tag (restriction sites listed in Table 3.1). The ORF24^{R328A} mutant was

made by site directed mutagenesis using the Platinum Pfx DNA polymerase (Invitrogen). For the construction of the K8.1 late gene promoter luciferase reporter, phosphorylated primers containing the 23 bases centered on the K8.1 TATT-box sequence were ligated into the pGL4.16 vector (Promega). All construct sequences were verified by sequencing. Primer sequences are listed in Table 3.1.

Cell culture

HEK293T (ATCC), the Caki-1 kidney carcinoma cell line iSLK puro (Sturzl et al., 2013) containing a doxycycline-inducible copy of the major KSHV lytic transactivator RTA (ORF50), and other iSLK puro-derived cell lines were cultured in Dulbecco's modified Eagle medium (DMEM; Invitrogen) with 10% fetal bovine serum (FBS; Invitrogen, HyClone) and penicillin-streptomycin (Invitrogen).

Split luciferase protein-protein interaction assay

The GLuc split luciferase assay was used to assess vTF interactions as previously described (Muller and Demeret, 2012). Briefly, wildtype ORF24 or ORF24^{R328A} in pSPICA-N2 was co-transfected with each of the other five pSPICA-N1 vTFs at a 1:1 ratio into HEK293T cells in a 24-well plate for a total of 400 ng of DNA. After 24 hr, cells were lysed and luciferase levels were determined (*Renilla* Luciferase Assay System, Promega). The luciferase values for the pSPICA-N1 and pSPICA-N2 co-transfected cells were normalized to the sum of the luciferase values for cells individually transfected with each pSPICA vTF.

KSHV BAC mutagenesis

ORF34 and ORF24 with premature stop mutations in their coding regions (KSHV.ORF34.Stop and KSHV.ORF24.Stop, respectively), and ORF24 with the R328A amino acid substitution were engineered using the KSHV BAC16 scarless recombination mutagenesis system (Brulois et al., 2012). Purified BAC DNA was transfected into iSLK puro cells using PolyJet (SignaGen) and selected with 400 μg/mL of hygromycin. Individual clones were amplified and tested for their ability to produce early and late gene products.

Affinity purification and Western blotting

Cell lysate was collected and analyzed as previously described (Davis et al., 2015). Briefly, HEK293T cells were transfected with expression constructs using Polyjet (SignaGen) and lysates were harvested 24 hr later. For the affinity purification, lysate was added to pre-washed Strep-Tactin (IBA) or FLAG (Sigma) beads. After washing, SDS-sample buffer was added to samples to elute bound proteins. Western blots were carried out with the following antibodies: rabbit α-FLAG (1:5000; Sigma), mouse α-FLAG (clone M2; 1:5000; Sigma), mouse α-strep (1:3000; Qiagen), rabbit α-RPB1 (RNAP II) (N20; 1:5000; Santa Cruz Biotechnologies), rabbit α-K8.1 (1:5000, antibody generated for this study), rabbit α-ORF59 (1:10000, antibody generated for this study), and goat α-mouse and goat

α-rabbit HRP secondary antibodies (1:5000; Southern Biotech). 6xHis-MBP-K8.1 and 6xHis-MBP-ORF59 protein was purified from E. coli and SF-9 insect cells, respectively (Labo et al., 2014) and was injected into rabbits to generate polyclonal antibodies (Pocono Rabbit Farm and Laboratory). Crude sera was used for Western blotting.

Supernatant transfer assay

Supernatant transfer assays were performed as previously described (Davis et al., 2015). Briefly, iSLK puro-derived cell lines were reactivated with sodium butyrate (1 mM) and doxycycline (1 μg/mL) to induce the viral lytic cycle. After three days, viral supernatant was collected, filtered, and added to uninfected HEK293T cells. Following 24 hr of incubation with the viral supernatant, cells were assessed for the successful transfer of infectious virions (indicated by GFP expressing cells harboring the GFP-positive KSHV BAC16 virus) by flow cytometry (LSR Fortessa, BD Biosciences) and confocal microscopy (LSM 710 laser scanning confocal microscope, Zeiss).

DNA replication

Viral and human genomic DNA was isolated from reactivated iSLK puro-derived cell lines using the DNeasy Blood and Tissue Kit (Qiagen). For each sample, the number of viral genomes were measured by qPCR and normalized to the levels of human GAPDH DNA.

Promoter activation assay

HEK293T cells plated in a 24-well plate were transfected with equal amounts of the K8.1 late promoter pGL4.16 firefly luciferase reporter along with combinations of the strep-tagged vTFs, and an SV-40 promoter-driven *Renilla* luciferase construct as a transfection control for a total of 400 ng of DNA. After 24 hr, cells were lysed and luciferase levels were determined (Dual–Luciferase Reporter Assay System, Promega). Firefly luciferase levels were normalized to the *Renilla* luciferase plasmid transfection control.

Table 3.1 Primer sequences used in this study.

SDM: site directed mutagenesis.

Primer	Sequence (5'-3')	Orientation F: Forward R: Reverse
ORF18 Gateway	ggggacaactttgtacaaaaaagttggcATGCTCGGAAAATACGTGTGTGAGA	F
ORF18 Gateway	ggggacaactttgtacaagaaagttgggtaTTAAACCGCGTTGTTGTTAAACGCA	R
ORF24 Gateway	ggggacaactttgtacaaaaaagttggcATGGCAGCGCTCGAGGG	F

ORF24 Gateway	ggggacaactttgtacaagaaagttgggtattaGACCAGCGGACGGACGC	R
ORF30 Gateway	ggggacaactttgtacaaaaaagttggcATGGGTGAGCCAGTGGATCC	F
ORF30 Gateway	ggggacaactttgtacaagaaagttgggtaTCATTTCGCACCGGTGTCTAGG	R
ORF31 Gateway	ggggacaactttgtacaaaaaagttggcATGTCACAAAACAGAAAGACTCTGCC	F
ORF31 Gateway	ggggacaactttgtacaagaaagttgggtaCTACGTATCTTTCGTTGATAGCATGCG	R
ORF34 Gateway	ggggacaactttgtacaaaaaagttggcATGTTTGCTTTGAGCTCGCTCG	F
ORF34 Gateway	ggggacaactttgtacaagaaagttgggtaTTAGAGTTGGTTGAGTCCATTCTCCTT	R
ORF66 Gateway	ggggacaactttgtacaaaaaagttggcATGGCCCTGGATCAGCGC	F
ORF66 Gateway	ggggacaactttgtacaagaaagttgggtaTCAGGAGGAACACTTCCCGC	R
ORF24^{R328A} SDM	ccatacacatgacgagtgcgtagatggccggtgtgc	F
ORF24^{R328A} SDM	gcacaccggccatctacgcactcgtcatgtgtatgg	R
ORF24^{R328A} BAC mutant	tacctactatctagtgtatccgggcacaccggccatctacgcactcgtcatgtgtatggcagAGGATGACGACGATAAGTAGGG	F
ORF24^{R328A} BAC mutant	gtggccgatgcagtctgccactgccatacacatgacgagtgcgtagatggccggtgtgcccgAACCAATTAACCAATTCTGATTAG	R
ORF34.Stop BAC mutant	acggctgagcagcatgtttgctttgagctcgctcgtgtcctagggtgacccggaggtgaccagAGGATGACGACGATAAGTAGGG	F
ORF34.Stop BAC mutant	gtacgcccttgacgtacctactggtcacctccgggtcaccttaggacacgagcgagctcaaagAACCAATTAACCAATTCTGATTAG	R
K8.1 pr pGL4.16 (KpnI-HindIII)	cCCGGCAGCAATATTAAAGGGACCa	F
K8.1 pr pGL4.16 (KpnI-HindIII)	agcttGGTCCCTTTAATATTGCTGCCGGgtac	R
ORF24 N-term (BamHI)	gctcggatccATGGCAGCGCTCGAGGG	F
ORF24 N-term (NotI)	tcgagcggccgcttGCTGCCCAGAGTCCGCG	R
ORF24 TBP (BamHI)	gctcggatccATGTCTACAGTGTCAGACATGCTGGAAC	F
ORF24 TBP (NotI)	tcgagcggccgcttGTTTCTGTGCTCAAGTAGGGAGAATATTCT	R
ORF24 C-term	gctcggatccatgCGTTCCCAAATACAGACGCTACACA	F

(BamHI)		
ORF24 C-term (NotI)	tcgagcggccgcAAGACCAGCGGACGGACG	R
ORF24 N-term truncations (BamHI)	gctcggatccATGGCAGCGCTCGAGGG	F
ORF24 1-288aa (NotI)	TCGAGCGGCCGCttCACGAGCTGGTTTAGAGCCAG	R
ORF24 1-305aa (NotI)	TCGAGCGGCCGCttGTATGGCTTTGGTCCGGGG	R
ORF24 1-322aa (NotI)	TCGAGCGGCCGCttGCCCGGATACACTAGATAGTAGGTACAA	R
ORF24 1-339aa (NotI)	TCGAGCGGCCGCttGATGCAGTCTGCCACTGCC	R
ORF24 1-356aa (NotI)	TCGAGCGGCCGCttGTGGGTGCCTAAAAAGTTTGCG	R
ORF24 1-373aa (NotI)	TCGAGCGGCCGCttCGGCGCGTACCGGATTC	R
ORF24^{R328A} SDM	Ccatacacatgacgagtgcgtagatggccggtgtgc	F
ORF24^{R328A} SDM	Gcacaccggccatctacgcactcgtcatgtgtatgg	R

Acknowledgements

We would like to thank Dr. Denise Whitby for the generous gift of K8.1 and ORF59 protein for the generation of rabbit polyclonal antibodies (Labo et al., 2014); members of the Britt Glaunsinger and Laurent Coscoy labs for reagents and advice, in particular, Mandy Muller for her help with the split luciferase assay; and Yves Jacob and Caroline Demeret for the split luciferase constructs.

Chapter 4 - Perspectives and Conclusions

How does DNA replication "license" the expression of late genes?

While this is an area of great interest to us, we are still at the stage of debating hypotheses around this question. We have two main hypotheses here: first, that DNA replication opens chromatin, and helps to relieve repression by granting access to repressed genomic regions (likely through epigenetic changes). The second main contender for how this regulatory axis functions is that the DNA replication machinery brings essential factors for late gene expression to the DNA, as seen for T4 phage (Geiduschek and Kassavetis, 2010). In differentiating between these possibilities or deciphering a different regulatory model, we do not know if late genes are transcribed from newly synthesized genomes, or if existing episomes can be converted over to late gene expression. There is also the possibility that what we consider "late" genes actually consists of genes regulated in multiple ways. For instance, some "late" genes may be activated by epigenetic remodeling triggered by DNA replication while others require the replication machinery to bring factors in, while still others may only be able to be transcribed from newly synthesized genomes. As we learn more about what actually regulates this class of gene expression, we may discover that this class is in fact a grouping of several distinct subclasses.

Nevertheless, the former possibility seems less likely. Late genes are not localized to one region of the viral genome; rather, they are sprinkled throughout and are found between early and delayed early genes. This results in many early and delayed early genes having at least some of their coding regions or 5' or 3' UTRs (very common) overlapping with the coding regions, promoters and untranslated regions of late genes. To analyze this further, I will use the example of the ORF23 through ORF25 gene locus in KSHV. ORF23 and ORF23 are both delayed early genes expressed prior to genome replication. ORF25, however, is a late gene expressed from the positive strand (Figure 4.1). For RNAP II to recognize and begin transcription of ORF24, it must access the promoter located in the ORF25 coding region. Further, the promoter of ORF25 is located in the coding region of ORF24. Because ORF24 is successfully made prior to genome replication, we know that RNAP II would be able to access the site of the ORF25 late promoter. Thus, the idea that epigenetic remodeling occurring with DNA replication is essential to open the chromatinized genome to allow late gene expression does not seem consistent with our knowledge that at the exact location of the ORF25 promoter, the DNA must be in an open, accessible state for the production of ORF24.

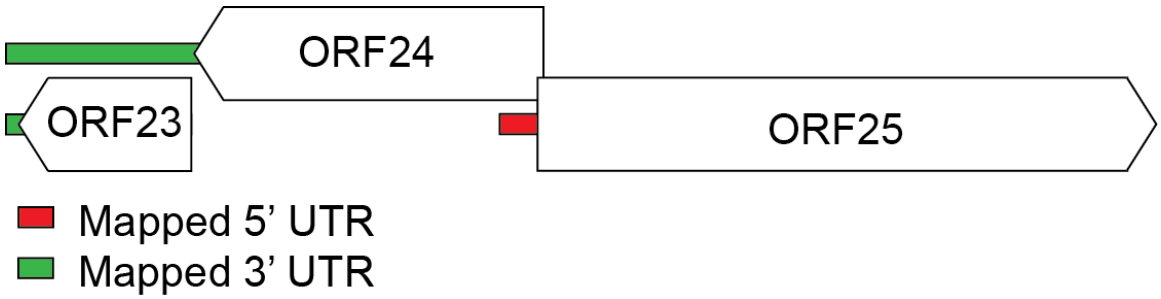

Figure 4.1: The genomic context of the ORF25 late promoter.
ORF23 and ORF24 are transcribed from the negative strand of the viral genome and are both delayed early genes. ORF25 is located on the positive strand and is a strictly late gene. How is ORF24 produced prior to genome replication while ORF25 is not?

In the absence of TBP, how are GTFs recruited to late PICs?

Contrary to canonical transcription, late gene promoters are not bound by TBP and instead rely on ORF24 to nucleate PIC assembly. The absence of TBP suggests that PIC organization at late gene promoters must occur in a manner distinct from the majority of previously described promoters, as TBP often serves as a critical nucleation factor for the recruitment of general transcription factors (GTFs) and other transcriptional machinery (Thomas and Chiang, 2006). While ORF24 contains a TBP like domain, this homology is largely structural – the two proteins only share ~10% sequence identity. Thus, ORF24 is likely unable to interact with and recruit the exact same group of proteins as TBP. Indeed, TBP does not interact directly with RNAPII (Murakami et al., 2013), and we have failed to detect an interaction between ORF24 and any of the GTFs besides RNAP II (Davis et al., 2015). Nevertheless, late gene PICs contain at least a subset of GTFs (Davis et al., 2015). This could suggest a potential role for some of the vTFs in the recruitment of GTFs to late promoters in the absence of TBP. We have not observed direct interactions between the other vTFs and cellular GTFs (Davis et al., 2015). We therefore hypothesize that recruitment of additional transcriptional machinery may occur in a regulated manner on late promoters in the context of the six vTF complex. This could further explain why disruption of the ORF24-ORF34 interaction is detrimental to late gene transcription.

Why do β- and γ-herpesviruses use the vPIC for late gene transcription?

The β- and γ-herpesviruses appear to have adopted a complex strategy relying on many viral proteins to facilitate late gene expression, raising the question of what advantage this may serve. The appropriation of limited viral coding capacity to the tight control and robust expression of late genes further bolsters the importance of regulating this class of genes. Intracellular TBP concentrations can be limiting for transcription (Bush et al., 2008; Paulson et al., 2002), especially given the need for

robust production of late proteins towards the end of the viral life cycle. The abundance of TBP as well as other GTFs may be particularly limiting during γ-herpesvirus infection as a consequence of virus-induced shutoff of host gene expression, which occurs via cytoplasmic mRNA degradation induced by the viral SOX protein (Covarrubias et al., 2011; Glaunsinger and Ganem, 2004). Indeed, the TBP mRNA is robustly targeted for degradation by the SOX orthologs from KSHV and MHV68 (Clyde and Glaunsinger, 2011), and we have observed decreased TBP protein abundance in cells with lytic KSHV infection (Davis et al., 2015)(data not shown). In contrast, in MHV68, five of the six vTF mRNAs escape SOX-mediated host shutoff (Abernathy et al., 2014). Thus, we hypothesize that relying on viral proteins may have been a successful strategy for the virus because it provided a way to ensure the availability of transcription machinery for late genes at late times in infection when cellular components are limited.

Why does late gene expression need to be so extensively regulated?

The virus has evolved a complex strategy for late gene expression. Why is this necessary and presumably advantageous for the viral lifecycle? Again, we are left here to speculate, but several ideas come forth as possibilities. One potential reason is that the class of late genes, including as they do the viral envelope glycoproteins, are the targets of the adaptive immune response in a previously exposed host. In a latently infected cell, there is thus a survival imperative to prevent the expression of late genes until the cell is fully "committed" to the lytic cycle. Spurious production of late gene proteins could alert the immune system to the presence of the latent virus, and target the latent cell for elimination. While early gene products may pose some of the same concerns, the fact that late genes are surface antigens makes them ideal targets for inactivation by the immune system.

A further reason for tightly regulated late gene expression could be due to an increased incidence in the formation of double-stranded RNA during late gene expression. This could be due to the possibility of forming double-stranded RNA products once condensed viral genomes are being fully transcribed (all kinetic classes), producing mRNAs from both DNA strands with overlapping UTRs and coding regions. In fact, in the poxvirus vaccinia, double-stranded RNA production has been shown to depend on late gene expression (Colby et al., 1971). The double-stranded RNA that can form is a major indicator of infection that could lead to recognition of the virus by the immune sensors in the cell and lead to global shutoff of mRNA and protein synthesis. Thus, late gene expression must be faithfully controlled to prevent spurious transcription.

Chapter 5 - References

Abada, R., Dreyfuss-Grossman, T., Herman-Bachinsky, Y., Geva, H., Masa, S.R., and Sarid, R. (2008). SIAH-1 interacts with the Kaposi's sarcoma-associated herpesvirus-encoded ORF45 protein and promotes its ubiquitylation and proteasomal degradation. Journal of virology *82*, 2230-2240.

Abernathy, E., Clyde, K., Yeasmin, R., Krug, L.T., Burlingame, A., Coscoy, L., and Glaunsinger, B. (2014). Gammaherpesviral gene expression and virion composition are broadly controlled by accelerated mRNA degradation. PLoS pathogens *10*, e1003882.

Anders, D.G., Kerry, J.A., and Pari, G.S. (2007). DNA synthesis and late viral gene expression. In Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis, A. Arvin, G. Campadelli-Fiume, E. Mocarski, P.S. Moore, B. Roizman, R. Whitley, and K. Yamanishi, eds. (Cambridge).

Arias, C., Weisburd, B., Stern-Ginossar, N., Mercier, A., Madrid, A.S., Bellare, P., Holdorf, M., Weissman, J.S., and Ganem, D. (2014). KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. PLoS pathogens *10*, e1003847.

Arumugaswami, V., Wu, T.T., Martinez-Guzman, D., Jia, Q., Deng, H., Reyes, N., and Sun, R. (2006). ORF18 is a transfactor that is essential for late gene transcription of a gammaherpesvirus. Journal of virology *80*, 9730-9740.

Aubry, V., Mure, F., Mariame, B., Deschamps, T., Wyrwicz, L.S., Manet, E., and Gruffat, H. (2014). Epstein-barr virus late gene transcription depends on the assembly of a virus-specific preinitiation complex. Journal of virology *88*, 12825-12838.

Barbera, A.J., Chodaparambil, J.V., Kelley-Clarke, B., Joukov, V., Walter, J.C., Luger, K., and Kaye, K.M. (2006). The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. Science *311*, 856-861.

Bauer-Mehren, A., Bundschus, M., Rautschka, M., Mayer, M.A., Sanz, F., and Furlong, L.I. (2011). Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. PloS one *6*, e20284.

Boshoff, C. (2012). Ephrin receptor: a door to KSHV infection. Nature medicine *18*, 861-863.

Boyne, J.R., Jackson, B.R., Taylor, A., Macnab, S.A., and Whitehouse, A. (2010). Kaposi's sarcoma-associated herpesvirus ORF57 protein interacts with PYM to enhance translation of viral intronless mRNAs. The EMBO journal *29*, 1851-1864.

Bravender, T. (2010). Epstein-Barr virus, cytomegalovirus, and infectious mononucleosis. Adolesc Med State Art Rev *21*, 251-264, ix.

Broyles, S.S. (2003). Vaccinia virus transcription. The Journal of general virology *84*, 2293-2303.

Brulois, K., Wong, L.Y., Lee, H.R., Sivadas, P., Ensser, A., Feng, P., Gao, S.J., Toth, Z., and Jung, J.U. (2015). The association of Kaposi's sarcoma-associated herpesvirus ORF31 with ORF34 and ORF24 is critical for late gene expression. Journal of virology.

Brulois, K.F., Chang, H., Lee, A.S., Ensser, A., Wong, L.Y., Toth, Z., Lee, S.H., Lee, H.R., Myoung, J., Ganem, D.*, et al.* (2012). Construction and manipulation of a new

Kaposi's sarcoma-associated herpesvirus bacterial artificial chromosome clone. Journal of virology *86*, 9708-9720.
Burysek, L., Yeow, W.S., Lubyova, B., Kellum, M., Schafer, S.L., Huang, Y.Q., and Pitha, P.M. (1999). Functional analysis of human herpesvirus 8-encoded viral interferon regulatory factor 1 and its association with cellular interferon regulatory factors and p300. Journal of virology *73*, 7334-7342.
Bush, S.D., Richard, P., and Manley, J.L. (2008). Variations in intracellular levels of TATA binding protein can affect specific genes by different mechanisms. Molecular and cellular biology *28*, 83-92.
Bushnell, D.A., Westover, K.D., Davis, R.E., and Kornberg, R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. Science *303*, 983-988.
Butler, L.M., Dorsey, G., Hladik, W., Rosenthal, P.J., Brander, C., Neilands, T.B., Mbisa, G., Whitby, D., Kiepiela, P., Mosam, A.*, et al.* (2009). Kaposi sarcoma-associated herpesvirus (KSHV) seroprevalence in population-based samples of African children: evidence for at least 2 patterns of KSHV transmission. The Journal of infectious diseases *200*, 430-438.
Cai, Q.L., Knight, J.S., Verma, S.C., Zald, P., and Robertson, E.S. (2006). EC5S ubiquitin complex is recruited by KSHV latent antigen LANA for degradation of the VHL and p53 tumor suppressors. PLoS pathogens *2*, e116.
Calderwood, M.A., Venkatesan, K., Xing, L., Chase, M.R., Vazquez, A., Holthaus, A.M., Ewence, A.E., Li, N., Hirozane-Kishikawa, T., Hill, D.E.*, et al.* (2007). Epstein-Barr virus and virus human protein interaction maps. Proceedings of the National Academy of Sciences of the United States of America *104*, 7606-7611.
Carroll, K.D., Khadim, F., Spadavecchia, S., Palmeri, D., and Lukac, D.M. (2007). Direct interactions of Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8 ORF50/Rta protein with the cellular protein octamer-1 and DNA are critical for specifying transactivation of a delayed-early promoter and stimulating viral reactivation. Journal of virology *81*, 8451-8467.
Carrozza, M.J., and DeLuca, N.A. (1996). Interaction of the viral activator protein ICP4 with TFIID through TAF250. Molecular and cellular biology *16*, 3085-3093.
Cassonnet, P., Rolloy, C., Neveu, G., Vidalain, P.O., Chantier, T., Pellet, J., Jones, L., Muller, M., Demeret, C., Gaud, G.*, et al.* (2011). Benchmarking a luciferase complementation assay for detecting protein complexes. Nature methods *8*, 990-992.
Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E.*, et al.* (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer discovery *2*, 401-404.
Chang, Y., Cesarman, E., Pessin, M.S., Lee, F., Culpepper, J., Knowles, D.M., and Moore, P.S. (1994). Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science *266*, 1865-1869.
Chang, Y., Moore, P.S., Talbot, S.J., Boshoff, C.H., Zarkowska, T., Godden, K., Paterson, H., Weiss, R.A., and Mittnacht, S. (1996). Cyclin encoded by KS herpesvirus. Nature *382*, 410.

Chen, D., Cousins, E., Sandford, G., and Nicholas, J. (2012). Human herpesvirus 8 viral interleukin-6 interacts with splice variant 2 of vitamin K epoxide reductase complex subunit 1. Journal of virology *86*, 1577-1588.
Choi, H., Larsen, B., Lin, Z.Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., Gingras, A.C., and Nesvizhskii, A.I. (2011). SAINT: probabilistic scoring of affinity purification-mass spectrometry data. Nature methods *8*, 70-73.
Clauser, K.R., Baker, P., and Burlingame, A.L. (1999). Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. Analytical chemistry *71*, 2871-2882.
Clyde, K., and Glaunsinger, B.A. (2011). Deep sequencing reveals direct targets of gammaherpesvirus-induced mRNA decay and suggests that multiple mechanisms govern cellular transcript escape. PloS one *6*, e19655.
Colby, C., Jurale, C., and Kates, J.R. (1971). Mechanism of synthesis of vaccinia virus double-stranded ribonucleic acid in vivo and in vitro. Journal of virology *7*, 71-76.
Covarrubias, S., Gaglia, M.M., Kumar, G.R., Wong, W., Jackson, A.O., and Glaunsinger, B.A. (2011). Coordinated destruction of cellular messages in translation complexes by the gammaherpesvirus host shutoff factor and the mammalian exonuclease Xrn1. PLoS pathogens *7*, e1002339.
D'Costa, S.M., Bainbridge, T.W., Kato, S.E., Prins, C., Kelley, K., and Condit, R.C. (2010). Vaccinia H5 is a multifunctional protein involved in viral DNA replication, postreplicative gene transcription, and virion morphogenesis. Virology *401*, 49-60.
Dai-Ju, J.Q., Li, L., Johnson, L.A., and Sandri-Goldin, R.M. (2006). ICP27 interacts with the C-terminal domain of RNA polymerase II and facilitates its recruitment to herpes simplex virus 1 transcription sites, where it undergoes proteasomal degradation during infection. Journal of virology *80*, 3567-3581.
Davis, Z.H., Verschueren, E., Jang, G.M., Kleffman, K., Johnson, J.R., Park, J., Von Dollen, J., Maher, M.C., Johnson, T., Newton, W.*, et al.* (2015). Global mapping of herpesvirus-host protein complexes reveals a transcription strategy for late genes. Molecular cell *57*, 349-360.
Dellis, S., Strickland, K.C., McCrary, W.J., Patel, A., Stocum, E., and Wright, C.F. (2004). Protein interactions among the vaccinia virus late transcription factors. Virology *329*, 328-336.
Dorjsuren, D., Lin, Y., Wei, W., Yamashita, T., Nomura, T., Hayashi, N., and Murakami, S. (1998). RMP, a novel RNA polymerase II subunit 5-interacting protein, counteracts transactivation by hepatitis B virus X protein. Molecular and cellular biology *18*, 7546-7555.
Dutta, D., Chakraborty, S., Bandyopadhyay, C., Valiya Veettil, M., Ansari, M.A., Singh, V.V., and Chandran, B. (2013). EphrinA2 regulates clathrin mediated KSHV endocytosis in fibroblast cells by coordinating integrin-associated signaling and c-Cbl directed polyubiquitination. PLoS pathogens *9*, e1003510.
Engelhardt, O.G., Smith, M., and Fodor, E. (2005). Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II. Journal of virology *79*, 5812-5818.
Fanis, P., Gillemans, N., Aghajanirefah, A., Pourfarzad, F., Demmers, J., Esteghamat, F., Vadlamudi, R.K., Grosveld, F., Philipsen, S., and van Dijk, T.B. (2012). Five friends of methylated chromatin target of protein-arginine-methyltransferase[prmt]-1

(chtop), a complex linking arginine methylation to desumoylation. Molecular & cellular proteomics : MCP *11*, 1263-1273.
Finkbeiner, E., Haindl, M., and Muller, S. (2011). The SUMO system controls nucleolar partitioning of a novel mammalian ribosome biogenesis complex. The EMBO journal *30*, 1067-1078.
Flint, S.J., and American Society for Microbiology. (2009). Principles of virology, 3rd edn (Washington, DC: ASM Press).
Freiman, R.N., Albright, S.R., Chu, L.E., Zheng, S., Liang, H.E., Sha, W.C., and Tjian, R. (2002). Redundant role of tissue-selective TAF(II)105 in B lymphocytes. Molecular and cellular biology *22*, 6564-6572.
Friborg, J., Jr., Kong, W., Hottiger, M.O., and Nabel, G.J. (1999). p53 inhibition by the LANA protein of KSHV protects against cell death. Nature *402*, 889-894.
Ganem, D. (2010). KSHV and the pathogenesis of Kaposi sarcoma: listening to human biology and medicine. J Clin Invest *120*, 939-949.
Gao, M., and Knipe, D.M. (1991). Potential role for herpes simplex virus ICP8 DNA replication protein in stimulation of late gene expression. Journal of virology *65*, 2666-2675.
Geiduschek, E.P., and Kassavetis, G.A. (2010). Transcription of the T4 late genes. Virol J *7*, 288.
Genomes Project, C., Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061-1073.
Germain, M.A., Chatel-Chaix, L., Gagne, B., Bonneil, E., Thibault, P., Pradezynski, F., de Chassey, B., Meyniel-Shicklin, L., Lotteau, V., Baril, M.*, et al.* (2013). Elucidating Novel Hepatitis C Virus/Host Interactions Using Combined Mass Spectrometry and Functional Genomics Approaches. Molecular & cellular proteomics : MCP.
Glaunsinger, B., and Ganem, D. (2004). Lytic KSHV infection inhibits host gene expression by accelerating global mRNA turnover. Molecular cell *13*, 713-723.
Gong, D., Wu, N.C., Xie, Y., Feng, J., Tong, L., Brulois, K.F., Luan, H., Du, Y., Jung, J.U., Wang, C.Y.*, et al.* (2014). Kaposi's sarcoma-associated herpesvirus ORF18 and ORF30 are essential for late gene expression during lytic replication. Journal of virology *88*, 11369-11382.
Gonzalez, C.M., Wang, L., and Damania, B. (2009). Kaposi's sarcoma-associated herpesvirus encodes a viral deubiquitinase. Journal of virology *83*, 10224-10233.
Gregory, S.M., Davis, B.K., West, J.A., Taxman, D.J., Matsuzawa, S., Reed, J.C., Ting, J.P., and Damania, B. (2011). Discovery of a viral NLR homolog that inhibits the inflammasome. Science *331*, 330-334.
Gruffat, H., Kadjouf, F., Mariame, B., and Manet, E. (2012). The Epstein-Barr virus BcRF1 gene product is a TBP-like protein with an essential role in late gene expression. Journal of virology *86*, 6023-6032.
Guasparri, I., Wu, H., and Cesarman, E. (2006). The KSHV oncoprotein vFLIP contains a TRAF-interacting motif and requires TRAF2 and TRAF3 for signalling. EMBO reports *7*, 114-119.
Guglielmi, B., La Rochelle, N., and Tjian, R. (2013). Gene-specific transcriptional mechanisms at the histone gene cluster revealed by single-cell imaging. Molecular cell *51*, 480-492.

Gwack, Y., Nakamura, H., Lee, S.H., Souvlis, J., Yustein, J.T., Gygi, S., Kung, H.J., and Jung, J.U. (2003). Poly(ADP-ribose) polymerase 1 and Ste20-like kinase hKFC act as transcriptional repressors for gamma-2 herpesvirus lytic replication. Molecular and cellular biology *23*, 8282-8294.
Heinzelmann, K., Scholz, B.A., Nowak, A., Fossum, E., Kremmer, E., Haas, J., Frank, R., and Kempkes, B. (2010). Kaposi's sarcoma-associated herpesvirus viral interferon regulatory factor 4 (vIRF4/K10) is a novel interaction partner of CSL/CBF1, the major downstream effector of Notch signaling. Journal of virology *84*, 12255-12264.
Herrera, F.J., Yamaguchi, T., Roelink, H., and Tjian, R. (2014). Core promoter factor TAF9B regulates neuronal gene expression. eLife *3*, e02559.
Hirsch, A.J. (2010). The use of RNAi-based screens to identify host proteins involved in viral replication. Future microbiology *5*, 303-311.
Homa, F.L., Glorioso, J.C., and Levine, M. (1988). A specific 15-bp TATA box promoter element is required for expression of a herpes simplex virus type 1 late gene. Genes & development *2*, 40-53.
Isomura, H., Stinski, M.F., Murata, T., Yamashita, Y., Kanda, T., Toyokuni, S., and Tsurumi, T. (2011). The human cytomegalovirus gene products essential for late viral gene expression assemble into prereplication complexes before viral DNA replication. Journal of virology *85*, 6629-6644.
Jackson, B.R., Boyne, J.R., Noerenberg, M., Taylor, A., Hautbergue, G.M., Walsh, M.J., Wheat, R., Blackbourn, D.J., Wilson, S.A., and Whitehouse, A. (2011). An interaction between KSHV ORF57 and UIF provides mRNA-adaptor redundancy in herpesvirus intronless mRNA export. PLoS pathogens *7*, e1002138.
Jager, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., Shales, M., Mercenne, G., Pache, L., Li, K.*, et al.* (2011a). Global landscape of HIV-human protein complexes. Nature *481*, 365-370.
Jager, S., Gulbahce, N., Cimermancic, P., Kane, J., He, N., Chou, S., D'Orso, I., Fernandes, J., Jang, G., Frankel, A.D.*, et al.* (2011e). Purification and characterization of HIV-human protein complexes. Methods *53*, 13-19.
Jarousse, N., Chandran, B., and Coscoy, L. (2008). Lack of heparan sulfate expression in B-cell lines: implications for Kaposi's sarcoma-associated herpesvirus and murine gammaherpesvirus 68 infections. Journal of virology *82*, 12591-12597.
Jia, Q., Wu, T.T., Liao, H.I., Chernishof, V., and Sun, R. (2004). Murine gammaherpesvirus 68 open reading frame 31 is required for viral replication. Journal of virology *78*, 6610-6620.
Kanno, T., Sato, Y., Sata, T., and Katano, H. (2006). Expression of Kaposi's sarcoma-associated herpesvirus-encoded K10/10.1 protein in tissues and its interaction with poly(A)-binding protein. Virology *352*, 100-109.
Kim, K.Y., Huerta, S.B., Izumiya, C., Wang, D.H., Martinez, A., Shevchenko, B., Kung, H.J., Campbell, M., and Izumiya, Y. (2013). Kaposi's sarcoma-associated herpesvirus (KSHV) latency-associated nuclear antigen regulates the KSHV epigenome by association with the histone demethylase KDM3A. Journal of virology *87*, 6782-6793.
Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. (1993). Crystal structure of a yeast TBP/TATA-box complex. Nature *365*, 512-520.

Kliche, S., Nagel, W., Kremmer, E., Atzler, C., Ege, A., Knorr, T., Koszinowski, U., Kolanus, W., and Haas, J. (2001). Signaling by human herpesvirus 8 kaposin A through direct membrane recruitment of cytohesin-1. Molecular cell *7*, 833-843.
Krithivas, A., Fujimuro, M., Weidner, M., Young, D.B., and Hayward, S.D. (2002). Protein interactions targeting the latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus to cell chromosomes. Journal of virology *76*, 11596-11604.
Kuang, E., Tang, Q., Maul, G.G., and Zhu, F. (2008). Activation of p90 ribosomal S6 kinase by ORF45 of Kaposi's sarcoma-associated herpesvirus and its role in viral lytic replication. Journal of virology *82*, 1838-1850.
Kuang, E., Wu, F., and Zhu, F. (2009). Mechanism of sustained activation of ribosomal S6 kinase (RSK) and ERK by kaposi sarcoma-associated herpesvirus ORF45: multiprotein complexes retain active phosphorylated ERK AND RSK and protect them from dephosphorylation. The Journal of biological chemistry *284*, 13958-13968.
Kugel, J.F., and Goodrich, J.A. (2002). Translocation after synthesis of a four-nucleotide RNA commits RNA polymerase II to promoter escape. Molecular and cellular biology *22*, 762-773.
Kumar, G.R., Shum, L., and Glaunsinger, B.A. (2011). Importin alpha-mediated nuclear import of cytoplasmic poly(A) binding protein occurs as a direct consequence of cytoplasmic mRNA depletion. Molecular and cellular biology *31*, 3113-3125.
Labo, N., Miley, W., Marshall, V., Gillette, W., Esposito, D., Bess, M., Turano, A., Uldrick, T., Polizzotto, M.N., Wyvill, K.M.*, et al.* (2014). Heterogeneity and breadth of host antibody response to KSHV infection demonstrated by systematic analysis of the KSHV proteome. PLoS pathogens *10*, e1004046.
Lee, J.H., You, J., Dobrota, E., and Skalnik, D.G. (2010). Identification and characterization of a novel human PP1 phosphatase complex. The Journal of biological chemistry *285*, 24466-24476.
Lee, S., Salwinski, L., Zhang, C., Chu, D., Sampankanpanich, C., Reyes, N.A., Vangeloff, A., Xing, F., Li, X., Wu, T.T.*, et al.* (2011). An integrated approach to elucidate the intra-viral and viral-cellular protein interaction networks of a gamma-herpesvirus. PLoS pathogens *7*, e1002297.
Lehman, I.R., and Boehmer, P.E. (1999). Replication of herpes simplex virus DNA. The Journal of biological chemistry *274*, 28059-28062.
Li, H., Wang, H., and Nicholas, J. (2001). Detection of direct binding of human herpesvirus 8-encoded interleukin-6 (vIL-6) to both gp130 and IL-6 receptor (IL-6R) and identification of amino acid residues of vIL-6 important for IL-6R-dependent and -independent signaling. Journal of virology *75*, 3325-3334.
Liang, Y., Chang, J., Lynch, S.J., Lukac, D.M., and Ganem, D. (2002). The lytic switch protein of KSHV activates gene expression via functional interaction with RBP-Jkappa (CSL), the target of the Notch signaling pathway. Genes & development *16*, 1977-1989.
Lim, C., Gwack, Y., Hwang, S., Kim, S., and Choe, J. (2001). The transcriptional activity of cAMP response element-binding protein-binding protein is modulated by the

latency associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus. The Journal of biological chemistry *276*, 31016-31022.
Lim, C., Lee, D., Seo, T., Choi, C., and Choe, J. (2003). Latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus functionally interacts with heterochromatin protein 1. The Journal of biological chemistry *278*, 7397-7405.
Lin, Q., Lo, C.G., Cerione, R.A., and Yang, W. (2002). The Cdc42 target ACK2 interacts with sorting nexin 9 (SH3PX1) to regulate epidermal growth factor receptor degradation. The Journal of biological chemistry *277*, 10134-10138.
Lin, R., Genin, P., Mamane, Y., Sgarbanti, M., Battistini, A., Harrington, W.J., Jr., Barber, G.N., and Hiscott, J. (2001). HHV-8 encoded vIRF-1 represses the interferon antiviral response by blocking IRF-3 recruitment of the CBP/p300 coactivators. Oncogene *20*, 800-811.
Listerman, I., Sapra, A.K., and Neugebauer, K.M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nature structural & molecular biology *13*, 815-822.
Liu, L., Eby, M.T., Rathore, N., Sinha, S.K., Kumar, A., and Chaudhary, P.M. (2002). The human herpes virus 8-encoded viral FLICE inhibitory protein physically associates with and persistently activates the Ikappa B kinase complex. The Journal of biological chemistry *277*, 13745-13751.
Maher, M.C., and Hernandez, R.D. (2014). A MOSAIC of methods: Improving ortholog detection through integration of algorithmic diversity. J arXiv preprint *arXiv: 1309.2319*.
Malik, P., Blackbourn, D.J., and Clements, J.B. (2004). The evolutionarily conserved Kaposi's sarcoma-associated herpesvirus ORF57 protein interacts with REF protein and acts as an RNA export factor. The Journal of biological chemistry *279*, 33001-33011.
Malik, P., Tabarraei, A., Kehlenbach, R.H., Korfali, N., Iwasawa, R., Graham, S.V., and Schirmer, E.C. (2012). Herpes simplex virus ICP27 protein directly interacts with the nuclear pore complex through Nup62, inhibiting host nucleocytoplasmic transport pathways. The Journal of biological chemistry *287*, 12277-12292.
Maston, G.A., Zhu, L.J., Chamberlain, L., Lin, L., Fang, M., and Green, M.R. (2012). Non-canonical TAF complexes regulate active promoters in human embryonic stem cells. eLife *1*, e00068.
Mavankal, G., Ignatius Ou, S.H., Oliver, H., Sigman, D., and Gaynor, R.B. (1996). Human immunodeficiency virus type 1 and 2 Tat proteins specifically interact with RNA polymerase II. Proceedings of the National Academy of Sciences of the United States of America *93*, 2089-2094.
McCormick, C., and Ganem, D. (2005). The kaposin B protein of KSHV activates the p38/MK2 pathway and stabilizes cytokine mRNAs. Science *307*, 739-741.
Mesri, E.A., Feitelson, M.A., and Munger, K. (2014). Human Viral Oncogenesis: A Cancer Hallmarks Analysis. Cell host & microbe *15*, 266-282.
Mita, P., Savas, J.N., Djouder, N., Yates, J.R., 3rd, Ha, S., Ruoff, R., Schafler, E.D., Nwachukwu, J.C., Tanese, N., Cowan, N.J.*, et al.* (2011). Regulation of androgen receptor-mediated transcription by RPB5 binding protein URI/RMP. Molecular and cellular biology *31*, 3639-3652.

Moore, P.S., Boshoff, C., Weiss, R.A., and Chang, Y. (1996). Molecular mimicry of human cytokine and cytokine response pathway genes by KSHV. Science *274*, 1739-1744.
Muller, M., and Demeret, C. (2012). The HPV E2-Host Protein-Protein Interactions: A Complex Hijacking of the Cellular Network. The open virology journal *6*, 173-189.
Muller, M., Jacob, Y., Jones, L., Weiss, A., Brino, L., Chantier, T., Lotteau, V., Favre, M., and Demeret, C. (2012). Large scale genotype comparison of human papillomavirus E2-host interaction networks provides new insights for e2 molecular functions. PLoS pathogens *8*, e1002761.
Murakami, K., Elmlund, H., Kalisman, N., Bushnell, D.A., Adams, C.M., Azubel, M., Elmlund, D., Levi-Kalisman, Y., Liu, X., Gibbons, B.J.*, et al.* (2013). Architecture of an RNA polymerase II transcription pre-initiation complex. Science *342*, 1238724.
Murakami, Y., Yamagoe, S., Noguchi, K., Takebe, Y., Takahashi, N., Uehara, Y., and Fukazawa, H. (2006). Ets-1-dependent expression of vascular endothelial growth factor receptors is activated by latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus through interaction with Daxx. The Journal of biological chemistry *281*, 28113-28121.
Myoung, J., and Ganem, D. (2011). Generation of a doxycycline-inducible KSHV producer cell line of endothelial origin: maintenance of tight latency with efficient reactivation upon induction. Journal of virological methods *174*, 12-21.
Navratil, V., de Chassey, B., Combe, C.R., and Lotteau, V. (2011). When the human viral infectome and diseasome networks collide: towards a systems biology platform for the aetiology of human diseases. BMC systems biology *5*, 13.
Ohno, S., Steer, B., Sattler, C., and Adler, H. (2012). ORF23 of murine gammaherpesvirus 68 is non-essential for in vitro and in vivo infection. The Journal of general virology *93*, 1076-1080.
Oshiro, N., Takahashi, R., Yoshino, K., Tanimura, K., Nakashima, A., Eguchi, S., Miyamoto, T., Hara, K., Takehana, K., Avruch, J.*, et al.* (2007). The proline-rich Akt substrate of 40 kDa (PRAS40) is a physiological substrate of mammalian target of rapamycin complex 1. The Journal of biological chemistry *282*, 20329-20339.
Osterberg, S., del Peso-Santos, T., and Shingler, V. (2011). Regulation of alternative sigma factor use. Annual review of microbiology *65*, 37-55.
Ottinger, M., Christalla, T., Nathan, K., Brinkmann, M.M., Viejo-Borbolla, A., and Schulz, T.F. (2006). Kaposi's sarcoma-associated herpesvirus LANA-1 interacts with the short variant of BRD4 and releases cells from a BRD4- and BRD2/RING3-induced G1 cell cycle arrest. Journal of virology *80*, 10772-10786.
Ouyang, P., Rakus, K., van Beurden, S.J., Westphal, A.H., Davison, A.J., Gatherer, D., and Vanderplasschen, A.F. (2014). IL-10 encoded by viruses: a remarkable example of independent acquisition of a cellular gene by viruses and its subsequent evolution in the viral genome. The Journal of general virology *95*, 245-262.
Paudel, N., Sadagopan, S., Balasubramanian, S., and Chandran, B. (2012). Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen and angiogenin interact with common host proteins, including annexin A2, which is essential for survival of latently infected cells. Journal of virology *86*, 1589-1607.

Paulson, M., Press, C., Smith, E., Tanese, N., and Levy, D.E. (2002). IFN-Stimulated transcription through a TBP-free acetyltransferase complex escapes viral shutoff. Nature cell biology *4*, 140-147.
Peng, L., Yuan, Z., Ling, H., Fukasawa, K., Robertson, K., Olashaw, N., Koomen, J., Chen, J., Lane, W.S., and Seto, E. (2011). SIRT1 deacetylates the DNA methyltransferase 1 (DNMT1) protein and alters its activities. Molecular and cellular biology *31*, 4720-4734.
Perng, Y.C., Campbell, J.A., Lenschow, D.J., and Yu, D. (2014). Human cytomegalovirus pUL79 is an elongation factor of RNA polymerase II for viral gene transcription. PLoS pathogens *10*, e1004350.
Pichlmair, A., Kandasamy, K., Alvisi, G., Mulhern, O., Sacco, R., Habjan, M., Binder, M., Stefanovic, A., Eberle, C.A., Goncalves, A.*, et al.* (2012). Viral immune modulators perturb the human molecular network by common and unique strategies. Nature *487*, 486-490.
Platt, G.M., Cannell, E., Cuomo, M.E., Singh, S., and Mittnacht, S. (2000). Detection of the human herpesvirus 8-encoded cyclin protein in primary effusion lymphoma-derived cell lines. Virology *272*, 257-266.
Platt, G.M., Simpson, G.R., Mittnacht, S., and Schulz, T.F. (1999). Latent nuclear antigen of Kaposi's sarcoma-associated herpesvirus interacts with RING3, a homolog of the Drosophila female sterile homeotic (fsh) gene. Journal of virology *73*, 9789-9795.
Radkov, S.A., Kellam, P., and Boshoff, C. (2000). The latent nuclear antigen of Kaposi sarcoma-associated herpesvirus targets the retinoblastoma-E2F pathway and with the oncogene Hras transforms primary rat cells. Nature medicine *6*, 1121-1127.
Rajagopala, S.V., Casjens, S., and Uetz, P. (2011). The protein interaction map of bacteriophage lambda. BMC microbiology *11*, 213.
Rappocciolo, G., Jenkins, F.J., Hensler, H.R., Piazza, P., Jais, M., Borowski, L., Watkins, S.C., and Rinaldo, C.R., Jr. (2006). DC-SIGN is a receptor for human herpesvirus 8 on dendritic cells and macrophages. J Immunol *176*, 1741-1749.
Rice, S.A., and Knipe, D.M. (1990). Genetic evidence for two distinct transactivation functions of the herpes simplex virus alpha protein ICP27. Journal of virology *64*, 1704-1715.
Rozenblatt-Rosen, O., Deo, R.C., Padi, M., Adelmant, G., Calderwood, M.A., Rolland, T., Grace, M., Dricot, A., Askenazi, M., Tavares, M.*, et al.* (2012). Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. Nature *487*, 491-495.
Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). CORUM: the comprehensive resource of mammalian protein complexes--2009. Nucleic acids research *38*, D497-501.
Sathish, N., Zhu, F.X., and Yuan, Y. (2009). Kaposi's sarcoma-associated herpesvirus ORF45 interacts with kinesin-2 transporting viral capsid-tegument complexes along microtubules. PLoS pathogens *5*, e1000332.
Seo, T., Lee, D., Shim, Y.S., Angell, J.E., Chidambaram, N.V., Kalvakolanu, D.V., and Choe, J. (2002). Viral interferon regulatory factor 1 of Kaposi's sarcoma-associated

herpesvirus interacts with a cell death regulator, GRIM19, and inhibits interferon/retinoic acid-induced cell death. Journal of virology *76*, 8797-8807.
Shamay, M., Liu, J., Li, R., Liao, G., Shen, L., Greenway, M., Hu, S., Zhu, J., Xie, Z., Ambinder, R.F.*, et al.* (2012). A protein array screen for Kaposi's sarcoma-associated herpesvirus LANA interactors links LANA to TIP60, PP2A activity, and telomere shortening. Journal of virology *86*, 5179-5191.
Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics *27*, 431-432.
Song, E.J., Werner, S.L., Neubauer, J., Stegmeier, F., Aspden, J., Rio, D., Harper, J.W., Elledge, S.J., Kirschner, M.W., and Rape, M. (2010). The Prp19 complex and the Usp4Sart3 deubiquitinating enzyme control reversible ubiquitination at the spliceosome. Genes & development *24*, 1434-1447.
Song, M.J., Hwang, S., Wong, W.H., Wu, T.T., Lee, S., Liao, H.I., and Sun, R. (2005). Identification of viral genes essential for replication of murine gamma-herpesvirus 68 using signature-tagged mutagenesis. Proceedings of the National Academy of Sciences of the United States of America *102*, 3805-3810.
Song, M.S., Salmena, L., Carracedo, A., Egia, A., Lo-Coco, F., Teruya-Feldstein, J., and Pandolfi, P.P. (2008). The deubiquitinylation and localization of PTEN are regulated by a HAUSP-PML network. Nature *455*, 813-817.
Sowa, M.E., Bennett, E.J., Gygi, S.P., and Harper, J.W. (2009). Defining the human deubiquitinating enzyme interaction landscape. Cell *138*, 389-403.
Stefansson, B., and Brautigan, D.L. (2006). Protein phosphatase 6 subunit with conserved Sit4-associated protein domain targets IkappaBepsilon. The Journal of biological chemistry *281*, 22624-22634.
Stern-Ginossar, N., Weisburd, B., Michalski, A., Le, V.T., Hein, M.Y., Huang, S.X., Ma, M., Shen, B., Qian, S.B., Hengel, H.*, et al.* (2012). Decoding human cytomegalovirus. Science *338*, 1088-1093.
Sturzl, M., Gaus, D., Dirks, W.G., Ganem, D., and Jochmann, R. (2013). Kaposi's sarcoma-derived cell line SLK is not of endothelial origin, but is a contaminant from a known renal carcinoma cell line. International journal of cancer Journal international du cancer *132*, 1954-1958.
Swift, S., Lorens, J., Achacoso, P., and Nolan, G.P. (2001). Rapid production of retroviruses for efficient gene delivery to mammalian cells using 293T cell-based systems. Current protocols in immunology / edited by John E Coligan [et al] *Chapter 10*, Unit 10 17C.
Szpiech, Z.A., and Hernandez, R.D. (2014). selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. Molecular biology and evolution.
Takada, K., and Ono, Y. (1989). Synchronous and sequential activation of latently infected Epstein-Barr virus genomes. Journal of virology *63*, 445-449.
Takramah, D., Seiffert, B.M., Schaller, S., Vigneron, M., and Hacker, G. (2003). Baculovirus P35 interacts with a subunit of human RNA polymerase II and can enhance promoter activity in human cells. The Journal of general virology *84*, 3011-3019.

Tanaka, K., Kondo, T., Torigoe, S., Okada, S., Mukai, T., and Yamanishi, K. (1994). Human herpesvirus 7: another causal agent for roseola (exanthem subitum). J Pediatr *125*, 1-5.
Tang, S., Yamanegi, K., and Zheng, Z.M. (2004). Requirement of a 12-base-pair TATT-containing sequence and viral lytic DNA replication in activation of the Kaposi's sarcoma-associated herpesvirus K8.1 late promoter. Journal of virology *78*, 2609-2614.
Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Critical reviews in biochemistry and molecular biology *41*, 105-178.
Thome, M., Schneider, P., Hofmann, K., Fickenscher, H., Meinl, E., Neipel, F., Mattmann, C., Burns, K., Bodmer, J.L., Schroter, M.*, et al.* (1997). Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors. Nature *386*, 517-521.
Uetz, P., Dong, Y.A., Zeretzke, C., Atzler, C., Baiker, A., Berger, B., Rajagopala, S.V., Roupelieva, M., Rose, D., Fossum, E.*, et al.* (2006). Herpesviral protein networks and their interaction with the human proteome. Science *311*, 239-242.
Vance, R.E., Isberg, R.R., and Portnoy, D.A. (2009). Patterns of pathogenesis: discrimination of pathogenic and nonpathogenic microbes by the innate immune system. Cell host & microbe *6*, 10-21.
Verschueren, E., Von Dollen, J., Cimermancic, P., Gulbahce, N., Sali, A., and Krogan, N.J. (2014). Analyzing AP-MS data with Mass Spectrometry Interaction Statistics (MIST). Current Protocols *In press*.
Verschuren, E.W., Jones, N., and Evan, G.I. (2004). The cell cycle and how it is steered by Kaposi's sarcoma-associated herpesvirus cyclin. The Journal of general virology *85*, 1347-1361.
Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS biology *4*, e72.
Wagner, E., Clement, S.L., and Lykke-Andersen, J. (2007). An unconventional human Ccr4-Caf1 deadenylase complex in nuclear cajal bodies. Molecular and cellular biology *27*, 1686-1695.
Wald, A., and Corey, L. (2007). Persistence in the population: epidemiology, transmission. In Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis, A. Arvin, G. Campadelli-Fiume, E. Mocarski, P.S. Moore, B. Roizman, R. Whitley, and K. Yamanishi, eds. (Cambridge).
Wong, E., Wu, T.T., Reyes, N., Deng, H., and Sun, R. (2007). Murine gammaherpesvirus 68 open reading frame 24 is required for late gene expression after DNA replication. Journal of virology *81*, 6761-6764.
Wong-Ho, E., Wu, T.T., Davis, Z.H., Zhang, B., Huang, J., Gong, H., Deng, H., Liu, F., Glaunsinger, B., and Sun, R. (2014). Unconventional Sequence Requirement for Viral Late Gene Core Promoters of Murine Gammaherpesvirus 68. Journal of virology.
Wu, T.T., Park, T., Kim, H., Tran, T., Tong, L., Martinez-Guzman, D., Reyes, N., Deng, H., and Sun, R. (2009). ORF30 and ORF34 are essential for expression of late genes in murine gammaherpesvirus 68. Journal of virology *83*, 2265-2273.
Wyrwicz, L.S., and Rychlewski, L. (2007). Identification of Herpes TATT-binding protein. Antiviral research *75*, 167-172.

Yamanishi, K., Okuno, T., Shiraki, K., Takahashi, M., Kondo, T., Asano, Y., and Kurata, T. (1988). Identification of human herpesvirus-6 as a causal agent for exanthem subitum. Lancet *1*, 1065-1067.
Zhou, C., and Knipe, D.M. (2002). Association of herpes simplex virus type 1 ICP8 and ICP27 proteins with cellular RNA polymerase II holoenzyme. Journal of virology *76*, 5893-5904.
Zhou, H., Kaplan, T., Li, Y., Grubisic, I., Zhang, Z., Wang, P.J., Eisen, M.B., and Tjian, R. (2013). Dual functions of TAF7L in adipocyte differentiation. eLife *2*, e00170.
Zhu, F.X., King, S.M., Smith, E.J., Levy, D.E., and Yuan, Y. (2002). A Kaposi's sarcoma-associated herpesviral protein inhibits virus-mediated induction of type I interferon by blocking IRF-7 phosphorylation and nuclear accumulation. Proceedings of the National Academy of Sciences of the United States of America *99*, 5573-5578.