

CHARACTERIZATION OF THE INTERACTIONS BETWEEN KAPOSI'S  
SARCOMA-ASSOCIATED HERPESVIRUS ORF57  
AND ITS RNA PARTNERS

APPROVED BY SUPERVISORY COMMITTEE

---

Nicholas K. Conrad, Ph.D.

---

David Corey, Ph.D.

---

Ivan D'Orso, Ph.D.

---

Beatriz Fontoura, Ph.D.

---

## DEDICATION

First, I would like to thank Dr. Nick Conrad for all the great experiences and loud laughs I've had in the lab. You are a great scientist, but an even more patient man. So, thank you. I would like to thank our collaborators Dr. Yang Xie and her graduate student Tao Wang, without them none of the HITS-CLIP bioinformatic analysis would be possible. To the members of my committee: Drs. David Corey, Ivan D'Orso, Beatriz Fontoura, and Wade Winkler. I am grateful for all their feedback and kind comments throughout the struggle that was my project. Even though Dr. Wade Winkler only formed part of my committee for a short time, he did persuade me to join the biochemistry program, which ended up being a great piece of advice.

For those who haven't met the Conrad Lab members, being in our lab is definitely an experience. To Brooke, Denish, Xiangmei, Sarah, Olga, Stefan, Kathryn, and all of our rotation and summer students... It has been a lot of fun. I hope you get bored once I graduate, but also that you maintain the long-standing lab tradition of pranking Nick. Don't disappoint me.

To René, Kalos, Mario, Raúl, Chikio, Walo, Totochi, Grissel, & Majo: it's been more than 15 years of continuous and uninterrupted friendship. I could not have asked for a more interesting, yet weird group of friends... you guys complete me.

A mis padres, les agradezco todos los años de esfuerzo y sacrificio que han tenido para que nosotros, sus hijos, gocemos de comodidades y oportunidades que simplemente, no merecemos. Gracias a ustedes he vivido una vida extraordinaria: Karate, Hawaiano, Tahitiano, Jazz, Ballet, Flamenco, piano, competencias académicas, atletismo, campamentos... ustedes me han enseñado el placer de vivir y poder compartir todos esos momentos en familia. Kengi, de tí he aprendido que la vida es una gran aventura. Sayu... la mayoría de la gente tiene la fortuna de conocerte, pero yo tengo el placer de tenerte como mi hermana.

Pon, sin duda alguna, eres el hombre más paciente del mundo. Gracias por todos esos momentos de silencio en los que me dejas desahogarme sabiendo que tú tienes la razón. Gracias por asegurarte de que cumpla todas mis metas, sin importar que tan ridículas sean. Eso es amor del bueno.

CHARACTERIZATION OF THE INTERACTIONS BETWEEN KAPOSI'S  
SARCOMA-ASSOCIATED HERPESVIRUS ORF57  
AND ITS RNA PARTNERS

by

EMI SEI

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August, 2014

Copyright

by

Emi Sei, 2014

All Rights Reserved

CHARACTERIZATION OF THE INTERACTIONS BETWEEN KAPOSI'S  
SARCOMA-ASSOCIATED HERPESVIRUS ORF57  
AND ITS RNA PARTNERS

Emi Sei, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2014

Supervising Professor: Nicholas K. Conrad, Ph.D.

Kaposi's sarcoma-associated herpesvirus (KSHV; HHV-8) is a human gammaherpesvirus and the etiological agent of Kaposi's sarcoma (KS), the most common AIDS-associated malignancy. Like all herpesviruses, KSHV has evolved mechanisms to modulate both host and viral gene expression. The essential and multifunctional KSHV ORF57 protein has been reported to enhance viral gene expression at multiple levels including transcription, splicing, mRNA export, RNA stability, and translation. At least in some cases, direct interactions between ORF57 and its target RNAs are necessary for ORF57-mediated upregulation of viral gene expression. The work highlighted in this

document reveals our current efforts to study the elements driving ORF57's binding specificity. We started by studying a known ORF57 target: the KSHV polyadenylated nuclear (PAN) RNA, a nuclear non-coding transcript of unknown function that is highly expressed during lytic stage. We first devised an *in vitro* binding assay to identify the regions in PAN RNA that were bound by ORF57. These PAN RNA fragments were also inserted into the 3' UTR of an intronless  $\beta$ -globin reporter to test ORF57 responsiveness *in vivo*. Our analyses revealed an ORF57 responsive element (ORE) at the 5' end of PAN RNA that we hypothesize functions as a high-affinity binding site to recruit ORF57. Next, we optimized the high-throughput sequencing of RNAs isolated by crosslinking and immunoprecipitation (HITS-CLIP) protocol to identify novel host and viral targets of ORF57 in the context of viral infection. Bioinformatic analysis of potential host ORF57 targets reveals that ORF57 binding is enriched near the 5' end of the transcripts and often close to the first exon-intron junction. Preferential binding at the 5' end is also seen for PAN RNA. However, our data suggests that ORF57 binding to other viral genes can be promiscuous and that, in some cases, binding can occur at multiple sites across the target RNAs. Through these studies, we hope to provide further insight into the requirements for ORF57 binding and potentially shed light into the mechanisms controlling gene expression of this oncogenic virus.

## TABLE OF CONTENTS

ABSTRACT .....	v
PRIOR PUBLICATIONS .....	IX
LIST OF FIGURES .....	X
LIST OF TABLES .....	XII
LIST OF APENDICES .....	XIII
LIST OF DEFINITIONS .....	XV
CHAPTER ONE: INTRODUCTION.....	1
CHAPTER TWO: REVIEW OF THE LITERATURE.....	4
KAPOSI'S SARCOMA-ASSOCIATED HERPESVIRUS (KSHV).....	4
KSHV GENOME AND GENE EXPRESSION.....	5
KSHV ORF57.....	8
ORF57 PROTEIN MOTIFS AND FUNCTIONS.....	8
ORF57 IS INVOLVED IN SPLICING, EXPORT AND TRANSLATION .....	11
ORF57 PLAYS A PIVOTAL ROLE IN KSHV GENE EXPRESSION AND RNA STABILITY .....	13
HOST SHUTOFF .....	14
KSHV PAN RNA .....	15
CHAPTER THREE: METHODOLOGY .....	18
PLASMIDS.....	18
CELL CULTURE AND TRANSFECTIONS .....	19
ANTIBODIES .....	20

LOW THROUGHPUT UV CROSSLINKING AND IMMUNOPRECIPITATION ....	21
ISOLATION OF NEWLY MADE TRANSCRIPTS .....	22
QRT-PCR.....	23
LABEL TRANSFER ASSAYS.....	24
WESTERN BLOTTING.....	26
ORF57 HITS-CLIP.....	27
HITS-CLIP BIOINFORMATIC ANALYSIS .....	27
CHAPTER FOUR: DELINIATION OF A CORE RNA ELEMENT REQUIRED FOR KAPOSI'S SARCOMA-ASSOCIATED HERPESVIRUS ORF57 BINDING AND ACTIVITY .....	
INTRODUCTION .....	32
RESULTS .....	34
DISCUSSION .....	50
CHAPTER FIVE: IDENTIFICATION OF KAPOSI'S SARCOMA-ASSOCIATED HERPESVIRUS ORF57 RNA TARGETS USING HITS-CLIP .....	
INTRODUCTION .....	55
RESULTS .....	58
DISCUSSION .....	84
CHAPTER SIX: CONCLUSIONS AND FUTURE DIRECTIONS .....	88
REFERENCES .....	135

## PRIOR PUBLICATIONS

**Sei, E.; Conrad, N. K.,** UV Cross-Linking of Interacting RNA and Protein in Cultured Cells.

*Methods in Enzymology.* **2014**, *539*, 53-66.

**Hunter, O.V.; Sei, E.; Richardson, R.B.; Conrad, N.K.** ChIP-chip analysis suggests

functional cooperation between ORF57 and K-bZIP. *Journal of Virology.* **2013**, *87*(7), 4005-

16.

**Sei, E.; Conrad, N. K.,** Delineation of a core RNA element required for Kaposi's sarcoma-

associated herpesvirus ORF57 binding and activity. *Virology* **2011**, *419* (2), 107-16.

## LIST OF FIGURES

FIGURE 1 – DIAGRAM OF THE KSHV LIFE CYCLE.....	5
FIGURE 2 – PUTATIVE ORF57 FUNCTIONAL MOTIFS .....	9
FIGURE 3 – ORF57 BINDS THE 5’ END OF PAN RNA IN LYTICALLY REACTIVATED CELLS .....	36
FIGURE 4 – ORF57 BINDS DIRECTLY TO THE ORE <i>IN VITRO</i> .....	38
FIGURE 5 – THE FIRST 79 NT OF PAN RNA ARE SUFFICIENT FOR ORE ACTIVITY IN A HETEROLOGOUS TRANSCRIPT .....	41
FIGURE 6 – A 30 NT STEM-LOOP STRUCTURE IS SUFFICIENT FOR ORE ACTIVITY AND BINDING .....	43
FIGURE 7 – THE CORE ORE IS NECESSARY FOR ORF57 RESPONSIVENESS IN PAN RNA .....	46
FIGURE 8 – POINT MUTATIONS IN THE LOOP PORTION OF SL2-T ABROGRATE ORF57 BINDING AND RESPONSE .....	49
FIGURE 9 – EXPERIMENTAL OUTLINE OF ORF57 HITS-CLIP .....	59
FIGURE 10 – REPRESENTATIVE ORF57 HITS-CLIP RESULTS.....	62
FIGURE 11 – RNA FRAGMENTS RECOVERED FROM NITROCELLULOSE MEMBRANE .....	63
FIGURE 12 – NORTHERN BLOTS OF RNAS EXTRACTED FROM THE NITROCELLULOSE MEMBRANE .....	64
FIGURE 13 – ILUMINA MAPPING STATISTICS .....	66

FIGURE 14 – ORF57 HITS-CLIP UNVEILS A CONSIDERABLE NUMBER OF BOUND VIRAL RNAS.....	67
FIGURE 15 – SCATTER PLOT AND PEARSON’S CORRELATION COEFFICIENTS FOR INPUT AND PELLET (HITS-CLIP) SAMPLES .....	70
FIGURE 16 – ENRICHED HUMAN CLUSTERS.....	71
FIGURE 17 – IGV TRACES OF ORF57-BOUND RNAS .....	75
FIGURE 18 – AVERAGE EXON COUNT IN ORF57-BOUND TRANSCRIPTS .....	76
FIGURE 19 – STEADY-STATE RNA LEVELS .....	77
FIGURE 20 – REPRESENTATIVE WESTERN BLOT FROM SAMPLES IN FIGURE 19 .....	79
FIGURE 21 – NEWLY MADE TRANSCRIPTS IN TREX BCBL1-RTA CELLS .....	82
FIGURE 22 – STEADY-STATE RNA LEVELS IN HEK293 CELLS.....	83

LIST OF TABLES

TABLE 1 – OLIGONUCLEOTIDES USED FOR PLASMID CONSTRUCTION ..... 30

## LIST OF APPENDICES

APPENDIX A – HITS-CLIP RECIPES .....	94
HOW TO MAKE <sup>32</sup> P LABELED DNA LADDER.....	96
APPENDIX B – PREPARATION OF CELLS FOR HITS-CLIP .....	97
APPENDIX C– HITS-CLIP PROTOCOL FOR ILLUMINA SEQUENCING.....	99
CELL LYSIS .....	99
DNASE AND MNASE TREATMENT .....	100
IMMUNOPRECIPITATION.....	101
PNK LABELING OF PELLETS.....	103
VISUALIZATION OF PROTEIN-RNA COMPLEXES .....	105
PROK TREATMENT.....	107
DNASE TREATMENT OF PELLETS .....	109
RNA SIZE SELECTION OF PELLETS .....	110
DNASE TREATMENT OF INPUTS.....	113
RIBO-ZERO MAGNETIC KIT PROTOCOL FOR INPUTS .....	115
MNASE TREATMENT OF INPUTS .....	118
PNK LABELING OF INPUTS .....	119
RNA SIZE SELECTION OF INPUTS.....	120
ILLUMINA LIBRARY SAMPLE PREPARATION FOR INPUTS AND PELLETS.	120
MODIFIED ILLUMINA TRUSEQ STRANDED MRNA SAMPLE PREPARATION	
GUIDE .....	122
1 <sup>ST</sup> CDNA STRAND .....	123

2 <sup>ND</sup> CDNA STRAND .....	124
ADENYLATE 3' ENDS.....	126
LIGATE ADAPTERS .....	127
ENRICH DNA FRAGMENTS .....	131
AGILENT BIOANALYZER –TESTING LIBRARY SIZE AND YIELD .....	133
APPENDIX D – HITS-CLIP WORKFLOW CHART FOR ILLUMINA SEQUENCING	134

## LIST OF DEFINITIONS

KS – Kaposi sarcoma

KSHV – Kaposi's sarcoma-associated herpesvirus

ORF57 – Open reading frame 57

EBV – Epstein-Barr virus

HVS – Herpes virus saimiri

HSV – Herpes simplex virus

HIV – Human immunodeficiency virus

UTR – Untranslated region

EJC – Exon-junction complex

hTREX – human transcription-export complex

ORE – ORF57 Responsive element

PAN RNA – Polyadenylated nuclear RNA

MNase – Micrococcal Nuclease

ProK – Proteinase K

RNAP – RNA polymerase

CBC – Cap binding complex

CBP80 – cap-binding complex protein

# CHAPTER ONE

## Introduction

Kaposi's sarcoma-associated herpesvirus (KSHV) is largely known for its association with Kaposi Sarcoma (KS), the most common AIDS-associated cancer [1]. KSHV is capable of establishing lifelong infections in its host, primarily in cells of endothelial and B-cell lineage [2-4]. The KSHV life cycle includes both a dormant (latent) and a productive (lytic) state. During latency, expression of a few viral genes is enough to maintain the viral genome. On the other hand, during lytic reactivation, KSHV orchestrates the ordered synthesis of numerous viral products that enable assembly of virion particles. This assortment of viral factors works in unison to manipulate the host cell and favor viral gene expression.

Two lytic factors have been studied for their involvement in viral gene expression: the ORF57 protein and PAN RNA, a long non-coding RNA. ORF57 is a multifunctional activator of gene expression with effects that extend from transcription to translation [5]. The importance of ORF57's function is accentuated by the presence of homologous proteins in every herpesvirus [6]. Multiple laboratories, including our own, have reported the contributions of this RNA binding protein at multiple stages of mRNA maturation. By the time an mRNA is exported, the newly made transcript associates with multi-protein complexes that mediate its splicing, polyadenylation, export, and translation. Hijacking these pathways is crucial to herpesviruses because they rely on host cellular machinery to express

their genes. Interestingly, a large portion of the KSHV genes produce intronless transcripts [7-9]. This presents a challenge to the virus given that splicing is often coupled with higher expression levels and export [10, 11]. Perhaps this is one of the reasons why ORF57 is essential to the virus, because it can preferentially upregulate intronless transcripts thereby bypassing the need for splicing [12-14].

The work highlighted in this document reveals our current efforts to study the elements driving ORF57's binding specificity. Previous work published by our laboratory demonstrates that ORF57 binds PAN RNA directly and increases its stability [12]. Additionally, we devised an *in vitro* binding assay to validate and study potential ORF57 RNA targets using whole cell extracts [15]. We identified a 30nt ORF57-responsive element (ORE) at the 5' end of PAN RNA that is both necessary and sufficient for ORF57 binding and up-regulation (Chapter Four). In our most current work, we seek to elucidate the interactions between ORF57 and novel RNA targets through High-Throughput Sequencing of RNAs isolated by Crosslinking Immunoprecipitation (HITS-CLIP) (Chapter Five). To do so, we induced lytic replication in KSHV-infected cells and exposed them to UV light to covalently crosslink ORF57-bound RNAs. Upon cell lysis, we treated the extracts with micrococcal nuclease (MNase) to partially digest the RNA (average size: ~50-300nt), immunoprecipitated ORF57 and constructed a strand-specific library of the crosslinked RNA targets for high-throughput sequencing. The HITS-CLIP protocol not only unveils a vast number of potential ORF57 targets during lytic infection, but it provides additional information about putative binding sites on the different target RNAs (due to the partial MNase digestion step). Our analysis of the immunoprecipitated host RNAs reveals CLIP tag

clusters are enriched at the 5' end of transcripts. A similar 5' end binding bias was observed when we analyzed a well-known viral target of ORF57: PAN RNA, which is consistent with previously published reports [12, 15]. More detailed mapping of the enriched 5' end clusters highlights that the great majority of these host sequences are located at the first exon-intron junction. We then monitored the RNA levels of a select number of potential ORF57 targets at different time points following lytic induction. Interestingly, when compared to the rapid disappearance of GAPDH and  $\beta$ -ACTIN controls, the levels of these ORF57-bound pre-mRNAs persist longer [16, 17]. By analyzing the RNAs that bind to ORF57, we hope to gain insight into the mechanisms driving ORF57 binding to its target RNAs and, at the same time, characterize the role this protein plays in KSHV pathogenesis.

## CHAPTER TWO

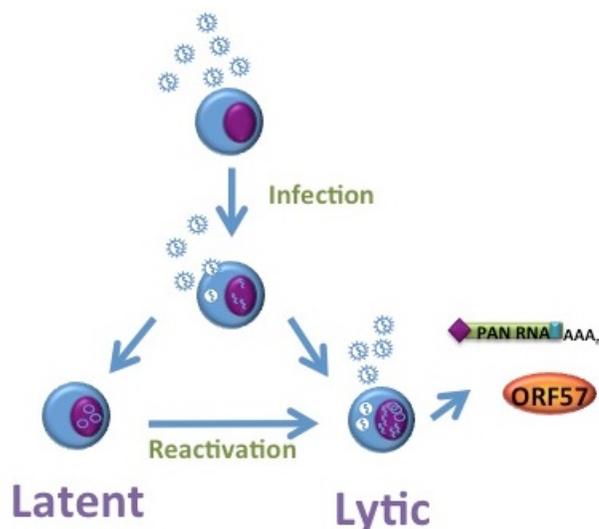
### Review of the Literature

#### **Kaposi's Sarcoma-Associated Herpesvirus (KSHV)**

Kaposi sarcoma (KS) lesions were first described by Moritz Kaposi [18]. KS is a vascular tumor present predominantly at mucocutaneous sites or on the skin [19, 20]. The incidence of KS in the general population is of 1 in 100,000, but it increases to 1 in 20 in human immunodeficiency virus (HIV-1) infected individuals [21, 22]. Nearly 120 years after the initial description of KS lesions, Chang and Moore identified the DNA virus responsible now known as KS-associated herpesvirus (KSHV; HHV-8) [1]. DNA sequence analysis classifies KSHV as a gammaherpesvirus, a subfamily known for their cell transforming capacity [7]. Additionally, KSHV is the etiological agent of other lymphoproliferative diseases such as primary effusion lymphoma (PEL) and some cases of multicentric Castleman's disease (MCD) [23, 24]. Efforts towards understanding the molecular mechanisms driving KSHV pathogenesis have not been without complications. In culture, KSHV can infect a wide variety of adherent cells such as endothelial cells, epithelial cells, and fibroblasts. However, B cell lines, which are thought to be the natural KSHV reservoir in infected patients, exhibit poor or low infectivity *in vitro* [2-4, 25]. This highlights the complexity of KSHV histopathology, which is characterized by heterogeneous cell composition [26, 27].

Target cell infection by KSHV is a multifarious process requiring carefully orchestrated interactions with multiple host factors. Unfortunately, most of the molecular details from these processes are vastly understudied. Viral entry is mediated by interactions among viral envelope glycoproteins and host cell receptors [28]. The viral capsid is then transported through the cytoplasm and docked at the nuclear pore. Once injected to the nuclear compartment, the naked genome (~165kb) is circularized and undergoes rapid chromatinization with help from host cellular factors [26, 28-31]. The circular KSHV genome is also called an episome or minichromosome. Establishing latency is not the only challenge encountered by the virus. Once inside the host cell, the virus has to evade the immune system, express its genes, generate multiple copies of its genome, and package them into virions in order to propagate into new host cells.

### KSHV Genome and Gene Expression



**Figure 1.** Diagram of the KSHV life cycle. KSHV virions infect B-cells and endothelial cells. Upon cell entry, the virus can enter the latency or the lytic state. *In vitro* a KSHV infected cell is capable of undergoing reactivation to switch from the latent to the lytic phase through expression of viral proteins (e.g. ORF50) or through chemical induction by agents like histone deacetylase inhibitors & phorbol esters.

KSHV successfully establishes lifelong infections in humans by alternating between its latent and lytic cycles (Figure 1). Latency is the predominant transcriptional program for the virus in cell culture, with ~1-5% of the cell population undergoing spontaneous or abortive lytic reactivation [3]. During latency, transcription is restricted to a handful of viral genes including: LANA/ORF73, viral cyclin D/ORF72, viral Fllice-inhibitory protein (FLIP)/ORF71, Kaposin/K12, a miRNA cluster and a few others. Efficient conservation and replication of the episome is due in part to latency-associated nuclear antigen (LANA). LANA binds directly to the terminal repeats (TR) in the viral genome and tethers it to the host chromosome [32, 33]. The TR region, also known as the origin of latent plasmid replication (ori-P), is both necessary and sufficient for episome maintenance during latency [34, 35].

Entrance to the lytic reactivation phase is required for production of viral particles. In fact, although most KSHV infected cells remain latent, continuous expression of lytic factors in a portion of the population seems to be required for disease progression [36, 37]. How exactly these factors function during pathogenesis remains debatable. However, once released from the lytic cell they are thought to act in a paracrine fashion promoting growth and angiogenesis. The expression cascade of the remaining viral genes is classified into: immediate early, delayed early and late [26]. Transcription of the immediate early genes occurs independently of *de novo* protein synthesis. The classic example for this category is the ORF50 gene. ORF50 (also known as replication and transcription activator, RTA) is the molecular switch responsible for virus reactivation and it is also important for viral DNA replication [38, 39]. Delayed early genes are involved in nucleic acid metabolism,

modulation of cellular functions and triggering lytic DNA replication [26]. The KSHV genome contains two functional origins of lytic DNA replication (oriLyt) [40, 41]. The one on the left portion of the genome (oriLyt-L) is situated between the K4.2 and K5 genes, while the one on the right (oriLyt-R) is an inverted repeat of oriLyt-L and falls between K12 and ORF71/vFLIP. Efficient lytic DNA replication requires recruitment of proteins including ORF50 and K8/K-bZIP [41-46]. Data from our laboratory suggests that ORF57 associates with the oriLyt-L, however, the contributions of this protein towards lytic DNA replication remain to be elucidated [47]. Lastly, genes that are classified as late code for virion structural proteins necessary for assembly and production of infectious viral particles.

Comprehensive KSHV gene expression profiles have surpassed the scope of traditional *in silico* gene predictions [48-50]. Recent technology developments in next generation sequencing coupled with the creation of cell lines tightly regulating latent expression have enabled further and more detailed analysis of KSHV gene expression. The Ganem laboratory performed RNA-seq and ribosome profiling on an iSLK.219 cell line [48, 49]. The iSLK.219 cells are an endothelial cell line latently infected with KSHV and expressing a doxycycline-inducible ORF50 transgene. The recombinant virus constitutively expresses GFP and also expresses RFP under the control of the PAN RNA promoter, which enable tracking of latent virus as well as the activation of lytic replication, respectively. In Arias *et al.*, iSLK.219 cells were used to study KSHV gene expression and authors found multiple sequencing reads mapping outside the standard genomic annotations [48, 51] suggesting that transcription during the lytic cycle is highly permissive. Some of their observations are consistent with a previous report comparing tiling arrays with a high-

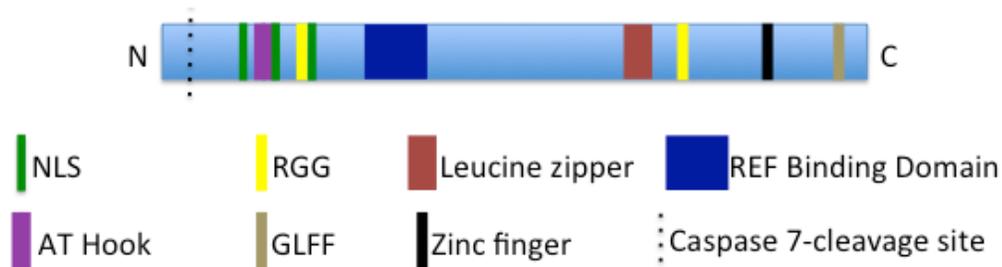
throughput proteomic analysis [51]. Several antisense transcripts were identified around ORFs like: vIL-6, ORFK5/K6, ORF58/59, ORF50, and PAN RNA. Genome annotation information including the location of RNA editing events, small ORFs (sORFs), and upstream ORFs (uORFs) were confirmed in these publications [48, 52, 53]. The presence of uORFs near the 5' UTR of the bicistronic transcript ORF35-ORF36 is consistent with previous findings [54]. These elements promote translation of the 3' gene via a continuous scanning mechanism. Overall, high-throughput technologies employed in current publications have identified novel sequence elements in the viral genome that expand previously reported genomic annotations (AF148805 and U75698) while, at the same time, increasing our understanding of KSHV gene expression.

## **KSHV ORF57**

### *ORF57 protein motifs and functions*

The open reading frame 57 (ORF57) encodes a monocistronic transcript containing a single short intron close to the 5' end of the transcript [6, 48]. This gene encodes for a 51kDa predominantly nuclear protein [also known as mRNA transcript accumulation (Mta) or KS-SM] that is conserved among all herpesviruses [6]. Homologs of this protein include better-characterized members such as HVS ORF57, EBV SM, hCMV UL69 and HSV-1 ICP27, all of which are distinguished by their multifunctional properties. ORF57 is involved in multiple aspects of viral gene expression ranging from transcription to translation; some of these are explained in greater detail in the next section [55-59]. In spite of the lack of structural conformation data for ORF57, computational predictions have been used to assist

in identification of functional features [60]. Computer modeling suggests the N-terminal region of ORF57 is mostly unstructured whereas the C-terminus is enriched with alpha helices. Multiple laboratories have identified and tested functional motifs in ORF57 including: three nuclear localization signals (NLSs), two RGG boxes, an AT-Hook, a leucine zipper domain, a Zn-finger domain, and a GLFF motif (Figure 2) [61, 62].



**Figure 2.** Putative ORF57 functional motifs: nuclear localization signals (NLSs, green), AT Hook (purple), Arg-Gly-Gly (RGG, yellow) boxes, REF binding domain (navy), leucine zipper (brown), zinc finger (black), and Gly-Leu-Phe-Phe (GLFF, tan). Dotted black line denotes caspase 7- cleavage site.

The three NLSs are clustered together at the N-terminal portion of ORF57 (Figure 2) [62]. The cellular distribution of ORF57 is predominantly nuclear, with high local concentrations in areas like nuclear speckles and the nucleolus [57, 61, 63]. Nuclear speckles are nuclear domains enriched with splicing factors [64]. The nucleolus is the site of rRNA synthesis and ribosome assembly, but some of its most highly expressed protein components (nucleolin and nucleophosmin) have been implicated in a plethora of cellular functions extending from chromatin remodeling to trafficking. In recent years, the number of reports describing how a broad spectrum of viruses are capable of manipulating nucleolar

components have increased significantly even though the functional significance of these interactions remains obscure [59, 65, 66].

ORF57 can function as a transcriptional transactivation factor, probably through a combination of direct and indirect associations with the DNA and/or DNA binding proteins [55, 67, 68]. The AT-Hook and leucine zipper domains present in ORF57 are commonly associated with DNA binding and are often found in transcription factors [69, 70]. An extensive amount of research has been done to study KSHV transcription factors like K8 and ORF50 and their role in lytic DNA replication [41-46]. ChIP-chip and co-immunoprecipitation experiments published from our laboratory show ORF57's association with the oriLyt-L, K8/K-bZIP promoter and K8 protein, adding a new layer of complexity to the previously described ORF50-K8-OriLyt interactions [41, 45, 47].

The ORF57 protein is capable of homomultimerization; even so, the contributions of multimerization in ORF57's functions remain unclear [58, 60, 71]. The presence of RGG boxes in RNA binding proteins has been previously documented [72, 73]. Deletion of this motif in ORF57 impairs target RNA binding, accumulation, and potentially even ORF57 multimerization. The remaining two motifs in ORF57 are not well characterized in KHSV. Zinc finger domains have been implicated in DNA and RNA binding and, in the case of herpesviruses, are conserved throughout all ORF57 homologs [74]. On the other hand, the GLFF motif is conserved only in gammaherpesviruses and in HVS plays a role in transactivation [74, 75]. One of the most noticeable host-induced ORF57 modifications occurs roughly 24hrs post-induction. Apoptotic signaling cascades activated by the host cell during lytic replication stimulate caspase 7-dependent cleavage of a <sup>30</sup>DETD<sup>33</sup> site at the N-

terminal region of ORF57 (Figure 2) [76]. The cleaved product is localized in the cytoplasm and results in a protein fragment unable to enhance viral gene expression.

*ORF57 is involved in splicing, export and translation*

In humans, as an RNA is being transcribed, RNA polymerase II orchestrates the recruitment and dissociation of protein factors involved in mRNA processing, many of which have a synergistic effect on gene expression [10]. Shortly after the mRNA is capped, the transcript is spliced and a protein complex is deposited next to the excision site; this set of proteins is called the exon-junction complex (EJC) [77]. Capping, association of the cap binding complex (CBC), and EJC deposition seem to be necessary for recruitment of the transcription-export (TREX) complex, which promotes export of the mRNA [78, 79]. Successful RNA processing results in a mature messenger ribonucleoprotein particle (mRNP) that is protected from RNA degradation pathways and is also export and translation competent [80]. In most cases, intron-containing cellular transcripts are expressed at significantly higher levels than their corresponding cDNA counterparts. In the recent decades, it has become evident that the presence of an intron and the act of undergoing splicing significantly enhance many stages of RNA metabolism including transcription, mRNA 3' end cleavage, polyadenylation, export, mRNA stability and translation [11].

There are inherent differences between the structure of human and herpes genes that may have driven herpesvirus evolution. The most striking difference being that most of the KSHV genes encode intronless transcripts [7-9]. This means that herpesviruses have to compensate for the absence of pre-mRNA splicing. ORF57 is utilized by KSHV to overcome

this roadblock. In order to stabilize and facilitate export of intronless viral mRNAs, ORF57 interacts with cellular RNA export adapter proteins REF/Aly and UIF – both individually capable of recruiting the TREX complex, which in turn enable export through the primary mRNA export receptor, TAP [12, 81-87]. The mechanism driving ORF57 binding to its target RNAs remains elusive, but our current efforts to address this question are discussed in chapters below.

There are multiple pieces of evidence suggesting ORF57 plays an important role in KSHV gene expression. ORF57 expression is essential for productive viral replication [83]. Experiments with ORF57-null mutants exhibit severely impaired lytic gene expression and are incapable of virion production [13, 14]. KSHV also intervenes in splicing through ORF57. Transfection experiments with an intron-containing K8 reporter have been used to show that ORF57 promotes removal of introns with weak splice sites [57]. Similar results were observed when intron-containing sections of human genes (e.g.  $\beta$ -globin) were cloned into the 3' end of a GFP reporter. *In vitro* binding experiments show the ORF57-pre-mRNA interaction occurs only in the presence of nuclear extracts, which indicates that the presence of host factors might be required for this interaction. Regardless of how ORF57 is initially recruited onto the RNAs, these results suggest that some of ORF57's functions are not restricted to viral transcripts, but could in fact extend to host transcripts as well.

In addition to its nuclear functions, once in the cytoplasm, ORF57 continues to recruit host factors to the intronless mRNP complex. In this cellular compartment, ORF57 interacts with PYM, a protein that associates with the small (40S) ribosomal subunit [88]. In the case of spliced host mRNAs two EJC proteins (Y14 and Magoh) remain bound to the mRNA after

export and subsequently recruit PYM [89]. Therefore, the ORF57-PYM interaction allows the intronless mRNP complex to bypass the need for EJC deposition and promotes its translation.

*ORF57 plays a pivotal role in KSHV gene expression and RNA stability*

The exact mechanisms governing RNA upregulation in the presence of ORF57 remains to be solved, but ORF57's ability to bind its targets correlates with RNA upregulation [12, 15]. Also, it is important to mention that the RNA stability effects observed with ORF57 are export-, promoter- and transcription-independent [12, 58, 81, 90]. The effect that ORF57 can exert on RNAs varies among transcripts, as not all RNAs are upregulated by this protein [55]. Data shows ORF57 preferentially upregulates intronless transcripts [12, 83, 90]. To date, there are several RNAs bound by ORF57 including intronless reporters and several viral transcripts. To name a few, the list of bound RNAs includes: chloramphenicol acetyltransferase (CAT),  $\beta$ -globin, ORF59, ORF56, HVS gB, PAN RNA, ORF50 and others [12, 15, 58, 62, 90, 91]. Additionally, transfection experiments of reporter RNA constructs demonstrate increased stability in the presence of ORF57, even after actinomycin D treatment (an inhibitor of transcription) [68]. The fact that ORF57 alone is sufficient to upregulate multiple heterologous transcripts strongly suggests this viral factor plays a key role in the regulation of KSHV gene expression.

In some cases, ORF57 recruitment is mediated by particular sequence elements on the RNAs, but analysis of a larger transcript pool is necessary to make any generalized conclusions [15, 92, 93]. Research from our laboratory has shown ORF57 preferentially

binds to the 5' end of PAN RNA [15]. In this region, we identified an ORF57 responsive element (ORE) that we hypothesize works as a high-affinity binding site for ORF57 (see Chapter Four) [12, 15]. Independent research from another laboratory identified a similar element in a second viral transcript: interleukin-6 (vIL-6) [93], suggesting that in some cases ORF57 recruitment might be favored or enhanced by the presence of particular sequence elements like the ORE (for comparison: loop sequence from PAN RNA ORE 5'-CCUAUGGGAUUUUG-3', see Chapter Four). Interestingly, studies involving ORF57 homologs show these related proteins might share conserved elements driving their RNA binding specificity. The EBV SM protein shows preferential binding to the 5' end of at least two of its target transcripts [94]. However, no analysis has been done to characterize these sequences. Systematic evolution of ligands by exponential enrichment (SELEX) was used to identify sequences preferentially bound by HVS ORF57, which revealed a purine rich sequence (5'-GGAG(A/G)G- 3') [92]. Ongoing experiments in our laboratory using high-throughput sequencing will aid in the classification of ORF57 ligands and the identification of elements recruiting this protein (Chapter Five).

### **Host Shutoff**

Shortly after lytic induction, KSHV targets a significant number of host mRNAs for degradation; this phenomena is known as the host shutoff [36, 95]. Host shutoff in KSHV infected cells is employed by the virus to prevent cellular antiviral responses as well as to favor viral gene expression. Experiments show host shutoff starts as early as ~8-12hrs postreactivation [37]. Nearly 95% of the host mRNAs are either moderately or severely

downregulated [36]. Remarkably, ~2% of the host mRNAs are strongly upregulated, but have no obvious commonalities or discernible patterns among them; these genes are often called escapees. Only a small number of these escapees can be attributed to the indirect effects of Kaposin B, which stabilizes AU-rich element (ARE)-containing mRNAs [96]. Nonetheless, the mechanism that targets host mRNAs for degradation and allows the remaining RNAs to escape remains to be determined.

The viral factor responsible for the host shutoff is the SOX (shutoff and exonuclease) protein, which has both DNase and endoribonuclease activities [37, 97, 98]. In the nucleus, its DNase function is essential for lytic viral DNA replication and it aids in the processing and packaging of newly replicated genome into the capsid [99]. In the cytoplasm, SOX cleaves target mRNAs internally creating an RNA fragment susceptible to Xrn1-mediated degradation [100]. How some of these targets escape shutoff or the exact mechanism guiding most host RNA transcripts to their inevitable destruction remains unknown. During Herpes simplex virus (HSV-1) infection, host mRNAs decrease through the action of two proteins: 1) vhs - a SOX homolog that degrades mRNAs and 2) ICP27 – an ORF57 homolog that inhibits splicing [101-104]. These observations are interesting because, in contrast to ICP27, ORF57 has been shown to promote viral mRNA splicing [14, 57]. Further analysis will be required in order to determine whether ORF57 shares some of ICP27's inhibitory functions.

### **KSHV PAN RNA**

Polyadenylated nuclear (PAN) RNA (also known as nut-1 or T1.1) is a target of ORF57. This non-coding RNA of unknown function displays characteristics of an mRNA: it

is transcribed by RNA polymerase II, has a m<sup>7</sup>G 5' cap, and is polyadenylated. PAN RNA is not spliced or exported, it accumulates in the nucleus during lytic replication and comprises up to ~80% of the total polyadenylated RNA (roughly 5x10<sup>5</sup> copies per cell) [105-108]. PAN RNA contains a 79-nt element near the 3' end of the transcript called the ENE (for expression and nuclear retention element). Deletion of the ENE destabilizes PAN RNA. Strikingly, insertion of the ENE into intronless  $\beta$ -globin results is sufficient to increase nuclear levels of this reporter [109, 110]. The ENE serves as a cis-acting stability element, where a U-rich sequence in the ENE interacts with the poly(A) tail to form a triple helix that protects the 3' end of the RNA from deadenylation and decay [111, 112]. Multiple ENE-like elements have been identified in DNA and RNA viruses of divergent lineages, as well as in human long non-coding RNAs [113, 114]. Their overall nucleotide sequences, ENE-like copy numbers, and position with respect to the poly(A) tail vary among the different organisms, but their characteristic accumulation due to protection from cellular decay pathways and the presence of U-rich sequences remains unanimous.

Multiple proteins have been found to promote PAN RNA abundance, among them are ORF57, PABPC1 and SOX. In transient transfection experiments, increasing amounts of ORF57 result in higher nuclear PAN RNA concentrations [12]. Similarly, transient transfection experiments with an ORF57-REF binding mutant (Figure 2) show that this upregulation is mediated through REF [81]. Research from our laboratory has shown both ORF57 and REF preferentially bind to the 5' end of PAN RNA in order to upregulate this transcript [15, 81]. Similar upregulation of PAN RNA is observed when SOX is transiently transfected into 293 tet-on cells [16, 115]. Albeit no direct association between SOX and

PAN RNA has been reported, the effects of SOX on PAN RNA levels are probably due to the change in PABPC1 localization induced by the host shutoff. During shutoff, PABPC1 relocates into the nucleus where it can then bind PAN RNA directly. Presumably, once in the nucleus, PABPC1 stabilizes PAN RNA by binding to its poly(A) tail. Nevertheless, through an unknown mechanism, PAN RNA is required for expression of late viral genes and virus production [16, 116].

## CHAPTER THREE

### Methodology

#### Plasmids

The  $\beta\Delta 1,2$ , CMV-WT, CMV- $\Delta 1$ , PAN $\Delta 79$ , PAN-WT and pcFl-ORF57II expression constructs were described previously [12, 110]. The PAN RNA inserts for the  $\beta$ -globin reporter assay were generated by PCR amplification of the corresponding sequences using the primers listed in the Supplementary Content with PAN-WT as a template. The PAN $\Delta 79$ - $\Delta$ SL2 construct (see below) was used as a PCR template for the  $\Delta$ SL2 insert. The resulting PCR products were digested with NotI and XhoI and inserted into the  $\beta\Delta 1,2$  plasmid cut with the same restriction enzymes. Smaller inserts were created by annealing DNA oligonucleotides with the appropriate NotI and XhoI overhangs. In this case, each oligonucleotide was 5' phosphorylated with T4 polynucleotide kinase prior to annealing using standard procedures.

Prior to making the CMV-driven PAN RNA expression constructs, the deletions in the ORE were first introduced into PAN $\Delta 79$  by SOEing PCR [117]. The primer sets used are listed in the Supplementary Content. Products were digested with PflMI and NcoI and inserted into the same sites on PAN $\Delta 79$ . NcoI was substituted by AfeI where appropriate. The CMV-driven PAN RNA deletion plasmids were constructed by PCR amplification of PAN RNA sequences from these PAN $\Delta 79$  derivatives using primers NC872 (5' agtctcAAGCTTactgggactgcccagtcacc 3') and NC10 (5' gggggcccgtcacatttagggcaaagtgg 3').

The PCR products were digested with HindIII and XbaI and ligated into the CMV-WT plasmid cut with the same enzymes. Point mutants in CMV-WT were created using the same SOEing approach and the primers are listed in the supplemental content. The sequence of all PCR-generated inserts was verified.

### **Cell culture and transfections**

TREx BCBL1-Rta cells [118] were carried in RPMI-1640 media (Sigma) supplemented with 10% tetracycline-free FBS (Clontech), penicillin-streptomycin (Sigma), 2 mM L-glutamate, and 100  $\mu$ g/ml hygromycin (Sigma). HEK293 cells were grown as previously described in [12].

HEK293 cells were transfected using TransIT-293 (Mirus).  $\beta$ -globin reporter constructs were transfected in 12-well tissue culture plates with a total of 0.8  $\mu$ g plasmid DNA. A typical transfection contained 0.4  $\mu$ g of pcFl-ORF57II plus pcDNA3, 0.1  $\mu$ g of a control plasmid (mgU2-19/30) [12], and 0.3  $\mu$ g of the  $\beta$ -globin reporter. For the CMV-PAN derivatives,  $\beta$ -globin reporter DNA was substituted by 0.3  $\mu$ g of a PAN RNA expression construct. Total RNA was harvested 18-24 hrs after transfection using TRI Reagent (Molecular Research Center) and analyzed by northern blotting using standard procedures [109, 119]. Riboprobes against the endogenous 7SK RNA or the co-transfected control were used to normalize RNA recovery and loading.

For HEK 293 cells 2 hr pulse experiments, were transfected for a total of 48 hrs. Fresh media was added to 60 mm plates the night prior to collection. A total of 5 $\mu$ M 4-thiouridine (4SU, Sigma #T4509) was added to the media 2 hours prior to collection and total

RNA was isolated with TRI Reagent (Molecular Research Center). For 2 hr pulse experiments with TReX BCBL1-Rta cells, cells were diluted to  $5 \times 10^5$  cells/ml and either left untreated (uninduced) or induced with with 1  $\mu$ g/ml doxycycline (Fisher) and 3 mM sodium butyrate (Sigma) for 0, 12 or 24 hrs. Cells were incubated with a final 4SU concentration of 5 $\mu$ M for 2 hours prior to the 0, 12, or 24 hr collection time points (therefore at -2hrs, 10hrs, and 22hrs 4SU was added into the media).

### **Antibodies**

ORF57 antibodies were made in rabbits (Cocalico Biologicals) using a mixture of two bacterially expressed ORF57 polypeptide sequences, fm9 (amino acids 1-153) and fm7 (amino acids 171-455). Due to propensity of these proteins to precipitate, they were kept under denaturing conditions (0.5% SDS or 8M urea) for the entire preparation. ORF57 antibodies were affinity purified from rabbit serum using the Microlink Peptide Coupling kit (Pierce) as per the manufacturer's instructions. For HITS-CLIP, ORF57 antibodies were affinity purified from rabbit serum using the AminoLink Plus Immobilization kit (Pierce). The affinity-purified antibody recognized a single 51-kDa protein on western blots of lysate from infected or transfected ORF57-expressing cells, but not from that of cells that did not express ORF57 (data not shown). Pre-bleed antibodies were purified with Protein A beads (Pierce) using standard procedures.

### **Low throughput UV crosslinking and immunoprecipitation**

TREx BCBL1-Rta cells at  $5 \times 10^5$  cells/ml were reactivated by treatment with 1  $\mu\text{g/ml}$  doxycycline (Fisher) and 3 mM sodium butyrate (Sigma) for 18-20 hrs. A total of  $1 \times 10^7$  cells/condition were washed in 10 mls of phosphate-buffered saline (PBS; Sigma), centrifuged at 700xg for 3 min at 4°C and resuspended in 3 mls of PBS. The cell suspension was transferred to a 10 cm cell culture dish and irradiated on ice at 250  $\text{mJ/cm}^2$  ~2-3 cm from the UV source (Spectroline XL-1500, 254nm). After cross-linking, 7 mls of PBS were added to the dish. Cells were transferred to 15 ml conical tubes, and centrifuged at 700xg for 3 min at 4°C. Cell pellets were resuspended in 1ml PBS and centrifuged at 2400xg for 1 min at 4°C. PBS was removed and the pellets were frozen at -80°C [120]. Cell lysis and immunoprecipitation were performed essentially as previously described [120] except addition of poly(U) RNA was omitted and the QIAshredder step was substituted by digestion with 30  $\mu\text{l}$  of RQ1 DNase (Promega) for 15 min at 25°C. Subsequently, 10  $\mu\text{l}$  of 1 $\mu\text{g/ml}$  RNase A (Sigma) was added and incubated at 25°C for 10 min [121]. For the immunoprecipitation, approximately 8  $\mu\text{g}$  of antibody (pre-bleed or  $\alpha\text{ORF57}$ ) were added to the extracts, which were then nutated for 1 hr at 4°C. The extracts were then added to 20  $\mu\text{l}$  of protein-A agarose beads (Pierce) and nutated for an additional hour at 4°C. Washing of bound antibody complexes, elution and RNA recovery were performed exactly as described [120].

### Isolation of newly made transcripts

Roughly ~20ug of DNased RNA were used for this procedure. RNA was biotinylated in a 50  $\mu$ l reaction with 10mM Tris pH 7.5, 1mM EDTA, 0.1% SDS, 0.2mg/ml EZ-Link Biotin-HPDP (ThermoScientific#21341). The reaction was incubated for 3 hours at room temperature in the dark. The reaction was brought up to 250  $\mu$ l in water, extracted twice with chloroform and ethanol precipitated with ammonium acetate and 15 $\mu$ g of glycoblue (Ambion). Streptavidin selection was carried out with Dynabeads MyOne Streptavidin T1 (Invitrogen #65602). 20  $\mu$ l of the bead slurry were used per condition. Streptavidin beads were pre-washed in MPG1:10-I buffer (100mM NaCl, 1mM EDTA, 10mM Tris pH 7.5 and 0.1% Igepal). After the last wash, for each condition the beads were resuspended in 180ul of MPG1:10-I and preblocked with 0.1  $\mu$ g/ $\mu$ l polyA RNA, 0.1  $\mu$ g/ $\mu$ l salmon sperm DNA (ssDNA), 0.1 $\mu$ g/ $\mu$ l of Torula yeast RNA (Sigma). The biotinylated RNA was resuspended in 30  $\mu$ l of water and 3ul were collected as 10% input. RNA was heated at 65°C for 5 min. The RNA was then nutated with 170  $\mu$ l of the blocked Streptavidin beads for 30 min at RT. The binding reaction was then washed with the following buffers (~300 $\mu$ l per wash): MPG1:10-I, MPG1:10 (100mM NaCl, 1mM EDTA, and 10mM Tris pH 7.5) at 55°C, MPG1:10-I, MPG-I (1M NaCl, 10mM EDTA, 100mM Tris pH 7.5 and 0.1% Igepal), MPG-I, MPG1:10-I, MPG-I no salt (10mM EDTA, 100mM Tris pH 7.5 and 0.1% Igepal) and MPG1:10-I. The bound RNAs were then eluted in two steps. First, by incubating at RT for 5 min in ~200ul of MPG1:10-I with 5%  $\beta$ -mercaptoethanol and second by incubating at 65°C for 5 min in ~200ul of MPG1:10-I with 5%  $\beta$ -mercaptoethanol. Eluted RNA fractions were combined,

PCA extracted, chloroform extracted, and then ethanol precipitated with sodium acetate and 15µg of glycoblue.

### qRT-PCR

To analyze the RNAs recovered from the UV cross-linking protocol, the input and pellet samples were DNase treated for 1 hr at 37°C in a 20 µl reaction containing 2 µl of RQ1 DNase, 20 units of RNasin<sup>®</sup> Plus (Promega), and DNase Buffer (40 mM Tris pH 8.0, 10 mM MgSO<sub>4</sub>, 1 mM CaCl<sub>2</sub>). The reaction volume was brought up to 200 µl in G-50 Buffer (20 mM Tris pH 7.5, 0.25% SDS, 0.3 M sodium acetate, 2 mM EDTA) and then extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1; PCA). An aliquot of the resulting RNAs (~1/7) was ethanol precipitated with 15 µg of glycoblue (Ambion). Reverse transcription reaction was performed using SuperScript<sup>™</sup> II RT (Invitrogen) and random hexamers (Sigma) as per the manufacturer's protocol. No-RT controls were performed in parallel for each sample. Inputs and pellets were diluted 40-fold, 2 µl of which was used as template for real-time PCR as described [12]. Real-time primer sets were: A, NC702 (5' GCTCGCTGCTTGCCTTCTT 3') and NC703 (5' CCAAAGCGACGCAATCAA 3'); B, NC704 (5' CTTGCGGGTTATTGCATTGG 3') and NC705 (5' GACACGTTAAGTATCCTCGCATATCA 3'); C, NC706 (5' TTTTCCAGTGTAAGCAAGTCGATTT 3') and NC707 (5' TGTTCTTACACGACTTTGAACTTCTG 3'); D, NC708 (5' TTAACGTGCCTAGAGCTCAAATTAAC 3') and NC709 (5'

TTGACCTTTATTTATGTTGTAAAGTTGCATTA 3'). Efficiency of the A, B, C and D amplification was determined to be 92%, 91%, 82% and 83%, respectively [122]. Relative quantities (RQ) of Inputs and Pellets were determined based on their amplification efficiency and Ct value. The RQ value for the no-RT controls was subtracted from the plus RT samples. Immunoprecipitation efficiency was determined by calculating the pellet/input ratio of the background corrected values. To compare between experiments, these values were then normalized to the "+UV/ $\alpha$ ORF57" ratio for primer set A.

Analysis of RNAs recovered from TREx BCLB1-Rta cells was performed as mentioned above, except for minor modifications. DNase reactions were scaled up to 200  $\mu$ l and incubated for an average of 2 hours at 37°C. A total of 2  $\mu$ g of DNased RNA were used in the RT reaction and substituted SuperScript™ II RT for the M-MuLV RT (NEB# M0253). Primers used are listed in Table 1.

### **Label transfer assays**

DNA templates for *in vitro* transcription of each substrate were generated by PCR amplification using primers containing the T7 promoter sequence. In the case of  $\Delta$ SL2, template was amplified from PAN $\Delta$ 79- $\Delta$ SL2, while all others utilized PAN-WT as a template. The smallest T7-DNA templates were made by annealing oligonucleotides. Each oligonucleotide pair was heated at 90°C for 3 min in annealing buffer [10 mM Tris pH 7.5, 150 mM NaCl, 25 mM EDTA, 2.5  $\mu$ M each oligonucleotide] and then slowly equilibrated to room temperature. Body labeled RNA fragments were generated in a 20  $\mu$ l reaction containing 50 units of T7 RNA polymerase, 5  $\mu$ l  $\alpha$ -<sup>32</sup>P-UTP (800 Ci/mmol, 0.1 mCi/ $\mu$ l), 20U

RNasin<sup>®</sup> Plus and transcription buffer (40 mM Tris pH7.5, 6 mM MgCl<sub>2</sub>, 4 mM spermidine, 50 μM UTP, 1 mM ATP, 1 mM CTP, 1 mM GTP, 10 mM DTT). Transcription reaction was run through an illustra MicroSpin<sup>™</sup> G-25 column (GE), the RNA was ethanol precipitated with 1M ammonium acetate and glycoblue, and then separated on a urea-PAGE. Full length RNAs were excised from the gel and eluted in G-50 buffer overnight at room temperature. Eluted fractions were extracted with PCA, ethanol precipitated, and resuspended in water. RNA yield was approximated by scintillation (Cerenkov) counting.

Whole cell extracts were made by transfecting a 10 cm dish with 10 μg of either pcDNA3 or pcFl-ORF57II. After 48 hrs, cells were harvested, washed with PBS, and the cell pellets were collected by centrifugation at 2400xg for 1 min at 4°C. The pellets were subsequently resuspended in 300 μl of RSB100T-Plus [10mM Tris pH 7.5, 2.5 mM MgCl<sub>2</sub>, 100mM NaCl, 0.5% TritonX100, 1mM phenylmethanesulphonylfluoride (PMSF, Sigma), protease inhibitors (cocktail V, Calbiochem)], sonicated three times for 10 sec at 30% amplitude (SONICS Vibra-cell<sup>™</sup> VCX130 with a 6 mm probe), and then centrifuged at 16,000xg for 10 min at 4°C. After adding glycerol to 10%, the supernatants were stored at -80°C.

The binding reactions for label transfer assays were performed in a 40 μl reaction volume including 10 μl of whole cell extract, 10 μl RSB100T (10mM Tris pH 7.5, 2.5 mM MgCl<sub>2</sub>, 100mM NaCl, 0.5% Triton X-100), substrate (~15 nM), 500 μM MgCl<sub>2</sub>, 80 mM KCl, 1 mM ATP, 20 mM creatine phosphate, and 0.3 mg/ml Torula yeast RNA (Sigma). In some cases 20 μl of extract were used, however, higher amounts of extract did not affect the

ORF57 signal strength. Binding was allowed to reach equilibrium by incubating at 30°C for 15 min. The protein-RNA complexes were then cross-linked on ice at 860 mJ/cm<sup>2</sup> in a Spectroline XL-1500 (254 nm). Samples were then treated with RNase A at 0.25 mg/ml for 30 min at 25°C. For immunoprecipitation, 80 µl of RIPA buffer (50 mM Tris pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) were added along with ~4.5 µg of affinity purified ORF57 antibody and the extract was nutated for 1 hr at 4°C. This mixture was then transferred to 15 µl of protein-A agarose beads and nutated for an additional hour at 4°C. Beads were washed a total of five times in 1 ml of RIPA buffer. Proteins were eluted by boiling in SDS gel loading buffer. The samples were separated on a 10% SDS-PAGE, dried, and exposed on a Phosphorimager screen.

### **Western blotting**

To test the ORF57 specific bands in the HITS-CLIP membrane, samples were run in NUPAGE 4-12% gradient gels and transferred to a nitrocellulose membrane utilizing standard procedures. Primary antibodies against ORF57 were purified as mentioned above, and utilized for western blotting. The secondary antibody used was the CleanBlot IP detection Reagent (HRP, ThermoScientific #21230).

Western blots testing ORF57 expression in TREx BCLB1-Rta cells were run using conventional 10% SDS-PAGE gels. The β -Actin primary antibody used was from Abcam (ab6276). The secondary antibody was from Licore (Odyssey Gt anti-Ms IRDye 680LT #92668020). The secondary antibody against ORF57 was also from Licore (Odyssey Gt anti-Rbt IRDye 800CW #92632211).

## **ORF57 HITS-CLIP**

Detailed protocols are included in the appendices: Recipes (A), cell preparation (B) and complete protocols including library preparation instructions (C & D for Illumina). We incorporated some procedure modifications from a previously published UV-crosslinking protocol into the HITS-CLIP procedure [121, 123-128], our hope is that these changes will aid in the study of other “difficult to work with” proteins like ORF57. For Inputs, TRIzol extracted RNA was DNased (RQ1 DNase), rRNA depleted (Epicentre #MRZH116), digested with MNase (NEB M0247S), and end-labeled with <sup>32</sup>P (T4 PNK NEB M0201S). INPUT and PELLET RNA fragments were size selected in urea gels, eluted and then normalized according to their radioactive signals measured in the scintillation counter. An RNAClean XP bead purification step (Beckman Coulter #A63987) was included prior to library preparation. Libraries were prepared with minor modifications from the recommended Illumina’s TruSeq Stranded mRNA Sample Prep Kit (#RS-122-2101). The Illumina cDNA library sequencing run was paired end (100x100bp).

## **HITS-CLIP bioinformatic analysis**

Adaptors marked as N at the 3’ ends of the paired-end sequencing data were trimmed. Then Gsnap [129, 130] was used to align sequencing data with parameters “-A sam --maxsearch 1 -N 1 -t 4 -n 1.” Reads were aligned to a concatenated human+KSHV genome (Hg19+U75698). Reads mapping to splice junctions were discarded and Gsnap-unaligned reads were then mapped to the Hg19 transcriptome by Novoalign (Novocraft). Read pairs

overlapping by at least 1bp were merged. Merged reads with matching chromosome, strand, left-most mapping coordinate and right-most mapping coordinate were defined as PCR duplicates and collapsed to unique tags. Unique tag mutations reflect mapping mutations of the read with the highest sequencing quality. Tags from all the experimental conditions were simultaneously overlapped to identify CLIP clusters with at least 10 tags from any one or more conditions. CLIP clusters are binned (20bp/bin) and then for each experimental condition the tag and mutation counts for each bin are summed and recorded separately. Deletions and T→C mutations showed preferential enrichment in the HITS-CLIP reads when compared to the input reads and were therefore chosen as the characteristic mutations for ORF57 in these experimental conditions. The total tag count and characteristic mutation counts are summed with weights of 0.8 and 0.2 respectively to give the binding intensity in each bin and condition. DESeq [131] was applied to analyze the binding intensity data of the 3 HITS-CLIP (Pellet) and 3 control (Input) conditions. DESeq uses the median intensity value of each condition for normalization and employs a negative binomial test for identifying differentially bound regions. The analysis was conducted separately for human and KSHV binding sites. Bins that were more enriched in the HITS-CLIP conditions than in the controls were extracted and neighboring enriched bins were concatenated into continuous regions depicting ORF57 binding sites. To identify enriched clusters two data sets with  $pval < 0.001$  and  $pval < 0.0001$  were generated, but the former was used in our current analysis. The ORF57 binding sites are further filtered by repeating sequences including rRNAs, tRNAs, low complexity regions, LINES, SINES and simple repeats, regardless of strand. Single-nucleotide resolution mutations (DEL and T→C) were extracted from within the

remaining ORF57 binding sites. T tests with a p value cutoff of 0.05 were applied to test whether mutation rates occur at a significantly higher rate in mutant tags *vs.* total tags. The significant bases are retained, which indicate high-confidence ORF57 binding events in close proximity. 60bp long sequences surrounding these significant bases (+/-30bp) are extracted from the concatenated genome. MEME [132] and RNAfold [133] were used to analyze these sequences for *de novo* motif discoveries and secondary structure predictions, respectively.

**Table 1. qRT-PCR primers for HITS-CLIP**

Target	Oligo	Sequence (5'-3')
BTG1 mRNA Fwd	NC1780	GAGCTGCTGGCAGAACATTA
BTG1 mRNA Rev	NC1781	GAACAGCTCCTGACTGCTCA
BTG1 intron Fwd	NC1782	GTCACCGGCACAATTAACAG
BTG1 intron Rev	NC1783	TGCACACAATGGAGTTGATG
TNFSF9 exon Fwd	NC1829	GGCCTGAGCTACAAAGAGGA
TNFSF9 exon Rev	NC1830	CCGCAGCTCTAGTTGAAAGA
TNFSF9 intron Fwd	NC1845	CACAAGCTCTGCATCTCTGG
TNFSF9 intron Rev	NC1846	AGGCTCTTTGGGAGTTAGCA
ZFP36 mRNA Fwd	NC1794	ACTGCCATCTACGAGAGCCT
ZFP36 mRNA Rev	NC1795	GACTCAGTCCCTCCATGGTC
ZFP36 pre-mRNA Fwd	NC1798	GACTGCCATCTACGAGGTGA
ZFP36 pre-mRNA Rev	NC1799	AGTTTGCGGCGCTAGAGA
EGR1 mRNA Fwd	NC1786	CACCTGACCGCAGAGTCTT
EGR1 mRNA Rev	NC1787	AAGCGGCCAGTATAGGTGAT
EGR1 intron Fwd	NC1788	GCGTCAGCTGTTGTTGAAAT
EGR1 intron Rev	NC1789	CTACCATTGACTCCCCGAGGT
KRI1 mRNA Fwd	NC1802	GTCAGAGCAGGAGGAGCTG
KRI1 mRNA Rev	NC1803	CAGGTGTCTCTGCCTCTTCTC
KRI1 intron Fwd	NC1847	GCTTTCGCCCAGTCTAGAAG
KRI1 intron Rev	NC1848	CACACACCTGGCCTCATTT
7SK Fwd	NC1164	TAAGAGC TCGGATGTGAGGGCGATCTG
7SK Rev	NC1165	CGAATTCGGAGCGGTGAGGGAGGAAG
GAPDH mRNA Fwd	NC638	AGCCTCAAGATCATCAGCAATG
GAPDH mRNA Rev	NC639	ATGGACTGTGGTCATGAGTCCTT
GAPDH premRNA Fwd	NC1240	TCCCCTCCTCATGCCTTCTT
GAPDH premRNA Rev	NC1241	CCAGGCGCCCAATACG
PAN RNA 5' end Fwd	NC702	GCTCGCTGCTTGCCTTCTT
PAN RNA 5' end Rev	NC703	CCAAAAGCGACGCAATCAA
ORF59 Fwd	NC566	GAGTGCCACTAAAACCGGAGTAG
ORF59 Rev	NC567	TGAGGGTTGGCGTGCAA
ORF57 Fwd	NC1402	TCCTCCTCTGAGTTTGACGA
ORF57 Rev	NC1403	CAATTGCTCGTCTTCCAGTG
B-Actin exon Fwd	NC1230	CTTCAACACCCCAGCCATGT
B-Actin exon Rev	NC1231	CCAGAGGCGTACAGGGATAGC
B-Actin intron Fwd	NC1825	ATCAAGGTGGGTGTCTTCC
B-Actin intron Rev	NC1826	CAGGAGGAGCAATGATCTGA

B-Actin intron Fwd	NC1821	AGGGCTTCTTGCCTTCCT
B-Actin intron Rev	NC1822	CATAGGAATCCTTCTGACCCA

## CHAPTER FOUR

### **Delineation of a core RNA element required for Kaposi's sarcoma-associated herpesvirus ORF57 binding and activity**

#### **Introduction**

Kaposi's sarcoma-associated herpesvirus (KSHV) is the causative agent of Kaposi's sarcoma, primary effusion lymphoma (PEL), and some cases of multicentric Castleman's disease [21, 134, 135]. KSHV is a member of the gammaherpesvirus family, which, like all herpesviruses, is characterized by both latent and lytic phases of infection. During latent infection, the virus expresses only a small subset of its genes that control functions essential for preserving the viral genome and the cells harboring them. Such functions include viral genome maintenance, immune evasion, and control of host cell proliferation. No viral progeny are made during this phase. In contrast, lytic infection involves a regulated cascade of viral gene expression that leads to the production of infectious virions. Both latently and lytically expressed genes have been implicated in KSHV pathogenesis.

The herpesviridae have evolved sophisticated mechanisms to control viral and host gene expression utilizing both viral and cellular factors. Because KSHV has a nuclear dsDNA genome, it is not surprising that KSHV utilizes host factors for transcription and mRNA processing. However, the virus encodes its own proteins that regulate gene expression both transcriptionally and posttranscriptionally [56, 136-140]. The 51 kDa ORF57 protein (Mta, KS-SM) is a member of a family of proteins conserved throughout the herpesviridae and it is essential for KSHV replication [13, 14, 141-145]. While ORF57 has been implicated in the regulation of KSHV transcription [55, 67, 68], its post-transcriptional

activities are more extensively studied. ORF57 has been reported to affect nearly every stage of gene expression including RNA stability, pre-mRNA splicing, RNA export, and translation [12, 58, 68, 88].

ORF57 binds to RNA *in vitro* and *in vivo* [12, 58, 62, 83, 93], but it is currently unknown what dictates the specificity of ORF57 for its targets. Our recent work showed that ORF57 stabilizes PAN RNA (nut-1, T1.1), a nuclear non-coding polyadenylated RNA [106, 107]. PAN RNA accumulates to very high levels during lytic infection [107, 108] suggesting an important, yet unknown, function in viral replication. ORF57 enhances the abundance of PAN RNA in transfected cells [12, 55, 58] and it is essential for PAN RNA accumulation during viral infection [13, 14]. Moreover, ORF57 binds directly to PAN RNA in cultured cells and RNA-binding is essential for ORF57 activity [12]. A 300-nucleotide (nt) sequence, the ORF57-responsive element (ORE), in the 5'-end of PAN RNA is necessary for binding and for ORF57-responsiveness. However, the minimal ORE has not previously been delineated, nor has it been shown that the ORE is directly and specifically bound by ORF57.

In the present work, we show that ORF57 binds the ORE in infected cells and we identify a 30-nt core ORE consisting of a predicted stem-loop structure. The core ORE was defined based on its activity in three assays. Specifically, ORF57 binds the core ORE in an *in vitro* label transfer assay, it is sufficient to confer ORF57-responsiveness to an intronless  $\beta$ -globin reporter, and it is necessary for ORF57 responsiveness of PAN RNA. We additionally identified point mutations in the 30-nt core ORE that abrogate binding and are necessary for ORF57 response in PAN RNA. These analyses implicate a 9-nt sequence in an unstructured loop of the core ORE as a potential ORF57-binding sequence. Recognition of the core ORE

by ORF57 is likely sequence-specific rather than structure-based, because alterations in stems adjacent to the 9 nt sequence had little effect on ORF57 binding *in vitro*. Our analyses also reveal that, while the core ORE is necessary, additional sequences in the full-length ORE are required for robust ORF57-mediated enhancement of PAN RNA levels. These data are the first to clearly delineate an ORF57-interacting RNA sequence of a nuclear target of ORF57. In the long term, these data provide an important foundation to inform future investigations of ORF57 specificity and mechanisms.

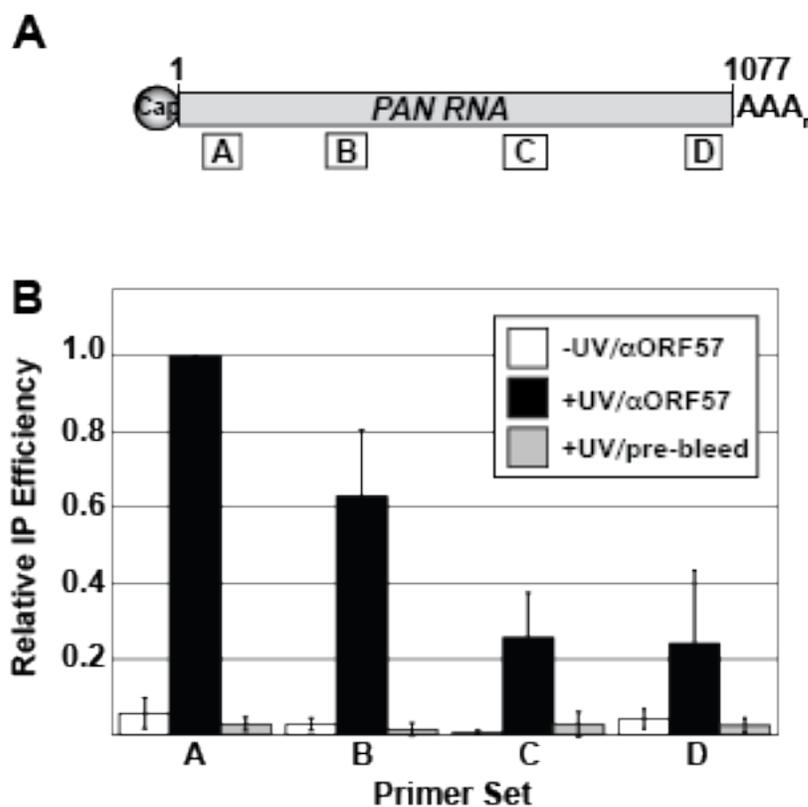
## **Results**

### *ORF57 binds to the 5' end of PAN RNA in lytically reactivated cells*

In transiently transfected cells, PAN RNA response to ORF57 is dictated by a 300-nt ORE at the 5' end of the transcript (nt 13-312), which will be referred to as the full-length ORE throughout this manuscript. Based on these observations, we hypothesized that the 5' end of PAN RNA contains a high-affinity binding site and that this site is important for ORF57 binding in infected cells [12]. We employed an ultraviolet (UV) cross-linking approach to examine whether ORF57 binds to the 5' end of PAN RNA in KSHV-infected cells [12, 120, 121]. In this assay, lytically reactivated cells are exposed to UV light, lysed, and a partial RNase digestion is performed to cleave the RNA into random fragments prior to immunoprecipitation of the RNA-protein complexes with ORF57 antibodies. The immunoprecipitated RNA fragments are then analyzed by quantitative reverse transcription PCR (qRT-PCR) with primers that span four different regions of PAN RNA (Figure 3A, boxes labeled A-D). Because UV light solely cross-links protein-RNA interactions in very

close proximity, the immunoprecipitated RNA fragments represent those that are directly bound by ORF57.

Examination of the relative immunoprecipitation efficiencies of each segment of PAN RNA shows a clear 5' bias (Figure 3B, black bars). Because viable cells are exposed to UV light in this procedure, demonstration of a UV-dependent cross-link reveals that the interaction occurs in living cells. Importantly, we see no immunoprecipitation of PAN RNA in the no-UV control (white bars), so we conclude that the interaction with ORF57 exists in lytic phase cells. Additionally, little to no immunoprecipitation was observed when pre-bleed serum was used in place of anti-ORF57 antibodies (gray bars). While we observe a 5' bias, comparison of the test samples to the no-UV and pre-bleed controls for primer sets B and C show statistically significant immunoprecipitation ( $p < 0.05$ , Student's t-test). One interpretation of this observation is that there are multiple binding sites throughout these regions of PAN RNA. Alternatively, it is possible that the RNA was not digested sufficiently for resolution of independent fragments. The values for primer set D at the 3' end of the transcript do not rise above statistical significance ( $p > 0.1$ ), so our data do not support the existence of an interaction between the 3' terminus of PAN RNA and ORF57. We conclude that ORF57 binds PAN RNA directly during KSHV lytic infection and that the predominant binding site lies at the 5' end of the transcript.



**Figure 3.** ORF57 binds the 5' end of PAN RNA in lytically reactivated cells. (A) Schematic diagram of PAN RNA with approximate positions of qRT-PCR amplicons shown below. Primer set A, B, C, and D amplify PAN RNA base pairs 50-124, 293-372, 642-728, and 994-1064, respectively. The numbering system for PAN RNA in this manuscript is relative to the start site as defined in Zhong et al. (1996). (B) Results from UV cross-linking immunoprecipitation experiments. The y-axis shows the immunoprecipitation efficiencies relative to that for primer set A (see Materials and Methods). The error bars are standard deviation ( $n=3$ ).

*ORF57 binds specifically and directly to the PAN ORE*

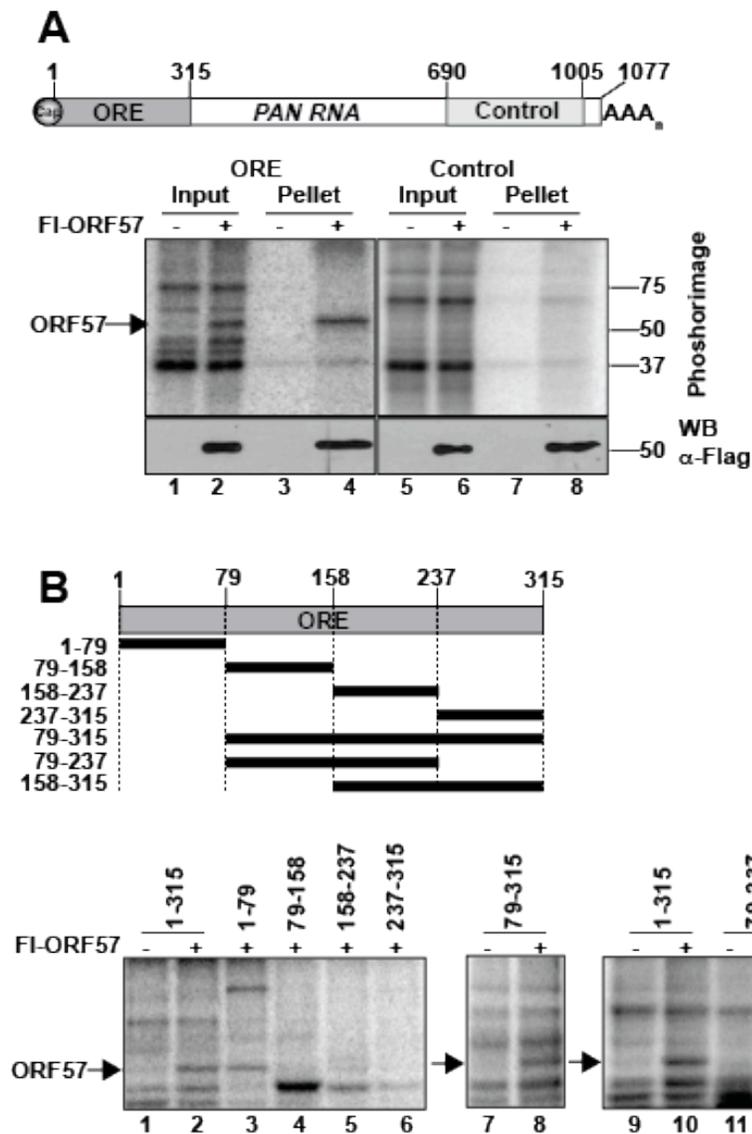
To further explore the interaction between ORF57 and the ORE, we employed an *in vitro* UV cross-linking, or label transfer, assay (Figure 4). In these assays, a radiolabeled RNA substrate is incubated in cell extract and exposed to UV light to covalently cross-link the proteins bound to the substrate RNA. After cross-linking, the RNAs are digested essentially to completion with RNase. However, small cross-linked RNA fragments (~1-10 nt) are protected from degradation due to the attached protein. The cross-linked protein-RNA complexes can be visualized by Phosphorimager analysis of protein gels by virtue of the radiolabeled RNA. We synthesized uniformly radiolabeled substrates containing the full-length ORE sequence and a control sequence lacking the ORE derived from the 3' end of PAN RNA (Figure 4A, top). We incubated these RNAs in whole cell lysate from cells expressing Flag-tagged ORF57 (FI-ORF57) or an empty vector control and performed the label transfer assay. The ORE substrate cross-links to a ~51 kDa protein in extract containing ORF57 but not in those extracts transfected with empty vector (lanes 1,2). The 51 kDa protein is immunoprecipitated with polyclonal antibodies specific for ORF57, confirming the identity of this protein as ORF57 (lanes 3,4). In contrast, the control substrate showed no binding to ORF57 in this assay (lanes 5-8). These experiments demonstrate that ORF57 binds specifically and directly to the ORE in whole cell extract and that those interactions reflect the binding patterns observed in infected cells (Figure 3).

We further exploited the label transfer assay to define the sequences necessary for ORF57 binding. Initially, we divided the full-length ORE into four non-overlapping 79/80-nt substrates (Figure 3B, top and lanes 1-6). The 5'-most of these substrates (nt 1-79) is

sufficient for efficient ORF57 cross-linking, while the other three do not bind ORF57. Next, we examined whether PAN RNA nt 1-79 are necessary for ORF57 interaction with the ORE, by deleting this sequence from the full-length ORE (nt 79-315). Surprisingly, this substrate binds to ORF57 even though its three constituent 79-nt fragments do not (lane 8 vs lanes 4-6). One possible explanation for this observation is that a second ORF57 binding site resides near nt 158 or nt 237, the endpoints of the shorter substrates. Therefore, we tested two overlapping fragments including the natural junctions between the individual 79-nt substrates (nt 79-237 and 158-315) and found that both of these fragments bind ORF57 (lanes 11-14). Two models are consistent with these data. It is possible that the RNA sequences surrounding nt 158 and 237 contain two independent ORF57 binding sites. However, based on observations in cultured cells (see below), we favor the model that the first 79 nt of PAN RNA contain a high-affinity ORF57 binding site, while nt 79-315 contain multiple lower-affinity binding sites.

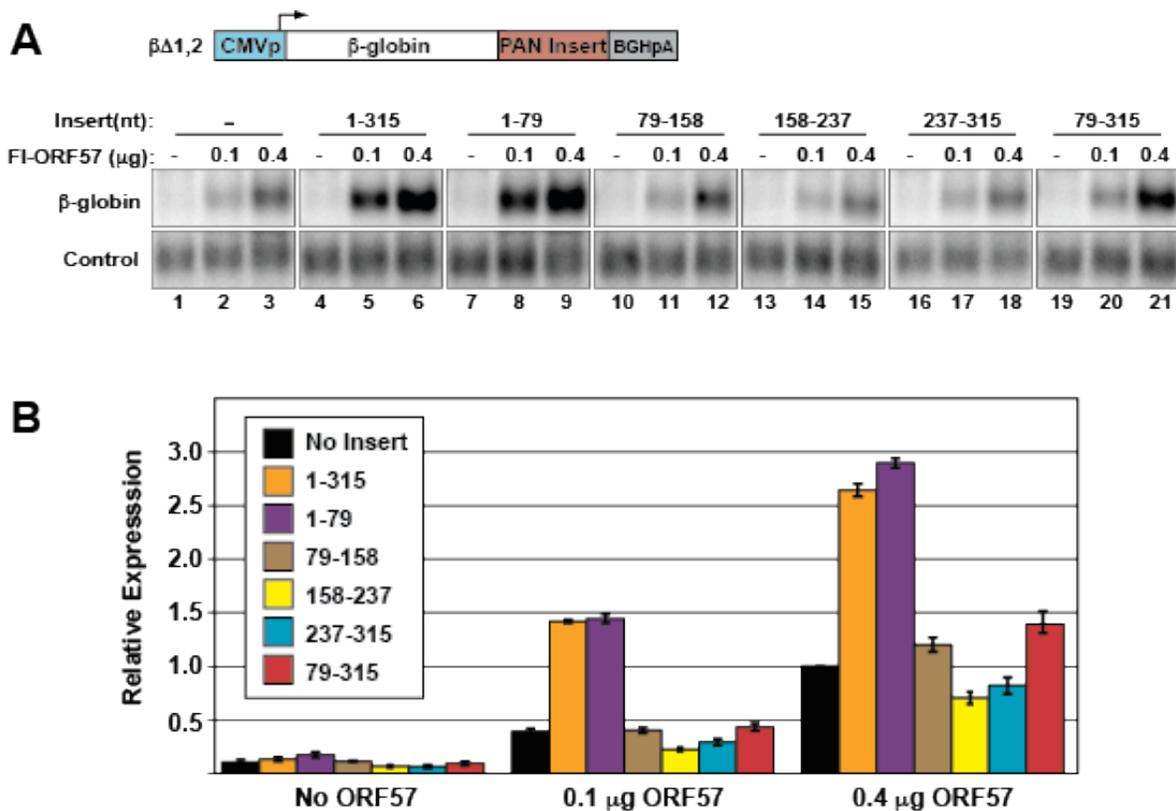
**Figure. 4.** ORF57 binds directly to the ORE *in vitro*. (A) Label transfer and immunoprecipitation assay. *Top*, Schematic diagram showing the positions of the full-length ORE or control substrate in PAN RNA. *Bottom*, Results from a representative label transfer assay. Extracts from cells expressing or not expressing flag-tagged ORF57 (F1-ORF57) were incubated with the indicated substrate as described in the Materials and Methods. The cross-linked, RNase-treated extracts were then immunoprecipitated using anti-flag antibodies; 10%

of input is shown. The bottom panels show an anti-flag western blot of the same samples demonstrating expression and immunoprecipitation of FI-ORF57. (B) Label transfer assays with substrates derived from the ORE region. *Top* Schematic showing the substrates and their position with respect to the full-length ORE. The bottom panels show the label transfer assay with extract from cells expressing FI-ORF57 or not as indicated above each lane. The position of ORF57 is indicated by the arrow.



*PAN RNA nt 1-79 are sufficient to confer ORF57-responsiveness to a heterologous transcript*

We next utilized a  $\beta$ -globin reporter assay [12] to test whether nt 1-79 are sufficient to confer ORF57-responsiveness in cells. We inserted ORE fragments into the 3' UTR of a  $\beta$ -globin reporter construct from which both  $\beta$ -globin introns had been removed ( $\beta\Delta_{1,2}$ ; Figure 5A). Our previous studies showed that co-transfection of ORF57 increases  $\beta\Delta_{1,2}$  mRNA levels in a dose-dependent fashion, even in the absence of the ORE [12]. However, inclusion of the full-length ORE increases the ORF57-specific response (Figure 5A, lanes 1-6; Figure 5B, black and orange bars). We placed the four 79/80-nt non-overlapping fragments described in Figure 2 into the 3'UTR of  $\beta\Delta_{1,2}$  and determined ORF57-responsiveness by northern blot (lanes 7-18). Quantification of the data shows that the first 79 nt of PAN RNA confer equivalent ORF57-responsiveness to  $\beta\Delta_{1,2}$  as the full-length ORE (Figure 5B, purple bars). Because nt 79-315 maintained binding to ORF57 in label-transfer assays (Figure 4), we also inserted this fragment into  $\beta\Delta_{1,2}$  (lanes 19-21). We observe a minimal ORF57 response with this insert (red bars). The effect is above background only at the highest concentration of ORF57 tested (compare red to black bars for 0.1  $\mu$ g and 0.4  $\mu$ g ORF57 samples). Taken with the *in vitro* results, these data support the model that nt 79-315 have low affinity for ORF57 and thus rely upon greater ORF57 levels to confer ORF57-dependent up-regulation. Most importantly, these data strongly support the conclusion that nt 1-79 contain a high-affinity ORF57 binding site, which is sufficient for ORF57-responsiveness in a heterologous context.



**Figure 5.** The first 79 nt of PAN RNA are sufficient for ORE activity in a heterologous transcript. (A) *Top*, Schematic diagram of the intronless  $\beta$ -globin reporter. Different portions of the ORE were placed into the 3' UTR ("PAN Insert") and tested for ORF57-responsiveness by northern blot. A representative northern blot with  $\beta$ -globin is shown below. The  $\beta$ -globin panels are from the same gel and are shown at the same exposure. The control lanes are probed for a co-transfected loading control. Amounts of co-transfected FI-ORF57 and the particular insert are given above each lane. (B) Quantification of the northern blot data; error bars are standard deviation ( $n=3$ ). Each value is relative to the no-insert control with 0.4  $\mu$ g of ORF57.

*A predicted stem-loop in PAN RNA is necessary for binding in vitro and sufficient to confer ORF57-responsiveness*

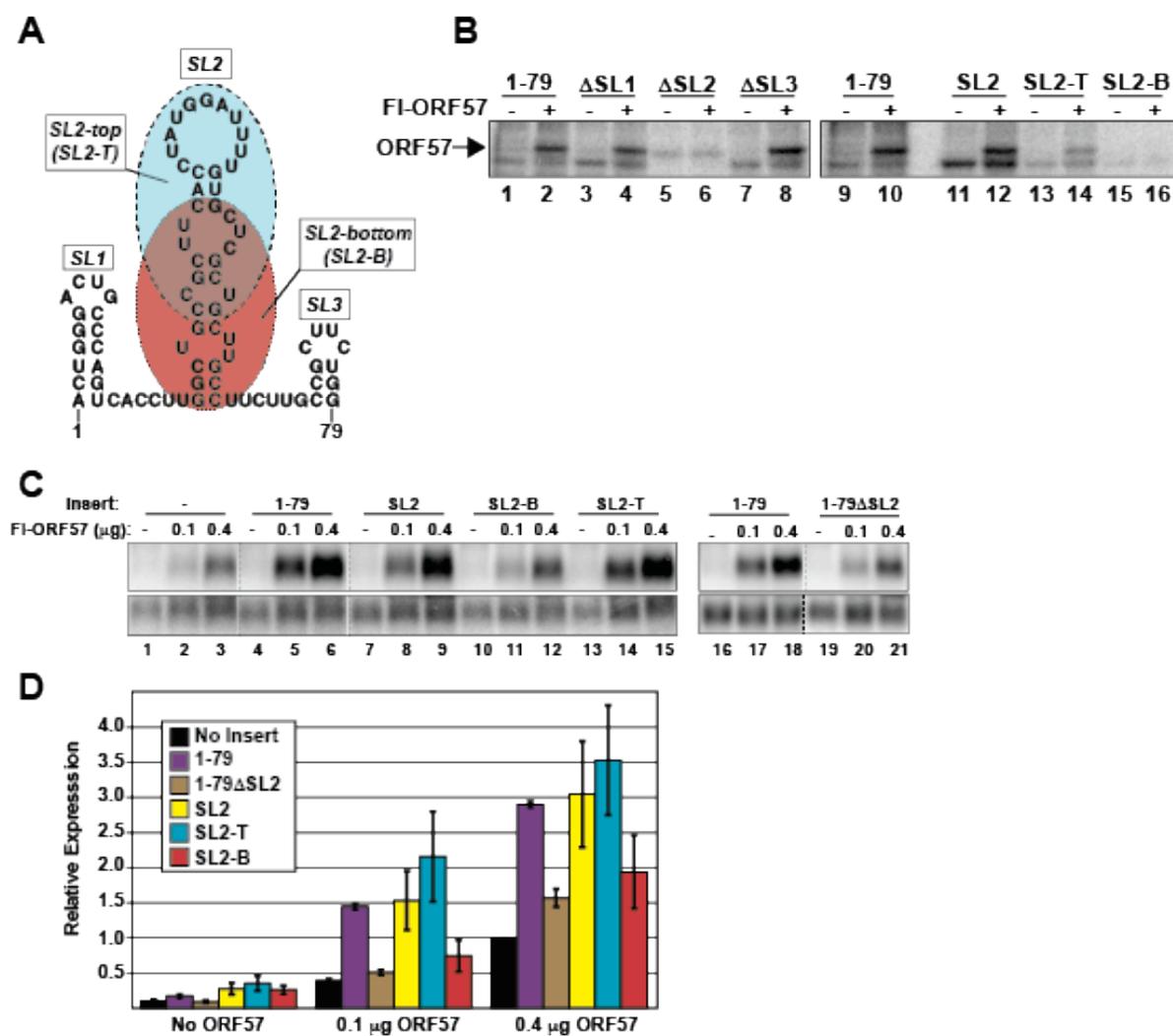
To determine a core sequence that is sufficient for ORF57 binding *in vitro* and for its activity in cells, we further dissected PAN RNA nt 1-79. The Mfold RNA secondary structure algorithm [146] predicted that PAN RNA contains three stem-loop structures within nt 1-79 (Figure 6A; SL1, SL2, SL3). We first determined whether any of these loops are necessary for ORF57 binding in cell extract (Figure 6B). Deletion of SL1 or SL3 from nt 1-79 have no effect on ORF57 binding (Figure 6B;  $\Delta$ SL1,  $\Delta$ SL3; lanes 3,4, 7,8), while ORF57 cross-linking is completely lost upon deletion of SL2 (lanes 5,6). The faint protein band that co-migrates with ORF57 in  $\Delta$ SL2 substrate lanes cannot be ORF57 because it is observed in the no ORF57 extract (lane 5). A substrate that contains the 40-nt SL2 without SL1 or SL3 efficiently cross-links to ORF57 (lanes 11, 12), demonstrating SL2 is sufficient for binding in whole cell extract. Further reduction of SL2 reveals that the cross-linking to ORF57 is driven by sequences in the top of the stem-loop structure. A 30-nt substrate derived from this portion (SL2-T, Figure 6A blue) efficiently cross-links ORF57 *in vitro* (lanes 13,14), while the bottom portion (Figure 6A; SL2-B) does not bind (lanes 15, 16). These results show that the 30-nt SL2-T sequence is necessary and sufficient to bind ORF57 in cell extract.

We next tested whether SL2 is sufficient to confer ORF57-responsiveness in a heterologous context by placing SL2, SL2-T, or SL2-B into the intronless  $\beta$ -globin reporter construct (Figure 6C and 6D). Consistent with the *in vitro* binding results, SL2 (yellow bars) and SL2-T (blue bars) are both sufficient to confer ORF57-responsiveness comparable to nt 1-79

(purple bars), but SL2-B is not (red bars). Additionally, deletion of SL2 from fragment 1-79 (1-79 $\Delta$ SL2, brown bars) abrogated ORF57-responsiveness. Both SL2-B and 1-79 $\Delta$ SL2 show limited, but above background, ORF57-responsiveness at high ORF57 levels (0.4  $\mu$ g). These data suggest that that the SL2-T region is a core ORE capable of conferring ORF57-binding and response. However, the *cis*-acting factors responsible for the control of RNA levels by ORF57 are likely significantly more complex than a single short RNA element.

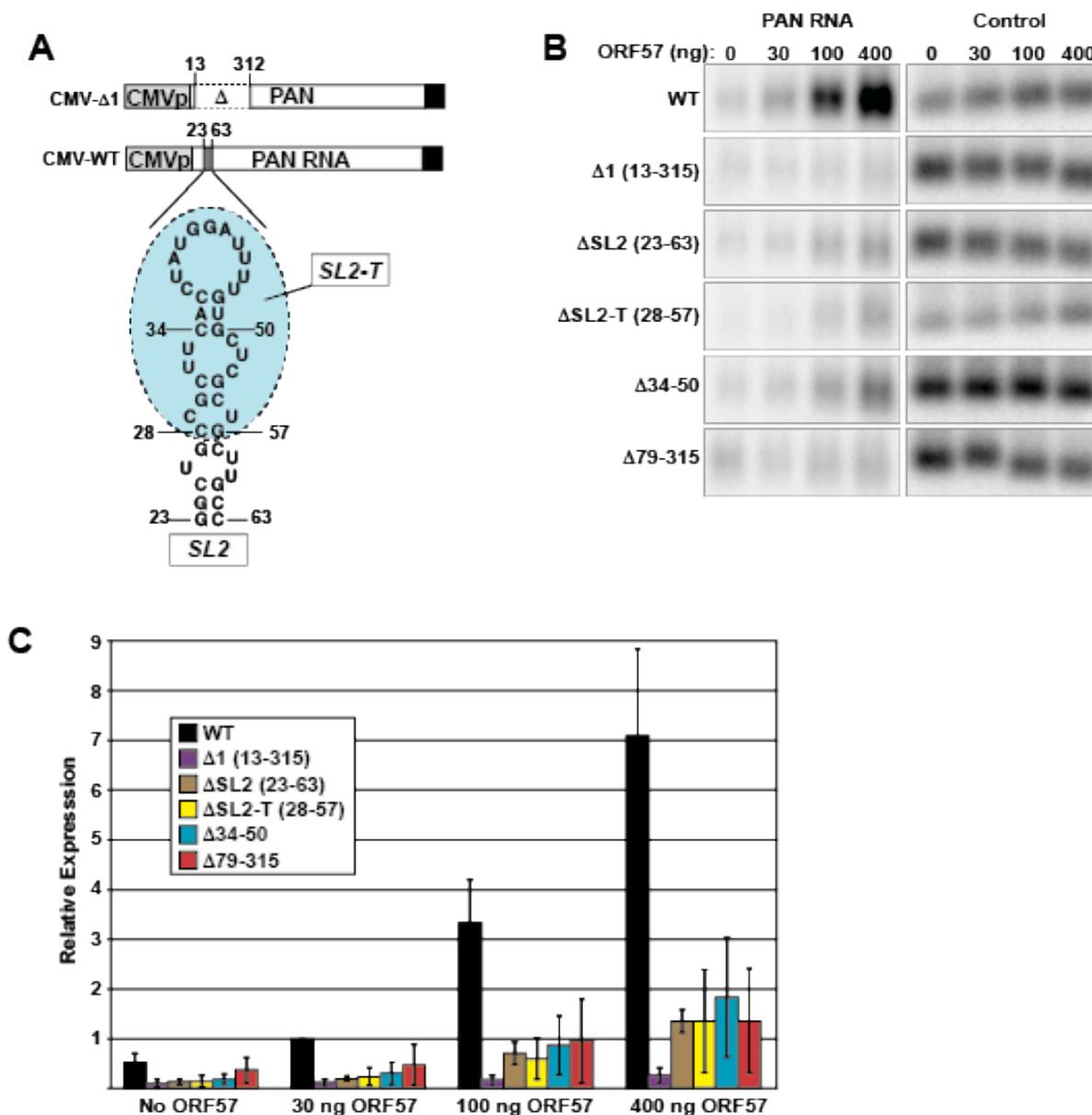
**Figure 6.** A 30 nt stem-loop structure is sufficient for ORE activity and binding. (A) Predicted secondary structure of PAN RNA nt 1-79. The three stem loops are labeled SL1, SL2, SL3 and the SL2 “top” (SL2-T) and “bottom” (SL2-B) portions are shown in ovals shaded with blue and red, respectively. (B) Label transfer assay with substrates that delete each of the stem loops (lanes 1-8) in the context of the nt 1-79 fragment. Lanes 11-16 show label transfer with SL2, SL2-T, or SL2-B alone. Cross-linking of each substrate is shown with extract either containing or lacking Fl-ORF57 as indicated. (C) Each of the indicated

fragments was inserted into the intronless  $\beta$ -globin construct and examined for ORF57 response by northern blot as described in Figure 5. The dashed lines represent positions where lanes were removed for presentation; the panels displayed are from the same blot at the same exposure. (D) Quantification of the northern blot results was performed as in Figure 3 ( $n=3$ ).



*SL2-T is necessary for ORE activity in PAN RNA*

The data presented above show that ORF57 binds to the 5' end of PAN RNA in lytically reactivated PEL cells and that the SL2-T region is sufficient for ORE activity and binding. If this element is indeed a core ORE, removal of SL2-T should decrease PAN RNA levels in the presence of ORF57. To test this, we generated ORE deletions in a previously described ORF57-responsive cytomegalovirus immediate early (CMV) promoter-driven PAN RNA [12] and examined their effects on PAN RNA accumulation in the presence of ORF57 (Figure 7A). We deleted the entire SL2 structure ( $\Delta$ SL2; nt 23-63), the SL2-T region ( $\Delta$ SL2-T; nt 28-57), or the predicted upper loop of SL2-T ( $\Delta$ 34-50). The 17 nt deleted in  $\Delta$ 34-50 bind ORF57 *in vitro*, but are not sufficient to confer ORF57-response in the  $\beta$ -globin reporter assay (data not shown). CMV- $\Delta$ 1, which lacks the full-length ORE, and CMV-WT are negative and positive controls, respectively. We co-transfected these constructs with increasing amounts of ORF57 (0, 30 ng, 100 ng, and 400 ng) and quantified PAN RNA production by northern blot (Figure 7B and 7C). As previously reported, CMV- $\Delta$ 1 is completely unresponsive to ORF57 (purple bars). As predicted, deletion of the SL2, SL2-T, or the upper loop of SL2 abrogate PAN RNA accumulation in the presence of ORF57 (brown, yellow, or blue bars, respectively). Thus, SL2, and in particular the SL2-T portion of the predicted stem loop contains a *cis*-acting activity necessary for the full ORF57 responsiveness in PAN RNA. Therefore, based on the three assays employed in this study, we conclude that the core ORE maps to the 30-nt SL2-T.



**Figure 7.** The core ORE is necessary for ORF57 responsiveness in PAN RNA. (A) Schematic representation of the CMV-driven PAN RNA expression constructs. The wild-type (WT) and  $\Delta 1$  constructs were previously described [12]. Constructs that delete nt 23-63, 28-57, and 34-50 were generated to make the CMV- $\Delta$ SL2, CMV- $\Delta$ SL2-T, and CMV- $\Delta$ 34-50 constructs, respectively. (B) Representative northern blot data analyzing the effects of the

indicated deletions on ORF57 responsiveness. Panels on the left were probed for PAN RNA, while those on the right were the same blot probed with an endogenous loading control. The PAN RNA panels are from the same gel and are shown at the same exposure. (C) Quantification of the northern blot data. Samples were normalized to the 30 ng ORF57 samples with CMV-WT; error bars represent standard deviation ( $n=4$ ).

*In vitro* binding and  $\beta$ -globin assays suggest that multiple potential binding sites are present in the ORE and that, at least in some contexts, they are sufficient for binding and limited activity. Moreover, deletion of the core ORE does not have as dramatic of an effect on ORE activity as deletion of the full-length ORE. Therefore, we tested a deletion construct ( $\Delta 79-315$ ), which leaves SL2 intact, but removes the proposed lower affinity binding sites for ORF57 (Figure 4). In this case, we observe a similar lack of response to ORF57 as we do for  $\Delta$ SL2 (red bars). One interpretation of this result is that the weak secondary binding sites observed in the *in vitro* binding experiments are necessary for full ORE activity in the context of PAN RNA. Thus, both the core ORE and secondary sites are necessary for full ORF57-responsiveness in PAN RNA. Alternatively, the core ORE may be sufficient in PAN RNA, however, the ORF57 binding site is masked by alterations in RNA secondary structure introduced by the deletions.

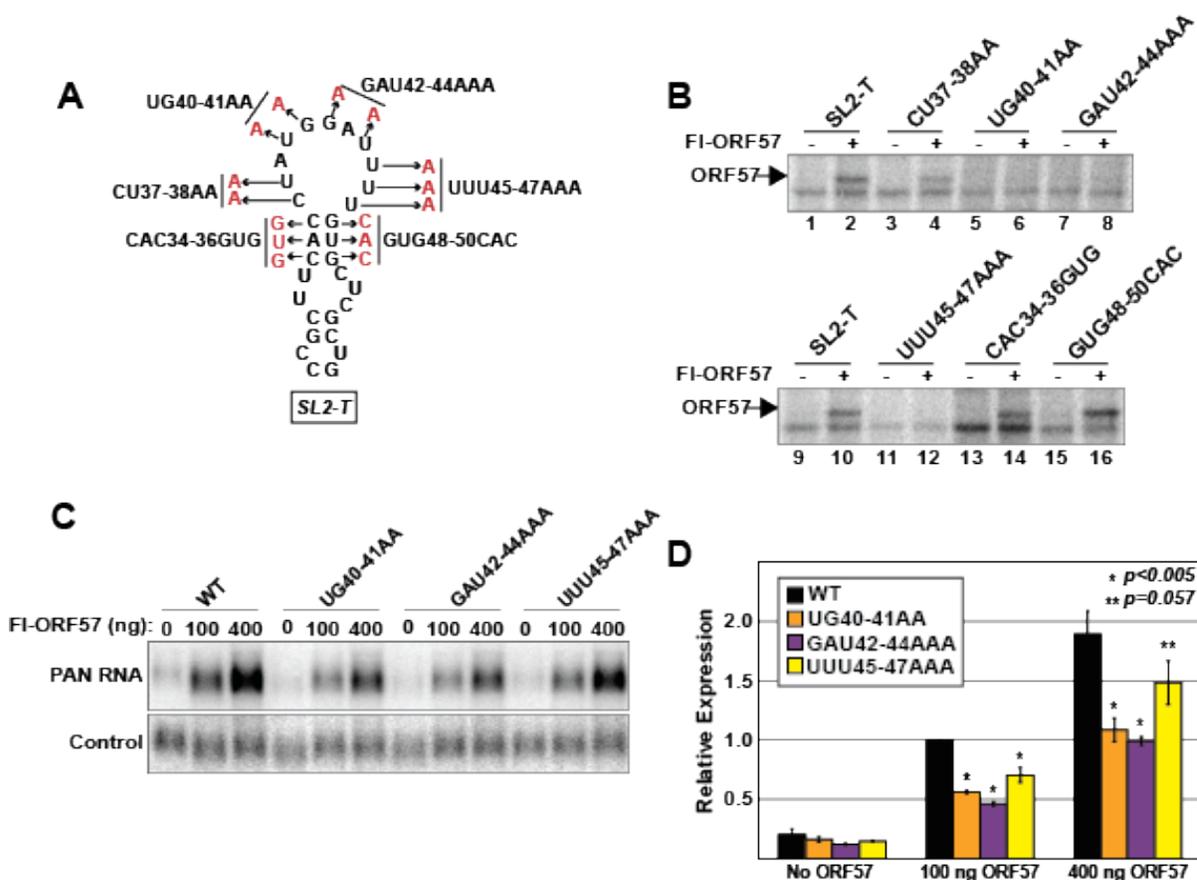
#### *Point mutations in the core ORE abrogate ORF57 binding and activity*

To examine activity of the core ORE at the nucleotide level, we generated a series of six point mutants that alter the 11-nt loop or the short stem of SL2-T (Figure 8A). Label

transfer assays were employed to test the ability of ORF57 to bind each of the mutant substrates (Figure 8B). Mutations in the first two nucleotides of the 11-nt loop have no effect on ORF57 binding (CU→AA<sub>37-38</sub> lanes 3,4). In contrast, ORF57 does not bind to any of the other three mutants tested in the SL2-T loop (lanes 5-8, 11,12). We also made mutations that disrupt the base pairing in the three nt stem structure immediately below the SL2-T loop. Neither of these mutations affect ORF57 binding (lanes 13-16), demonstrating that this structure is not necessary for ORF57 binding *in vitro*. We additionally generated a substrate containing mutations that restore the stem by combining CAC→GUG<sub>34-36</sub> with GUG→CAC<sub>48-50</sub>. As predicted from the lack of effect of the individual triple point mutants, these compensatory mutations had no effect on ORF57 binding *in vitro* (data not shown). Together, these data suggest that the sequences in nt 40-47 comprise a sequence-specific binding site for ORF57 and that the secondary structure contributes little to binding *in vitro*.

We next introduced the mutations that abrogate binding *in vitro* into the PAN RNA expression constructs and examined their effects on PAN RNA accumulation in the presence of ORF57 (Figure 8C, 8D). While the mutated constructs retain significant ORF57-responsiveness (Figure 8C), accumulation of PAN RNA is reduced by 2-fold in two of the mutants (UG→AA<sub>40-41</sub>, GAU→AAA<sub>42-44</sub>). The two-fold effect is observed both at 100 ng and 400 ng of ORF57 and is highly statistically significant ( $p < 0.005$ ), so we conclude that these specific nucleotides are central to ORF57 activity on PAN RNA. The third mutation tested, UUU→AAA<sub>45-47</sub>, reduces PAN RNA levels by a small (30%), but statistically significant ( $p < 0.005$ ), quantity when co-transfected with 100 ng ORF57. Upon transfection of four times as much ORF57, the difference still exists, but it is of questionable statistical significance

( $p=0.057$ ). Taken together, these data suggest that the nucleotide sequence AUGGAUUUU contributes significantly to the binding of ORF57 to PAN RNA, but also demonstrate that other binding sites contribute to the ORF57 response.



**Figure 8.** Point mutations in the loop portion of SL2-T abrogate ORF57 binding and response. (A) Schematic diagram of triple and double point mutations generated in SL2-T. The mutations introduced are shown in red adjacent to the name of the mutant. (B) Label transfer assays with mutant substrates. The mutant substrates were generated in the context of

the SL2-T sequence, which is used as a positive control (lanes 1, 2, 9, 10). Presence of FL-ORF57 in the extract is indicated above each lane and the position of ORF57 is given by the arrow. (C) Representative northern blot data analyzing the effects of the indicated mutations on ORF57 responsiveness probed for PAN RNA or a co-transfected loading control. The PAN RNA panels are from the same gel and are from the same exposure. (D) Quantification of northern blots. Values are relative to the CMV-WT at 100 ng transfection. The error bars are standard deviation ( $n=3$ ). The p-values are derived from a two-tailed unpaired Student's t-test comparing each sample to the WT.

## Discussion

The ORF57 protein is an essential KSHV factor with multiple roles in viral gene expression [56, 138, 141, 142]. While it likely plays a role in transcription, most of its proposed activities involve posttranscriptional regulation of RNA metabolism and its RNA-binding properties are important for these functions. However, the mechanisms ORF57 uses to recognize its ligand RNAs are unknown. Here, we delineate a core ORF57-responsive element in PAN RNA, a known ORF57 target in both cultured and infected cells [12-14, 55, 58]. Conservatively, we define the core ORE to be a 30-nt sequence, SL2-T, because it fits all three criteria tested in the present work. That is, SL2-T 1) binds to ORF57 *in vitro*, 2) enhances intronless  $\beta$ -globin RNA levels in the presence of ORF57, and 3) is necessary for ORF57-responsiveness of PAN RNA. Most likely, the core activity is contained in an even smaller region. Specifically, the nucleotides AUGGAUUUU in the unstructured loop at the

top of the core ORE are essential for binding and for full ORF57-responsiveness in PAN RNA. Thus, these nine nucleotides may constitute an ORF57 RNA binding motif.

ORF57 post-transcriptionally enhances the expression of many viral genes as well as various reporter constructs, so it may seem that its binding to RNA is relatively nonspecific. Indeed, in our  $\beta$ -globin assays (Figures 5 and 6) there is a consistent background of ~4-5-fold enhancement by ORF57 in the absence of any ORE. This relatively non-specific effect may be important for the biology of ORF57. In mammalian cells, the stages of nuclear pre-mRNA processing are tightly coupled, and abrogation of one activity often affects other processing steps [11, 147-149]. Pre-mRNA splicing is central to this coupling, but most KSHV transcripts lack introns [9]. As a result, KSHV genes are missing an important component of efficient expression. One ascribed function of ORF57 is to enhance the expression of intron-lacking viral mRNAs. As such, it seems likely that ORF57 would evolve a relatively promiscuous binding specificity to enable it to bind the large pool of intronless viral mRNAs expressed during lytic phase. Peculiarities of individual viral transcripts may require more robust binding by ORF57 to yield a higher activity. For example, the nuclear accumulation of PAN RNA appears to require a stable complex with ORF57 to protect it from nuclear decay enzymes. Perhaps in these cases, transcripts evolved *cis*-acting OREs like the one described here.

While this work was in preparation, Zheng and colleagues published a different ORF57-responsive element derived from the viral interleukin-6 (vIL-6) mRNA [93]. The vIL-6 MRE (Mta-responsive element) overlapped with a miRNA binding site and the authors proposed that binding of ORF57 antagonized the down-regulatory effects of the miRNA.

While this mechanism is unlikely to be responsible for ORF57-dependent stabilization of the nuclear PAN RNA, binding of ORF57 to the vIL-6 MRE or to the PAN ORE is essential for its regulatory function. Comparison of the core PAN ORE with the vIL-6 MRE reveals little overall sequence homology. However, there are two interesting similarities. Consistent with the idea that high-affinity ORF57 binding is driven by sequence rather than structure, both elements are predicted to fold into stem-loops, but neither shows particularly strong secondary structures. Even more compelling, the uppermost loops of both the vIL-6 MRE and the PAN ORE share the tetranucleotide GGAU. This element is necessary for ORE activity because mutation of this motif (UG→AA<sub>40-41</sub>, GAU→AAA<sub>42-44</sub>) abrogates binding *in vitro* and ORF57-responsiveness in cells. Thus, these data suggest that the GGAU motif may provide sequence specificity for ORF57 binding that is utilized in multiple transcripts. Interestingly, the herpesvirus saimiri ORF57 protein has been reported to preferentially bind a purine-rich GGAGR element [92], so there may be at least loose conservation of binding-site specificity with respect to the GGA trinucleotide. Identification and detailed examination of additional ORF57-responsive elements is necessary to test whether this sequence is a general feature of ORF57 regulated RNAs.

In the same study that identified the vIL-6 MRE [93], a nonbiased approach was taken to clone viral RNA fragments that interact with ORF57. Consistent with our results, they found that PAN RNA interacted with ORF57; 16 out of 91 PAN RNA clones (18%) overlapped the SL2-T core ORE sequence. Surprisingly, other clones spanned nearly every region of PAN RNA with close to 50% overlapping the ENE, a *cis*-acting stability element in PAN RNA [109, 110, 112]. No validation of the cloned PAN RNA fragments for their ability

to confer ORF57-responsiveness was performed in that study, but our data clearly show that ORF57 binding and response is conferred by sequences residing at the 5' end of PAN RNA. In fact, deletion of the ENE actually increases ORF57-responsiveness, presumably due to the fact that ORF57's stabilization effects are greater on an inherently less stable transcript [12]. Moreover, our *in vivo* and *in vitro* UV cross-linking results do not support a direct interaction between ORF57 and the 3' end of PAN RNA (Figures 3 and 4).

Our data are consistent, however, with the general model that the core ORE is not the only binding site in PAN RNA for ORF57. Sequences contained in nt 79-315 bind PAN RNA *in vitro* and deletion or mutation of the core ORE is not sufficient to completely abrogate ORF57-responsiveness (Figures 7 and 8). In contrast, when the full-length ORE is deleted, virtually no ORF57 response is observed. We speculate that the core ORE efficiently recruits ORF57, which has been shown to be in a homomultimer complex of unknown number[58]. Once one ORF57 molecule is bound to the ORE, the RNA binding domain(s) on the unbound ORF57 molecule(s) are present in a high local concentration driving binding to weaker sites in adjacent sequences. Similar cooperative models have been proposed for other RNA-binding proteins like hnRNP A1 and the HIV REV protein [150, 151]. The presence of high ORF57 levels and/or multiple weak ORF57-binding sites can drive this interaction in the absence of a high-affinity ORE leading to the observed ORF57 promiscuity.

The steady-state analyses of PAN RNA and  $\beta$ -globin used in this study monitored the accumulation of transcript as a measurement of ORF57 activity. These approaches are confounded by the presence of an uncharacterized enhancer of PAN RNA expression that overlaps the full-length ORE [12, 109]. When driven by the PAN RNA promoter, deletion of

the full-length ORE decreases PAN RNA levels by ~5-fold in the absence of ORF57. Use of the CMV promoter decreases the magnitude of the effect of this enhancer, but some up-regulation of RNA levels by this element is still observed. Close examination of the no ORF57 controls in this study suggest a subtle trend of increased RNA accumulation in the presence of the core ORE in the absence of ORF57 (Figures 5-8). To be sure, careful quantification of these data rarely yields a statistically significant effect (data not shown), so it is difficult to interpret the relevance of the trend. One formal possibility is that the core ORE does not drive ORF57-response directly, but rather it is an element that has a function in gene expression that lies upstream of ORF57 activity. However, this idea seems unlikely based on our *in vitro* binding and  $\beta$ -globin reporter assays. We favor the idea that the element may bind to cellular factors that act in concert with ORF57 to promote gene expression. In the absence of ORF57, the activity of the cellular factor is minimized, but in its presence the effect is enhanced. Consistent with this idea, *in vitro* binding of baculovirus-expressed ORF57 is enhanced by the presence of cellular extract [62]. Further experimentation is required to unravel the complex molecular mechanisms of these primary and secondary ORF57-responsive elements in PAN RNA.

ORF57 is a multifunctional regulator of gene expression that is essential for KSHV replication. In order to take a meaningful global approach to examine ORF57 targets, we must first employ reductionist techniques to define the requirements for ORF57-RNA interactions. Here we have identified a core element from a natural ORF57 target that is sufficient for ORF57 binding and response. These data provide a foundation to compare ORF57-responsive elements from novel viral and cellular targets as they are uncovered.

## CHAPTER FIVE

### **Identification of Kaposi's sarcoma-associated herpesvirus ORF57 RNA targets using HITS-CLIP**

#### **Introduction**

Kaposi's sarcoma-associated herpesvirus (KSHV; HHV-8) is a human gammaherpesvirus and the etiological agent associated with Kaposi's sarcoma (KS), primary effusion lymphoma (PEL) and some cases of multicentric Castleman's disease (MCD) [1, 21, 23, 24, 134, 135]. The KSHV life cycle includes both a latent and a lytic state. During latency, only a few viral genes are expressed. On the other hand, during lytic reactivation, KSHV orchestrates the ordered synthesis of numerous viral products that enable assembly of viral particles. This assortment of viral factors works in unison to manipulate the host cell and favor viral gene expression.

There are inherent differences between the structure of human and herpes genes, the most striking being that most of the KSHV genes encode intronless transcripts [7-9]. This observation represents an obstacle for KSHV given that the presence of an intron in a pre-mRNA and the act of splicing results in the deposition of multi-protein complexes that significantly enhance many stages of RNA metabolism including transcription, mRNA 3' end cleavage, polyadenylation, export, mRNA stability and translation [11]. One of the virally encoded proteins utilized by KSHV to overcome this roadblock is ORF57. ORF57 is essential for productive viral replication and it is involved in multiple aspects of viral gene expression ranging from transcription to translation [13, 14, 55-59, 83, 89]. A previous report

has shown that ORF57 promotes splicing of introns with weak splice sites [57]. Additionally, ORF57 can bypass the need for pre-mRNA splicing by recruiting cellular RNA export adapter proteins REF/Aly and UIF to intronless viral RNAs, both of which are individually capable of recruiting the hTREX complex and facilitating TAP-mediated export [78, 79, 81-87].

ORF57 also plays a role in RNA stability. The exact mechanism governing RNA upregulation in the presence of ORF57 remains to be solved, but ORF57's ability to bind its targets correlates with RNA upregulation [12, 15]. There are multiple publications documenting the effects of ORF57 on different RNAs [12, 15, 55, 58, 62, 68, 83, 90, 91]. In some cases, ORF57 recruitment is mediated by particular sequence elements on the RNAs [15, 92, 93], but analysis of a larger transcript pool is necessary to make any general conclusions. Work from our laboratory has shown ORF57 preferentially binds to the 5' end of PAN RNA [15]. In this region, we identified an ORF57 responsive element (ORE) that we hypothesize works as a high-affinity binding site for ORF57 [12, 15]. Independent research from the Zheng laboratory also immunoprecipitated RNA fragments mapping to the 5' end of PAN RNA [93]. Additionally, this laboratory identified an RNA sequence in the outer loop of vIL-6 that is bound by ORF57 [93]. These observations suggest that in some cases ORF57 recruitment might be favored or enhanced by the presence of particular sequence elements like the ORE.

Here, we set out to identify direct host and viral targets of ORF57 in the context of viral infection. Previous work from our laboratory has shown that ORF57-RNA target reassortment can occur during cell lysis [12]. Meaning that upon cell lysis, naturally

occurring RNA-protein complexes might dissociate (resulting in a false negative result) or also that RNA binding proteins might associate with RNAs that are not normally bound *in vivo* (resulting in a false positive result) [123, 124]. Therefore, we performed the high-throughput sequencing of RNAs isolated through crosslinking and immunoprecipitation (HITS-CLIP) procedure to identify direct ORF57 targets, as well as potential RNA binding sites for this multifunctional protein [121, 125-128, 152]. As predicted from previous reports [12, 15], we identified CLIP tags mapping to the 5' end of PAN RNA. Additionally, our data suggests that ORF57 binding to other viral genes can be promiscuous and that, in some cases, binding can occur at multiple sites across the target RNAs. Bioinformatic analysis of potential host ORF57 targets reveals that ORF57 binding is enriched near the 5' end of a subset of the transcripts and often close to the first exon-intron junction. We then monitored the RNA levels of a select number of potential ORF57 targets at different time points following lytic induction. Interestingly, the levels of the ORF57-bound pre-mRNAs persist while their corresponding mRNAs disappear rapidly, suggesting that ORF57 is inhibiting host splicing. In contrast, both pre-mRNAs and mRNAs from GAPDH and  $\beta$ -ACTIN transcripts rapidly disappear [16, 17]. These data provide further insight into ORF57 recruitment and binding and suggest a novel functional role in host RNA regulation.

## Results

### *ORF57 HITS-CLIP*

To identify host and viral RNA targets of ORF57 we performed HITS-CLIP [121, 125-127] in lytically reactivated TREx BCBL1-Rta cells [118] at ~20 hours post induction (hpi) (Appendix B). We modified the HITS-CLIP protocol to improve the solubility and recovery of ORF57-RNA complexes from lysates by 1) decreasing the UV-crosslinking energy, 2) incorporating an initial protein denaturation step from a previously published UV-crosslinking protocol [123, 124] and 3) lowering the temperature used for MNase treatment. Detailed protocols are included in Appendices A-D. The HITS-CLIP experimental procedure is outlined in Figure 9. TREx BCBL1-Rta cells were UV crosslinked, collected, and frozen at 20 hpi. Cell pellets were lysed as previously published in [123, 124], with minor changes noted in Appendix C. Lysates were first DNase treated and later incubated with MNase to partially digest the RNAs in the lysate. The soluble lysate fraction was used to immunoprecipitate covalently bound ORF57-RNA complexes. The immunoprecipitated RNAs were end-labeled and separated in a gradient NUPAGE gel. Protein-RNA complexes were transferred to a nitrocellulose membrane, ORF57-RNA complexes were dissected (see Figure 10) and then proteinase K treated to recover the bound RNAs. The isolated RNAs were run in a urea gel and the RNAs ranging from 50-300 nt were eluted from the gel. Eluted RNAs were then utilized for Illumina strand specific library preparations.

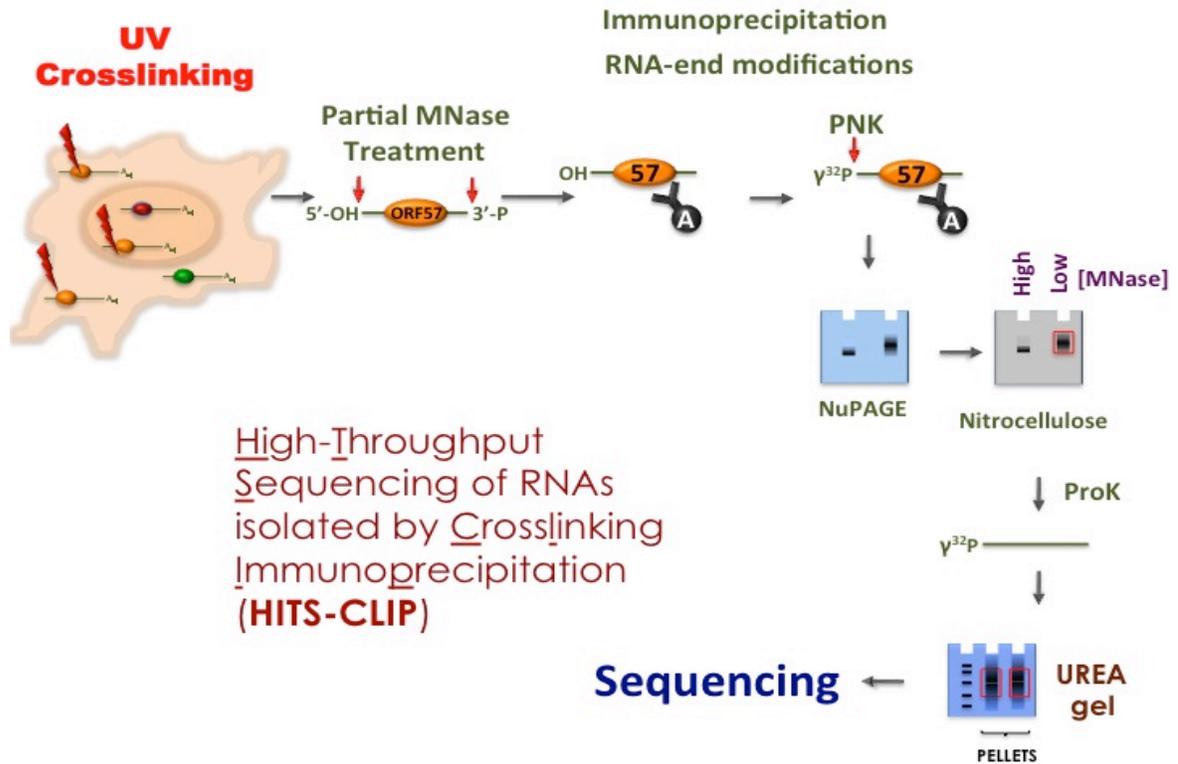


Figure 9. Experimental outline of ORF57 HITS-CLIP. See text for details. Image was modified from [121].

Prior to starting with library preparations, we ran a series of control experiments to assess the quality and specificity of the immunoprecipitated ORF57-RNA complexes. First we tested immunoprecipitation of ORF57-RNA complexes from TReX BCBL1-Rta cells under different experimental conditions (Figure 10A). As expected, the immunoprecipitation of ORF57-RNA complexes was undetectable when we used a non-specific antibody (lane 1), as well as when the protein-RNA complexes were not subjected to UV light. Uninduced samples retain a weak signal consistent with observations that a small population of infected cells undergoes spontaneous lytic reactivation (see Figure 10B lane 2 and 5) [3]. High concentrations of MNase result in a condensed protein-RNA complex band near 51kDa (Figure 10A lane 4) that shifts to a higher molecular weight (MW) when the MNase concentration is decreased (lane 5). In addition to the “50kDa” complex, we observed a signal from a lower MW complex (“37kDa”), one around 100kDa and one at 150kDa. We then used western blotting (Figure 10B) to confirm only the 50kDa, 100kDa and 150kDa signals are ORF57 specific. Notice the 100kDa and 150kDa protein bands are only present upon treatment with UV light. We therefore conclude that the 100kDa band likely represents two ORF57 protein molecules crosslinked to the same RNA. A similar phenomenon was observed with TDP-43, an RNA binding protein that when mutated causes amyotrophic lateral sclerosis (ALS) [153].

Finally, we assayed the length and origin of the RNAs present on each protein-RNA complex. RNAs purified from the 37, 50, and 100kDa complexes (Figure 11) show the average RNA fragment size increases according to the complex MW. This is consistent with previous observations suggesting these slow migrating complexes represent ORF57

homomultimers bound to the same RNA, thus resulting in longer RNA fragments. We rationalized that because ORF57 binds PAN RNA directly, we could use the presence of PAN RNA in the different MW complexes as an indirect measure of the location of biologically relevant ORF57-RNA complexes. We analyzed RNAs purified from the different MW complexes through northern blots (Figure 12). PAN RNA probe shows significant IP of PAN RNA fragments in the 50kDa and 100+150kDa bands, while a 5.8S rRNA probe shows rRNA signal concentrating in the 37kDa band. Thus, we concluded both the 50kDa and the 100+150kDa are the most likely to contain specific ORF57-RNA complexes, whereas the 37kDa band is likely a contaminating protein. We sequenced and performed preliminary analyses on RNAs from the “monomeric” (~50kDa band) and the slow migrating protein-RNA complexes (~100+150kDa bands) and found that both complexes contain similar RNA populations. While transcripts present on both complexes are similar, the overall number of CLIP tags was lower in the ~100+150kDa bands, as could be expected from on the weak RNA and protein signals we observed for this complex (Figure 10-12). Based on these observations (Figures 10-12), we chose the 50kDa ORF57-RNA complex band to continue with the Illumina library preparations (Figure 10A lane 5).

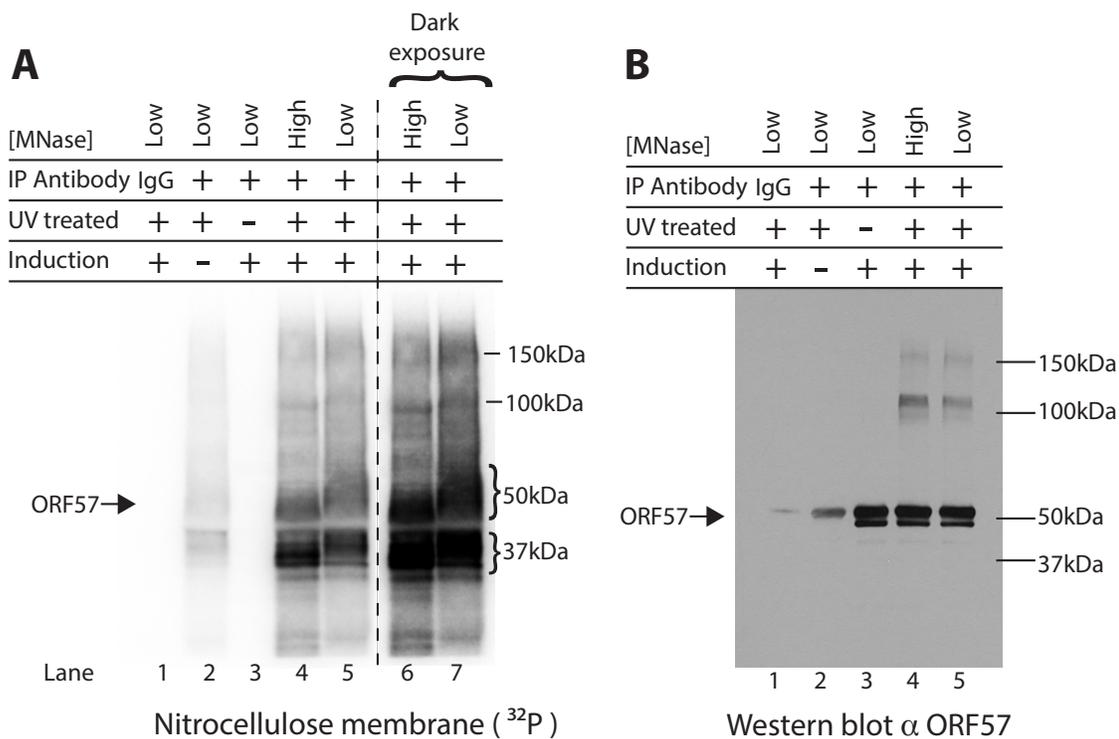


Figure 10. Representative ORF57 HITS-CLIP results. A) Phosphoimager picture depicting nitrocellulose membrane with ORF57-RNA complexes. Lanes 1-3 are negative controls: IP with a prebleed antibody (IgG control), uninduced cells (latent infection), and no UV-crosslinking controls, respectively. Lane 4 and 5 show ORF57-RNA complexes treated with high or low concentrations of MNase, respectively. Lanes 6 and 7 show a dark exposure of the previous two lanes. B) Western blot of HITS-CLIP samples. Samples were run in a NUPAGE gradient gel and loaded in the same order as Figure 10A. Affinity purified Rabbit $\alpha$ ORF57 was used to detect ORF57 specific protein bands. CleanBlot was used as the secondary antibody. Arrow on the left shows the location of ORF57 (~51kDa).

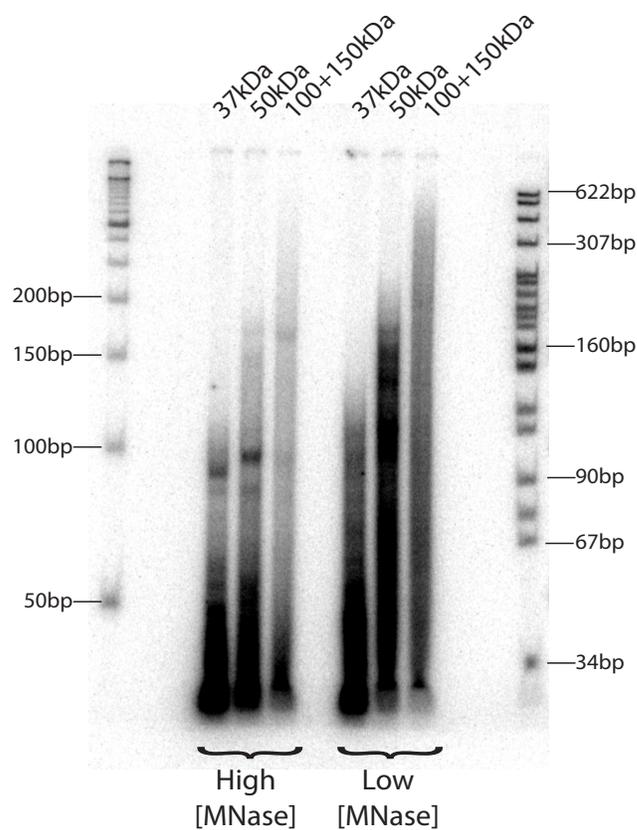


Figure 11. RNA fragments recovered from nitrocellulose membrane. The 37kDa, 50kDa and 100+150kDa protein-RNA complex bands were cut from the nitrocellulose membrane, proteinase K treated and the purified RNAs were then run in a urea gel. RNAs isolated from the condition treated with high MNase concentrations are loaded on the left, while the fragments digested with a low MNase concentrations are on the right.

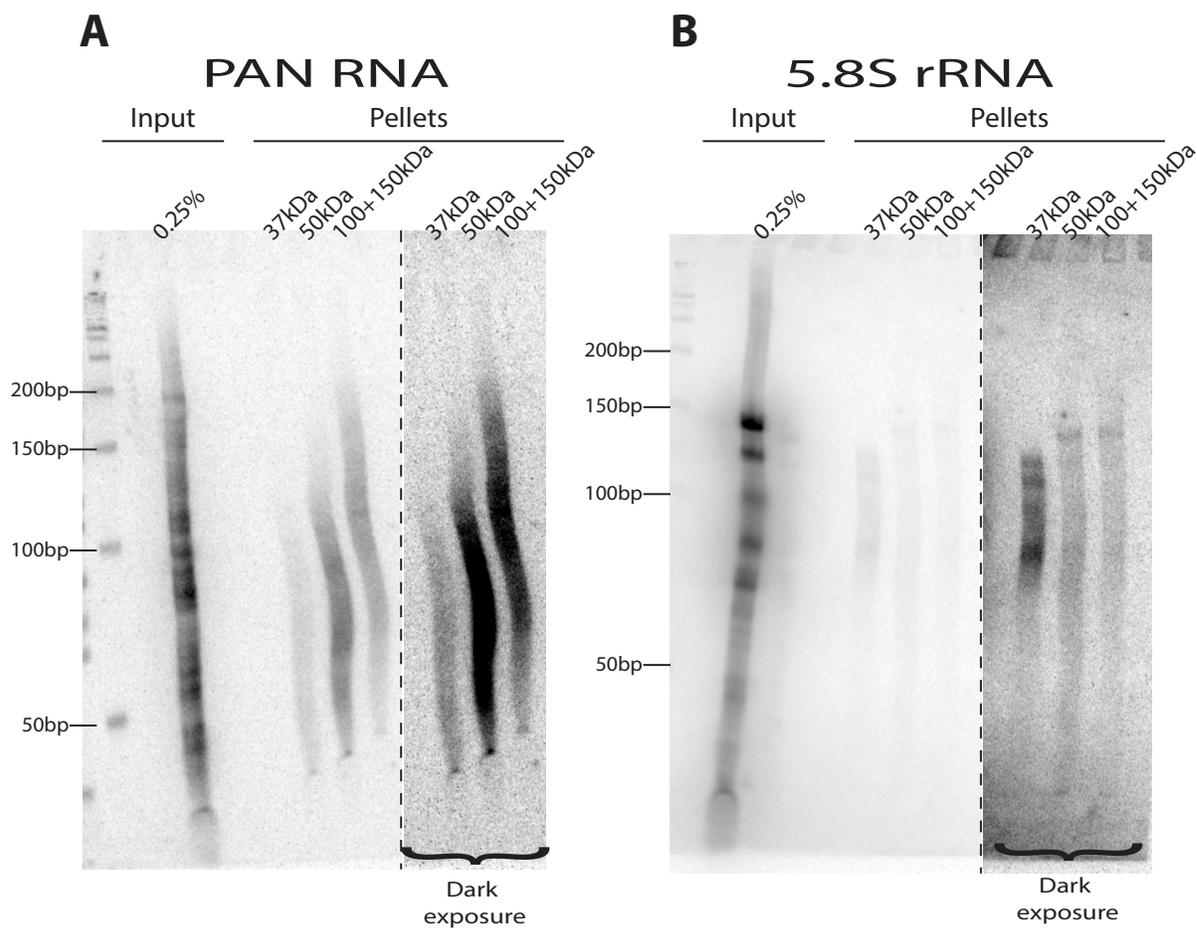


Figure 12. Northern blots of RNAs extracted from the nitrocellulose membrane. RNAs digested with a low MNase concentration were extracted from the 37kDa, 50kDa and 100+150kDa protein-RNA complex bands and used to run a northern blot. PAN RNA (A) and 5.8S rRNA (B) were detected using an oligo probes. DNA ladder was included on the left of each membrane.

PAN RNA 5' end oligo probe:

(5'-ATCGGCGGCACCAATGAAAACCAGAAGCGGCAAGAAGGCA-3').

5.8S rRNA oligo probe: 5' -GTGTCGATGATCAATGTGTCCTGCAATTCA - 3'.

The nomenclature and other bioinformatic analysis considerations are noted below. “Pellet” refers to the immunoprecipitated (ORF57-bound) RNA fragments that underwent the HITS-CLIP protocol. “Input” refers to RNAs that were isolated from the exact same time point as the pellets but did not undergo UV treatment or immunoprecipitation. CLIP tags are sequencing reads recovered after crosslinking and immunoprecipitation (CLIP). “Enriched clusters” are defined as continuous regions of the genome where CLIP tags are enriched in the pellets when compared to the equivalent region in the input samples. Mapping statistics for all the biological replicates are listed in Figure 13 and the general outline for bioinformatic analysis is shown in Figure 14A (for more details see Chapter three). Scatter plots and Pearson’s correlations show recovery of CLIP tags is highly reproducible and similar among the different biological replicates (Figure 15). HITS-CLIP analysis of all the recovered CLIP tags revealed a total of 2,448 enriched clusters. Gene ontology analysis of these clusters did not show significant representation of genes in any particular category [154, 155]. Furthermore, mutation analysis identified point mutants enriched in the immunoprecipitated tags. HITS-CLIP induced mutations represent ORF57 specific, single-nucleotide resolution mutations that occur at a higher rate in the pellet CLIP tags when compared to the input CLIP tags. These mutations represent high-confidence ORF57 binding events that likely reflect points of direct RNA-protein interaction [156].

## Mapping statistics

	Total Read	Mapped Read (hg19)	percentage Mapped (hg19)	Mapped Read (virus)	percentage (virus)	Insert size > 400nt	percentage (insert)	Introns in read	Percentage (intron)
I1	69047995	27933655	40.4%	22887292	33.1%	245265	0.35%	2972357	4.30%
I2	97841141	41396437	42.3%	35282358	36.0%	476072	0.48%	4029222	4.11%
I3	83944131	36403063	43.3%	29800739	35.5%	392984	0.46%	3316209	3.95%
P1	60549062	12368578	20.4%	1924779	3.17%	32704	0.05%	173816	0.28%
P2	41225319	13582522	32.9%	2241643	5.43%	24880	0.06%	213228	0.51%
P3	56835116	8441660	14.8%	1827394	3.21%	17742	0.03%	149388	0.26%

Mapping across spliced junctions (sample I1):

Mapped reads with introns + Mapped paired > 400 apart + Reads mapped to transcriptome  
Total mapped reads

$$= \frac{2972357 + 245265 + 118211}{27933655 + 22887292} = 6.56\%$$

Figure 13. Illumina Mapping Statistics. The read count and mapping statistics of 3 biological replicates. Samples labeled as “I” and “P” correspond to Input and Pellet (HITS-CLIP) samples, respectively. Reads spanning across spliced junctions were discarded from our analysis.

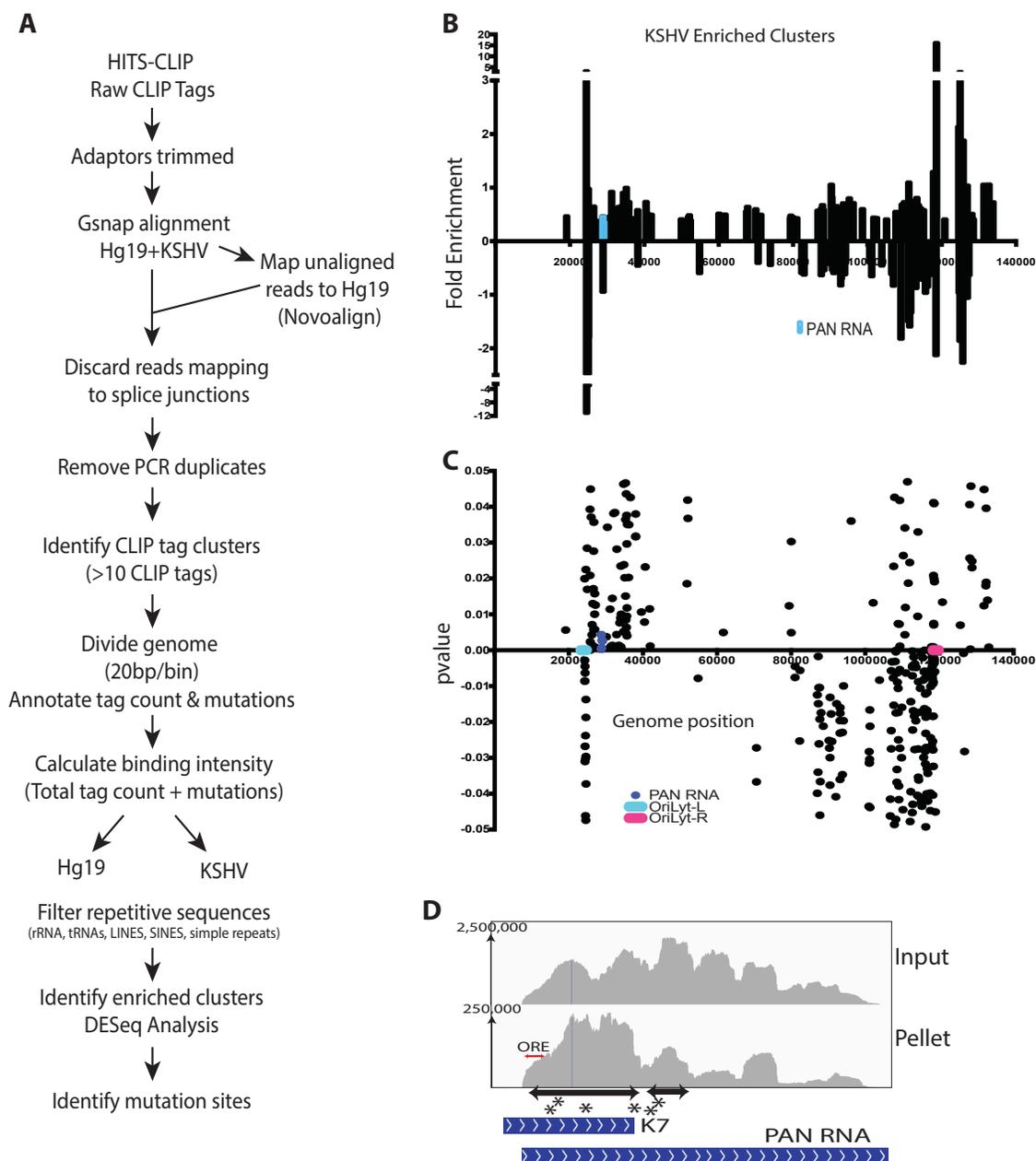


Figure 14. ORF57 HITS-CLIP unveils a considerable number of bound viral RNAs. A) Outline indicating how sequencing reads were processed. B) Genomic location of KSHV enriched clusters. The middle point of each cluster was used as the x-coordinate. The y-axis represents the fold enrichment of the Pellets over the Input samples. For display purposes, the

clusters mapping to the KSHV genome in the “right-to-left” orientation were arbitrarily given a negative fold enrichment to distinguish them from the clusters mapping in the “left-to-right” orientation. PAN RNA clusters are shown in blue. C) KSHV HITS-CLIP single nucleotide mutations. For display, p values of mutations from reads mapping in the “right-to-left” orientation were arbitrarily assigned negative values. PAN RNA mutations are shown in dark blue. oriLyt-L and oriLyt-R are shown in light blue and pink, respectively. D) IGV traces for PAN RNA. Numbers at the top left of each frame correspond to the scale of the y-axis and correlate to the total number of reads stacked at a given window. Black and red arrows emphasize enriched clusters and ORE position, respectively. (\*) represent HITS-CLIP induced point mutations.

*ORF57 HITS-CLIP reveals new ORF57-RNA interactions in KSHV*

Our analysis identified 219 enriched clusters mapping to the KSHV genome. The middle point of each enriched cluster identified is represented in Figure 14B with its corresponding fold enrichment value. Interestingly, the most highly enriched clusters and a subset of HITS-CLIP induced mutation sites in KSHV RNAs (Figure 14C) are concentrated near the origins of lytic replication (oriLyts). The ORF57-oriLyt-L interaction is consistent with previous observations made from ChIP-chip experiments [47] in which ORF57 was found to associate with OriLyt-L. The array omits regions of the genome with highly repetitive sequences, which may explain why we did not observe an ORF57-dependent signal close to the oriLyt-R. Integrative Genomics Viewer (IGV, [157, 158]) snapshots (Figure 14D) demonstrate CLIP tags mapping to PAN RNA. The 5' binding to PAN RNA occurs as expected from previous observations [12, 15, 81]. Our analysis identified two enriched clusters and three HITS-CLIP induced mutations at the 5' end of PAN RNA (Figure 14D). Our analysis also identifies HITS-CLIP induced point mutations on the antisense strands for several ORFs. We are currently trying to determine why these antisense mutations are only observed in the viral genome and therefore, further analysis on the viral clusters is currently pending.

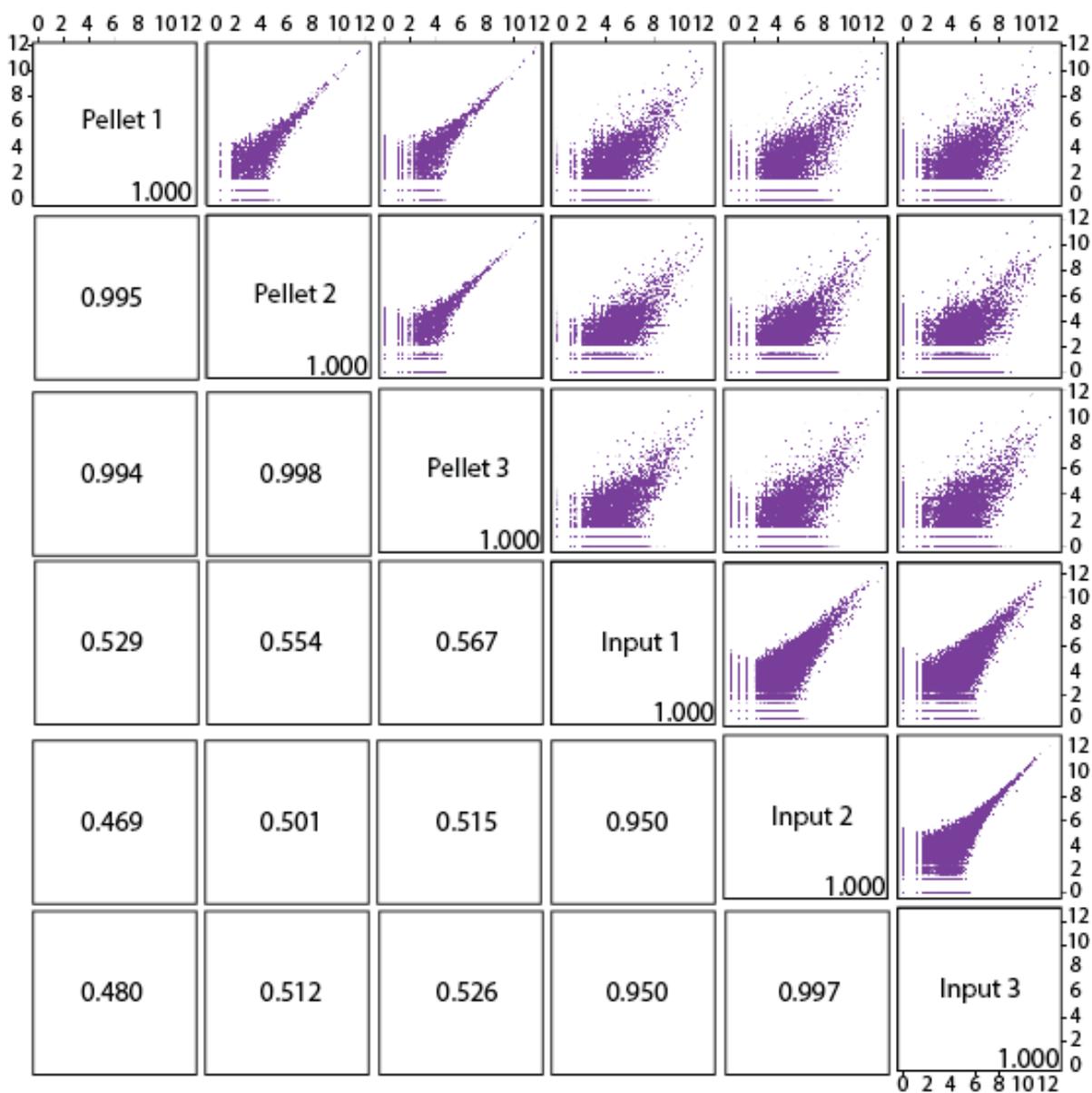


Figure 15. Scatter plot and Pearson's correlation coefficients for Input and Pellet (HITS-CLIP) samples. Scatter plots are shown on the upper right quadrant and Pearson's correlations are shown on the lower left quadrant. For scatter plots, the CLIP tags in the human and KSHV genomes were binned (20bp/bin) and the raw tag count data was log

transformed as follows:  $\log_e(\text{Raw tag count} + 1)$ . Pearson's correlations represent the correlation between raw tag counts of any 2 conditions.

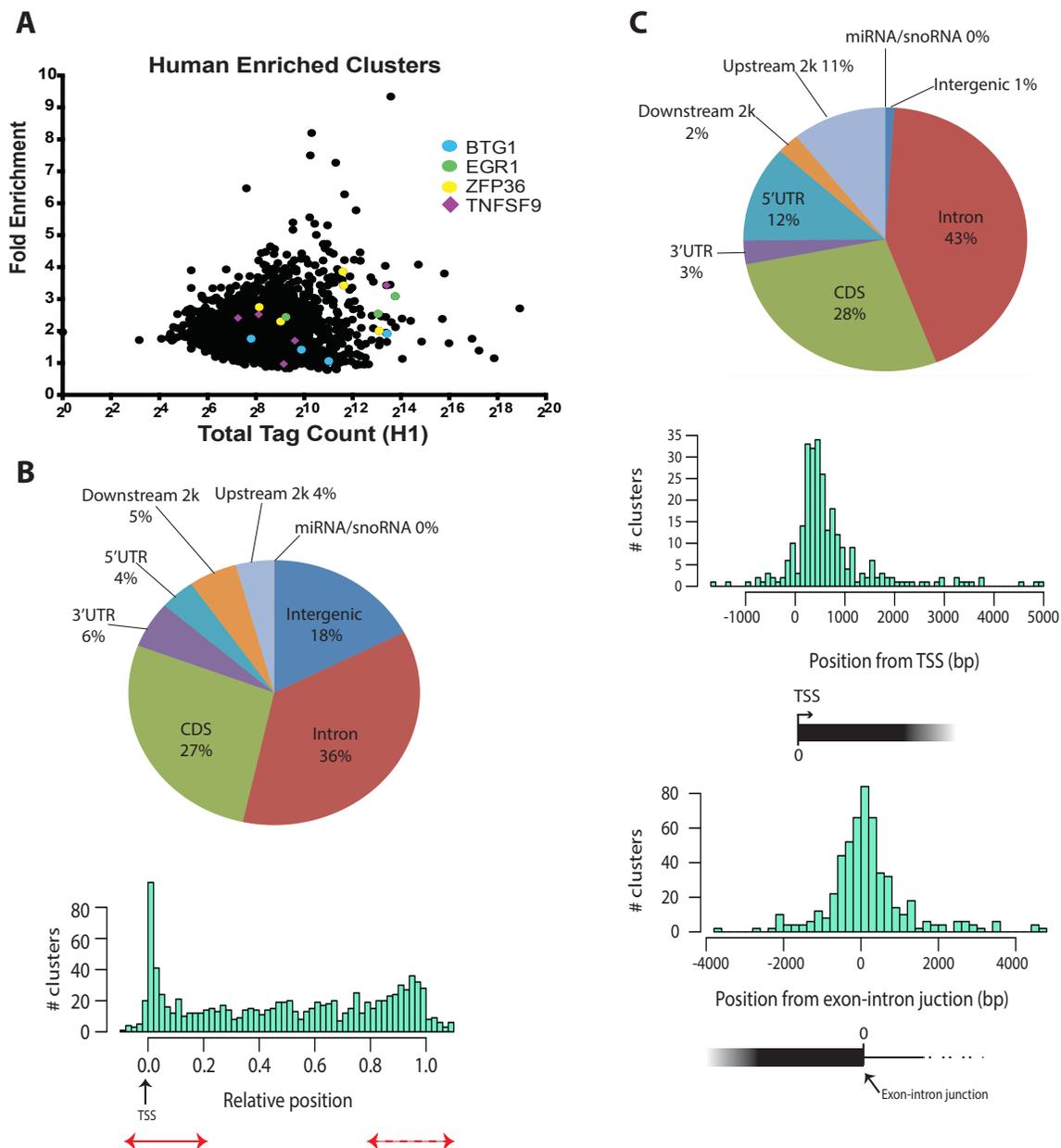


Figure 16. Enriched human clusters. A) Fold Enrichment vs. Total Tag Count. The fold enrichment values of the HITS-CLIP (Pellet) clusters over the Input clusters were calculated

considering the total tag count for all 3 biological replicates and constitute the y-axis. The x-axis values represent the total tag count for one biological replicate. A select number of ORF57-bound transcripts are shown in different colors: (●) represents single intron transcripts and (◆) represents transcripts with  $\geq 2$  introns. Note most clusters remain unannotated (in black). B) Enriched clusters (n=2,226) were classified based on their location (pie graph). The bar graph catalogs the position of the human clusters relative to the length of the closest annotated transcript. 0.0 and 1.0 correspond to the transcription start site (TSS) and the end of the transcript, respectively. Solid and dashed red arrows mark clusters enriched at 5' and 3' ends, respectively. The upstream 2k category includes clusters mapping up to 2000 bp prior to the TSS. The downstream 2k category includes clusters mapping up to 2000 bp after the end of the transcript. If a cluster is shared by  $\geq 2$  categories, the cluster is split among them. C) Enriched clusters mapping to the 5' end of the transcripts. Pie graph classifies the clusters (n=288) based on their relative position to the closest annotated gene. Top bar graph maps enriched 5' end clusters (in bp) relative to the TSS (shown as 0). Bottom bar graph maps the enriched 5' end clusters (in bp) relative to the first exon-intron junction (shown as 0).

*ORF57 binding is enriched at the 5' end of host RNAs*

A total of 2,229 enriched human clusters were identified in our assay and are represented in Figures 16 A and B. The identified clusters represent approximately 700 unique host genes that are bound by ORF57. A few human RNAs were selected for further analysis and are shown in different colors (Figure 16A). The highlighted RNAs contain multiple enriched clusters throughout the length of their transcript (also shown in Figure 17). More than half of the enriched human clusters map to coding sequences (CDS) or introns (Figure 16B). When arranged relative to their location to the closest annotated gene, a large portion of the enriched clusters are concentrated close to the 5' end of the transcript (Figure 16B). The clusters concentrated at the 5' end (Figure 16B, solid red arrow) were re-classified based on their genomic location (Figure 16C). As expected, this increases the percentage of clusters mapping to the 5' UTR and upstream 2K categories. The majority of these clusters map within the first kb after the TSS (Figure 16C, top bar graph).

When aligned based on their position to the first exon-intron boundary (Figure 16C, bottom bar graph), the cluster density peaks at the exon-intron junction. This observation is also evident in the IGV traces of some of these bound transcripts (Figure 17). We chose a few transcripts for further analysis based on their expression levels in our cell line and their potential biological relevance to KSHV pathogenesis. EGR1 (also known as ZNF225, KROX24 or AT225) is a potential cancer suppressor gene that functions as transcriptional regulator of genes involved multiple cellular processes such as inflammation and apoptosis [159]. ZFP36 (also known as TTP, NUP475, tristetraproline, GOS24 or TIS11A) belongs to a family of zinc finger proteins previously shown to bind ARE-mRNAs and is downregulated

in many cancers [160]. BTG1 expression inhibits cell proliferation and angiogenesis [161, 162]. TNFSF9 (also known as CD137L or 4-1BBL) is involved in T-cell-mediated immunity, as well as, the maturation and regulation of B cells [163, 164]. GAPDH and  $\beta$ -Actin CLIP-tag clusters are shown for comparison. The location of enriched clusters as well as HITS-CLIP induced point mutations identified is highlighted with black arrows and (\*), respectively. Consistent with our previous observations (Figure 16C), these transcripts display enrichment of CLIP tags close to the 5' end of the transcript and also at the exon-intron boundaries.

While sorting through the list of potential targets, we noticed several of the transcripts had a single intron. In addition, ORF57 selectively upregulates intronless transcripts [12, 83, 90]. We therefore asked whether genes with a few introns were overrepresented in the bound RNAs identified by our analysis (Figure 18). In contrast to our expectations, the average number of exons in the annotated genes represented by all of our enriched clusters (Figure 16B bar graph and Figure 18 yellow line), as well as the genes with clusters enriched at the 5' end of transcripts (Figure 16B solid red arrow and 16C; Figure 18 green line), does not deviate considerably from the average number of exons in the entire genome (Figure 18 red line). There is, however, a slight shift in the average exon number of clusters enriched at the 3' end of transcripts towards a larger number of introns (Figure 18 green line). At the moment, it is not clear whether this is of any biological significance. Further studies are necessary to determine if ORF57-dependent effects differ according to the location bound on the target RNA. Enriched clusters identified here (Figure 16 and 17) may be useful to guide

the experimental design of future experiments if we are to study how ORF57 binding patterns affect the fate of the bound transcript.

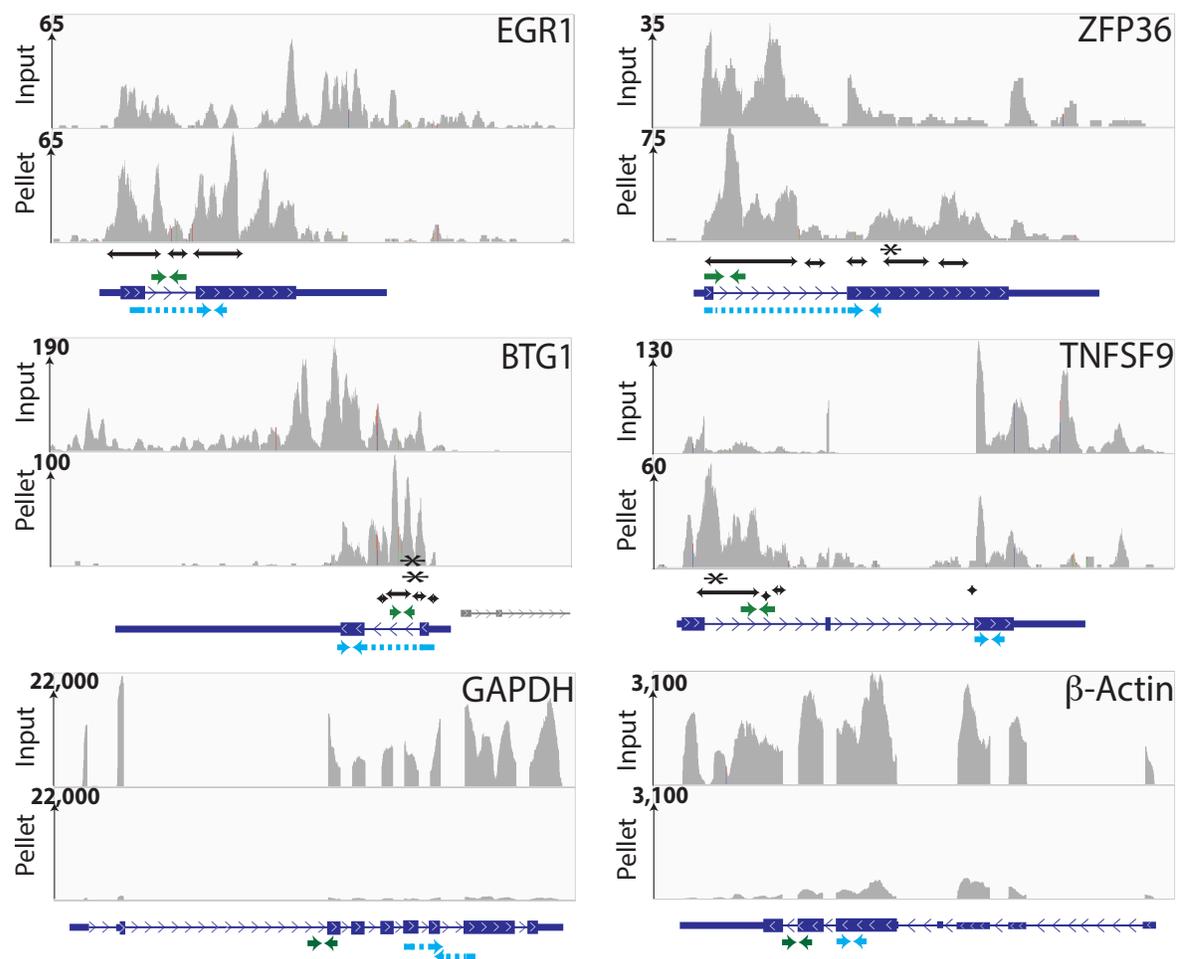


Figure 17. IGV traces of ORF57-bound RNAs. Numbers at the top left of each frame correspond to the scale of the y-axis and correlate to the total number of reads stacked at a given window. The x-axis for each transcript was scaled to fit in a single window. GAPDH and  $\beta$ -Actin were included for comparison. Black arrows emphasize identified enriched clusters. (\*) represent HITS-CLIP induced point mutations. Neighboring genes are shown in grey. qRT-PCR amplicons shown in light blue (mRNA) and green (intron/premRNA).

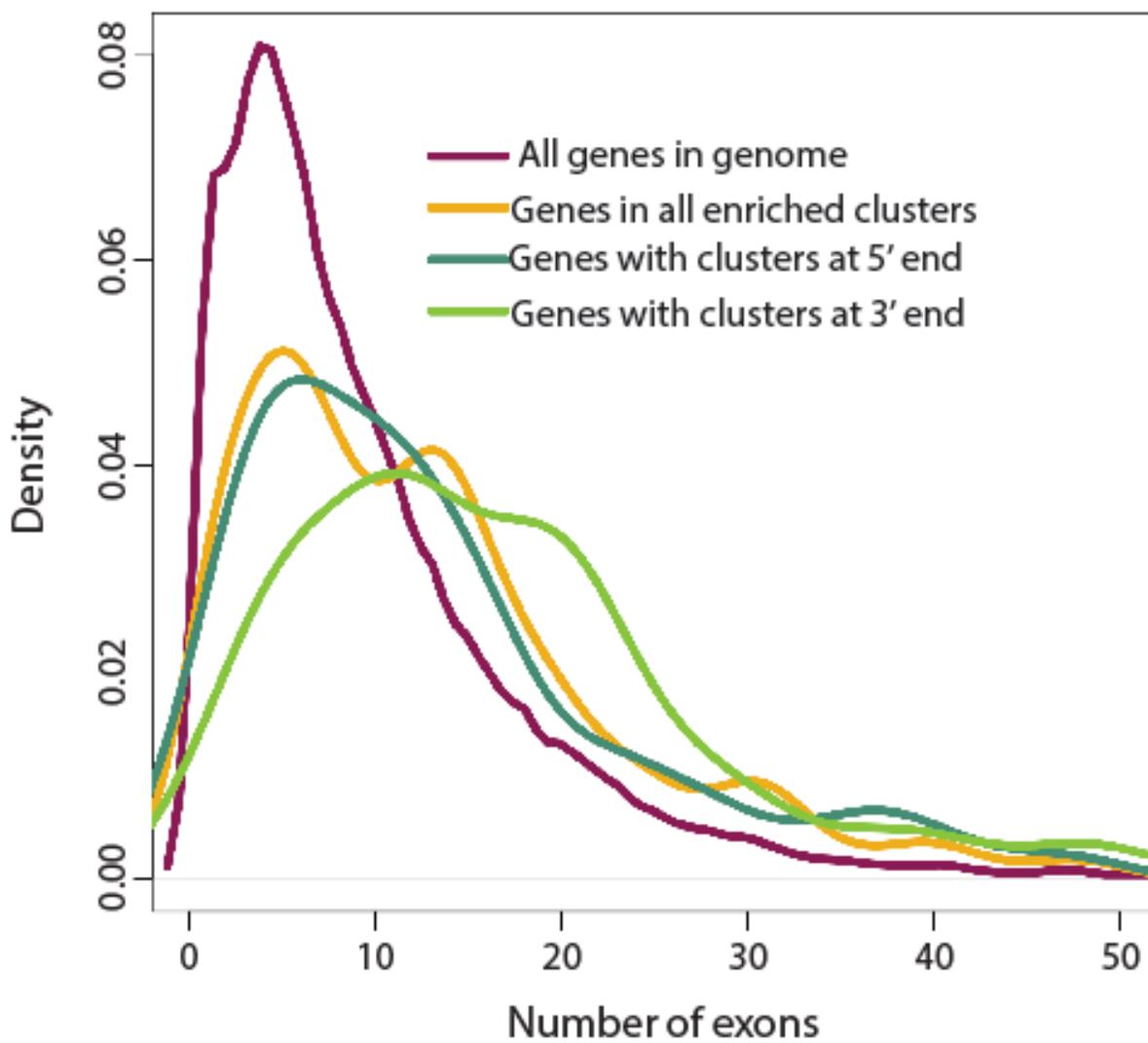


Figure 18. Average exon count in ORF57-bound transcripts. The number of exons in four different categories were quantified and graphed. The red line shows the average number of exons in the entire human genome. The “genes in all the enriched clusters” category (Figure 16B, bar graph) includes the annotated genes in all the enriched clusters identified by our analysis, here shown in yellow. The blue line (genes with clusters at 5' end) represents the annotated genes from clusters concentrated at the 5' end of transcripts (Figure 16B solid red arrow and Figure 16C). The light green line (genes with clusters at 3' end) represents the

annotated genes from clusters concentrated at the 3' end of transcripts (Figure 16B dashed red arrow).

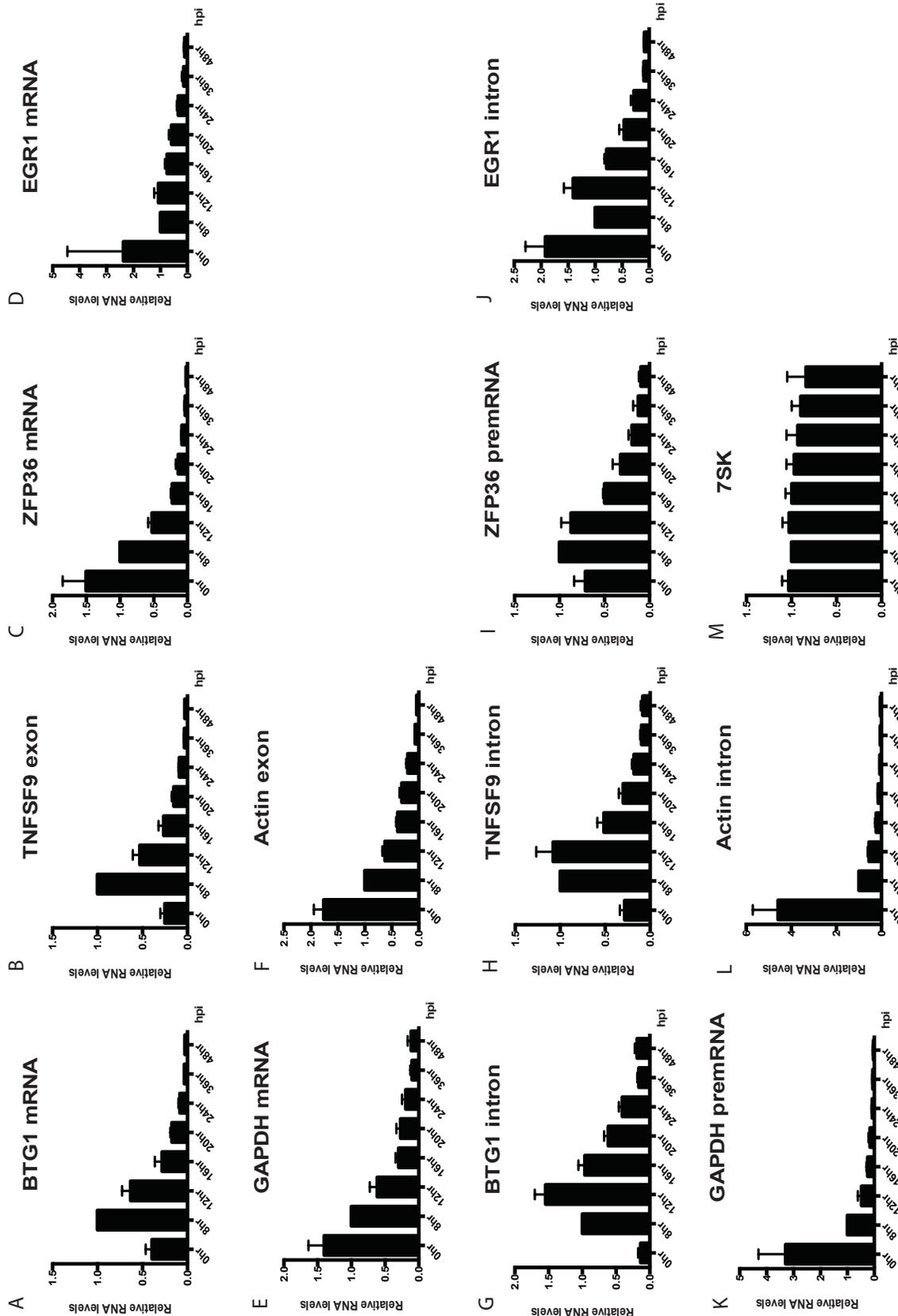


Figure 19. Steady-state RNA levels. RNA levels were monitored through qRT-PCR. 5ug of total RNA were used for each time point. RNA levels were first normalized to 7SK (M) and then to their corresponding 8 hpi time point. 7SK was normalized to the 8 hpi time point. Hours postinduction (hpi). Here, mRNA label next to transcript name denotes amplicons across exon-exon junctions. Figure 17 shows the location of the amplicons used here.

*pre-mRNAs bound by ORF57 peak early upon lytic reactivation*

We next wanted to see if some of the ORF57-bound transcripts identified here behaved differently when compared to unbound RNAs during lytic induction. We monitored steady-state levels of select transcripts at multiple time points following lytic induction (Figure 19), as well as ORF57 protein levels (Figure 20). As expected, host mRNAs rapidly decrease due to RNA decay by KSHV host shutoff (Figure 19 A-F) [17, 98]. IGV traces (Figure 17) showed significant number of CLIP tag clusters inside introns of these RNAs. Therefore, we decided to monitor the pre-mRNA abundance of some of these transcripts as well. Steady-state pre-mRNA levels of transcripts bound by ORF57 (Figure 19 G-L) followed a different pattern than GAPDH or  $\beta$ -Actin control pre-mRNAs (Figure 19 K and L). Control pre-mRNAs rapidly disappear, at 8 hpi pre-mRNA levels are reduced by ~75%. In contrast, the steady-state levels of pre-mRNAs bound by ORF57 peak around 12 hpi. In addition, these pre-mRNA peaks occur after the mRNA peaks observed at 8 hpi (Figure 19 A and B). We noticed the steady-state mRNA levels of BTG1 and TNFSF9 (Figure 19 A and B) are higher at 8 hpi than their corresponding uninduced controls (0 hpi). Previous publications have identified a small subset of host transcripts that escape the host shutoff and

are upregulated during infection, known as escapees [36, 165]. BTG1 and other TNF related genes have been previously identified as some of these escapees [36]. It is important to note that not all of the ORF57-bound transcripts are escapees, but it is certainly an observation that can be pursued in future experiments.

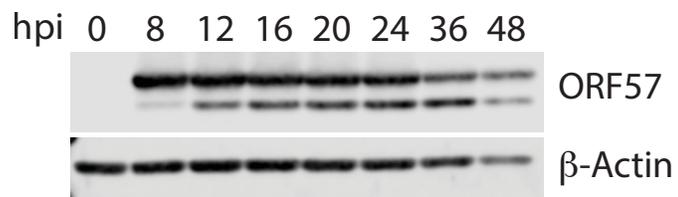


Figure 20. Representative western blot from samples in Figure 19. Western blot was probed using affinity purified Rabbit $\alpha$ ORF57 antibody and Mouse $\alpha$  $\beta$ -Actin (Abcam). Secondary antibodies: Goat $\alpha$ Rabbit 800CW and Goat $\alpha$ Mouse 680LT (Licore Odyssey).

To ensure transcripts made prior to induction were not influencing our results, we performed a 2 hr transcriptional pulse with 4-thiouridine (4SU) [166]. This uridine analogue is readily incorporated by all three RNA polymerases into nascent transcripts. We added 4SU into the cell culture media growing TREx BCBL1-Rta cells 2 hours prior to each collection time point and then extracted total RNA from uninduced cells (0 hpi), as well as from cells at 12 and 24 hpi. Only the RNAs made during this 2 hrs contain the uridine analogue. We can then attach a biotin label to the 4SU containing mRNAs and immunoprecipitate these transcripts using streptavidin-conjugated beads. Figure 21 shows the results from the 4SU experiments. Newly made BTG1 mRNA levels decrease from 0 to 12 hpi, which differs from steady-state observations (Figure 19 A). The overall increase in the steady-state RNA levels may be explained by a short burst of transcription shortly after lytic induction. Nonetheless, the downward trend of newly made BTG1 mRNA levels as lytic reactivation progresses is consistent with the decrease observed at steady-state. More importantly, newly made BTG1 pre-mRNA levels increase significantly from 0 to 12 hpi (Figure 21) and the GAPDH control shows a decrease on both the mRNA and the pre-mRNA populations. Interestingly, at 24 hpi, BTG1 pre-mRNA levels decrease significantly, as do all other RNA populations. Taken together, these set of experiments suggest a novel functional role for ORF57 where ORF57-binding to pre-mRNAs results in the inhibition of splicing.

*ORF57 is sufficient to upregulate steady-state pre-mRNA levels*

Because steady-state levels of these select transcripts were monitored in the context of viral infection, it is possible some of our observations might be due to the action of viral

proteins other than ORF57. To ensure these effects are ORF57-specific, we monitored the steady-state levels of some of these transcripts in HEK293 cells after a 24 hr transfection in the presence of ORF57 or an empty vector control (pcDNA3) (Figure 22). BTG1 and EGR1 pre-mRNA levels are significantly increased with ORF57 transfection. In contrast, GAPDH pre-mRNA decreases in the presence of ORF57. This indicates ORF57 alone is sufficient to upregulate pre-mRNA levels of at least two of the transcripts in this study.

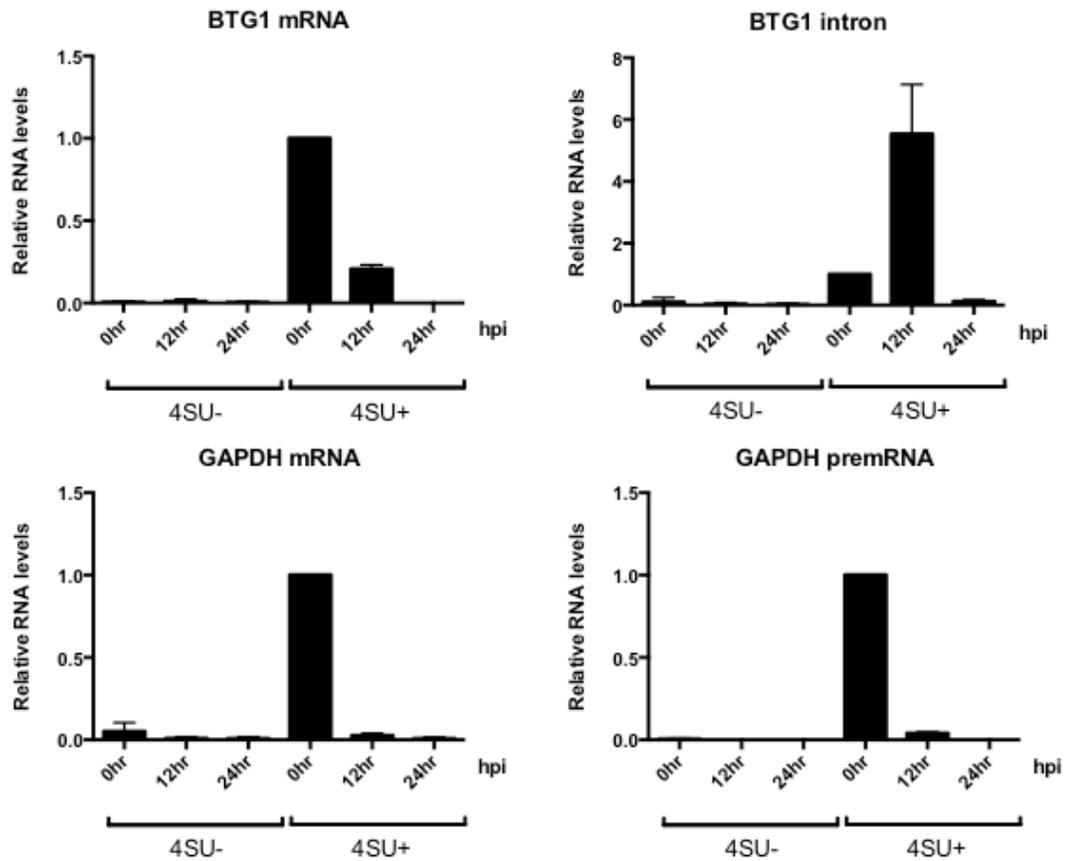


Figure 21. Newly made transcripts in TREx BCBL1-Rta cells. Cells were incubated with 4SU starting at -2, 10 and 22 hpi. Cells were collected and RNAs were extracted at 0, 12 and 24 hpi, respectively. RNA levels were monitored through qRT-PCR. RNAs from cells incubated with 4SU are on the right side of the graph (n=3). No 4SU controls are included on the left (n=2).

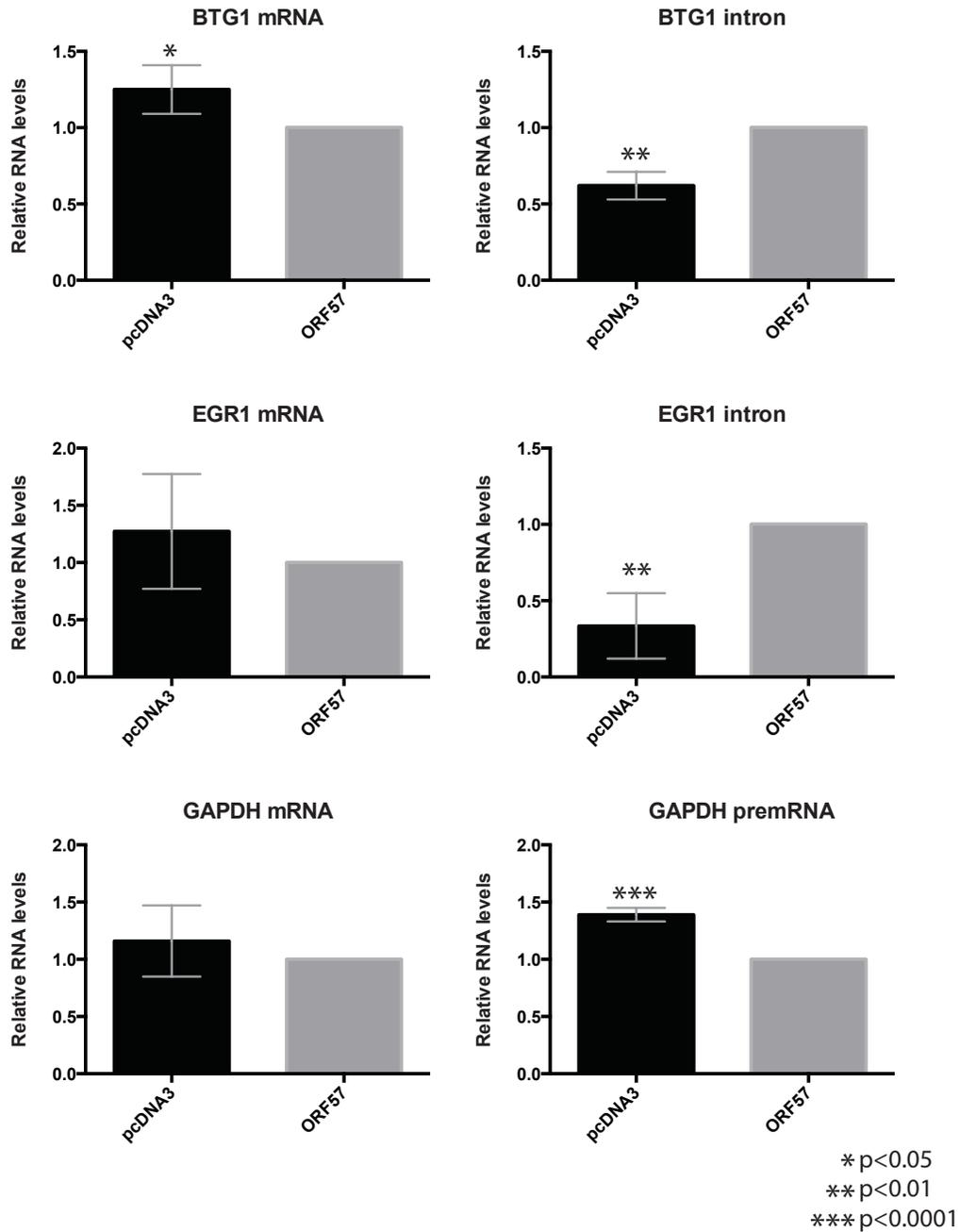


Figure 22. Steady-state RNA levels in HEK293 cells. HEK293 cells were transfected with ORF57 and total RNAs were collected after 24 hrs (n=3). RNA levels were normalized to ORF57 and 7SK.

## Discussion

In this study, we summarize how RNA binding sites can be used to generate insights into the function of RNA binding proteins such as ORF57. We performed HITS-CLIP on a latently infected cell line to identify novel ORF57 targets and their corresponding ORF57 binding sites. Recovered CLIP tags map to both human and viral genomes. Mapping of the enriched human CLIP tags reveals ORF57 binding is enriched at the 5' end of a subset of the bound transcripts (Figure 16A) and, to a lesser extent, the 3' end. More detailed mapping of the 5' end clusters, positions these sequences at the first exon-intron junction (Figure 16B). When we monitored the steady-state mRNA and pre-mRNA levels of these transcripts, the pre-mRNA levels persisted for longer time periods after lytic induction when compared to the their corresponding mRNAs and control pre-mRNAs (GAPDH and  $\beta$ -Actin). More importantly, analysis of newly made BTG1 pre-mRNAs shows an increase from 0 to 12 hpi. The increase in pre-mRNA levels observed (Figures 19, 21 and 22) does not translate to a proportional increase in the mRNA population; in fact they are inversely correlated. This has led us to propose that the ORF57-binding to the pre-mRNAs we describe here results in the inhibition of splicing.

A previous publication has shown that when ORF57 is expressed with a partially spliced K8 reporter gene in HELA and HEK293 cells, the production of the mature K8 mRNA and K-bZIP protein increase in a dose dependent manner [57]. Additionally, the same publication introduced intron containing  $\beta$ -globin fragments into the 3' end of a GFP reporter and observed an increase in the spliced form when ORF57 was present. On the other hand multiple publications studying HSV-1 ICP27, an ORF57 homolog, argue that this related

protein inhibits splicing [101-104]. Regardless of the role ORF57 plays in splicing, we are interested in dissecting the contributions of this protein during lytic infection.

On a separate note, the 4SU experiment shows an overall decrease in newly made RNA levels at 24 hpi. The simplest explanations for this observation are that either degradation rates increase or transcription rates decrease dramatically at 24 hpi. To monitor decay, we are currently working on 4SU pulse-chase experiments. To examine transcription rates, we will perform nuclear run-on (NRO) experiments. Pulse-chase and NRO experiments will help monitor the RNA stability and account for changes in transcription of these transcripts during different times following lytic induction. Previous publications using NRO dot blots to measure herpesvirus host shutoff, reported transcription levels of host transcripts (e.g.  $\beta$ -Actin and  $\beta$ -tubulin) remain unaffected [95, 101]. Therefore, any transcriptional effects observed in these proposed experiments will represent a novel finding.

We ran all of the enriched cluster sequences and the enriched clusters with mutations through MEME, a *de novo* motif discovery algorithm [132], but were unable to retrieve conserved elements for these sequence groups. Additionally, RNAfold [133] analysis of the sequences flanking HITS-CLIP induced mutations was performed. Preliminary observations of these 60 nt sequences suggests they can fold into one or two stem loops. Initial inspection of these fragments shows no obvious patterns as to whether certain nucleotides are exposed at loop or stem regions, but their analysis remains in progress.

The HITS-CLIP assay alone does not determine whether recovered CLIP tags represent the formation of functional/productive protein-RNA complexes, but the high local concentration of some of these clusters at the 5' end of transcripts does however raise an

important question regarding ORF57 recruitment. The positioning of enriched clusters at the 5' end suggests the interaction of ORF57 with host proteins normally present at the 5' end of the transcript, such as the one previously reported with cap-binding complex protein (CBP80), may affect the local concentration of our protein or perhaps be involved in the initial recruitment to its target RNAs [12, 81-87]. The ORF57-bound transcripts identified here will be useful tools to address some of these observations.

The inhibitory effects of ORF57 do not translate to all of its bound RNAs, as multiple publications have shown, this multifunctional RNA binding protein preferentially upregulates intronless transcripts and reporters [12, 15, 55, 58, 62, 68, 83, 90, 91]. Analysis of the viral enriched clusters is currently pending, but we expect observations made from the viral genome may shed light as to why ORF57 binding results in a different outcome. We are also interested in the characterization of enriched clusters close to the OriLyt-L, since previous work from our laboratory has shown that ORF57 associates with the DNA in this region [47]. Previous work from the Lieberman lab has shown that transcription from the OriLyt is necessary for lytic DNA replication in gammaherpesviruses, we are curious to determine the extent of ORF57's contributions to the RNA transcripts made from this location [43, 48, 51, 167]. The high density of HITS-CLIP induced mutations to the right of OriLyt-L (Figure 14D) suggests a strong association with RNAs in its vicinity.

Even though identification of enriched clusters does not depend on genomic annotations, our current analysis focuses on the host targets because readily available annotations facilitated the sorting and classification of the clusters identified. Additionally, multiple KSHV transcripts are bicistronic (such as ORF58-ORF59, Figure 14E) or

polycistronic, complicating the classification of some of these clusters [48, 51, 168]. In the viral genome we observe a significant number of CLIP tags and HITS-CLIP induced mutations mapping to the opposite strand of annotated ORFs and have yet to determine the relevance of these observations. Nonetheless, analysis of viral clusters is currently ongoing as closer inspection of these clusters is necessary. We have sorted through a lot of the HITS-CLIP data and still we have a lot more, but we are confident our data will help address some of the questions concerning ORF57 recruitment and binding requirements to its target RNAs.

## CHAPTER SIX

### Conclusions and Future Directions

#### Conclusions

Our work highlights the use of ORF57 RNA binding sites to gain insight into the mechanisms driving recruitment and functions of this essential viral protein. We started by studying the ORF57-PAN RNA interaction and were able to determine ORF57 binding to this RNA is direct and specific. We UV-crosslinked *in vivo* protein-RNA complexes in lytically reactivated cells and partially digested the RNAs to enrich for the RNA fragments covalently bound to the crosslinked proteins. Immunoprecipitation of covalently bound ORF57-RNA complexes helped us determine ORF57 preferentially binds the 5' end of PAN RNA (Figure 3). This observation was also confirmed *in vitro* through label transfer assays. Incubation of whole cell extracts with exogenous, body-labeled PAN RNA fragments pinpoint ORF57 binding to a 9 nt in a predicted loop (Figure 8). We cloned ORE-containing PAN RNA fragments of varying sizes at the 3' UTR of an intronless  $\beta$ -globin reporter and were able to identify a 30nt element that was sufficient to elicit an ORF57 response to similar levels as the full length ORE (~300nt, Figure 6). Deletion of this 30nt element from PAN RNA impairs ORF57 responsiveness (Figure 7). We defined this 30nt element as the core ORE because it is: 1) bound by ORF57 *in vitro*, 2) sufficient to observe ORF57-dependent upregulation when placed at the 3' UTR of intronless  $\beta$ -globin and 3) necessary for PAN RNA responsiveness to ORF57. From these observations, we concluded that ORF57 binding correlates with target transcript upregulation. We hypothesize the ORE functions as a high-

affinity binding site where ORF57 is initially recruited, which then enables recruitment of additional ORF57 molecules to the bound transcript. This is further supported by the fact that sequences adjacent to the ORE are sufficient to recruit ORF57 binding *in vitro* (Figure 4) and are also important for full ORE activity in the context of PAN RNA (Figure 7), but were not enough to elicit ORF57-mediated upregulation of our intronless reporter (Figure 5). In addition, the HITS-CLIP studies further corroborated this idea (see below).

In our most current work, ORF57 HITS-CLIP identified a large number of human and viral RNAs bound by ORF57. Through analyses of the recovered human CLIP tags we may have stumbled upon a novel ORF57 function, where ORF57 binding to pre-mRNAs results in splicing inhibition. We propose this new ORF57 role based on several experimental observations. CLIP tag mapping shows ORF57 binding is most enriched at the 5' end of multiple host transcripts. Further analysis of these 5' end clusters positions these sequences at the first exon-intron junction. When we monitored steady-state RNA levels of ORF57-bound RNAs at multiple time points during lytic induction, we noticed ORF57-bound pre-mRNAs behaved differently than GAPDH and  $\beta$ -Actin controls (Figure 19). GAPDH and  $\beta$ -Actin pre-mRNAs disappear rapidly. ORF57-bound mRNAs also disappear quickly, but their corresponding pre-mRNAs are lost considerably more slowly. We also 4SU labeled newly made transcripts for 2 hours prior to their collection at 0, 12, and 24 hpi. We measured newly made BTG1 (an ORF57-bound transcript) and GAPDH pre-mRNA levels (Figure 21). Newly made BTG1 pre-mRNA levels increase almost 6-fold from 0 to 12 hpi, while in contrast, GAPDH pre-mRNA is barely detected. Therefore, the analysis of newly made pre-mRNAs reinforces observations made during steady-state. Additionally, ORF57 transfections in HEK

293 cells demonstrate ORF57 is sufficient to upregulate some of these pre-mRNAs, even in the absence of other viral factors. Taken together, observations made from ORF57-bound RNAs point towards a novel role for ORF57, where it may be inhibiting splicing of host transcripts.

### **Future Directions**

Studies involving ICP27, an ORF57 homolog capable of splicing inhibition, have shown that ICP27 is sufficient to inhibit splicing *in vitro* and that it interacts with splicing regulatory factors to alter their localization [102, 103, 169, 170]. With the exception of one publication reporting that when ORF57 is expressed with a couple of intron-containing reporters the production of their corresponding mature mRNAs increases in a dose dependent manner, there is not much known about the role of this protein in splicing [57]. It would be interesting to compare ORF57 in similar experimental conditions as those used for ICP27, to determine if they inhibit host splicing in a similar manner.

To understand the effects ORF57 has on its bound RNAs, we will need to closely examine transcriptional and posttranscriptional changes on these transcripts during lytic replication. The overall newly made GAPDH RNA levels at 12 hpi are much lower than BTG1 RNA levels (Figure 21). Disparities in RNA levels may be explained by changes in the transcription or degradation rates of these transcripts. Previous publications have used nuclear run-on (NRO) dot blots to report that host shutoff does not affect transcription of host transcripts [95, 101]. However, if the effects we observed were exclusively due to increased decay rates, the decay pathway must be particularly rapid and efficient as nearly no newly

made transcripts are observed at 24 hpi. Therefore, we are interested in revisiting whether virus host shutoff also occurs at the transcriptional level. Additionally, we are planning on doing 4SU pulse-chase experiments to monitor the decay rates of some of these transcripts. At 24 hpi, the levels of all the 4SU labeled transcripts are close to the limit of detection (Figure 21). It is possible the uridine analogue is not being incorporated efficiently at the 24 hr time point. Anecdotally, changing the cell culture media a few hours prior to 4SU labeling stimulates transcription and therefore promotes incorporation of 4SU, which can be easily tested in future experiments.

Transcripts studied in Chapter five, have one or two introns and show CLIP tag enrichment at exon-intron junctions, but our analysis also identified different binding patterns in other RNAs bound by ORF57. Some of the transcripts bound are intronless, others show preferential binding at the 3' end of the transcript, others have CLIP tag distributions suggesting ORF57 is binding primarily to the mRNA, etc. Classification of some of these targets may help distinguish whether ORF57 binding patterns correlate with different ORF57 functions, such as effects in translation, transcript nucleocytoplasmic distribution, transcript stability, etc. If ORF57 binding to human transcripts results in different functional outcomes, can we find any commonalities among these groups?

Our analysis of the recovered CLIP tags resulted in identification of 2,448 enriched clusters mapping to both the human and viral genomes. Currently, we have evaluated the fate of a few host transcripts and have yet to analyze the viral clusters. The location of enriched human clusters at the 5' end of bound transcripts, together with the absence of discernible sequence elements in the enriched clusters, suggest interactions with host or viral factors at

the 5' end of the bound transcripts may play a more important role than previously anticipated in the recruitment of ORF57 [12, 81-87]. It would be interesting to perform the analogous analysis from Figure 16 on the enriched viral clusters. Unfortunately, we haven't had the chance to do the analysis because viral genome annotations are not readily available in a compatible file-format. This analysis might reveal whether there is preferential binding at certain locations in the viral transcripts. If these binding patterns are the same, then how is it that viral and human transcripts behave differently in the presence of ORF57? If they are different, does their location hint which proteins might be involved in ORF57 recruitment or functional outcome? Are there any sequence elements over represented in these regions? How is it that ORF57 binding results in upregulation of some of these viral transcripts? For instance, ORF59 is a known ORF57 target. ORF57 increases the nuclear and cytoplasmic ORF59 mRNA levels [58]. In contrast, PAN RNA is upregulated only in the nucleus, whereas other transcripts like HVS gB accumulate in the cytoplasm.

RNAs near the OriLyt represent ORF57-binding hot spots. Multiple publications have shown that several proteins are recruited to the OriLyt (e.g. K-bZIP, ORF50, ORF57, etc.) to promote viral DNA replication, therefore isolation of ORF57 contributions in this process will most likely be complicated. At the present, we know of the association of ORF57 with the OriLyt-L DNA, but have yet to determine what role does ORF57 play in lytic DNA replication. Curiously, there are no enriched clusters to the left of OriLyt-L, even though reads from the Input sample confirm these genes are being expressed. ChIP-chip experiments have also identified ORF57-DNA interactions at the ORF4 promoter [47, 171]. This promoter region is also bound by K-bZIP in KSHV infected cells [171]. The ORF4

promoter-ORF57 interaction might turn out to be useful in future studies dissecting ORF57's function in transcription.

### **Summary**

Our work studying ORF57-PAN RNA interactions has demonstrated that ORF57 binding correlates with transcript upregulation. This statement is consistent with observations made by multiple laboratories reporting the effects of ORF57 on its target RNAs and by ORF57's preferential upregulation of intronless transcripts. We used HITS-CLIP to identify global ORF57 RNA binding sites and may have identified a novel function for ORF57, inhibiting host pre-mRNA splicing. Viral CLIP tag analysis reveals clusters close to the OriLyt-L and OriLyt-R are among the most highly enriched and, similarly, contain the highest density of HITS-CLIP-induced point mutations, representing high confidence ORF57 binding events. Together, the observations presented here, highlight the involvement of ORF57 in key processes important for KSHV pathogenesis: it upregulates viral intronless transcripts, it may be involved in inhibiting splicing of host transcripts repressing cell proliferation, and it is concentrated in genomic locations involved in lytic DNA replication. All reasonable explanations as to why this protein is essential to this oncogenic virus.

## APPENDIX A HITS-CLIP Recipes

Notice that these recipes were obtained from the Darnell lab, our UV crosslinking protocol[123], or Invitrogen. Most solutions should be stored at room temperature unless otherwise noted.

<b>RIPA Correction</b>
1.25% NP40
0.625% Sodium deoxycholate
62.5 mM Tris pH 8.0
2.25 mM EDTA
187.5 mM NaCl
0.125 mg/ml Heparin
1mM DTT
*1mM PMSF
*1X PI

<b>SDS Lysis Buffer (Make Fresh)</b>
0.5% SDS
50mM Tris pH 6.8
1mM EDTA
0.125 mg/ml Heparin
1mM DTT
*1mM Phenylmethylsulfonyl fluoride (PMSF)
*1X Protease Inhibitor (PI)

Items marked (\*) should be added immediately before use

<b>RIPA Buffer</b>
1% NP40
0.5% Sodium deoxycholate
0.1% SDS
150 mM NaCl
50 mM Tris pH 8.0
2mM EDTA

<b>High Salt Buffer</b>
5X PBS
0.1% SDS
0.5% NP40

<b>IXPNK Buffer</b>
50 mM Tris pH 7.4
10 mM MgCl <sub>2</sub>
0.5% NP40

<b>IXPNK+EGTA Buffer</b>
50 mM Tris pH 7.4
0.5% NP40
20mM EGTA

<b>20X NuPAGE MOPS SDS Running Buffer @ 4°C</b>
1 M MOPS
1 M Tris Base
2% SDS
20 mM EDTA

<b>20X NuPAGE Transfer Buffer @ 4°C</b>
500 mM Bicine
500 mM Bis-Tris (free base)
20 mM EDTA
10% Methanol

<b>1X Proteinase K (PK) Buffer</b>
100 mM Tris pH 7.5
50 mM NaCl
10 mM EDTA

<b>Urea Dye</b>
8M Urea
2.5mM EDTA
0.15% Bromophenol blue
0.15% xylene cyanol
1X TBE

<b>1XPNK/7M Urea Buffer (Make fresh)</b>
100 mM Tris pH 7.5
50 mM NaCl
10 mM EDTA
7M Urea

<b>G50 Buffer</b>
0.25% SDS
20mM Tris pH7.5
0.3M Sodium Acetate pH5.2
2mM EDTA

## Phenol (SATURATED PHENOL)

Make the same way as you would make PCA (except without adding chloroform and isoamylalcohol)

## PCA

Mix 100ml phenol, 100ml chloroform, and 0.100g hydroquinoline.

Add 200ml 1M Tris pH 8.0. Mix

Let the aqueous and organic layer separate (>1hr)

Check pH or aqueous layer. It should be ~ 8.0. If not repeat by adding fresh 1M Tris pH 8.0

Aspirate/pipet out the aqueous layer.

Add 100ml 0.1M Tris pH 7.5. Mix

Let the aqueous and organic layer separate (>1hr)

Check pH or aqueous layer. It should be ~ 7.5. If not repeat by adding fresh 0.1M Tris pH 7.5

Add 2mls of isoamylalcohol. Mix and let the aqueous and organic layer separate overnight.

Store at 4 °C.

## How to make <sup>32</sup>P labeled DNA ladder

50bp DNA ladder (Invitrogen #10416-014); pBR322 DNA/MspI Digest ladder (NEB N3032S)

### 1. CIP treatment

20ul (20ug)	pBR322/MspI or 50bp DNA ladder
5ul	10X NEBuffer
3ul	Alkaline phosphatase (CIP) M0290S
22ul	H <sub>2</sub> O
50ul	

- Incubate at 37 °C for ~30min.
- PCR purify using a QIAGEN PCR purification kit. Elute in 40ul of EB buffer to yield a final concentration of 0.5ug/ul

### 2. PNK treatment

1ul	CIP treated DNA ladder (0.5ug/ul)
1ul	10X PNK Buffer
1ul	$\gamma^{32}\text{P}$ ATP (PE NEG035C001MC)
6ul	H <sub>2</sub> O
1ul	T4 PNK (NEB M0201S)
10ul	

- Incubate at 37 °C for ~30min.
- Bring volume up to 50ul and purify reaction with G-25 spin columns (Fisher #45-001-397).

## **APPENDIX B**

### **Preparation of cells for HITS-CLIP**

#### ***Cell culture:***

TREx BCBL1-Rta cells [118] were carried in RPMI-1640 media (Sigma) supplemented with 10% tetracycline-free FBS (Clontech), penicillin-streptomycin (Sigma), 2 mM L-glutamate, and 100 µg/ml hygromycin (Sigma). BCBL cells were induced with 1µl/ml DOX and 3mM sodium butyrate and collected after 20hrs. At the time of induction, cells were at a concentration of  $5 \times 10^5$  cells/ml. For high-throughput sequencing library preparation, each biological replicate utilized  $1.2 \times 10^8$  cells. A total of  $2 \times 10^7$  cells were used to isolate the “INPUT” RNAs (see Input RNA Samples section below). The remaining  $1.0 \times 10^8$  cells were washed, aliquoted, and frozen as highlighted below.

#### ***Collect and wash cells after the 20hr induction:***

Keep cells on ice at all times. Aliquot  $2 \times 10^7$  cells into each 50ml conical tube, spin at 700xg for 3min at 4°C. Decant supernatant. Add 10 ml of cold PBS (without Calcium & Magnesium, -/-). Spin at 700xg for 3min at 4°C. Decant supernatant. Add 3mls of PBS (-/-) and transfer to 10cm dish and place it on a Pyrex tray with ice (see left picture below). Make sure 10cm dish is leveled. Place two loading pipette tip boxes inside the Spectrolinker. Remove the lid from your 10cm plate, place Pyrex tray on top of the tip boxes in the Spectrolinker (see right picture below). Irradiate at  $125 \text{mJ/cm}^2$  with 254-nm wavelength bulbs.



After crosslinking, add 7ml of PBS (-/-) to each plate and transfer to 15 ml conical tubes (\*\*keep on ice\*\*). Spin at 700xg for 3min at 4°C. Decant supernatant. Add 1ml PBS (-/-) and transfer to 1.5ml eppendorf tubes. Spin at 2400xg for 1 min at 4°C. Pipet out the supernatant. Freeze cell pellets on dry ice (~5-15min) and then store them at -80°C indefinitely.

### ***Input RNA Samples***

RNAs for the input samples were recovered with TRI reagent (Molecular Research Center) immediately after a 20hr induction. Recovered RNAs were ethanol precipitated and kept at -20°C until ready to be used.

## **APPENDIX C**

### **HITS-CLIP Protocol for Illumina Sequencing**

For Illumina library preparation we combined the RNA from 5 cell pellets. Parallel to the previously mentioned pellets, we processed one high MNase control to aid in identification of the ORF57 protein-RNA complexes. If you are troubleshooting this protocol, I would suggest you start only with the high MNase concentration condition.

#### **Cell Lysis**

1. Thaw cell pellets on ice. (Do not leave them on ice for too long)
2. Vortex/flick to break the cell pellet. Vortex the pellet again immediately before the next step.
3. Add 140 ul of SDS Lysis Buffer per sample and quickly pipet up and down to lyse your cell pellet. You will only be able to pipet once or twice before the DNA makes your solution viscous.
4. Heat samples at 65°C for 5 min. Meanwhile, make your next buffer.
5. Immediately place your samples on ice (at least ~2min)
6. Add 560ul of RIPA Correction Buffer per sample. You will not be able to pipet the lysate very well at this point.
7. With a P1000 tip “pinch” some of your viscous pellet and decant your sample into a QIAGEN QIASHredder column.
8. Spin the column at 16,000xg for 1 min at 4°C.

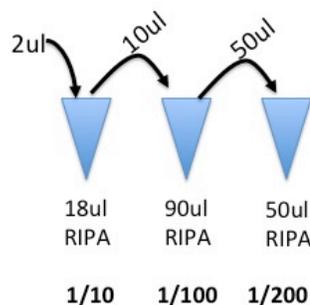
9. Resuspend the insoluble pellet, mix the lysate and run through the same column again. Spin the column at 16,000xg for 1 min at 4°C.
10. Discard the column and save the flowthrough.
11. Resuspend the insoluble pellet and transfer lysate into a clean 1.5ml tube. Avoid bubbles

### DNase and MNase Treatment

12. Add 3.5ul 1M CaCl<sub>2</sub> (Final Concentration of 5mM)
13. Add 30ul of RQ1 DNase (Promega). Pipet up and down slowly to mix.
14. Incubate at 25°C water bath for 15min.

15. Meanwhile make MNase dilutions

- a. NOTE: make dilutions immediately before use
- b. MNase (NEB M0247S)



16. Make a serial dilution to get 1/200 dilution of MNase stock in RIPA buffer at RT (see below)
  - a. NOTE: If you are trying to include a high MNase concentration to get “complete” digestion of the RNA just add 2-5ul of the MNase stock.
17. Add 5ul of the 1/200 MNase dilution to each sample. Pipet up and down slowly to mix the lysate.
18. Incubate sample in the 25°C water bath for EXACTLY 10min.

- a. NOTE: ORF57 will crash out of solution at higher temperatures (with RNase treatment)
19. Add 47ul 300mM EGTA (Final concentration of 20mM). Invert tubes ~10X to mix lysate.
20. Centrifuge at maximum speed (21130xg) for 10min at 4°C. Transfer the supernatant to a clean 1.5ml eppendorf tube.
21. Repeat step 20 two more times for a total of 3 times.
22. Meanwhile, aliquot Magnetic protein A Dynabeads (Invitrogen #1000020, 30mg/ml) and wash them 3 times in 1ml of RIPA buffer. For each sample you will need **two** 200ul slurry aliquots. The first to preclear the lysate, and the second to perform the actual IP. Leave beads in the last wash (do not let them dry).
23. When step 21 is done, if needed, collect an INPUT sample at this point and freeze it in dry ice.

### **Immunoprecipitation**

24. Remove the RIPA Buffer from the pre-washed Protein A Dynabeads and transfer the lysate onto the beads.
25. Nutate for 30min-1hr at 4°C.
26. At this time take out:
  - a. Novex Loading Buffer out from 4°C and let it warm up to room temperature.
  - b. Make 800ml 1X MOPS SDS Running Buffer and store at 4°C
  - c. Make 1L of 1X Transfer Buffer (with 10% methanol) and store at 4°C

- d. Do not add the Antioxidant yet
27. Place your tube in the magnetic stand and transfer the lysate to a clean 1.5ml eppendorf tube.
  28. Add 29ul (~48ug) of affinity purified ORF57 antibody (ES607, 1.66mg/ml). Nutate for 1hr at 4°C.
    - a. This step is highly optimized. We know that the IP of ORF57 works best with this lysate volume (vs a 200ul lysate) and it works best if you incubate the lysate with the antibody first (vs prebinding the antibody to the protein A beads).
  29. Remove the RIPA Buffer from the second set of pre-washed Protein A Dynabeads and transfer the lysate onto the beads.
  30. Nutate for 1hr at 4°C.
  31. Wash beads with **ice-cold** buffers: (~1ml/wash)
    - a. Twice with RIPA Buffer
    - b. Twice with High Salt Buffer
    - c. Twice with 1X PNK Buffer

Make sure to rinse the lid of the tube

**PNK Labeling of Pellets**

32. Treat each sample with the following mix:

1X Cocktail	
8ul	10X T4 PNK Buffer
1ul	$\delta^{32}\text{P}$ ATP (Perkin Elmer NEG035C001MC, ~0.15mCi/ul)
4ul	T4 PNK Enzyme (NEB M0201S)
1ul	RNASIN (Promega N261B)
66ul	H <sub>2</sub> O
80ul per condition	

33. Incubate at 37°C for 20min (mix by flicking the tube every 5min)

34. Add 10ul of 1mM rATP (mix with pipet).

35. Incubate at 37°C for ~30min. (mix by flicking the tube every 5min)

36. Pulse Spin ( $\leq 700\times g$ ) and/or place tube in magnetic stand. Remove supernatant

37. Wash three times at RT in 1XPnk buffer (~1ml each time).

a. Leave last wash on beads until ready to add the reagents on the next step.

38. Resuspend each pellet in the following:

1X Cocktail	
20ul	1X PNK Buffer
20ul	Novel LB @ RT (Invitrogen NP0007)
8ul	Reducing Agent (Invitrogen NP0009)
48ul per condition	

39. Setup the Novex NuPAGE Gel Apparatus.

a. Use NuPAGE gradient gels (Invitrogen NP0335Box). These work better than the 10% to visualize the ORF57-RNA complexes)

- b. Aliquot 200ml of 1X MOPS SDS Running buffer, add 500ul of Antioxidant (Invitrogen NP0005), and mix. Add the 200ml into the inner chamber of the NUPAGE apparatus.
- c. Clean the wells with a syringe.
40. Boil your samples at 100°C for 5min (not longer)
41. Load samples into the wells.
42. Pour the rest of the MOPS SDS Running Buffer on the outer chamber.
43. Start by running gel at 100V constant, you can then increase to 150-200V. Until the ladder 20kDa band is almost out of the gel (Use Bio-Rad's Precision Plus Protein All Blue ladder #161-0373)
44. Add 1ml of Antioxidant to 1L of 1X Transfer buffer (with 10% Methanol)
45. Once gel is done running, transfer proteins into a 100% nitrocellulose membrane (GE protran BA85 #10401196) using a wet protein transfer apparatus. The size of the filters and membrane is 8cmx6cm. Setup the transfer as follows:

Black →	Sponge
	Whatman filter paper
	Gel (flipped horizontally)
	100% Nitrocellulose Membrane
	Whatman filter paper
White/Red →	Sponge

46. Run at 100V (constant) for 2 hours in the cold room. Change the ice pack after 1 hour.

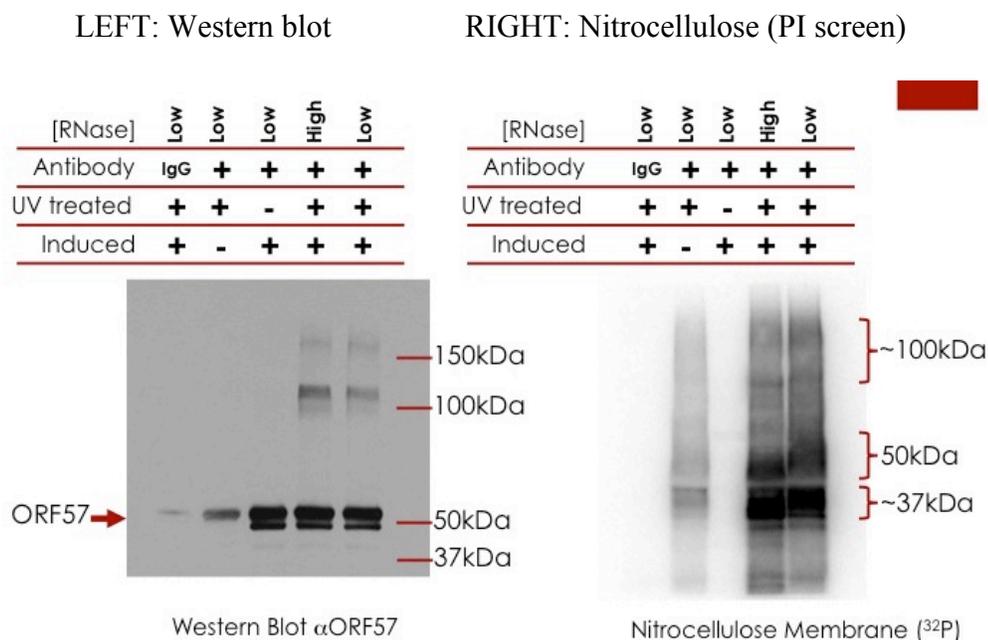
## Visualization of Protein-RNA Complexes

47. Once the transfer is done:

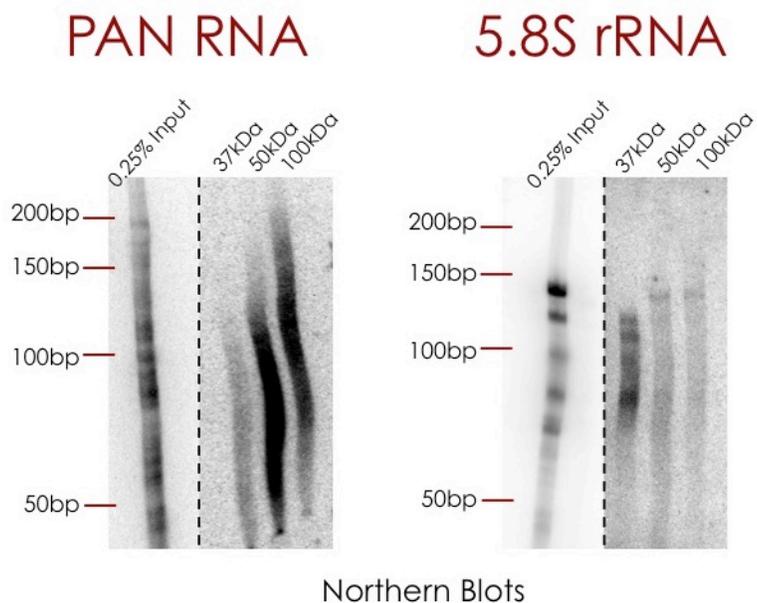
- Rinse the membrane with PBS (-/-) ~3 times
- Blot it dry with a clean kimwipe
- Wrap the membrane in cling wrap

48. To visualize ORF57-RNA complexes you can expose to film ~15-20min at RT (no need for enhancer screen) or you can expose for 2-5min in a PI screen. You will need the film exposure to help you cut the ORF57 specific bands.

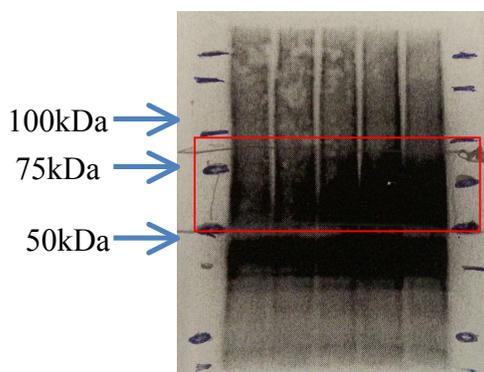
49. Here is what the PI screen exposure will look like:



NOTE that we know that the 37kDa band (see above) contains mainly rRNA, so we decided not to recover RNAs from that segment in the membrane. See below:



50. This is what the film will look like :



We defined the “**50kDa band**” as the band above 50kDa and below 100kDa

NOTE: For each biological replicate in our Illumina Sequencing experiment we isolated RNA from a total of 5 pellets treated with low MNase. See picture on the left.

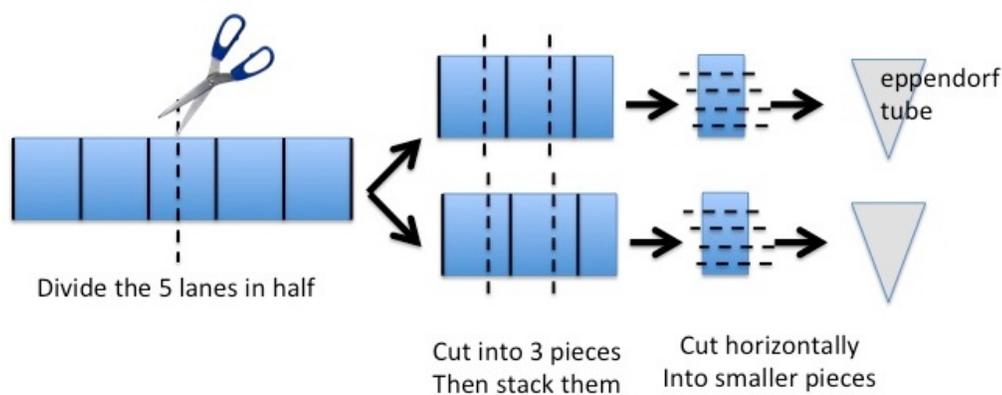
We purified RNAs from the “50kDa band” only.

51. To cut the desired PROTEIN-RNA complexes, place the membrane on a piece of plexiglass and set it in place with some tape. Place the film on top of your membrane and use that as a guide to cut your bands. (see step 53)

52. Make a 4mg/ml stock of Proteinase K (in 1X PK buffer) and incubate it at 37°C for ~20min. After the incubation is done, keep it on ice.

a. Make this fresh every time

53. Divide the pieces of nitrocellulose into two 2ml tubes (Place the equivalent of 2.5 lanes in each 1.5ml eppendorf tube)



### ProK Treatment

54. Add 350ul Proteinase K (4mg/ml) to each tube. The ProK solution should cover your pieces of membrane. VORTEX profusely.

55. Incubate at 37°C for 30 min. Vortex every 10 min.

56. Add 350ul of 1XPNK/7M Urea per tube.

100ul	1M Tris pH 7.5
12.5ul	4M NaCl
20ul	500mM EDTA
0.7ml	10M Urea
167.5ul	H <sub>2</sub> O
1ml	

57. Incubate at 37°C for 30 min. Vortex every 10min

58. Add 750ul (saturated) Phenol. Vortex
  59. Incubate at 37°C for 20min. Vortex every 10min.
  60. Vortex. Spin at maximum speed for 5min.
  61. Transfer aqueous supernatant to a clean 2ml tube.
  62. Add 750ul PCA. Vortex.
  63. Spin at maximum speed for 5min.
  64. Transfer aqueous supernatant to a clean 2ml tube.
  65. Add 750ul of chloroform. Vortex.
  66. Spin at maximum speed for 5min.
  67. Transfer aqueous supernatant to a clean 2ml tube.
  68. Combine both aqueous layers and measure the total volume.
  69. Add the following based on your measured volume:
    - a. Add 100% ETOH : 2-2.5X the volume of your aqueous layer
    - b. Add 3M Sodium acetate (pH5.2): 10% of your aqueous layer
    - c. Add 1ul of glycoblue per tube (Ambion AM9515)
- Note: You might need to use a 2ml tube or split it into 2.
70. Leave on dry ice overnight.
  71. Spin at 16,000xg for 10min at RT. Discard the supernatant.
  72. Wash pellet in 500ul of 70% ETOH (make fresh)
  73. Spin at 16,000xg for 5min at RT. Discard the supernatant.
  74. Resuspend the RNA pellets in 79ul H<sub>2</sub>O. Keep on ice ~2min.

**DNase Treatment of Pellets**

75. DNase the recovered RNAs:

For each pellet:

1X	
79ul	RNA/H <sub>2</sub> O
10ul	10X DNase Buffer
1ul	RNasin
10ul	RQ1 DNase
100ul per tube	

76. Incubate at 37°C for 1-2hrs.

77. Add 100ul H<sub>2</sub>O

20ul 10M Ammonium Acetate

78. Add 200ul phenol. Vortex.

79. Spin at maximum speed for 5min.

80. Transfer aqueous supernatant to clean tube.

81. Add 200ul PCA. Vortex.

82. Spin at maximum speed for 5min.

83. Transfer aqueous supernatant to clean tube.

84. Add 200ul of chloroform. Vortex.

85. Spin at maximum speed for 5min.

86. Transfer aqueous supernatant to clean tube.

87. Add the following to your aqueous layer:

1ul glycoblue

500ul 100% Ethanol

Vortex. Leave on dry ice overnight.

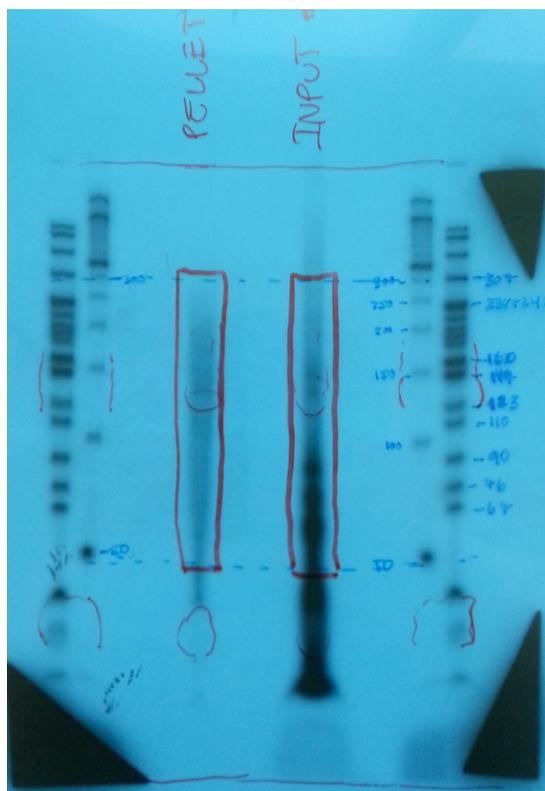
**RNA Size Selection of Pellets**

88. Spin at 16,000xg for 10min at RT. Discard the supernatant.
89. Wash pellet in 500ul of 70% ETOH (make fresh)
90. Spin at 16,000xg for 5min at RT. Discard the supernatant.
91. Prewarm the 6% urea gel by running it ~15min at 25Watts (constant) before loading your samples.
92. Resuspend the RNA pellet in 12ul urea dye. If you have several tubes combine them at this step (the final volume should be 12ul). Heat them at 90°C for 5min.
93. You will need to make a radiolabeled DNA ladder in advance.
  - a. To see the bands on a film after a ~30min exposure at RT you will use ~5ul of ladder+10ul urea dye (heat this at 100°C ~5min).
    - i. 50bp DNA ladder (Invitrogen #10416-014)
    - ii. pBR322 DNA/MspI Digest ladder (NEB N3032S)
94. Load samples in the 6% Urea gel. Make sure to clean the wells immediately before use.
95. Run the Urea gel and stop it 2-3cm before the bromophenol blue dye runs out of the gel.
96. Once the gel is done running:
  - a. Remove the short glass plate
  - b. Wrap the gel (now in the long glass plate) with cling wrap.
  - c. Put some RADTAPE (Diversified Biotech #RAD-10) on the corners
  - d. Place inside the Plexiglass box

e. Expose to film ~30min at RT

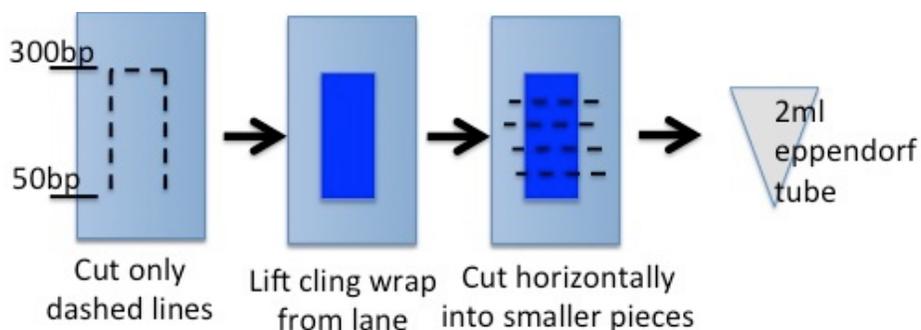
NOTE: Use a clean/dry glass plate to put pressure on your film.

97. See representative picture below. The xylene cyanol and bromophenol blue are shown in circles. The area that I cut to extract the RNA is highlighted in RED.



98. To maximize our yield on the Illumina TruSeq stranded mRNA Sample Prep kit (#RS-122-2101) we expanded the original range of 50-150nt to **50-300nt**

99. To cut the bands accurately, place the film below your glass plate and use a clean razor blade to cut the gel pieces. See diagram below.



100. Place the Urea gel pieces into ~800ul of G-50 buffer. Leave overnight at RT
101. Collect the G-50 supernatant (you will get ~600ul)
102. To the eluted RNAs in G-50:
  - a. Do a 1:1 phenol extraction (Aqueous layer volume:Phenol volume)
    - i. Vortex.
    - ii. Spin at maximum speed 5min at RT
    - iii. Transfer aqueous layer to clean tube
  - b. Do a 1:1 PCA extraction
    - i. Vortex.
    - ii. Spin at maximum speed 5min at RT
    - iii. Transfer aqueous layer to clean tube
  - c. Do a 1:1 chloroform extraction
    - i. Vortex.
    - ii. Spin at maximum speed 5min at RT
    - iii. Transfer aqueous layer to clean tube
103. To the remaining gel pieces:
  - a. Add additional ~700ul G-50.
  - b. Incubate at 37°C for 4 hours.
  - c. Heat at 65°C for 15min.
  - d. Collect the eluted RNAs and repeat step 102
104. Measure the aqueous layer volume.
105. Add the following based on your measured volume:

- a. Add 100% ETOH : 2-2.5X the volume of your aqueous layer
- b. Add 0.5ul of glycoblu per tube

Note: You might need to use 2ml tubes

106. Continue with the RNAClean XP purification steps

### DNase Treatment of Inputs

It is important to remove all traces of DNA from the RNAs so that we get the most efficient recovery of our RNAs using the RNAClean XP beads, and also to ensure strand specific amplification of our RNA templates using the Illumina TruSeq kit.

1. Spin at 16,000xg for 10min at RT. Discard the supernatant.
2. Wash pellet in 500ul of 70% ETOH (make fresh)
3. Spin at 16,000xg for 5min at RT. Discard the supernatant.
4. Resuspend the RNA pellet in 79ul H<sub>2</sub>O. Keep on ice ~2min.
5. DNase the recovered RNAs:

For each pellet:

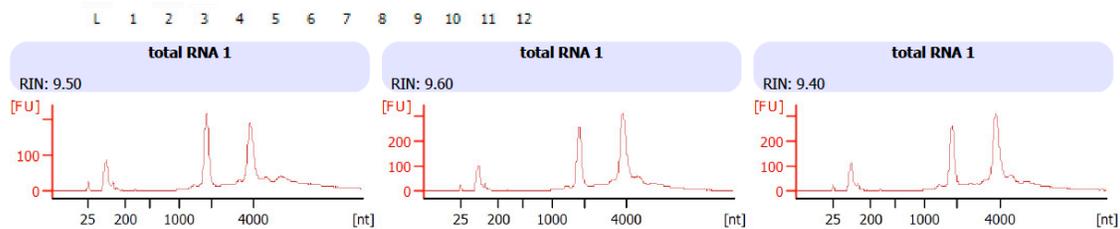
1X	
79ul	RNA/H <sub>2</sub> O
10ul	10X DNase Buffer
1ul	RNasin
10ul	RQ1 DNase
100ul per tube	

6. Incubate at 37C for 1-2hrs.

7. Add 100ul H<sub>2</sub>O  
20ul 10M Ammonium Acetate
8. Add 200ul phenol. Vortex.
9. Spin at maximum speed for 5min.
10. Transfer aqueous supernatant to clean tube.
11. Add 200ul PCA. Vortex.
12. Spin at maximum speed for 5min.
13. Transfer aqueous supernatant to clean tube.
14. Add 200ul of chloroform. Vortex.
15. Spin at maximum speed for 5min.
16. Transfer aqueous supernatant to clean tube.
17. Add the following to your aqueous layer:  
1ul glycoblue  
500ul 100% Ethanol  
Vortex. Leave on dry ice overnight.
18. Spin at 16,000xg for 10min at RT. Discard the supernatant.
19. Wash pellet in 500ul of 70% ETOH (make fresh)
20. Spin at 16,000xg for 5min at RT. Discard the supernatant.
21. Resuspend the RNA pellet in ~200ul H<sub>2</sub>O.
22. Read the RNA concentration using the nanodrop (~should be ~1mg/ml)
23. BEFORE the rRNA depletion, save ~1-4 ul of DNase TOTAL RNA and ask for a  
RNA Integrity Number (RIN) analysis from the sequencing core.

- a. This will tell you about the quality of your RNA before you started with the rRNA depletion.
- b. A “perfect” RNA prep, should have a RIN score of 10.0.

Example of RIN Analysis (Total RNA from BCBLTrex/RTA cells after 20hr induction)



### Ribo-Zero Magnetic Kit Protocol for Inputs

(rRNA depletion kit, # MRZH116) One round is enough to remove rRNA from 5ug of TOTAL RNA.

1. From previously DNased RNA, aliquot 5ug (the total volume of the RNA should be less than 26ul of H<sub>2</sub>O) and continue with the rRNA depletion.
2. Leave magnetic beads equilibrate at RT. Thaw the -80°C kit components at RT and then place them on ice.
3. Aliquot 225ul of magnetic beads into a 1.5ml tube.
4. Place on magnetic stand ~1-2min. Discard supernatant. Remove tubes from magnetic stand.
5. Add 225ul of H<sub>2</sub>O to each tube. Mix by pipetting or vortex at medium speed (speed 3).

6. Place on magnetic stand ~1-2min. Discard supernatant. Remove tubes from magnetic stand.
7. Repeat wash one more time (for a total of 2 washes). (Repeat steps 5-6).
8. After removing the water, remove the tube from the magnetic stand.
9. Add Magnetic Bead resuspension solution: 65ul per tube.
10. Mix by pipetting or vortexing at medium speed.
11. Add 1ul of RiboGuard RNase Inhibitor per tube. Vortex briefly.
12. Store tubes at Room Temperature (RT).
13. For 5ug RNA the total volume should be 26 ul H<sub>2</sub>O. Assemble as follows:

For Example:

5ul	Total RNA (DNased)
21ul	H <sub>2</sub> O (RNA+H <sub>2</sub> O volume = 26ul)
4ul	Ribo-Zero Reaction Buffer
10ul	RiboZero rRNA removal solution
40ul	

14. Pipet to mix the reaction.
15. Heat at 68°C for 10min
16. Remove tube from 68°C and place it at RT for 5min.
17. Continue with the following steps as quickly as possible:
  - a. Read steps 18-19 first and then continue
18. Transfer the 40ul reaction to the beads washed on step 12 and with the same tip pipet up and down ~10X quickly.
  - a. Never add beads to the RNA, it will cause clumps and inefficient recovery of your RNA.

19. Immediately afterwards, vortex ~10 sec at medium speed and place the tube at 25°C.

This will prevent clumps from forming.

20. Incubate at 25°C for 5min.

21. Vortex at medium speed ~10sec

22. Place at 50°C for 5min

23. Place on magnetic stand ~1min

24. Remove supernatant (~90ul) containing RNA and transfer to a clean 1.5m tube.

25. Place tube on magnetic stand, let it sit ~1min, and transfer the supernatant with your RNA to a clean tube.

- a. You want to make sure that there are no remaining magnetic beads on your supernatant, because that will result in rRNA contamination.

26. Place RNA on ice.

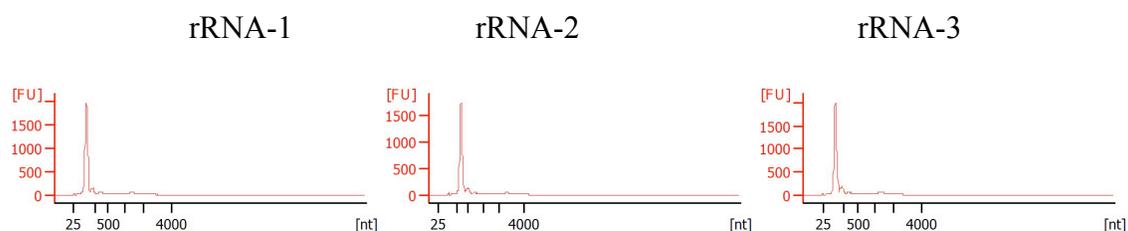
27. Add the following to each tube:

~90ul	RNA/Supernatant
90ul	H <sub>2</sub> O
18ul	3M Sodium Acetate
2ul	Glycoblue
600ul	100% ETOH

28. Vortex. Place on dry ice overnight.

29. Next day, resuspend pellet in 20ul H<sub>2</sub>O and save 1-4ul to run on the Bioanalyzer

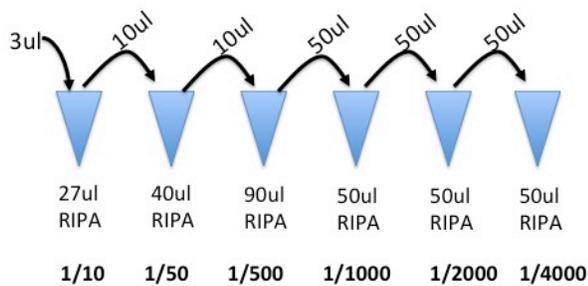
Example of Bioanalyzer results after ONE round of rRNA depletion.



### MNase Treatment of Inputs

From previously rRNA depleted RNA samples in ~20ul H<sub>2</sub>O

1. Heat RNA at 65°C for ~5min
2. Leave on ice ~2min
3. Add 330ul RIPA Buffer
4. Add 1.75ul CaCl<sub>2</sub>
5. Add 5ul of 1/4,000 MNase



a. NOTE: Make dilutions immediately before use

6. Incubate in 25°C water bath for exactly 10min
7. Add 23.3ul of 300mM EGTA
8. Add 400ul Phenol. Vortex. Spin at maximum speed for 5min.
9. Transfer aq. layer into a clean tube
10. Add 400ul PCA. Vortex. Spin at maximum speed for 5min.
11. Transfer aq. layer into a clean tube
12. Add 400ul Chloroform. Vortex. Spin at maximum speed for 5min.
13. Transfer aq. layer into a clean tube

14. Add 875ul 100% ETOH
15. Add 35ul 3M Sodium acetate pH 5.2
16. Add 1.0 ul glycobblue
17. Leave on dry ice overnight
18. Spin at 16,000xg for 10min at RT. Discard supernatant
19. Wash in ~150ul 70% ETOH. Spin 5 min at 16,000xg. Discard supernatant.

### PNK labeling of Inputs

20. Resuspend RNA pellet in 6.5ul H<sub>2</sub>O. Leave on ice ~2min

6.5ul	RNA/H <sub>2</sub> O
0.5ul	RNasin
1ul	10X PNK Buffer
1ul	$\gamma^{32}\text{P}$ ATP
1ul	T4 PNK enzyme
10ul	

21. Incubate at 37°C for 20min.
22. Add 1ul 10mM rATP
23. Incubate at 37°C for 30min
24. Add 300ul H<sub>2</sub>O
25. Add 400ul Phenol. Vortex. Spin at maximum speed for 5min.
26. Transfer aq. layer into a clean tube
27. Add 400ul PCA. Vortex. Spin at maximum speed for 5min.
28. Transfer aq. layer into a clean tube
29. Add 400ul Chloroform. Vortex. Spin at maximum speed for 5min.
30. Transfer aq. layer into a clean tube

31. Add 750ul 100% ETOH
32. Add 30ul 10M Ammonium acetate
33. Add 1.0ul glycoblue
34. Leave on dry ice overnight

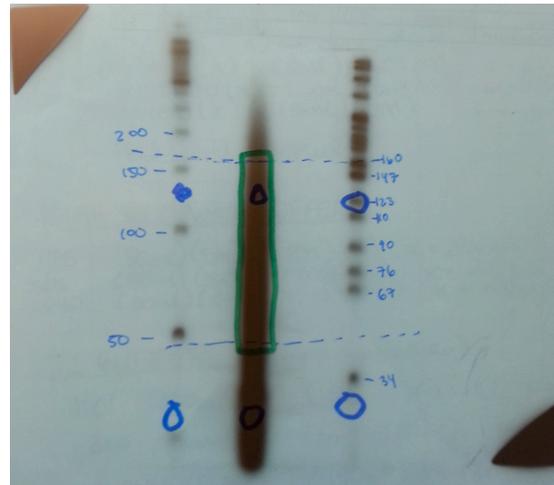
### **RNA Size Selection of Inputs**

35. Run the INPUTS in a 6% Urea gel (as mentioned above for the PELLETS)

To maximize our yield on the Illumina TruSeq stranded mRNA Sample Prep kit (#RS-122-2101) we expanded the range from 50-150nt, to **50-300nt**

Here is an example to the right.

36. Continue with the RNAClean XP purification steps



### **Illumina library sample preparation for Inputs and Pellets**

Remove traces of UREA using the RNAClean XP beads protocol (Beckman Coulter #A63987)

1. Spin at 16,000xg for 10min at RT. Discard supernatant
2. Wash in ~500ul 70% ETOH. Spin 5 min at 16,000xg. Discard supernatant.
3. Wash in ~500ul 70% ETOH. Spin 5 min at 16,000xg. Discard supernatant.

4. After gel elution, resuspend INPUT and Pellet samples in 102ul H<sub>2</sub>O.

NOTE: If you have multiple tubes after the gel purification step, resuspend and combine all the RNA pellets in a final total volume of 102ul

5. Take 1ul of each and put them in scintillation vials.

6. Compare the counts from INPUT and PELLETT (use scintillation counter; no need to add scintillation fluid) and use equivalent amounts of RNA to continue with the protocol.

7. For consistent results: Make sure to use these exact brands of ETOH (American Bioanalytical #AB00138-01000) and isopropanol (American Bioanalytical # A07015)

8. Vortex the RNAClean XP beads to resuspend the beads. Make sure the beads equilibrate at RT for  $\geq 30$ min.

9. Bring up the volume of your RNA to 150ul of H<sub>2</sub>O in a 1.5ml tube

10. Add 270ul of RNAClean XP beads to each tube

11. Add 270ul of isopropanol to each tube

12. Mix the reaction 10X until the solution is completely homogenous

13. Incubate at RT for 5min

14. Place tubes on magnetic stand and let them settle for 15min

15. Discard the supernatant. Keep tubes on magnetic stand

16. Add 700ul of 85% ETOH slowly

17. Let sit 30sec

18. Remove supernatant.

19. Add 700ul of 85% ETOH slowly

20. Let sit 30sec

21. Remove supernatant.
22. Air dry beads for 5min
23. Elute your sample with 200ul H<sub>2</sub>O
24. Remove tubes from magnetic stand.
25. Pipet up and down 10X to mix the solution
26. Let tubes sit at RT for 2min
27. Place tubes on magnetic stand for 2-5min
28. Collect the eluted RNAs in the supernatant
29. Ethanol precipitate the RNAs by adding the following:

200ul	H <sub>2</sub> O
500ul	100% ETOH
20ul	3M Sodium Acetate
1ul	glycoblue

- a. NOTE: Upon mixing, the PELLET tube will appear turbid/opaque and the INPUT will be clear.
30. Incubate on dry ice overnight

### **Modified Illumina TruSeq Stranded mRNA Sample Preparation Guide**

Manual#15031047 Rev E Oct 2013. (Omit all in-line controls)

1. Thaw the Fragment, Prime, Finish Mix (FPF) at RT (Stored at -20°C)
2. Spin RNA-ETOH mix at 16000xg for 10min at RT.
3. Wash RNA pellets in ~500ul 70% ETOH.
4. Spin at 16000xg for 5min at RT. Discard supernatant.
5. Wash RNA pellets in ~500ul 70% ETOH.

6. Spin at 16000xg for 5min at RT. Discard supernatant.
7. Resuspend the RNA pellet in 18ul of Fragment, Prime, Finish Mix. Pipet up & down 6-10X.
8. Store the Fragment, Prime, Finish Mix @20°C
9. Transfer solution to 0.2ml PCR tubes (keep on ice)
10. Check PCR program “65to4”
  - a. Make sure the lid matches the temperature of the PCR block
    - i. SETUP/LID/MODE/TRACKING/Lid offset 5°C
  - b. PCR Program:  $V_T=18\text{ul}$ 
    - i. 65°C for 5min
    - ii. 4°C for  $\infty$
11. Place tubes on PCR machine.
12. As soon as the PCR block reaches 4°C, remove the tube from the block.
13. Pulse spin. Leave at RT.

### **1<sup>st</sup> cDNA strand**

14. Thaw the First Strand Synthesis Act D (FSA) at RT (Stored at -20°C)
15. Once thawed, centrifuge FSA tube at 600xg 3sec
16. Make a master mix of the FSA mix with SuperScript II RT
  - a. The ratio is as follows: 9ul FSA for 1ul SuperScript II RT
  - b. However, you will only need 8ul of FSA+RT Mix per tube
17. Add 8ul of FSA+RT Mix to each 0.2ml tube. Mix by pipetting 6X

18. Store stock FSA mix at -20°C. Store any remaining FSA+RT mix at -20°C.
19. Select the PCR program “1CDNA”
  - a. Set lid to 100°C
    - i. SETUP/LID/MODE/CONSTANT/Lid target 100°C
  - b. PCR Program:  $V_T=26\text{ul}$ 
    - i. 25°C for 10min
    - ii. 42°C for 15min
    - iii. 70°C for 15min
    - iv. 4°C for  $\infty$
20. As soon as the PCR block reaches 4°C, remove the tube from the block.
21. Pulse spin. Leave at RT.

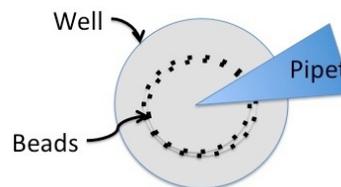
## **2<sup>nd</sup> cDNA strand**

22. Thaw at RT:
  - a. Resuspension Buffer (After initial thaw, store at 4°C)
  - b. Second Strand Marking Master Mix (Stored at -20°C)
  - c. AMPURE XP beads (Stored at 4°C, should equilibrate at RT  $\geq 30\text{min}$ )
  - d. Setup PCR at 16°C  $V_T=50\text{ul}$ 
    - i. SETUP/LID/MODE/CONSTANT/Lid target 30°C
    - ii. SETUP/LID/MIN/20°C
23. Pulse spin all reagent tubes at 600xg for 3sec
24. Add 5ul Resuspension Buffer to each tube.

25. Add 20ul of thawed Second Strand Marking Master Mix. Pipet up and down 6X.
26. Store Second Strand Marking Master Mix to -20°C
27. Place tube in preheated PCR block at 16C for 1hr
28. Remove tubes from 16°C, pulse spin, and place on bench
29. Vortex AMPURE XP beads until they are well dispersed. Make sure they have been sitting at RT for  $\geq 30$ min
30. Add 100ul of mixed AMPURE XP beads to the 50ul reaction
  - a. 2X ratio selects  $\geq 100$ bp fragments. 100bp is the lower end limit of the beads
31. Pipet up and down to mix 10X. Make sure solution is homogeneous
32. Transfer solution to a U-bottom 96well plate (Evergreen Scientific #290-8117-01R).

Skip wells and rows in between samples to prevent cross-contamination.
33. Leave at RT for 15min
34. Place on 96well magnet 5min. Keep plate on magnet until specified otherwise
35. Remove  $\sim 135$ ul supernatant from each well.

Aspirate the supernatant by putting your pipet in the middle of the well and slowly removing the liquid. See picture on right.
36. Add 200ul of 80% ETOH without disturbing the beads
37. Let sit 30sec. Remove and discard the supernatant
38. Wash one more time in 200ul of 80% ETOH (total of 2X)
39. Let sit 30sec. Remove and discard the supernatant
40. Let stand at RT 15min to dry.



41. Pulse spin Resuspension buffer at 600xg for 3sec
42. Add 17.5ul of Resuspension buffer to each well. Remove plate from magnet.
43. Pipet up and down 10X to mix thoroughly
44. Incubate at RT 2min
45. Place on magnetic stand 5min
46. Transfer 15ul of supernatant to a clean 0.2ml PCR tube (keep U-bottom plate for use later in the protocol)

### **Adenylate 3' ends**

47. Thaw at RT:
  - a. A Tailing Mixt (ATL) (Stored at -20°C)
  - b. Resuspension Buffer
  - c. Setup PCR program "ATAIL70"
48. Add 2.5ul of Resuspension buffer to each tube
49. Centrifuge A-Tailing mix at 600xg for 3sec
50. Add 12.5ul of A-Tailing mix to each tube. Pipet up and down 10X
51. Place tubes on PCR machine
52. Select PCR program: "ATAIL70"  $V_T=30\text{ul}$ 
  - a. SETUP/LID/MODE/CONSTANT/Lid target 100°C
  - b. PCR Program:
    - i. 37°C for 30min
    - ii. 70°C for 5min

iii. 4°C for ∞

53. As soon as the PCR block reaches 4°C, remove the tube from the block.

54. Pulse spin. Leave at RT

### Ligate Adapters

55. Thaw at RT the RNA Adapter Indices needed

- a. NOTE: For Illumina, color balancing is very important!
  - i. BEFORE you start your library, check with the sequencing core which set of RNA Adapter Indices you can use. It will depend on how they load your samples on the machine.
  - ii. For our samples:

SAMPLE	ADAPTER
P1	2
P2	4
P3	5
I1	6
I2	7
I3	12

Our lab has RNA adapter indexes 2, 4, 5, 6, 7, 12, 13, 14, 15, 16, 17, 18, & 19.

These are sensitive to multiple cycles of freezing and thawing.

- b. Resuspension buffer
- c. AMPURE XP beads (at RT ≥30min)
- d. Preheat PCR block to 30°C
- e. Preheat PCR LID at 100°C  $V_T=35\mu\text{l}$ 
  - i. SETUP/LID/MODE/CONSTANT/Lid target 100°C

56. Centrifuge RNA Adapter Index tubes at 600xg for 5sec
57. Pulse spin 0.2ml tubes
58. Add 2.5ul Resuspension buffer to each tube
59. Pulse spin Ligation Mix (LIG) at 600xg for 5sec (Stored at -20°C, take out of freezer immediately before use)
60. Add 2.5ul of Ligation Mix to each tube. Store Ligation mix at -20°C right away
61. Add 2.5ul of corresponding RNA adapter Index to each tube
62. Pipet 10X to mix. Pulse Spin.
63. Thaw Stop Ligation Buffer (STL) at RT (Stored at -20°C)
64. Place tubes on preheated 30°C block  $V_T=35\text{ul}$ 
  - a. PCR program “30C” with lid at 100°C
    - i. 30°C for 10min
    - ii. 25°C for  $\infty$
65. Remove tubes from block. Pulse spin
66. Add 5ul Stop Ligation Buffer
67. Pipet to mix 10X
68. Vortex AMPURE XP beads until they are well dispersed. Make sure they have been sitting at RT for  $\geq 30\text{min}$
69. Add 64ul of mixed AMPURE XP beads to the 40ul reaction
  - a. Adapters are 120bp
  - b. Our RNA fragments are 50bp (120bp adapters+50bpfragment=170bp)
  - c. Therefore a 1.6X ratio should select for  $\geq 150\text{bp}$

70. Pipet up and down to mix 10X. Make sure solution is homogeneous
71. Transfer solution to a U-bottom 96well plate. Skip wells and rows in between samples to prevent cross-contamination.
72. Leave at RT for 15min
73. Place on 96well magnet 5min. Keep plate on magnet until specified otherwise
74. Remove ~100ul supernatant from each well. Aspirate the supernatant by putting your pipet in the middle of the well and SLOWLY removing the liquid.
75. Add 200ul of 80% ETOH without disturbing the beads
76. Let sit 30sec. Remove and discard the supernatant
77. Wash one more time in 200ul of 80% ETOH (total of 2X)
78. Let sit 30sec. Remove and discard the supernatant
79. Let stand at RT 15min to dry.
80. Pulse spin Resuspension buffer at 600xg for 3sec
81. Add 52.5ul of Resuspension buffer to each well. Remove plate from magnet.
82. Pipet up and down 10X to mix thoroughly
83. Incubate at RT 2min
84. Place on magnetic stand 5min
85. Transfer 50ul of supernatant to a clean 96 well
86. Vortex AMPURE XP beads until they are well dispersed. Make sure they have been sitting at RT for  $\geq 30$ min
87. Add 80ul of mixed AMPURE XP beads to the 50ul reaction
  - a. Therefore a 1.6X ratio should select for  $\geq 150$ bp

88. Pipet up and down to mix 10X. Make sure solution is homogeneous
89. Transfer solution to a U-bottom 96well plate. Skip wells and rows in between samples to prevent cross-contamination.
90. Leave at RT for 15min
91. Place on 96well magnet 5min. Keep plate on magnet until specified otherwise
92. Remove ~100ul supernatant from each well. Aspirate the supernatant by putting your pipet in the middle of the well and slowly removing the liquid.
93. Add 200ul of 80% ETOH without disturbing the beads
94. Let sit 30sec. Remove and discard the supernatant
95. Wash one more time in 200ul of 80% ETOH (total of 2X)
96. Let sit 30sec. Remove and discard the supernatant
97. Let stand at RT 15min to dry.
98. Pulse spin Resuspension buffer at 600xg for 3sec
99. Add 22.5ul of Resuspension buffer to each well. Remove plate from magnet.
100. Pipet up and down 10X to mix thoroughly
101. Incubate at RT 2min
102. Place on magnetic stand 5min
103. Transfer 20ul of supernatant to a clean 0.2ml tube (keep U-bottom plate for use later in the protocol)

**Enrich DNA Fragments**

104. Thaw at RT:
  - a. PCR Master Mix (Stored at -20°C)
  - b. PCR Primer Cocktail (Stored at -20°C)
  - c. Resuspension Buffer
  - d. AMPURE XP Beads (at RT  $\geq$ 30min)
  - e. PCR lid SETUP/LID/MODE/CONSTANT/Lid target 100°C
  - f. Program “CLIPPCR”
105. Add 5ul of PCR Primer Cocktail to each tube
106. Add 25ul PCR Master Mix to each tube. Pipet to mix 10X
107. Setup PCR Machine:
  - a. SETUP/LID/MODE/CONSTANT/Lid target 100°C
  - b. PCR Program “CLIPPCR”:  $V_T=50\text{ul}$ 
    - i. 1) 98°C for 30sec
    - ii. 2) 98°C for 10sec
    - iii. 3) 60°C for 30sec
    - iv. 4) 72°C for 30sec
    - v. 5) Go to 2) 11X TOTAL of 12 cycles
    - vi. 6) 72°C for 5min
    - vii. 7) 4°C for  $\infty$
    - viii. 8) END
108. Pulse spin tubes.

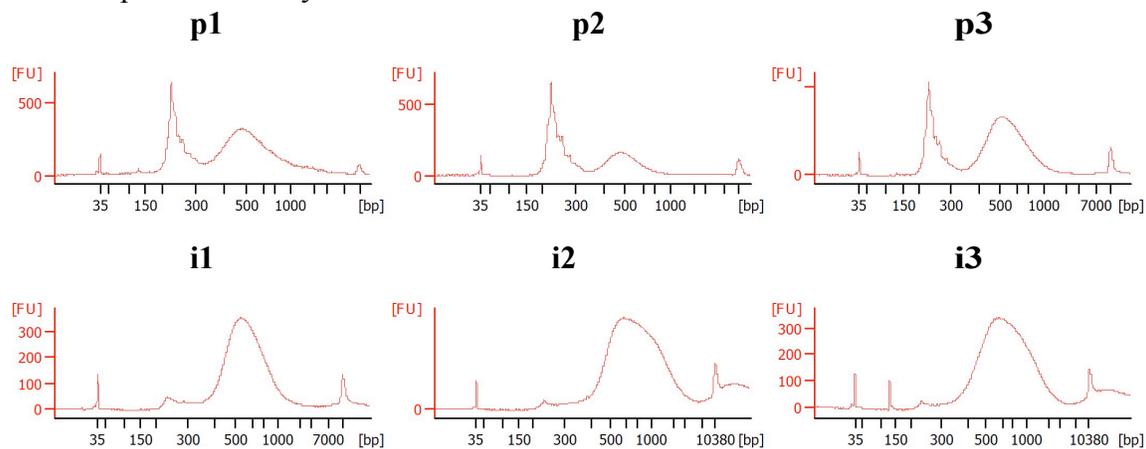
109. Vortex AMPURE XP beads until they are well dispersed. Make sure they have been sitting at RT for  $\geq 30$ min
110. Add 80ul of mixed AMPURE XP beads to the 50ul reaction
  - a. Therefore a 1.6X ratio should select for  $\geq 150$ bp
111. Pipet up and down to mix 10X. Make sure solution is homogeneous
112. Transfer solution to a U-bottom 96well plate. Skip wells and rows in between samples to prevent cross-contamination.
113. Leave at RT for 15min
114. Place on 96well magnet 5min. Keep plate on magnet until specified otherwise
115. Remove  $\sim 100$ ul supernatant from each well. Aspirate the supernatant by putting your pipet in the middle of the well and slowly removing the liquid.
116. Add 200ul of 80% ETOH without disturbing the beads
117. Let sit 30sec. Remove and discard the supernatant
118. Wash one more time in 200ul of 80% ETOH (total of 2X)
119. Let sit 30sec. Remove and discard the supernatant
120. Let stand at RT 15min to dry.
121. Pulse spin Resuspension buffer at 600xg for 3sec
122. Add 32.5ul of Resuspension buffer to each well. Remove plate from magnet.
123. Pipet up and down 10X to mix thoroughly
124. Incubate at RT 2min
125. Place on magnetic stand 5min

126. Transfer 30ul of supernatant to a clean 1.5ml tube

### Agilent Bioanalyzer – Testing library size and yield

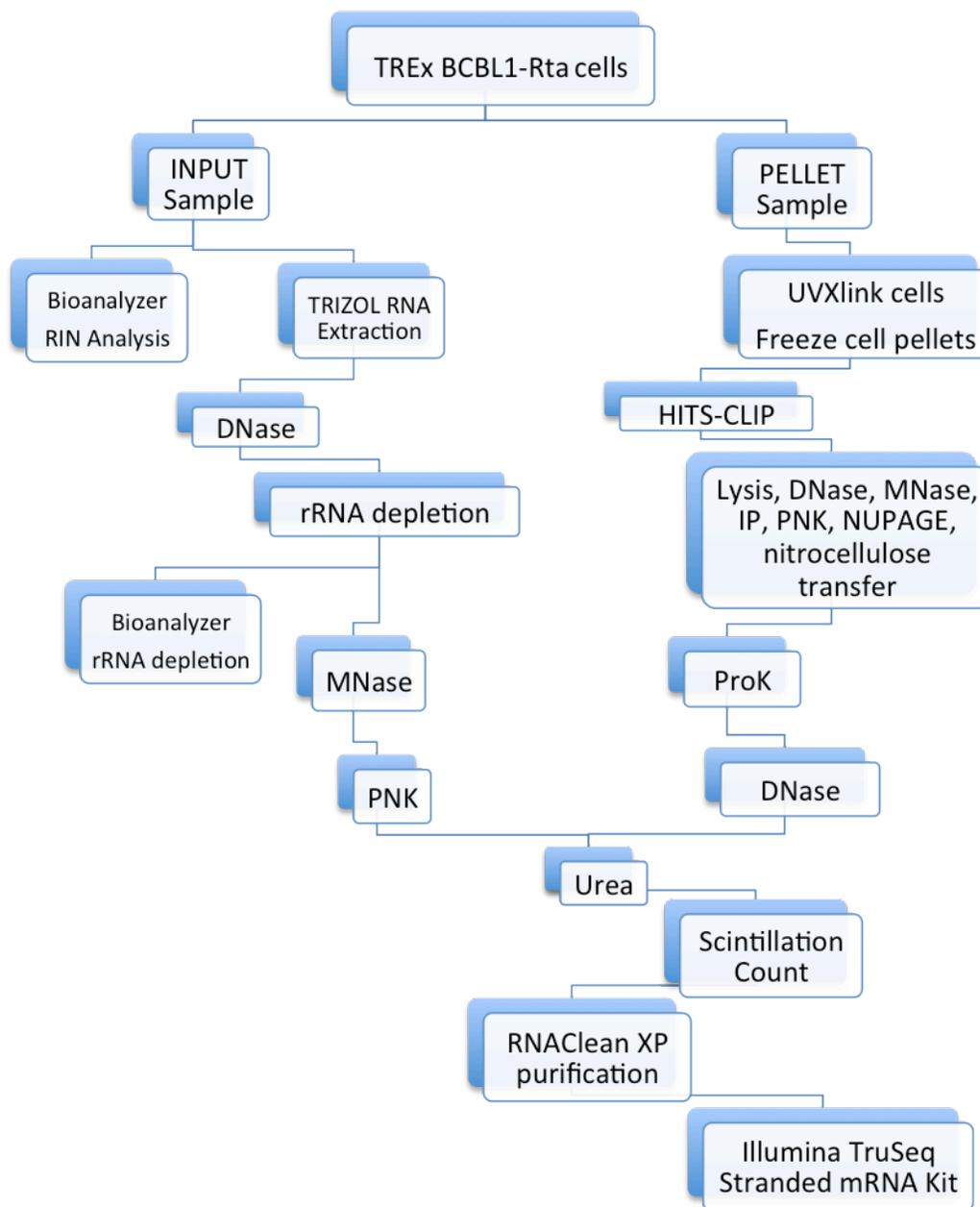
127. Analyze the yield of your cDNA library through Bioanalyzer.

Example of Bioanalyzer Results



## APPENDIX D

### HITS-CLIP workflow chart for Illumina Sequencing



## References:

1. Chang, Y., et al., *Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma*. Science, 1994. **266**(5192): p. 1865-9.
2. Bechtel, J.T., et al., *Host Range of Kaposi's Sarcoma-Associated Herpesvirus in Cultured Cells*. Journal of Virology, 2003. **77**(11): p. 6474-6481.
3. Speck, S.H. and D. Ganem, *Viral latency and its regulation: lessons from the gamma-herpesviruses*. Cell Host Microbe, 2010. **8**(1): p. 100-15.
4. Renne, R., et al., *Limited transmission of Kaposi's sarcoma-associated herpesvirus in cultured cells*. J Virol, 1998. **72**(6): p. 5182-8.
5. Majerciak, V. and Z.M. Zheng, *Kaposi's sarcoma-associated herpesvirus ORF57 in viral RNA processing*. Front Biosci (Landmark Ed), 2009. **14**: p. 1516-28.
6. Bello, L.J., et al., *The human herpesvirus-8 ORF 57 gene and its properties*. J Gen Virol, 1999. **80** ( Pt 12): p. 3207-15.
7. Russo, J.J., et al., *Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8)*. Proc Natl Acad Sci U S A, 1996. **93**(25): p. 14862-7.
8. Yakushko, Y., et al., *Kaposi's sarcoma-associated herpesvirus bacterial artificial chromosome contains a duplication of a long unique-region fragment within the terminal repeat region*. J Virol, 2011. **85**(9): p. 4612-7.
9. Zheng, Z.M., *Split genes and their expression in Kaposi's sarcoma-associated herpesvirus*. Rev Med Virol, 2003. **13**(3): p. 173-84.
10. Maniatis, T. and R. Reed, *An extensive network of coupling among gene expression machines*. Nature, 2002. **416**(6880): p. 499-506.
11. Le Hir, H., A. Nott, and M.J. Moore, *How introns influence and enhance eukaryotic gene expression*. Trends Biochem Sci, 2003. **28**(4): p. 215-20.
12. Sahin, B.B., D. Patel, and N.K. Conrad, *Kaposi's sarcoma-associated herpesvirus ORF57 protein binds and protects a nuclear noncoding RNA from cellular RNA decay pathways*. PLoS Pathog, 2010. **6**(3): p. e1000799.
13. Han, Z. and S. Swaminathan, *Kaposi's sarcoma-associated herpesvirus lytic gene ORF57 is essential for infectious virion production*. J Virol, 2006. **80**(11): p. 5251-60.
14. Majerciak, V., et al., *Targeted disruption of Kaposi's sarcoma-associated herpesvirus ORF57 in the viral genome is detrimental for the expression of ORF59, K8alpha, and K8.1 and the production of infectious virus*. J Virol, 2007. **81**(3): p. 1062-71.
15. Sei, E. and N.K. Conrad, *Delineation of a core RNA element required for Kaposi's sarcoma-associated herpesvirus ORF57 binding and activity*. Virology, 2011. **419**(2): p. 107-16.
16. Borah, S., et al., *A viral nuclear noncoding RNA binds re-localized poly(A) binding protein and is required for late KSHV gene expression*. PLoS Pathog, 2011. **7**(10): p. e1002300.

17. Covarrubias, S., et al., *Host shutoff is a conserved phenotype of gammaherpesvirus infection and is orchestrated exclusively from the cytoplasm*. J Virol, 2009. **83**(18): p. 9554-66.
18. Kaposi, M., *Idiopathisches multiples Pigmentsarkom der Haut*. Arch Dermatol Syph, 1872. **4**: p. 265-273.
19. Radu, O. and L. Pantanowitz, *Kaposi sarcoma*. Arch Pathol Lab Med, 2013. **137**(2): p. 289-94.
20. Fatahzadeh, M. and R.A. Schwartz, *Oral Kaposi's sarcoma: a review and update*. Int J Dermatol, 2013. **52**(6): p. 666-72.
21. Mesri, E.A., E. Cesarman, and C. Boshoff, *Kaposi's sarcoma and its associated herpesvirus*. Nat Rev Cancer, 2010. **10**(10): p. 707-19.
22. Gallo, R.C., *The enigmas of Kaposi's sarcoma*. Science, 1998. **282**(5395): p. 1837-9.
23. Cesarman E, C.Y., Moore PS, Said JW, Knowles DM., *Kaposi's sarcoma-associated herpesvirus-like DNA sequences in AIDS-related body-cavity-based lymphomas*. N Engl J Med, 1995. **332**(18): p. 1186-91.
24. Soulier J, G.L., Oksenhendler E, Cacoub P, Cazals-Hatem D, Babinet P, d'Agay MF, Clauvel JP, Raphael M, Degos L, and Sigaux F., *Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease*. Blood, 1995. **86**(4): p. 1276-80.
25. Chen, L. and M. Lagunoff, *Establishment and maintenance of Kaposi's sarcoma-associated herpesvirus latency in B cells*. J Virol, 2005. **79**(22): p. 14383-91.
26. Cai, Q., et al., *Molecular biology of Kaposi's sarcoma-associated herpesvirus and related oncogenesis*. Adv Virus Res, 2010. **78**: p. 87-142.
27. Duprez, R., et al., *Evidence for a multiclonal origin of multicentric advanced lesions of Kaposi sarcoma*. J Natl Cancer Inst, 2007. **99**(14): p. 1086-94.
28. Chandran, B., *Early events in Kaposi's sarcoma-associated herpesvirus infection of target cells*. J Virol, 2010. **84**(5): p. 2188-99.
29. Lieberman, P.M., *Keeping it quiet: chromatin control of gammaherpesvirus latency*. Nat Rev Microbiol, 2013. **11**(12): p. 863-75.
30. Toth, Z., K. Brulois, and J.U. Jung, *The chromatin landscape of Kaposi's sarcoma-associated herpesvirus*. Viruses, 2013. **5**(5): p. 1346-73.
31. Renne, R.L., M.; Zhong, W.; Ganem, D., *The size and conformation of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) DNA in infected cells and virions*. Journal of Virology, 1996. **70**(11): p. 8151-4.
32. Ballestas, M.E. and K.M. Kaye, *The latency-associated nuclear antigen, a multifunctional protein central to Kaposi's sarcoma-associated herpesvirus latency*. Future Microbiol, 2011. **6**(12): p. 1399-413.
33. Verma SC, L.K., Robertson E., *Structure and function of latency-associated nuclear antigen*. Curr Top Microbiol Immunol, 2007. **312**: p. 101-136.
34. Ballestas, M.E. and K.M. Kaye, *Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen 1 mediates episome persistence through cis-acting terminal repeat (TR) sequence and specifically binds TR DNA*. J Virol, 2001. **75**(7): p. 3250-8.

35. Ballestas, M.E., *Efficient Persistence of Extrachromosomal KSHV DNA Mediated by Latency-Associated Nuclear Antigen*. Science, 1999. **284**(5414): p. 641-644.
36. Chandriani, S. and D. Ganem, *Host transcript accumulation during lytic KSHV infection reveals several classes of host responses*. PLoS One, 2007. **2**(8): p. e811.
37. Glaunsinger, B. and D. Ganem, *Lytic KSHV infection inhibits host gene expression by accelerating global mRNA turnover*. Mol Cell, 2004. **13**(5): p. 713-23.
38. Lukac, D.M., J.R. Kirshner, and D. Ganem, *Transcriptional activation by the product of open reading frame 50 of Kaposi's sarcoma-associated herpesvirus is required for lytic viral reactivation in B cells*. J Virol, 1999. **73**(11): p. 9348-61.
39. Xu, Y., et al., *A Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8 ORF50 deletion mutant is defective for reactivation of latent virus and DNA replication*. J Virol, 2005. **79**(6): p. 3479-87.
40. AuCoin, D.P., et al., *Kaposi's Sarcoma-Associated Herpesvirus (Human Herpesvirus 8) Contains Two Functional Lytic Origins of DNA Replication*. Journal of Virology, 2002. **76**(15): p. 7890-7896.
41. Lin, C.L., et al., *Kaposi's sarcoma-associated herpesvirus lytic origin (ori-Lyt)-dependent DNA replication: identification of the ori-Lyt and association of K8 bZip protein with the origin*. J Virol, 2003. **77**(10): p. 5578-88.
42. AuCoin, D.P., et al., *Amplification of the Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8 lytic origin of DNA replication is dependent upon a cis-acting AT-rich region and an ORF50 response element and the trans-acting factors ORF50 (K-Rta) and K8 (K-bZIP)*. Virology, 2004. **318**(2): p. 542-55.
43. Wang, Y., et al., *Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: cis-acting requirements for replication and ori-Lyt-associated RNA transcription*. J Virol, 2004. **78**(16): p. 8615-29.
44. Wang, Y., et al., *Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: dual role of replication and transcription activator*. J Virol, 2006. **80**(24): p. 12171-86.
45. Chen, J., et al., *Genome-wide identification of binding sites for Kaposi's sarcoma-associated herpesvirus lytic switch protein, RTA*. Virology, 2009. **386**(2): p. 290-302.
46. Rossetto, C., I. Yamboliev, and G.S. Pari, *Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8 K-bZIP modulates latency-associated nuclear protein-mediated suppression of lytic origin-dependent DNA synthesis*. J Virol, 2009. **83**(17): p. 8492-501.
47. Hunter, O.V., et al., *Chromatin immunoprecipitation and microarray analysis suggest functional cooperation between Kaposi's Sarcoma-associated herpesvirus ORF57 and K-bZIP*. J Virol, 2013. **87**(7): p. 4005-16.
48. Arias, C., et al., *KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features*. PLoS Pathog, 2014. **10**(1): p. e1003847.

49. Myoung, J. and D. Ganem, *Generation of a doxycycline-inducible KSHV producer cell line of endothelial origin: maintenance of tight latency with efficient reactivation upon induction*. J Virol Methods, 2011. **174**(1-2): p. 12-21.
50. Vieira, J. and P.M. O'Hearn, *Use of the red fluorescent protein as a marker of Kaposi's sarcoma-associated herpesvirus lytic gene expression*. Virology, 2004. **325**(2): p. 225-40.
51. Dresang, L.R., et al., *Coupled transcriptome and proteome analysis of human lymphotropic tumor viruses: insights on the detection and discovery of viral genes*. BMC Genomics, 2011. **12**: p. 625.
52. Basrai, M.A., P. Hieter, and J.D. Boeke, *Small open reading frames: beautiful needles in the haystack*. Genome Res, 1997. **7**(8): p. 768-71.
53. Vattem, K.M. and R.C. Wek, *Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells*. Proc Natl Acad Sci U S A, 2004. **101**(31): p. 11269-74.
54. Kronstad, L.M., et al., *Dual short upstream open reading frames control translation of a herpesviral polycistronic mRNA*. PLoS Pathog, 2013. **9**(1): p. e1003156.
55. Kirshner, J.R., et al., *Kaposi's sarcoma-associated herpesvirus open reading frame 57 encodes a posttranscriptional regulator with multiple distinct activities*. J Virol, 2000. **74**(8): p. 3586-97.
56. Swaminathan, S., *Post-transcriptional gene regulation by gamma herpesviruses*. J Cell Biochem, 2005. **95**(4): p. 698-711.
57. Majerciak, V., et al., *Kaposi's sarcoma-associated herpesvirus ORF57 functions as a viral splicing factor and promotes expression of intron-containing viral lytic genes in spliceosome-mediated RNA splicing*. J Virol, 2008. **82**(6): p. 2792-801.
58. Nekorchuk, M., et al., *Kaposi's sarcoma-associated herpesvirus ORF57 protein enhances mRNA accumulation independently of effects on nuclear RNA export*. J Virol, 2007. **81**(18): p. 9990-8.
59. Boyne, J.R. and A. Whitehouse, *Nucleolar trafficking is essential for nuclear export of intronless herpesvirus mRNA*. Proc Natl Acad Sci U S A, 2006. **103**(41): p. 15190-5.
60. Taylor, A., et al., *Mutation of a C-terminal motif affects Kaposi's sarcoma-associated herpesvirus ORF57 RNA binding, nuclear trafficking, and multimerization*. J Virol, 2011. **85**(15): p. 7881-91.
61. Jackson, B.R., M. Noerenberg, and A. Whitehouse, *The Kaposi's Sarcoma-Associated Herpesvirus ORF57 Protein and Its Multiple Roles in mRNA Biogenesis*. Front Microbiol, 2012. **3**: p. 59.
62. Majerciak, V., et al., *Structural and functional analyses of Kaposi sarcoma-associated herpesvirus ORF57 nuclear localization signals in living cells*. J Biol Chem, 2006. **281**(38): p. 28365-78.
63. Boyne, J.R. and A. Whitehouse, *Nucleolar disruption impairs Kaposi's sarcoma-associated herpesvirus ORF57-mediated nuclear export of intronless viral mRNAs*. FEBS Lett, 2009. **583**(22): p. 3549-56.

64. Spector, D.L. and A.I. Lamond, *Nuclear speckles*. Cold Spring Harb Perspect Biol, 2011. **3**(2).
65. Greco, A., *Involvement of the nucleolus in replication of human viruses*. Rev Med Virol, 2009. **19**(4): p. 201-14.
66. Nemeth, A. and G. Langst, *Genome organization in and around the nucleolus*. Trends Genet, 2011. **27**(4): p. 149-56.
67. Malik, P., et al., *Functional co-operation between the Kaposi's sarcoma-associated herpesvirus ORF57 and ORF50 regulatory proteins*. J Gen Virol, 2004. **85**(Pt 8): p. 2155-66.
68. Palmeri, D., et al., *Promoter- and cell-specific transcriptional transactivation by the Kaposi's sarcoma-associated herpesvirus ORF57/Mta protein*. J Virol, 2007. **81**(24): p. 13299-314.
69. Aravind, L. and D. Landsman, *AT-hook motifs identified in a wide variety of DNA-binding proteins*. Nucleic Acids Res, 1998. **26**(19): p. 4413-21.
70. Vinson, C., A. Acharya, and E.J. Taparowsky, *Deciphering B-ZIP transcription factor interactions in vitro and in vivo*. Biochim Biophys Acta, 2006. **1759**(1-2): p. 4-12.
71. Malik, P. and J.B. Clements, *Protein kinase CK2 phosphorylation regulates the interaction of Kaposi's sarcoma-associated herpesvirus regulatory protein ORF57 with its multifunctional partner hnRNP K*. Nucleic Acids Res, 2004. **32**(18): p. 5553-69.
72. Kiledjian, M. and G. Dreyfuss, *Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box*. EMBO J, 1992. **11**(7): p. 2655-64.
73. Darnell, J.C., et al., *Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function*. Cell, 2001. **107**(4): p. 489-99.
74. Goodwin, D.J., et al., *The carboxy terminus of the herpesvirus saimiri ORF 57 gene contains domains that are required for transactivation and transrepression*. J Gen Virol, 2000. **81**(Pt 9): p. 2253-65.
75. Ruvolo, V., et al., *Functional analysis of Epstein-Barr virus SM protein: identification of amino acids essential for structure, transactivation, splicing inhibition, and virion production*. J Virol, 2004. **78**(1): p. 340-52.
76. Majerciak, V., et al., *Caspase-7 cleavage of Kaposi sarcoma-associated herpesvirus ORF57 confers a cellular function against viral lytic gene expression*. J Biol Chem, 2010. **285**(15): p. 11297-307.
77. Le Hir, H., et al., *The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions*. EMBO J, 2000. **19**(24): p. 6860-9.
78. Cheng, H., et al., *Human mRNA export machinery recruited to the 5' end of mRNA*. Cell, 2006. **127**(7): p. 1389-400.
79. Strasser, K., et al., *TREX is a conserved complex coupling transcription with messenger RNA export*. Nature, 2002. **417**(6886): p. 304-8.
80. Carmody, S.R. and S.R. Wentz, *mRNA nuclear export at a glance*. J Cell Sci, 2009. **122**(Pt 12): p. 1933-7.

81. Stubbs, S.H., et al., *Viral factors reveal a role for REF/Aly in nuclear RNA stability*. Mol Cell Biol, 2012. **32**(7): p. 1260-70.
82. Malik, P., D.J. Blackbourn, and J.B. Clements, *The evolutionarily conserved Kaposi's sarcoma-associated herpesvirus ORF57 protein interacts with REF protein and acts as an RNA export factor*. J Biol Chem, 2004. **279**(31): p. 33001-11.
83. Boyne, J.R., K.J. Colgan, and A. Whitehouse, *Recruitment of the complete hTREX complex is required for Kaposi's sarcoma-associated herpesvirus intronless mRNA nuclear export and virus replication*. PLoS Pathog, 2008. **4**(10): p. e1000194.
84. Jackson, B.R., et al., *An interaction between KSHV ORF57 and UIF provides mRNA-adaptor redundancy in herpesvirus intronless mRNA export*. PLoS Pathog, 2011. **7**(7): p. e1002138.
85. Boyne, J.R., B.R. Jackson, and A. Whitehouse, *ORF57: Master regulator of KSHV mRNA biogenesis*. Cell Cycle, 2010. **9**(14): p. 2702-3.
86. Chen, I.H., et al., *ICP27 recruits Aly/REF but not TAP/NXF1 to herpes simplex virus type 1 transcription sites although TAP/NXF1 is required for ICP27 export*. J Virol, 2005. **79**(7): p. 3949-61.
87. Chen, I.H., K.S. Sciabica, and R.M. Sandri-Goldin, *ICP27 interacts with the RNA export factor Aly/REF to direct herpes simplex virus type 1 intronless mRNAs to the TAP export pathway*. J Virol, 2002. **76**(24): p. 12877-89.
88. Boyne, J.R., et al., *Kaposi's sarcoma-associated herpesvirus ORF57 protein interacts with PYM to enhance translation of viral intronless mRNAs*. EMBO J, 2010. **29**(11): p. 1851-64.
89. Diem, M.D., et al., *PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs*. Nat Struct Mol Biol, 2007. **14**(12): p. 1173-9.
90. Gupta, A.K., et al., *The human herpesvirus 8 homolog of Epstein-Barr virus SM protein (KS-SM) is a posttranscriptional activator of gene expression*. J Virol, 2000. **74**(2): p. 1038-44.
91. Majerciak, V., K. Yamanegi, and Z.M. Zheng, *Gene structure and expression of Kaposi's sarcoma-associated herpesvirus ORF56, ORF57, ORF58, and ORF59*. J Virol, 2006. **80**(24): p. 11968-81.
92. Colgan, K.J., J.R. Boyne, and A. Whitehouse, *Identification of a response element in a herpesvirus saimiri mRNA recognized by the ORF57 protein*. J Gen Virol, 2009. **90**(Pt 3): p. 596-601.
93. Kang, J.G., et al., *Kaposi's sarcoma-associated herpesvirus ORF57 promotes escape of viral and human interleukin-6 from microRNA-mediated suppression*. J Virol, 2011. **85**(6): p. 2620-30.
94. Han, Z., et al., *General and target-specific RNA binding properties of Epstein-Barr virus SM posttranscriptional regulatory protein*. J Virol, 2009. **83**(22): p. 11635-44.

95. Glaunsinger, B. and D. Ganem, *Lytic KSHV Infection Inhibits Host Gene Expression by Accelerating Global mRNA Turnover*. Mol Cell, 2004. **13**: p. 713–723.
96. McCormick, C. and D. Ganem, *The kaposin B protein of KSHV activates the p38/MK2 pathway and stabilizes cytokine mRNAs*. Science, 2005. **307**(5710): p. 739-41.
97. Bagneris, C., et al., *Crystal structure of a KSHV-SOX-DNA complex: insights into the molecular mechanisms underlying DNase activity and host shutoff*. Nucleic Acids Res, 2011. **39**(13): p. 5744-56.
98. Glaunsinger, B., L. Chavez, and D. Ganem, *The exonuclease and host shutoff functions of the SOX protein of Kaposi's sarcoma-associated herpesvirus are genetically separable*. J Virol, 2005. **79**(12): p. 7396-401.
99. Wilkinson, D.E. and S.K. Weller, *The role of DNA recombination in herpes simplex virus DNA replication*. IUBMB Life, 2003. **55**(8): p. 451-8.
100. Covarrubias, S., et al., *Coordinated destruction of cellular messages in translation complexes by the gammaherpesvirus host shutoff factor and the mammalian exonuclease Xrn1*. PLoS Pathog, 2011. **7**(10): p. e1002339.
101. Hardwicke, M.A. and R.M. Sandri-Goldin, *The herpes simplex virus regulatory protein ICP27 contributes to the decrease in cellular mRNA levels during infection*. J Virol, 1994. **68**(8): p. 4797-810.
102. Smith, R.W., P. Malik, and J.B. Clements, *The herpes simplex virus ICP27 protein: a multifunctional post-transcriptional regulator of gene expression*. Biochem Soc Trans, 2005. **33**(Pt 3): p. 499-501.
103. Lindberg, A. and J.P. Kreivi, *Splicing inhibition at the level of spliceosome assembly in the presence of herpes simplex virus protein ICP27*. Virology, 2002. **294**(1): p. 189-98.
104. Hardy, W.R. and R.M. Sandri-Goldin, *Herpes simplex virus inhibits host cell splicing, and regulatory protein ICP27 is required for this effect*. J Virol, 1994. **68**(12): p. 7790-9.
105. Zhong, W. and D. Ganem, *Characterization of ribonucleoprotein complexes containing an abundant polyadenylated nuclear RNA encoded by Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8)*. J Virol, 1997. **71**(2): p. 1207-12.
106. Zhong, W., et al., *Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma*. Proc Natl Acad Sci U S A, 1996. **93**(13): p. 6641-6.
107. Sun, R., et al., *Polyadenylylated nuclear RNA encoded by Kaposi sarcoma-associated herpesvirus*. Proc Natl Acad Sci U S A, 1996. **93**(21): p. 11883-8.
108. Song, M.J., et al., *Transcription activation of polyadenylated nuclear rna by rta in human herpesvirus 8/Kaposi's sarcoma-associated herpesvirus*. J Virol, 2001. **75**(7): p. 3129-40.
109. Conrad, N.K. and J.A. Steitz, *A Kaposi's sarcoma virus RNA element that increases the nuclear abundance of intronless transcripts*. EMBO J, 2005. **24**(10): p. 1831-41.

110. Conrad, N.K., et al., *Identification of a rapid mammalian deadenylation-dependent decay pathway and its inhibition by a viral RNA element*. Mol Cell, 2006. **24**(6): p. 943-53.
111. Mitton-Fry, R.M., et al., *Poly(A) tail recognition by a viral RNA element through assembly of a triple helix*. Science, 2010. **330**(6008): p. 1244-7.
112. Conrad, N.K., et al., *Mutational analysis of a viral RNA element that counteracts rapid RNA decay by interaction with the polyadenylate tail*. Proc Natl Acad Sci U S A, 2007. **104**(25): p. 10412-7.
113. Tycowski, K.T., et al., *Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses*. Cell Rep, 2012. **2**(1): p. 26-32.
114. Brown, J.A., et al., *Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MENbeta noncoding RNAs*. Proc Natl Acad Sci U S A, 2012. **109**(47): p. 19202-7.
115. Lee, Y.J. and B.A. Glaunsinger, *Aberrant herpesvirus-induced polyadenylation correlates with cellular messenger RNA destruction*. PLoS Biol, 2009. **7**(5): p. e1000107.
116. Rossetto, C.C. and G. Pari, *KSHV PAN RNA associates with demethylases UTX and JMJD3 to activate lytic replication through a physical interaction with the virus genome*. PLoS Pathog, 2012. **8**(5): p. e1002680.
117. Horton, R.M., *PCR-mediated recombination and mutagenesis. SOEing together tailor-made genes*. Mol Biotechnol, 1995. **3**(2): p. 93-9.
118. Nakamura, H., et al., *Global changes in Kaposi's sarcoma-associated virus gene expression patterns following expression of a tetracycline-inducible Rta transactivator*. J Virol, 2003. **77**(7): p. 4205-20.
119. Church, G.M. and W. Gilbert, *Genomic sequencing*. Proc Natl Acad Sci U S A, 1984. **81**(7): p. 1991-5.
120. Conrad, N.K., *Chapter 15. Co-immunoprecipitation techniques for assessing RNA-protein interactions in vivo*. Methods Enzymol, 2008. **449**: p. 317-42.
121. Ule, J., et al., *CLIP: a method for identifying protein-RNA interaction sites in living cells*. Methods, 2005. **37**(4): p. 376-86.
122. Pfaffl, M.W., *A new mathematical model for relative quantification in real-time RT-PCR*. Nucleic Acids Res, 2001. **29**(9): p. e45.
123. Conrad, N.K., *Chapter 15 Co-Immunoprecipitation Techniques for Assessing RNA-Protein Interactions In Vivo*. 2008. **449**: p. 317-342.
124. Sei, E. and N.K. Conrad, *UV Cross-Linking of Interacting RNA and Protein in Cultured Cells*. Methods Enzymol, 2014. **539**: p. 53-66.
125. Licatalosi, D.D., et al., *HITS-CLIP yields genome-wide insights into brain alternative RNA processing*. Nature, 2008. **456**(7221): p. 464-9.
126. Darnell, R.B., *HITS-CLIP: panoramic views of protein-RNA regulation in living cells*. Wiley Interdiscip Rev RNA, 2010. **1**(2): p. 266-86.
127. Ule, J., et al., *CLIP identifies Nova-regulated RNA networks in the brain*. Science, 2003. **302**(5648): p. 1212-5.

128. Chi, S.W., et al., *Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps*. Nature, 2009. **460**(7254): p. 479-86.
129. Wu, T.D. and S. Nacu, *Fast and SNP-tolerant detection of complex variants and splicing in short reads*. Bioinformatics, 2010. **26**(7): p. 873-81.
130. Wu, T.D. and C.K. Watanabe, *GMAP: a genomic mapping and alignment program for mRNA and EST sequences*. Bioinformatics, 2005. **21**(9): p. 1859-75.
131. Anders, S. and W. Huber, *Differential expression analysis for sequence count data*. Genome Biol, 2010. **11**(10): p. R106.
132. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching*. Nucleic Acids Res, 2009. **37**(Web Server issue): p. W202-8.
133. Lorenz, R., et al., *ViennaRNA Package 2.0*. Algorithms Mol Biol, 2011. **6**: p. 26.
134. Cai, Q., et al., *Molecular biology of Kaposi's sarcoma-associated herpesvirus and related oncogenesis*. Advances in Virus Research, 2010. **78**: p. 87-142.
135. Dourmishev, L.A., et al., *Molecular genetics of Kaposi's sarcoma-associated herpesvirus (human herpesvirus-8) epidemiology and pathogenesis*. Microbiol Mol Biol Rev, 2003. **67**(2): p. 175-212.
136. Deng, H., Y. Liang, and R. Sun, *Regulation of KSHV lytic gene expression*. Curr Top Microbiol Immunol, 2007. **312**: p. 157-83.
137. Staudt, M.R. and D.P. Dittmer, *The Rta/Orf50 transactivator proteins of the gamma-herpesviridae*. Curr Top Microbiol Immunol, 2007. **312**: p. 71-100.
138. Conrad, N.K., *Posttranscriptional gene regulation in Kaposi's sarcoma-associated herpesvirus*. Adv Appl Microbiol, 2009. **68**: p. 241-61.
139. Glaunsinger, B.A. and D.E. Ganem, *Messenger RNA turnover and its regulation in herpesviral infection*. Adv Virus Res, 2006. **66**: p. 337-94.
140. Sinclair, A.J., *bZIP proteins of human gammaherpesviruses*. J Gen Virol, 2003. **84**(Pt 8): p. 1941-9.
141. Boyne, J.R. and A. Whitehouse, *gamma-2 Herpes virus post-transcriptional gene regulation*. Clin Microbiol Infect, 2006. **12**(2): p. 110-7.
142. Majerciak, V. and Z.M. Zheng, *Kaposi's sarcoma-associated herpesvirus ORF57 in viral RNA processing*. Front Biosci, 2009. **14**: p. 1516-28.
143. Ote, I., J. Piette, and C. Sadzot-Delvaux, *The Varicella-Zoster virus IE4 protein: a conserved member of the herpesviral mRNA export factors family and a potential alternative target in antiherpetic therapies*. Biochem Pharmacol, 2010. **80**(12): p. 1973-80.
144. Toth, Z. and T. Stamminger, *The human cytomegalovirus regulatory protein UL69 and its effect on mRNA export*. Front Biosci, 2008. **13**: p. 2939-49.
145. Sandri-Goldin, R.M., *The many roles of the regulatory protein ICP27 during herpes simplex virus infection*. Front Biosci, 2008. **13**: p. 5241-56.
146. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction*. Nucleic Acids Res, 2003. **31**(13): p. 3406-15.
147. Reed, R., *Coupling transcription, splicing and mRNA export*. Curr Opin Cell Biol, 2003. **15**(3): p. 326-31.
148. Pandit, S., D. Wang, and X.D. Fu, *Functional integration of transcriptional and RNA processing machineries*. Curr Opin Cell Biol, 2008. **20**(3): p. 260-5.

149. Vinciguerra, P. and F. Stutz, *mRNA export: an assembly line from genes to nuclear pores*. *Curr Opin Cell Biol*, 2004. **16**(3): p. 285-92.
150. Daugherty, M.D., I. D'Orso, and A.D. Frankel, *A solution to limited genomic capacity: using adaptable binding surfaces to assemble the functional HIV Rev oligomer on RNA*. *Mol Cell*, 2008. **31**(6): p. 824-34.
151. Zhu, J., A. Mayeda, and A.R. Krainer, *Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins*. *Mol Cell*, 2001. **8**(6): p. 1351-61.
152. Haecker, I., et al., *Ago HITS-CLIP expands understanding of Kaposi's sarcoma-associated herpesvirus miRNA function in primary effusion lymphomas*. *PLoS Pathog*, 2012. **8**(8): p. e1002884.
153. Polymenidou, M., et al., *Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43*. *Nat Neurosci*, 2011. **14**(4): p. 459-68.
154. Eden, E., et al., *Discovering motifs in ranked lists of DNA sequences*. *PLoS Comput Biol*, 2007. **3**(3): p. e39.
155. Eden, E., et al., *GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists*. *BMC Bioinformatics*, 2009. **10**: p. 48.
156. Zhang, C. and R.B. Darnell, *Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data*. *Nat Biotechnol*, 2011. **29**(7): p. 607-14.
157. Robinson, J.T., et al., *Integrative genomics viewer*. *Nat Biotechnol*, 2011. **29**(1): p. 24-6.
158. Thorvaldsdottir, H., J.T. Robinson, and J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration*. *Brief Bioinform*, 2013. **14**(2): p. 178-92.
159. Chu, L., et al., *Activation of Egr-1 in human lung epithelial cells exposed to silica through MAPKs signaling pathways*. *PLoS One*, 2013. **8**(7): p. e68943.
160. Selmi, T., et al., *ZFP36 expression impairs glioblastoma cell lines viability and invasiveness by targeting multiple signal transduction pathways*. *Cell Cycle*, 2012. **11**(10): p. 1977-87.
161. Zhao, Y., et al., *BTG1 expression correlates with the pathogenesis and progression of ovarian carcinomas*. *Int J Mol Sci*, 2013. **14**(10): p. 19670-80.
162. Zhu, R., et al., *BTG1 inhibits breast cancer cell growth through induction of cell cycle arrest and apoptosis*. *Oncol Rep*, 2013. **30**(5): p. 2137-44.
163. Zhao, S., et al., *CD137 ligand is expressed in primary and secondary lymphoid follicles and in B-cell lymphomas: diagnostic and therapeutic implications*. *Am J Surg Pathol*, 2013. **37**(2): p. 250-8.
164. Gullo, C., et al., *Inhibition of proliferation and induction of apoptosis in multiple myeloma cell lines by CD137 ligand signaling*. *PLoS One*, 2010. **5**(5): p. e10845.
165. Clyde, K. and B.A. Glaunsinger, *Deep sequencing reveals direct targets of gammaherpesvirus-induced mRNA decay and suggests that multiple mechanisms govern cellular transcript escape*. *PLoS One*, 2011. **6**(5): p. e19655.

166. Zeiner, G.M., et al., *RNA analysis by biosynthetic tagging using 4-thiouracil and uracil phosphoribosyltransferase*. *Methods Mol Biol*, 2008. **419**: p. 135-46.
167. Rennekamp, A.J. and P.M. Lieberman, *Initiation of Epstein-Barr virus lytic replication requires transcription and the formation of a stable RNA-DNA hybrid molecule at OriLyt*. *J Virol*, 2011. **85**(6): p. 2837-50.
168. Bai, Z., et al., *Genomewide mapping and screening of Kaposi's sarcoma-associated herpesvirus (KSHV) 3' untranslated regions identify bicistronic and polycistronic viral transcripts as frequent targets of KSHV microRNAs*. *J Virol*, 2014. **88**(1): p. 377-92.
169. Sandri-Goldin, R.M. and G.E. Mendoza, *A herpesvirus regulatory protein appears to act post-transcriptionally by affecting mRNA processing*. *Genes Dev*, 1992. **6**(5): p. 848-63.
170. Sandri-Goldin, R.M., M.K. Hibbard, and M.A. Hardwicke, *The C-terminal repressor region of herpes simplex virus type 1 ICP27 is required for the redistribution of small nuclear ribonucleoprotein particles and splicing factor SC35; however, these alterations are not sufficient to inhibit host cell splicing*. *J Virol*, 1995. **69**(10): p. 6063-76.
171. Ellison, T.J., et al., *A comprehensive analysis of recruitment and transactivation potential of K-Rta and K-bZIP during reactivation of Kaposi's sarcoma-associated herpesvirus*. *Virology*, 2009. **387**(1): p. 76-88.