DEREGULATION OF HUMAN LONG NONCODING RNAS BY ONCOGENIC GAMMA-HERPESVIRAL MICRORNAS

By

SUNANTHA SETHURAMAN

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2017
To my parents, S Vijaya and K Sethuraman, and my brother S Srinivasan.
ACKNOWLEDGMENTS

I would like to acknowledge and thank all the people who have contributed directly and indirectly to my doctoral education and the work presented in this thesis.

I would like to thank the members of the Renne laboratory, Dr. Rolf Renne, Dr. Peter Turner, Dr. Vaibhav Jain, Lauren Gay, Jackie Serfecz, Natalie Martinez and Merin Thomas, for making the lab an inspiring and fun environment. I am extremely thankful to my advisor Dr. Rolf Renne for being a mentor, a friend and a critique through the four years I have worked with him. I am also thankful to Dr. Peter Turner for carefully reviewing my manuscripts and this thesis. I would like to acknowledge the past members of the Renne laboratory, Dr. Irina Haecker and Dr. Hong Seok Choi. Dr. Irina Haecker is one of the most inspirational scientists I have met, who mentored me as a rotation student on the project that has now developed into my dissertation work. Dr. Hong Seok Choi, Dr. Vaibhav Jain and Lauren Gay have guided me with my experiments and made me a better scientist through the many scientific discussions. Jackie, Natalie and Merin have been amazing friends, and they make even a stressful day in the lab seem easy.

I extend my thanks to the members of my doctoral committee, Dr. James Flanagan, Dr. Alfred Lewin, Dr. Andrew Rosalsky and Dr. Maurice Swanson. Their enthusiasm, encouragement and constructive criticism have been crucial in shaping up my dissertation project. Discussing my work with them has inspired me to think critically and view my project in new perspectives.

I am deeply grateful to my family, my parents and my brother, for their endless love, unwavering support and willingness to sacrifice many comforts to make me the person I am today. Their countless blessings and prayers, and unyielding confidence in me have seen me through some of the challenging days of graduate school. Finally, I would like to thank my
friends, here in Gainesville and back home in India, for their constant support and motivation. I am thankful to Radhika Sampath, Varun Venkata Kuppili, and Aditya Kumar Kasinadhuni for being the family away from home. Special thanks to Aditya for his constant encouragement and help in putting this dissertation together.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>8</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>9</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>10</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>17</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>19</td>
</tr>
<tr>
<td>Kaposi’s Sarcoma-Associated Herpesvirus (KSHV)</td>
<td>19</td>
</tr>
<tr>
<td>KSHV-Driven Cancers</td>
<td>19</td>
</tr>
<tr>
<td>KSHV Biphasic Life-Cycle</td>
<td>20</td>
</tr>
<tr>
<td>KSHV Latency-Associated Region (KLAR)</td>
<td>21</td>
</tr>
<tr>
<td>MicroRNAs (miRNAs)</td>
<td>22</td>
</tr>
<tr>
<td>MicroRNA Biogenesis</td>
<td>23</td>
</tr>
<tr>
<td>MicroRNA Targeting and Mechanisms of Gene Silencing</td>
<td>24</td>
</tr>
<tr>
<td>KSHV MicroRNA Targets</td>
<td>26</td>
</tr>
<tr>
<td>Ribonomics Techniques to Identify MicroRNA Targets</td>
<td>27</td>
</tr>
<tr>
<td>Long Noncoding RNAs (lncRNAs)</td>
<td>31</td>
</tr>
<tr>
<td>Diversity of LncRNA Localization and Function</td>
<td>35</td>
</tr>
<tr>
<td>Long Noncoding RNAs in Cancer</td>
<td>36</td>
</tr>
<tr>
<td>Long Noncoding RNAs in Viral Infections</td>
<td>37</td>
</tr>
<tr>
<td>Regulatory Networks Involving MicroRNAs and LncRNAs</td>
<td>39</td>
</tr>
<tr>
<td>2 MICRONONCODED AND INDEPENDENT DEREGULATION OF LONG NONCODING RNAs BY AN ONCOGENIC HERPESVIRUS</td>
<td>44</td>
</tr>
<tr>
<td>Summary</td>
<td>44</td>
</tr>
<tr>
<td>Significance</td>
<td>44</td>
</tr>
<tr>
<td>Introduction</td>
<td>45</td>
</tr>
<tr>
<td>Results</td>
<td>47</td>
</tr>
<tr>
<td>KSHV Deregulates Host LncRNAs</td>
<td>47</td>
</tr>
<tr>
<td>Mature Viral MicroRNAs and Ago2 are Present in the Nuclei of KSHV-Infected Cells</td>
<td>48</td>
</tr>
<tr>
<td>KSHV MicroRNAs Directly Target Host LncRNAs</td>
<td>49</td>
</tr>
<tr>
<td>Latent KSHV Deregulates LncRNAs Aberrantly Expressed In Cancer</td>
<td>51</td>
</tr>
<tr>
<td>Viral MicroRNAs Downregulate Tumor Suppressor LncRNA MEG3</td>
<td>52</td>
</tr>
<tr>
<td>Viral MicroRNAs and Latency Proteins Both Target ANRIL</td>
<td>52</td>
</tr>
</tbody>
</table>
MicroRNA-Independent Deregulation of Host LncRNA UCA1 Promotes Proliferation and Migration...54
Discussion...................................................................................................................56
Material and Methods ................................................................................................59

3 COMPUTATIONAL ANALYSIS OF RIBONOMICS DATASETS IDENTIFIES LONG NONCODING RNA TARGETS OF GAMMA-HERPESVIRAL MICRORNAS ...76

Summary.....................................................................................................................76
Significance ................................................................................................................76
Introduction................................................................................................................77
Results.......................................................................................................................79
  PIPE-CLIP Analysis Identifies Putative LncRNA Targets of KSHV and EBV MicroRNAs ........................................................................................................79
KSHV and EBV MicroRNAs Target Common LncRNAs ........................................81
Some LncRNA Targets of KSHV and EBV MicroRNAs are Aberrantly Expressed in Cancer ..............................................................................................................82
Modified CLASH Identifies KSHV MicroRNA Targets in Endothelial Cells ..........83
A Subset of LncRNA Targets are Exclusively Nuclear .............................................85
Comparative Analysis of MicroRNA Interaction with LncRNAs versus MRNAs ......86
3’ Ligated MicroRNAs Preferentially Target the 3’ End of LncRNAs .......................87
LncRNAs are Often Targeted by Non-Canonical Base-Pairing that is Stabilized by Compensatory 3’ Base Interactions ..................................................................................................................88
MicroRNAs Have Unique Binding Profiles along Their Length ..............................89
Discussion..................................................................................................................90
Material and Methods ..............................................................................................93

4 CONCLUSIONS AND FUTURE DIRECTIONS .........................................................108

Major Conclusions ...................................................................................................108
Future Directions ......................................................................................................111
Immediate Relevance ..............................................................................................114
Broader Impact .........................................................................................................114

LIST OF REFERENCES ...............................................................................................118

BIOGRAPHICAL SKETCH ..........................................................................................146
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Summary</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Summary of deregulated IncRNAs from the microarray analysis</td>
<td>73</td>
</tr>
<tr>
<td>2-2</td>
<td>Examples of oncogenic and tumor-suppressor IncRNAs deregulated by KSHV</td>
<td>74</td>
</tr>
<tr>
<td>2-3</td>
<td>Oligonucleotides used in this study</td>
<td>75</td>
</tr>
<tr>
<td>3-1</td>
<td>Summary of CLIP-seq studies of KSHV and/or EBV infected lymphoma cells</td>
<td>103</td>
</tr>
<tr>
<td>3-2</td>
<td>Number of clusters containing 7mer or 6mer seed match for viral miRNAs identified from CLIP-seq datasets</td>
<td>104</td>
</tr>
<tr>
<td>3-3</td>
<td>Percentage overlap between targets identified using CLIP-seq studies of KSHV and/or EBV infected lymphoma cells</td>
<td>105</td>
</tr>
<tr>
<td>3-4</td>
<td>Number of miRNA-IncRNA hybrids found in the CLASH experiment</td>
<td>106</td>
</tr>
<tr>
<td>3-5</td>
<td>Examples of cancer-associated IncRNAs that interact with KSHV miRNAs</td>
<td>107</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>An outline of the KSHV genome.</td>
<td>41</td>
</tr>
<tr>
<td>1-2</td>
<td>An outline of the miRNA biogenesis pathway.</td>
<td>42</td>
</tr>
<tr>
<td>1-3</td>
<td>Schematic outline of HITS-CLIP, PAR-CLIP and CLASH ribonomics protocols.</td>
<td>43</td>
</tr>
<tr>
<td>2-1</td>
<td>Expression profiling of wt-KSHV and Δcluster-KSHV infected endothelial cells.</td>
<td>65</td>
</tr>
<tr>
<td>2-2</td>
<td>KSHV miRNAs and Ago2 are partially localized in the nuclei of latently infected cells.</td>
<td>66</td>
</tr>
<tr>
<td>2-3</td>
<td>KSHV miRNAs directly bind to and downregulate host lncRNAs.</td>
<td>67</td>
</tr>
<tr>
<td>2-4</td>
<td>Tumor suppressor lncRNA MEG3 is targeted by KSHV miRNAs.</td>
<td>68</td>
</tr>
<tr>
<td>2-5</td>
<td>LncRNA ANRIL is targeted by both KSHV miRNAs and latency proteins.</td>
<td>69</td>
</tr>
<tr>
<td>2-6</td>
<td>UCA1 is upregulated by KSHV in a miRNA-independent manner.</td>
<td>71</td>
</tr>
<tr>
<td>3-1</td>
<td>Distribution of viral miRNA targets identified from all CLIP datasets.</td>
<td>96</td>
</tr>
<tr>
<td>3-2</td>
<td>Venn diagrams of overlapping targets between different CLIP datasets.</td>
<td>97</td>
</tr>
<tr>
<td>3-3</td>
<td>Sub-cellular distribution of lncRNA targets of cellular and viral miRNAs.</td>
<td>98</td>
</tr>
<tr>
<td>3-4</td>
<td>miRNAs ligate more frequently to the 3’ end of target lncRNAs than to the 3’ end of target mRNAs.</td>
<td>99</td>
</tr>
<tr>
<td>3-5</td>
<td>miRNA binding distribution along the length of the target RNA.</td>
<td>100</td>
</tr>
<tr>
<td>3-6</td>
<td>miRNAs often bind lncRNAs via non-canonical base pairing.</td>
<td>101</td>
</tr>
<tr>
<td>3-7</td>
<td>Percentage binding frequencies along the length of the miRNAs.</td>
<td>102</td>
</tr>
<tr>
<td>4-1</td>
<td>Examples of co-regulation of lncRNAs and neighboring mRNAs.</td>
<td>117</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

4SU 4-thiouridine
6G 6-thioguanine
AFAP1-AS1 actin filament associated protein 1- antisense 1
Ago Argonaute
AIDS acquired immunodeficiency syndrome
ANRIL antisense noncoding RNA in the INK4 locus
ANXA2 Annexin A2
APOBEC apolipoprotein B mRNA editing enzyme catalytic subunit
ARF alternative reading frame
BACH1 BTB domain and CNC homolog 1
BC-1 body cavity-1
BC-3 body cavity-3
BCBL-1 body cavity based lymphoma-1
BCC basal cell carcinoma
BCLAF1 BCL2 –associated transcription factor 1
BED browser extensible data
BIC B cell receptor inducible
C/EBP CCAAT/enhancer binding proteins
CDK cyclin-dependent kinase
CDKN1A cyclin-dependent kinase inhibitor 1A
CDKN2B-AS cyclin dependent kinase inhibitor 2B - antisense
cDNA complementary DNA
CDR1-AS CDR1 antisense RNA
CDS coding sequence
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ceRNA</td>
<td>competitive endogenous RNA</td>
</tr>
<tr>
<td>ChIRP</td>
<td>chromatin isolation by RNA purification</td>
</tr>
<tr>
<td>circRNA</td>
<td>circular RNA</td>
</tr>
<tr>
<td>CLASH</td>
<td>cross-linking ligation and sequencing of hybrids</td>
</tr>
<tr>
<td>CLIP</td>
<td>cross-linking immunoprecipitation</td>
</tr>
<tr>
<td>CNS</td>
<td>central nervous system</td>
</tr>
<tr>
<td>CRISPR</td>
<td>clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>CRNDE</td>
<td>colorectal neoplasia differentially expressed</td>
</tr>
<tr>
<td>DAPI</td>
<td>4’,6-diamidino-2-phenylindole</td>
</tr>
<tr>
<td>DCP</td>
<td>decapping enzyme</td>
</tr>
<tr>
<td>DGCR8</td>
<td>DiGeorge syndrome critical region gene 8</td>
</tr>
<tr>
<td>DLEU2</td>
<td>deleted in lymphocytic leukemia 2</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DR</td>
<td>direct repeats</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double stranded DNA</td>
</tr>
<tr>
<td>dsRNA</td>
<td>double stranded RNA</td>
</tr>
<tr>
<td>EBER</td>
<td>Epstein-Barr virus encoded small RNA</td>
</tr>
<tr>
<td>EBV</td>
<td>Epstein-Barr virus</td>
</tr>
<tr>
<td>ENCODE</td>
<td>encyclopedia of DNA elements</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>FANTOM</td>
<td>functional annotation of the mammalian genome</td>
</tr>
<tr>
<td>FIRRE</td>
<td>functional intergenic repeating RNA element</td>
</tr>
<tr>
<td>FN1</td>
<td>fibronectin 1</td>
</tr>
<tr>
<td>GAPDH</td>
<td>glyceraldehyde-3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>GEMIN8</td>
<td>gem nuclear organelle associated protein 8</td>
</tr>
</tbody>
</table>
GENCODE  genes of ENCODE
GR        glucocorticoid receptor
GTP       guanine tri-phosphate
HAART     highly active anti-retroviral therapy
HBV       hepatitis B virus
HCC       hepatocellular carcinoma
HCV       hepatitis C virus
HEIH       highly expressed in hepatocellular carcinoma
HIF1-α     hypoxia inducible factor 1- α
HITS-CLIP high throughput sequencing- CLIP
HIV       human immunodeficiency virus
HLA       histocompatibility complex
HMGA      high mobility group proteins characterized by an AT-hook
HOTAIR    HOX transcript antisense RNA
HOTAIRM1  HOXA transcript antisense RNA, myeloid-specific 1
HOTTIP    HOXA transcript at distal tip
HOX       homeobox
HRE       hormone response element
HSUR      herpesvirus Saimiri U-RNA genes
HSV       herpes simplex virus
HULC      highly upregulated in liver cancer
HUVEC     human umbilical cord vein endothelial cells
HVS       herpesvirus saimiri
IFA       immunofluorescence assay
IL-6      interleukin-6
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>INK4</td>
<td>inhibits CDK4</td>
</tr>
<tr>
<td>IRF2BP2</td>
<td>interferon regulatory factor 2 binding protein 2</td>
</tr>
<tr>
<td>kb</td>
<td>kilo base-pairs</td>
</tr>
<tr>
<td>KICS</td>
<td>KSHV inflammatory cytokine syndrome</td>
</tr>
<tr>
<td>KLAR</td>
<td>KSHV latency-associated region</td>
</tr>
<tr>
<td>KS</td>
<td>Kaposi’s sarcoma</td>
</tr>
<tr>
<td>KSHV</td>
<td>Kaposi’s sarcoma-associated herpesvirus</td>
</tr>
<tr>
<td>LANA</td>
<td>latency-associated nuclear antigen</td>
</tr>
<tr>
<td>LAT</td>
<td>latency associated transcript</td>
</tr>
<tr>
<td>LCL</td>
<td>lymphoblastoid cell line</td>
</tr>
<tr>
<td>LEC</td>
<td>lymphatic endothelial cells</td>
</tr>
<tr>
<td>lincRNA</td>
<td>long intergenic noncoding RNA</td>
</tr>
<tr>
<td>IncRNA</td>
<td>long (or large) noncoding RNA</td>
</tr>
<tr>
<td>LRRC8D</td>
<td>leucine rich repeat containing 8 family member D</td>
</tr>
<tr>
<td>LSD1</td>
<td>lysine demethylase 1</td>
</tr>
<tr>
<td>MAF</td>
<td>musculoaponeurotic fibrosarcoma</td>
</tr>
<tr>
<td>MALAT1</td>
<td>metastasis associated lung adenocarcinoma transcript 1</td>
</tr>
<tr>
<td>MAPK</td>
<td>mitogen-activated protein kinases</td>
</tr>
<tr>
<td>MCD</td>
<td>multicentric Castleman’s disease</td>
</tr>
<tr>
<td>MEG3</td>
<td>maternally expressed gene 3</td>
</tr>
<tr>
<td>MICB</td>
<td>MHC class I polypeptide-related sequence B</td>
</tr>
<tr>
<td>MIR17HG</td>
<td>micro RNA 17 host gene</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNA</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>NCBI</td>
<td>national center for biotechnology information</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>ncRNA</td>
<td>noncoding RNA</td>
</tr>
<tr>
<td>NEAT1</td>
<td>nuclear enriched abundant transcript 1</td>
</tr>
<tr>
<td>NF-κB</td>
<td>nuclear factor-κ B subunit</td>
</tr>
<tr>
<td>NHP2L1</td>
<td>NHP2-Like Protein 1</td>
</tr>
<tr>
<td>NORAD</td>
<td>noncoding RNA activated by DNA damage</td>
</tr>
<tr>
<td>NRED</td>
<td>noncoding RNA expression database</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PABPC</td>
<td>poly A binding protein complex</td>
</tr>
<tr>
<td>PAN RNA</td>
<td>polyadenylated nuclear RNA</td>
</tr>
<tr>
<td>PANDA</td>
<td>p21 associated ncRNA DNA damage activated</td>
</tr>
<tr>
<td>PAR-CLIP</td>
<td>photoactivatable ribonucleoside-CLIP</td>
</tr>
<tr>
<td>PCA3</td>
<td>prostate cancer associated 3</td>
</tr>
<tr>
<td>PcG</td>
<td>polycomb group</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PEL</td>
<td>primary effusion lymphoma</td>
</tr>
<tr>
<td>PLAC2</td>
<td>placenta-specific 2</td>
</tr>
<tr>
<td>PRC</td>
<td>polycomb repressive complex</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>quantitative reverse transcription – polymerase chain reaction</td>
</tr>
<tr>
<td>RBL2</td>
<td>Retinoblastoma-Like Protein 2</td>
</tr>
<tr>
<td>RBP</td>
<td>RNA binding protein</td>
</tr>
<tr>
<td>REV-T</td>
<td>Reticuloendotheliosis virus-T</td>
</tr>
<tr>
<td>RIP</td>
<td>RNA immunoprecipitation</td>
</tr>
<tr>
<td>RISC</td>
<td>RNA-induced silencing complex</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RNP</td>
<td>ribonucleoprotein</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>SARS-CoV</td>
<td>severe acute respiratory syndrome - coronavirus</td>
</tr>
<tr>
<td>Scr</td>
<td>scrambled</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate – polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>siRNA</td>
<td>small (or short) inhibitory RNA</td>
</tr>
<tr>
<td>SLA</td>
<td>Src like adaptor</td>
</tr>
<tr>
<td>SMAD</td>
<td>small / mothers against decapentaplegic</td>
</tr>
<tr>
<td>snoRNA</td>
<td>small nucleolar RNA</td>
</tr>
<tr>
<td>snRNA</td>
<td>small nuclear RNA</td>
</tr>
<tr>
<td>SRA</td>
<td>sequence read archive</td>
</tr>
<tr>
<td>STAT</td>
<td>signal transducer and activator of transcription</td>
</tr>
<tr>
<td>TERC</td>
<td>telomerase RNA component</td>
</tr>
<tr>
<td>TGF-β</td>
<td>transforming growth factor β</td>
</tr>
<tr>
<td>THBS1</td>
<td>thrombospondin 1</td>
</tr>
<tr>
<td>TIME</td>
<td>telomerase immortalized microvascular endothelial</td>
</tr>
<tr>
<td>TIVE</td>
<td>telomerase immortalized vein endothelial cells</td>
</tr>
<tr>
<td>TIVE-LTC</td>
<td>TIVE-long term culture</td>
</tr>
<tr>
<td>TP53INP1</td>
<td>tumor protein p53 inducible nuclear protein 1</td>
</tr>
<tr>
<td>TPD52</td>
<td>tumor protein D52</td>
</tr>
<tr>
<td>TRBP</td>
<td>TAR RNA binding protein</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>TWEAKR</td>
<td>TNF-like weak inducer of apoptosis</td>
</tr>
<tr>
<td>TβRII</td>
<td>TGF-β receptor II</td>
</tr>
</tbody>
</table>
UCA1  urothelial cancer associated 1
UTR  untranslated region
UV  ultraviolet
vCyclin  viral Cyclin
vFLIP  viral FLICE interacting protein
VIN  virus inducible lincRNA
wt  wild type
XIC  X inactivation center
XIST  X-inactive specific transcript
XRN  exoribonucleases
YWHAE  tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein epsilon
Kaposi’s sarcoma-associated herpesvirus (KSHV), an oncogenic virus, is the etiological agent of Kaposi’s sarcoma, primary effusion lymphoma (PEL), and certain forms of multicentric Castleman’s disease (MCD). KSHV persists as a latent episome with restricted gene expression in all tumor cells. KSHV primarily encodes four proteins and twenty-five mature microRNAs (miRNAs) during latency, and these are crucial for tumorigenesis. Multiple studies have reported the mRNA targets of KSHV miRNAs. Recent studies have suggested that there are at least twice as many long noncoding RNAs (lncRNAs) encoded in the cells as mRNAs. Diverse regulatory roles are played by lncRNAs in human cells, and their aberrant expression is associated with almost all human cancers. However, whether miRNAs target lncRNAs remains an open question. In this work, we demonstrate that human and viral miRNAs target cellular lncRNAs. We identified putative lncRNA targets of KSHV miRNAs in endothelial cells using microarray analysis. Using biochemical methods, we confirmed that miRNAs directly bind to and downregulate lncRNAs, especially two cancer-related lncRNAs, MEG3 and ANRIL. Aside from miRNAs, KSHV latency-associated proteins also contribute to lncRNA deregulation. UCA1, a lncRNA often upregulated in multiple cancers, is strongly upregulated by viral proteins.
Knockdown of UCA1 using siRNAs inhibits the cancer-associated phenotypes of proliferation and migration in endothelial cells.

Epstein-Barr virus (EBV), which causes Burkitt’s lymphoma, non-Hodgkin’s lymphoma and nasopharyngeal carcinoma, is another oncogenic human herpesvirus. High-throughput ribonomics experiments have been previously reported to identify mRNA targets of KSHV and EBV miRNAs in lymphoma cells. We reinvestigated the data from these studies to identify and catalog putative IncRNA targets of viral miRNAs in lymphomas caused by these viruses. Crosslinking ligation and sequencing of hybrids (CLASH) is a ribonomics experiment that directly investigates RNA-RNA interaction. We performed bioinformatics analysis on CLASH data from KSHV infected endothelial cells to identify thousands of direct miRNA-IncRNA interactions. Using in silico analysis, we compared miRNA-IncRNA interactions to miRNA-mRNA interactions and identified that they are largely similar with small yet significant differences. In contrast to the current opinion in the field, we identified that the majority of miRNA targeting occurs via seed-sequence independent base-pairing between the miRNA and the target RNA. These results serve as an important first step towards understanding miRNA-mediated IncRNA regulation. Together, the data presented in this thesis provides evidence for direct miRNA-mediated IncRNA regulation using genetic, biochemical and bioinformatics approaches, and highlights the potential importance of IncRNAs in KSHV pathogenesis and tumorigenesis.
CHAPTER 1
INTRODUCTION

Kaposi’s Sarcoma-Associated Herpesvirus (KSHV)

KSHV-Driven Cancers

Kaposi’s sarcoma (KS) was first described by Dr. Moritz Kaposi, an Austro-Hungarian dermatologist, in 1872 [1]. Kaposi’s sarcoma-associated herpesvirus (KSHV) was later identified in 1994 as the etiological agent of KS [2]. In the following years, primary effusion lymphoma (PEL) and certain forms of multicentric Castleman’s disease (MCD) were added to the list of human diseases caused by KSHV [3, 4]. A new disease called KSHV inflammatory cytokine syndrome (KICS) was described in 2010 [5, 6]. While MCD and KICS involve systemic inflammatory symptoms, KS and PEL are two major KSHV-driven lethal malignancies [6, 7]. Although a large percentage of the world population is KSHV positive, only a few individuals go on to develop KS [8]. Immunosuppression is the primary trigger that allows for advancement of KSHV-driven cancers, because in normal individuals the immune system effectively restricts KSHV replication [9]. Acquired immunodeficiency syndrome (AIDS) patients and recipients of organ transplants have a high risk for incidence of KS or PEL [9]. The KS that arises due to immunosuppression is called “Iatrogenic KS”. There also exist other clinical variants of KSHV that are endemic to Sub-Saharan Africa, and are more aggressive in infiltrating the skin. The KS caused by these variants is called “endemic KS” [8]. KS is the third leading cancer type and the fourth leading cause of death among sub-Saharan African males, with almost 22,400 reported cases per year as of 2008 [10].

KS initiates as visible lesions in the dermis with pronounced neovascularization and angiogenesis from the early stages [9]. These lesions then form plaques and eventually nodules,
which are predominantly comprised of spindle-shaped endothelial cells [9]. KS lesions in a patient are often multi-centric, that is they are independent occurrences and did not spread from a primary lesion [11]. Mortality due to KS is usually a result of visceral KS lesions that cause gastrointestinal bleeding or respiratory failure [9]. In contrast, PEL does not involve any tumor masses, but instead is characterized by lymphomatous effusions in body cavities, commonly the pleural, peritoneal and pericardial cavities [12]. Lymphoma progression is the primary cause of death in PEL patients.

Current approaches to control KS and PEL are primarily general cancer treatments like chemotherapy and stem cell transplantation [13, 14]. In cases where AIDS is the trigger, Anti-Retroviral Therapy (ART) used to restrict HIV replication is partially effective in controlling KS or PEL [14, 15]. However, there are no existing KSHV-specific treatment options which can directly address the root cause of KS and PEL. Development of KSHV-specific treatments requires a detailed understanding of the molecular biology of the virus and its interactions with host cellular defense mechanisms. The last two decades have seen extensive research in this direction; however, the exact molecular mechanisms by which KSHV drives cancer remain to be uncovered.

**KSHV Biphasic Life-Cycle**

While the mode of KSHV transmission remains debatable, it is well established that KSHV replicates in both B cells and endothelial cells within the human body [16]. KSHV, like other herpesviruses, uses both lytic and latent modes of replication. Lytic replication facilitates the spread of infection from one individual to another and also between different cell types within an infected individual. Latent replication allows for persistence of the virus in the human body for the lifespan of the individual, as inactive viral reservoirs. The lytic phase of the KSHV
life-cycle involves replication of viral episomes within the host cell by a viral DNA polymerase, and concludes with release of infectious viral particles and necrosis of the host cell due to viral egress [17]. In contrast, latent replication is a process coupled with host cell replication where the human DNA polymerase also replicates the episome. In the event of immunosuppression, latent KSHV drives rapid proliferation of human cells, and all tumors harbor latent viral episomes [18]. During latent infection, only a small subset of the viral genes is expressed [19, 20]. A graphical representation of KSHV genome is shown in Figure 1-1. Most of the latency-associated genes are clustered towards the 3’ end of the genome, in an area commonly referred to as the KSHV Latency-Associated Region (KLAR), which is described below [21]. Since continual viral persistence as latent episomes requires active cell division of host cells, latency associated proteins and noncoding RNAs (ncRNAs) play different roles in ensuring cell survival and proliferation [7].

**KSHV Latency-Associated Region (KLAR)**

A graphical outline of KLAR is indicated in Figure 1-1 and presented in more detail later in Figure 2-1. KLAR includes four protein coding genes which are described below.

**Latency Associated Nuclear Antigen (LANA)** is encoded by ORF 73. LANA is a multi-functional protein, whose primary role is to tether viral episomes to human chromosomes and thus ensure episome replication, segregation and hence persistence [22-25]. LANA has also been shown to inactivate p53 and Rb tumor suppressor pathways [26, 27]. LANA is a transcriptional regulator of several human and viral promoters, including its own promoter [28-30].

**Viral-Cyclin (vCyclin)** is encoded by ORF 72. vCycl is a viral homolog of the human Cyclin-D protein, which phosphorylates Rb protein and enhances S-phase entry during cell cycle
It also overrides the inhibitory effect of two major Cyclin Dependent Kinase (CDK) inhibitors p21 and p27, thus promoting progression through the cell cycle [33-36].

**Viral-FLIP (vFLIP)** is encoded by ORF 71. vFLIP is a viral homolog of the cellular FLICE-Inhibitory Protein (cFLIP) [37]. vFLIP plays a role in cell survival by constitutively activating NF-κB and upregulating transcription of anti-apoptotic genes like Bcl-2 and BCL-X_L [38-40]. It also promotes cell survival by interfering with CD95/Fas signaling [39].

**Kaposin** is encoded by ORF 70. Kaposin gene encodes 3 protein variants namely Kaposin A, Kaposin B and Kaposin C [41]. Kaposin A is encoded from the canonical ORF 70 start codon. Upstream of the Kaposin A start codon are two GC-rich direct repeat sequences, DR1 and DR2. Kaposin B and Kaposin C are products of alternative translation initiation from these direct repeats using a non-canonical CUG codon for initiation. Kaposin B, the predominant variant in PEL cells, is involved in activation of the p38/MAPK pathway and stabilization of cytokine mRNAs [42].

Apart from the four major latency proteins, KLAR also encodes 12 pre-microRNAs (miRNAs) which eventually get processed into 25 mature miRNAs [43-45]. The biogenesis pathway of miRNAs and their functions are described in future sections. These miRNA genes are arranged such that 10 of them form what is called a ‘miRNA-cluster’, while the other two are within the ORF of the Kaposin gene (Figures 1-1 and 2-1) [46]. All of the cluster miRNAs are processed from the intron of the same primary transcript of the Kaposin gene which uses a promoter located within ORF73 [47].

**MicroRNAs (miRNAs)**

miRNAs are small ncRNAs, typically 21-24 nucleotides (nt) long, and they function as post-transcriptional regulators of gene expression [48]. *Lin-4*, the first described miRNA, was
identified in *C. elegans* by the Ambros lab [49]. Functionally, miRNAs form a sub-class of RNA interference (RNAi) pathway molecules. RNAi is a process in which small RNA molecules, such as miRNAs or siRNAs (small inhibitory RNAs), inhibit translation or gene expression by neutralizing mRNAs [50]. RNAi can be broadly viewed as two major steps: specific targeting of mRNAs and inhibition of mRNA expression. siRNAs, also 20-23 nt long, target any mRNA with which they achieve 100% base complementarity, and the inhibition often proceeds via mRNA cleavage and degradation. However, siRNAs are not common in mammalian systems, and few reported examples pertain to germline cells [51]. In contrast, miRNAs play important functional roles in almost all cell types. They function based on partial complementarity with the target mRNA, and miRNA-dependent RNAi involves silencing of mRNA expression through translational inhibition and mRNA destabilization [52]. miRNA biogenesis and targeting have been shown to be slightly different between plants, insects and humans [53]. The descriptions presented below are based on mammalian miRNAs.

**MicroRNA Biogenesis**

miRNAs are encoded as separate genes, or within the introns of coding and noncoding genes, and are often transcribed by RNA Polymerase II [54, 55] or in rare cases by RNA Polymerase III [45, 56-58]. Multiple miRNAs are often found in close proximity, thus forming a poly-cistronic transcription unit [59]. A graphical outline of the miRNA biogenesis pathway is shown in Figure 1-2. After transcription, miRNAs are in their primary form, typically longer than 1 kb, and are called pri-miRNAs. The Drosha/DGCR8 complex, also called as the microprocessor complex, is a nuclear protein complex which cleaves pri-miRNA to obtain pre-miRNA using the RNase III-type endonuclease function of Drosha [60, 61]. The pre-miRNA is approximately 65 nt long and has a staggered end with a 2-nt 3’ overhang. The pre-miRNA gets
exported to the cytoplasm via Exportin 5 transporters in a GTP-hydrolysis dependent step [62, 63]. In the final maturation step, Dicer, along with a key co-factor TRBP, cleaves the loop part of the pre-miRNA leaving a double stranded molecule with 2 nt 3’ overhangs at either end [64-66]. One strand of the duplex is called the guide strand, which is the functional mature miRNA, and the other strand, referred to as the passenger strand and often denoted as the * strand, is usually degraded [67, 68]. The guide strand gets loaded into the RNA-induced silencing complex (RISC) to make functionally active ternary ribonucleoprotein (RNP) complexes called miRISC [61]. The primary effector protein of miRISC that directly binds both the miRNA and the mRNA is called Argonaute (Ago) [69-71]. In mammalian cells, there are four different types of Ago proteins: Ago1, Ago2, Ago3, and Ago4. Over 40% of RISCs contain Ago2, the only Ago that can perform endonucleolytic cleavage of target mRNA, and this is commonly referred to as the ‘slicer activity’ of Ago2 [72].

**MicroRNA Targeting and Mechanisms of Gene Silencing**

Once within the RISC, the miRNA base-pairs with the target mRNA specifically at nucleotides 2-8 from the 5’ end [73-76]. This 7-nt long region is called the ‘seed sequence’ of the miRNA and is a key determinant of miRNA targeting. However, 6-nt and 8-nt seed sequences are not uncommon [77]. mRNA targets of miRNA which require base-pairing at the seed sequence are referred to as the canonical targets of that miRNA. There also exist non-canonical targets that bypass the seed sequence base-pairing with the help of 3’ compensatory base-pairing [78]. Current *in silico* target-identification programs such as TargetScan and PicTar use a prediction algorithm based on identifying sequences complementary to miRNA seed sequences in the 3’ untranslated regions (3’UTRs) of mRNAs [79, 80]. These complementary sequences are referred to as ‘seed matches’ for the rest of this thesis. The emphasis for identifying seed
matches in 3’UTRs is due to the fact that several initial studies found functional miRNA binding sites in the 3’UTRs of target mRNAs [49, 81-83]. Moreover, independent studies discovered that miRNA binding sites in the coding sequence (CDS) are functional but not as effective as those found in 3’UTRs [84, 85]. However, a recent study showed that miRNA binding sites in CDS are more effective in translational inhibition, while those in 3’UTRs are effective in accelerating mRNA degradation, thus concluding that they contribute differentially to the two key events in miRNA-mediated repression [86].

Genome-wide studies that measure the effect of miRNA on mRNA and protein expression, along with ribosome-profiling experiments, suggest that mRNA degradation is the primary effect and accounts for 66-90% of the observed repression [87-89]. To bring about mRNA degradation, Ago of the miRISC interacts with the GW182 protein, which in turn interacts with Poly A Binding Protein Complex (PABPC) and polyadenylases such as CCR4-NOT and PAN2-PAN3 complexes which deadenylate the target mRNAs [90, 91]. Deadenylated mRNAs are then decapped by DCP1 and DCP2 enzymes in the cell, which is followed by 5’ to 3’ exonucleolytic degradation by XRN1 exonucleases [89].

In comparison with mRNA degradation, pure translational inhibition (excluding the inhibition which is a consequence of mRNA degradation) has been shown to contribute to only about 6-26% of the observed repression [88, 89]. Work from several labs has identified distinct yet non-mutually exclusive mechanisms of how this inhibition proceeds, and at present the exact sequence of molecular events remains debatable [92]. Broadly, the proposed mechanisms support either inhibition of translation initiation [93-96] or inhibition of elongation including release of the peptide molecule [97-99].
**KSHV MicroRNA Targets**

KSHV encoded miRNAs are transcribed by RNA polymerase II and are processed via the same pathway detailed above [21]. Concerted efforts from several labs to identify mRNA targets of KSHV miRNAs have led to identification of several targets such as BACH1 [100], THBS1 [101], BCLAF1 [102], MICB [103], SMAD5 [104], MAF [105], TWEAKR [106], and C/EBPβ [107]. KSHV miRNA targets play important roles in cellular pathways such as cell cycle arrest, apoptosis, and innate immune responses [108]. For example, BCLAF1, encoding Bcl2-associated factor (a pro-apoptotic factor), is targeted by miR-K12-5, miR-K12-9 and miR-K12-10b [102]. BACH-1 is a transcription factor which co-regulates transcription along with MAF (musculoaponeurotic fibrosarcoma oncogene homolog) proteins [109]. MAF plays an important role in tissue specification and terminal differentiation of Lymphatic Endothelial Cells (LECs) [110]. While BACH-1 is targeted by miR-K12-11 [100], MAF is targeted by both miR-K12-11 and miR-K12-6 [105]. Such redundant targeting of different proteins of the same cellular pathways by KSHV has been reported for several pathways. For instance, KSHV miRNAs target multiple TGF-β signaling pathway molecules. THBS1 (thrombospondin 1) is a tumor suppressor protein and anti-angiogenic factor which acts by activating TGF-β signaling [111, 112]. KSHV miRNAs miR-K12-1, miR-K12-3-3p, miR-K12-6-3p, and miR-K12-11 target THBS1 and thus decrease the expression of TGF-β pathway genes such as SBE-4 and MMP-9 [101]. SMAD5, a key mediator of TGF-β signaling, was shown to be downregulated by miR-K12-11 in a TGF-β-sensitive cell line that also resulted in increased cell proliferation upon TGF-β treatment [104]. Moreover, a variant of miR-K12-10 referred to as miR-K10a_+1_5 downregulates TGF-β type II receptor (TβRII) in telomerase-immortalized human dermal microvascular endothelial (TIME) cells [113]. In fact, latency proteins also contribute to such cooperative targeting of cellular
pathways. For instance, vCyclin and vFLIP upregulate a cellular miRNA cluster called the miR-17/92 cluster in endothelial cells, which downregulates SMAD2, yet another important mediator of TGF-β signaling [114]. In addition to specific targeting, Haecker et al. also suggested that the high expression levels of viral miRNAs could lead to global derepression of cellular miRNA targets by increasing the competition for available RISCs in the cell [115].

**Ribonomics Techniques to Identify MicroRNA Targets**

Identifying the targets of KSHV miRNAs is crucial to understanding how the virus manipulates the host cell functions to its advantage. Prediction strategies and biochemical assays developed to identify cellular miRNA targets also apply to viral miRNA targets. As mentioned previously, most prediction algorithms rely on seed sequence dependent base-pairing and focus on identifying seed matches in the 3’ UTR of mRNAs [77]. However, biochemical experiments that assay the Ago-bound RNA population allow for identification of all targets without any bias based on the location of the binding site within the mRNA sequence. Ribonomics refers to high-throughput investigation of RNA bound RNA-binding proteins (RBPs) [116]. Ribonomics experiments of Ago proteins allow for analysis of all cellular miRNA targets simultaneously. Some of the ribonomics methods employed to study KSHV miRNA targets are described below and the graphical representation for three of them is provided in Figure 1-3.

**RNA Immunoprecipitation followed by microarray (RIP-Chip).** Ago RIP-Chip involves immunoprecipitation of Ago proteins from cell lysates followed by reverse transcription of Ago-bound RNAs and a microarray analysis. Doelken et al. performed Ago2 RIP-Chip analysis in BCBL-1 cells (a PEL cell line) to identify the targets of KSHV miRNAs [117]. Their study also included a similar analysis to identify the targets of Epstein-Barr virus (EBV) miRNA targets. They identified 114 KSHV miRNA targets and 44 EBV miRNA targets, and
independently validated some of the targets such as LRRC8D, GEMIN8 and NHP2L1. However, the number of targets identified was far fewer than expected based on the reasoning that every miRNA targets hundreds of mRNAs. The primary drawback of this approach is the potential for false binding caused by lysis. Riley and Steitz describe this as the ‘observer effect’ in ribonomics [118]. Advanced ribonomics methods included a cross-linking step to overcome this observer effect. Cross-linking covalently binds the RNA and the protein which avoids artefactual RNA-Ago interactions and prevents the loss of legitimate targets during washing steps.

**High Throughput Sequencing – Cross-linking Immunoprecipitation (HITS-CLIP).**

HITS-CLIP was developed by the Darnell laboratory in 2008 [119, 120]. Briefly, HITS-CLIP involves cross-linking cells, and hence RNA-protein complexes, using UV radiation with a wavelength of 254 nm. UV radiation, unlike formaldehyde or glutaraldehyde based cross-linking, does not create covalent bonds amongst proteins, thus ensuring targeted cross-linking of RNA only to the directly bound protein partner. The cross-linked RBPs are then isolated from lysed cells through immunoprecipitation, and the ternary complexes pulled down are treated with RNase to trim the mRNAs extending outside the complex. The complexes are processed through SDS-PAGE and western blotting, and the Ago-bound RNA is extracted from the nitrocellulose membrane. This radiolabeled RNA is then reverse transcribed and sequenced. Ago HITS-CLIP allows global identification of all Ago-bound miRNAs and mRNAs, however, it does not provide information on which mRNA was bound by which miRNA. This relationship is reconstructed using bioinformatics, by screening the Ago footprints on the mRNA sequence for specific miRNA seed matches. Haecker et al. performed Ago HITS-CLIP on KSHV positive PEL cell lines, BCBL-1 and BC-3, and identified 552 and 413 high stringency (3/3 bioreplicates) targets in BCBL-1 and BC-3 cells, respectively, and 1170 and 950 medium stringency
(2/3 bio-replicates) targets in BCBL-1 and BC-3 cells, respectively [115]. Their study confirmed several previously published targets such as BACH1, FOS, SLA and C/EBPβ, and identified hundreds of novel targets for KSHV miRNAs, some of which, such as ANXA2, C/EBPα, HLA-C, HMGA1, IRF2BP2, TP53INP1, TPD52, and YWHAE were validated by luciferase assays. A similar HITS-CLIP analysis of EBV infected Jijoye cells was performed by Riley et al. and they identified that EBV miRNAs target 161 3’ UTRs alone and 1503 3’ UTRs along with cellular miRNAs [121].

**Photoactivatable Ribonucleoside Enhanced Cross-linking Immunoprecipitation (PAR-CLIP).** PAR-CLIP was adapted from HITS-CLIP to increase the cross-linking efficiency and was developed in the Tuschl laboratory in 2010 [122]. For PAR-CLIP, cells are grown in the presence of a non-toxic photoactivatable ribonucleoside which the cells readily take-up. The ribonucleoside most commonly used for PAR-CLIP is 4-thiouridine (4SU) and occasionally 6-thioguanine (6SG). UV radiation is used at a wavelength of 365 nm to cross-link the RNA molecules, which now contain the variant ribonucleosides, and this increases the RNA recovery by 100- to 1000-fold. Further, during reverse transcription C is incorporated opposite 4SU instead of a T, and this T to C conversion in the sequencing reads allows for high confidence identification of cross-linking sites. Gottwein et al. performed PAR-CLIP analysis of two PEL cell lines, BC-3 (KSHV+), and BC-1 (KSHV+ and EBV+) [123]. They identified 1741 mRNA targets for KSHV miRNAs in BC-1 cells and 1409 targets in BC-3 cells, and also confirmed published targets such as BACH1, FOS, CDKN1A (p21), TWEAKR, RAD21 and RBL2 mRNAs. PAR-CLIP analysis of EBV infected Lymphoblastoid cell lines (LCL) identified close to 630 targets for EBV miRNAs [124].
HITS-CLIP and PAR-CLIP of KSHV infected PEL cells has led to the identification of several novel miRNA targets [125]. However, the overlap between the target RNAs identified by these techniques is only 42%, suggesting that there exist technique-dependent biases in identification [115]. Moreover, as noted by Gottwein et al., sequencing depth limitations could have contributed to incomplete identification of targets by one individual method [123]. Additionally, the western blotting, followed by extraction of radiolabeled RNA from the membrane, is an extremely inefficient process resulting in loss of legitimate targets. Although HITS-CLIP and PAR-CLIP allow for simultaneous identification of all miRNA targets in the cell, the major disadvantage of both of these techniques is their dependence on bioinformatics and seed sequence based interactions to identify the miRNA that targets a given mRNA.

**Cross-linking, ligation, and sequencing of hybrids (CLASH).** CLASH was developed in the Tollervey laboratory in 2011 [126]. CLASH is built upon HITS-CLIP to identify specific RNA-RNA interactions in the context of a particular RBP. It includes an RNA ligation step (Figure 1-3), which allows RNA molecules in close proximity to ligate and form hybrid molecules. The immunoprecipitation and processing steps are otherwise similar to HITS-CLIP analysis. Identification of these hybrid molecules in the sequencing data requires a specialized bioinformatics pipeline called Hyb [127]. The key advantage is that every hybrid identified provides strong evidence of proximity. Thus in Ago-CLASH, all hybrid miRNA-mRNAs identified provide high-confidence targets. However, the RNA ligation step is extremely inefficient, rendering the false negative discovery rates high. In other words, many legitimate targets could be missed using this approach.

Recently, our lab performed Ago-CLASH analysis on KSHV infected endothelial cells after making significant changes to the protocol to minimize procedural losses. Some of the
results from this study are briefly described in Chapter 3 and the majority of this work will be reported in detail elsewhere (unpublished work).

**Limitations of ribonomics methods.** CLIP-based methods such as HITS-CLIP, PAR-CLIP and CLASH have revolutionized the study of protein-RNA complexes. Nevertheless, as with other biochemical techniques, these also suffer from certain forms of bias. UV crosslinking is more efficient between certain amino acids and nucleotides than others and this could preferentially enrich targets that undergo efficient crosslinking [128]. Specifically in case of PAR-CLIP, since crosslinking happens only at 4SU positions, there could be a bias towards U-rich targets [122]. It has been shown that the nucleotide composition of RNA adaptors ligated to the miRNAs for PCR and high-throughput sequencing could influence the representation of miRNAs in small RNA libraries prepared from HITS-CLIP and PAR-CLIP experiments [129]. RNA ligase is an enzyme with low efficiency and this could be a factor that contributes to significant bias in CLASH experiments [130]. Moreover, Kishore et al. demonstrated that the RNases used for trimming, RNase A and RNase T1, significantly influence the distribution of target RNA fragment sizes obtained [131]. Hence, it is often recommended to perform these experiments in multiple biological replicates to overcome these biases, and reduce any background from the extensive procedure of immunoprecipitation [118]. In spite of how useful these techniques are in identifying putative targets of miRNAs, it is important to remember that they are often not comprehensive because the above mentioned biases.

**Long Noncoding RNAs (IncRNAs)**

An integrated world-wide effort to catalog all the functional elements of the human genome was undertaken in the year 2003 and was called the Encyclopedia of DNA elements (ENCODE) project [132]. Under the ENCODE project, a sub-project called GENCODE was
launched to characterize specifically the protein-coding genes of the human genome [133].

However, with progress of these and similar such studies, it was found that although over 90% of the human genome is transcribed, only about 2-3% codes for protein coding genes [134-136]. The non-protein coding transcriptome is collectively referred to as noncoding RNAs (ncRNAs). A significant subset of the ncRNAs that make up the transcriptome is comprised of previously known small RNAs such as the ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small-nucleolar RNAs (snoRNAs) and small-nuclear RNAs (snRNAs) which contribute to cellular functions such as protein translation and mRNA splicing. Other small noncoding RNAs such as miRNAs and piRNAs perform regulatory functions in the cells. Typically, all these above mentioned RNAs with well-understood functional roles are less than 200 nt in length [137]. The novel class of ncRNAs that became known through large-scale projects aimed at characterizing the entire transcriptome are ncRNAs longer than 200 nt in length with no well-defined functions. All ncRNAs longer than 200 nt were classified as ‘long noncoding RNAs’. By virtue of this length-based definition, IncRNAs is a very heterogeneous class of RNAs [138, 139]. Cech and Steitz present a detailed review of how ncRNAs have revolutionized our understanding of a variety of biological functions [137].

The study of IncRNAs as a class has seen exponential growth in the last decade; however, certain IncRNAs were reported in the literature long before the term ‘IncRNA’ was coined [139]. For example, H19 is a lncRNA which was first described in 1990 to function as a noncoding RNA since it lacked a detectable protein coding sequence [140]. XIST is another lncRNA that was functionally described as early as 1997 [141]. The recent version of GENCODE (Release 25) reports a total of 23898 RNAs in the lncRNA category. Cabili et al. cataloged more than 8000 RNAs to a specific sub-class of lncRNAs called long intergenic noncoding RNAs.
(lincRNAs) which do not overlap with any protein coding genes [142]. This study also noted the tissue-specific expression of several lincRNAs, and that lincRNAs are often co-expressed with their neighboring genes. In a recent study, Iyer et al. interrogated 7256 RNA-seq libraries and cataloged 58648 lncRNA genes [143]. Many other independent studies have also cataloged lncRNAs through high-throughput sequencing or tiling array analysis [144-146]. Some of the publicly available databases that comprehensively report most of the lncRNAs known to date include RefSeq, Ensembl, GENCODE, UCSC Known genes dataset, LncRNAdb, and NRED.

Many studies have reported the large number of transcribed lncRNAs, and yet their functional characterization has been a rather slow process and often low-throughput [147]. Four non-mutually exclusive themes of lncRNA function are commonly discussed based on our current understanding of lncRNAs:

**Signaling.** LncRNAs can have specific expression patterns based on cell type, developmental stage or differentiation stage or even the allele of origin [142]. Spatial and temporal expression of lncRNAs associated with the Homeobox (HOX) transcription factor genes during development is a good example of the signaling function of lncRNAs. HOTAIR and Frigidair are two lncRNAs expressed from the HOX C cluster. While HOTAIR expression signals distal and posterior positional identities, Frigidair expression is associated with anterior positions [148, 149]. Kcnq1ot1 and Air are two imprinted lncRNAs, which are transcribed from the paternal allele, and mediate repressive histone modification of their target promoters thus facilitating allele specific signaling [150, 151].

**Decoys.** Some lncRNAs can act as molecular sinks for RNA-binding proteins (RBP) and this binding can be specific or nonspecific. In both cases, the lncRNA titrates away the RBP thus preventing the RBP from carrying out its original function. Gas5 is an example of a decoy that
titrates away glucocorticoid receptors (GR) in growth-arrested cells, and it co-migrates to the nucleus with GRs. Gas5 achieves this by forming a stem-loop structure that mimics the hormone-response elements (HREs) of glucocorticoid-responsive genes [152]. PANDA, a p53-dependent IncRNA, is another example of a decoy that promotes cell cycle arrest while delaying apoptosis following DNA damage. PANDA sequesters NF-YA, which is a key transcription factor required for DNA damage dependent initiation of apoptosis [153]. The name ‘decoy’ is used to refer to sequestration of RBPs, while a very similar sequestration of miRNAs has been termed ‘sponging’ [154]. Several pseudogenes and circular RNAs (circRNAs) have been shown to act as miRNA sponges.

**Scaffold.** Scaffolding function, unlike other functions of IncRNAs, relies heavily on a well-defined 3D structure. Traditionally, this function has been associated with proteins for the same reason. Recent studies have uncovered IncRNAs such as TERC and ANRIL that can serve as scaffolds. TERC is the telomerase-associated IncRNA that helps in assembling the telomerase and serves as the RNA template for reverse transcription and addition of repeats to chromosome ends [155]. ANRIL in transcribed from the INK4 locus in a direction antisense to the tumor suppressor genes p14 and p15. ANRIL recruits PRC2 and PRC1 complexes by acting as a scaffold and silences the INK4b/ARF/INK4a locus [156, 157]. Disruption of this ANRIL-dependent scaffolding inhibits the silencing of this locus [156].

**Guides.** LncRNAs can draw in and recruit proteins to specific regulatory domains on the chromosome like a promoter, and thus act as a guide. This regulation can happen in *cis*, where the guiding function of the lncRNA is localized at the site of its transcription, or in *trans*, where the lncRNA functions away from where it is transcribed [139]. Recruitment of PcG proteins by XIST to the XIC (X inactivation center) is an example of a guiding function in *cis* [158]. Other
PRC2 binding lncRNAs such as Jpx, lincRNA-p21 and HOTAIR have been suggested to act in \textit{trans} [159].

The list of potential lncRNA functions continues to grow as more of these candidates are identified and characterized. Interestingly, recent studies have indicated that some of the lncRNAs may encode short proteins called micropeptides [160-162]. However, the coding ability that a subset of lncRNAs might possess does not preclude a noncoding RNA-based function for the same molecule.

\textbf{Diversity of LncRNA Localization and Function}

The majority of lncRNAs are pol II transcripts, and most of them are spliced, capped and polyadenylated [149]. Just like proteins, the function of any lncRNA is dependent on its subcellular localization. While some lncRNAs like MALAT1 and ANRIL have a distinct nuclear localization, many lncRNAs have a range of subcellular distributions, as was shown recently by \textit{in situ} hybridization [163]. LncRNAs that are localized in the nucleus can be broadly viewed as two kinds: lncRNAs that localize in \textit{cis} and lncRNAs that localize in \textit{trans} [164]. ANRIL, XIST, FIRRE are some examples of lncRNAs that localize in \textit{cis} [159, 165]. These function by recruiting regulatory proteins, often histone modifying enzymes or chromatin remodelers, to their sites of transcription. HOTAIR works \textit{in trans} by recruiting PRC complexes to several different loci in cancer cells [159, 164]. NEAT1, which interacts with paraspeckle-localized proteins, is another example of a lncRNA localized in \textit{trans} [166]. Some examples of lncRNAs that function in the cytoplasm include linc-p21, NORAD, and CDR1-AS [167-169]. However, identification of lncRNAs strictly localized to the cytoplasm can be challenging because of the ambiguity involved in distinguishing freshly transcribed lncRNAs in the nucleus from mature functional lncRNAs in the cytoplasm. Furthermore, some lncRNAs like Gas5 translocate
between the cytoplasm and the nucleus in response to a stimulus such as starvation [152]. The factors that govern nuclear retention of lncRNAs are unclear. Dynamic assembly of RBPs on lncRNA during lncRNA transcription is one of the many proposed theories for how nuclear lncRNAs are retained [164].

**Long Noncoding RNAs in Cancer**

Cancer is a disease of the genome, i.e., genetic alterations disrupt gene expression and regulatory networks, which leads to loss of cellular homeostasis. Deregulation of the expression levels of certain lncRNAs can cause or contribute to cancer phenotypes [170]. Based on studies involving several cancer types, various lncRNAs were identified as oncogenes or tumor-suppressors and some others were downstream targets of oncogenic or tumor-suppressive proteins [170]. The first lncRNA to be associated with cancer was PCA3, which is aberrantly expressed in prostate cancer and is currently used as a biomarker for prostate cancer [171, 172]. MALAT1, an intergenic lncRNA, was identified as a prognostic marker for lung adenocarcinomas and has since been found to be upregulated in several cancers including breast, colon and liver cancers [173, 174]. MALAT1 plays an important role in alternative splicing and binds active chromatin sites [175, 176]. ANRIL is transcribed antisense to the INK4 locus and it epigenetically silences the INK4b/ARF/INK4a tumor-suppressor locus by recruiting PRC complexes [156, 157, 177]. Breast, ovarian, and a multitude of other cancers have upregulated expression of ANRIL compared to normal tissue [178]. Tumor-suppressor lncRNAs are usually silenced by hyper-methylation or are deleted in cancer cells. Based on such evidence lncRNAs like MEG3 and DLEU2 were identified to have cancer-restrictive properties [179-181]. MEG3 activates p53-dependent transcription of genes involved in cell-cycle arrest and over-expression of MEG3 drives cells into senescence and apoptosis [182].
New developments in microarray technology to include probes for IncRNAs and extensive RNA sequencing of patient-derived tumor samples have helped establish that most tumor tissues have an aberrant IncRNA profile [170]. Validation and functional characterization of the roles of deregulated IncRNAs require loss-of-function and gain-of-function experiments. Both transient approaches, such as RNAi to knockdown IncRNAs, and long-term approaches, such as CRISPR/Cas9 based genome editing, are used to investigate the functions of these IncRNAs [147, 183-186].

**Long Noncoding RNAs in Viral Infections**

Aside from cancer, IncRNA deregulation occurs in several diseases including neurological and developmental disorders and infectious diseases [187, 188]. HULC is a IncRNA whose expression is restricted to liver tissues, and is strongly upregulated during hepatocellular carcinoma [189]. High levels of HULC drive cell proliferation and promote hepatoma by inhibiting p18, which is a tumor suppressor [190]. Evidence suggests that hepatitis B virus (HBV) protein HBx is responsible for upregulation of HULC in HCCs caused by HBV [190]. Another IncRNA called IncRNA-HEIH is upregulated in other hepatocellular carcinomas caused by HBV [191]. IncRNA-HEIH is involved in G0/G1 transition and is associated with the EZH2 protein of PRC2 complexes, thus influencing the transcription profile of the cell. During HIV infection, the IncRNA 7SL plays an important role in viral packaging into virions [192]. Interestingly, 7SL also functions as a cofactor for innate antiviral cytidine deaminases such as APOBEC3G and APOBEC3F [193]. HIV infection also alters the expression level of NEAT1, a paraspeckle-associated IncRNA [194]. Knockdown of NEAT1 allows for enhanced virion export by increasing Rev-dependent mRNA export from the nucleus. NEAT1 expression is also upregulated in mouse CNS during Rabies or Japanese encephalitis virus infections [195].
Oncogenic viruses such as EBV, HCV, and Reticuloendotheliosis virus strain T (REV-T) upregulate BIC, a lncRNA that serves as a precursor for the oncogenic miRNA miR-155 [196-198]. Involvement of host IncRNAs in viral biology also extends to RNA virus infections such as Influenza and SARS-CoV. VIN, a lincRNA, was induced strongly during infection with different strains of influenza and its expression was correlated with viral progeny production [199]. Another study performed a high-throughput screen for IncRNAs with altered expression during either influenza or SARS-CoV infection in mice [200]. They reported that while some IncRNAs are altered in an infection-specific manner, a significant subset was induced upon any viral infection, suggesting they play a role in host cell response to viral infection. The involvement of IncRNAs in innate immune response and inflammation has been reviewed in detail [201]. Host IncRNAs whose expression is altered during viral infection have been cataloged for different viruses such as Influenza, SARS-CoV, Zika and HBV [200, 202-204].

In addition to promoting differential expression of host IncRNAs, viruses encode their own IncRNAs [205]. For example, KSHV encodes the PAN RNA, which is highly upregulated during lytic replication of KSHV [206]. EBV encodes IncRNAs called EBERs and herpesvirus salimiri (HVS) encodes HSURs [205]. Detailed annotation of the EBV genome using different RNA-seq techniques identified pervasive transcription of the viral genome suggesting that more viral IncRNAs exist than were previously described [207, 208]. Herpes simplex virus 1 (HSV1) also encodes a noncoding RNA called LAT, which is formed as a stable intron lariat during latency [209, 210]. There are more examples of viral IncRNAs which extend outside the herpesviridae family [205]. However, the work presented in this thesis focuses exclusively on understanding the regulation of host IncRNAs in the context of γ-herpesviral infection.
Regulatory Networks Involving MicroRNAs and LncRNAs

miRNAs and lncRNAs have emerged as two important classes of regulatory RNAs in the last two decades, and numerous lines of evidence support their importance in the regulation of gene expression, especially in a broad spectrum of cancers [211]. Many of the precursor RNAs from which miRNAs are processed are lncRNAs by definition. Some RNAs such as BIC, H19, DLEU2 and Rncr3 identified as lncRNAs have been shown to be miRNA precursors [212-215]. The inter-regulation of these two RNA families is a novel area of research. LncRNAs can inhibit miRNAs from silencing their target mRNAs by sequestering (‘sponging’) them [154]. Various classes of lncRNAs such as pseudogenes, ceRNAs and circRNAs act as effective miRNA sponges [154]. In fact, HVS encoded viral lncRNAs HSUR1 and HSUR2 sponge miRNA-27 in T-lymphocytes, suggesting that such interactions are at least as old as virus-host interactions [216]. Surprisingly, there are very few reported examples demonstrating the regulation of lncRNAs by miRNAs. The few examples thus far include downregulation of MALAT1 by miR-9 in the nucleus of Hodgkin’s lymphoma and glioblastoma cells; and repression of UCA1 by miR-1 in bladder cancer [217, 218].

In this dissertation work, the two primary goals were: (1) To catalog human lncRNAs deregulated by latent KSHV infection in a KS model system, and (2) To investigate whether KSHV encoded miRNAs play a role in deregulation of human lncRNAs.

Accordingly, in Chapter 2 we describe our use of microarray analysis to identify and catalog lncRNAs deregulated by KSHV. We also present genetic and biochemical evidence for regulation of host lncRNAs by KSHV miRNAs. We extended our knowledge from the data gathered, to present some preliminary evidence on how latency associated proteins of KSHV also influence the lncRNA expression of infected endothelial cells. In addition, we investigated
publicly available Ago HITS-CLIP and Ago PAR-CLIP datasets for KSHV and EBV infected lymphoma cells. The results from this bioinformatics-based study are reported in Chapter 3. While the data from lymphoma cells were obtained by HITS-CLIP or PAR-CLIP, our lab also recently generated a CLASH dataset from an endothelial cell line representative of KS. The IncRNA targets identified from this dataset are described in Chapter 3.
Figure 1-1: An outline of the KSHV genome. KSHV latency-associated region is indicated by a blue line.
Figure 1-2: An outline of the miRNA biogenesis pathway [219].
Figure 1-3: Schematic outline of HITS-CLIP, PAR-CLIP and CLASH ribonome protocols. The steps are shown from UV irradiation of cells through to sequencing library construction, with the differences between HITS-CLIP, PAR-CLIP and CLASH indicated. HITS-CLIP and PAR-CLIP also involve isolation and sequencing of Ago-miRNA complexes; this has been omitted in the outline for the sake of clarity.
CHAPTER 2
MICRORNA DEPENDENT AND INDEPENDENT DEREGLULATION OF LONG NONCODING RNAS BY AN ONCOGENIC HERPESVIRUS

Summary

Kaposi’s sarcoma (KS) is a highly prevalent cancer in AIDS patients, especially in sub-Saharan Africa. Kaposi’s sarcoma-associated herpesvirus (KSHV) is the etiological agent of KS and other cancers like primary effusion lymphoma (PEL). In KS and PEL, all tumors harbor latent KSHV episomes and express latency-associated viral proteins and microRNAs (miRNAs). The exact molecular mechanisms by which latent KSHV drives tumorigenesis are not completely understood. Recent developments have highlighted the importance of aberrant long noncoding RNA (lncRNA) expression in cancer. Deregulation of lncRNAs by miRNAs is a newly described phenomenon. We hypothesized that KSHV-encoded miRNAs deregulate human lncRNAs to drive tumorigenesis. We performed lncRNA expression profiling of endothelial cells infected with wild type (wt) and miRNA-deleted KSHV and identified 126 lncRNAs as putative viral miRNA targets. Here we show that KSHV deregulates host lncRNAs in both a miRNA-dependent fashion by direct interaction and in a miRNA-independent fashion through latency-associated proteins. Several lncRNAs that were previously implicated in cancer, including MEG3, ANRIL and UCA1, are deregulated by KSHV. Our results also demonstrate that KSHV-mediated UCA1 deregulation contributes to increased proliferation and migration of endothelial cells.

Significance

KS is the most prevalent cancer associated with AIDS in sub-Saharan Africa, and is also common in males not affected by AIDS. KSHV manipulates human cells by targeting protein-
coding genes and cell signaling. Here we show that KSHV alters the expression of hundreds of human lncRNAs, a broad class of regulatory molecules involved in a variety of cellular pathways including cell cycle and apoptosis. KSHV uses both latency proteins and miRNAs to target lncRNAs. miRNA-mediated targeting of lncRNAs is a novel regulatory mechanism of gene expression. Given that most herpesviruses encode miRNAs, this mechanism might be a common theme during herpesvirus infections. Understanding lncRNA deregulation by KSHV will help decipher the important molecular mechanisms underlying viral pathogenesis and tumorigenesis.

Introduction

Kaposi’s sarcoma-associated herpesvirus (KSHV) is an opportunistic human oncovirus, which causes Kaposi’s sarcoma (KS), primary effusion lymphoma (PEL) and multicentric Castleman’s disease (MCD) in immunocompromised individuals, primarily AIDS patients and organ-transplant recipients [220]. KSHV uses the lytic mode of replication for spread of infection, and latency for persistence in the host. All tumor cells isolated from KS patients test positive for latent viral episomes [220]. Latent KSHV expresses only 10% of its 140-kb dsDNA genome, encoding primarily four latency proteins (Kaposin, vFLIP, vCyclin and LANA) and 25 mature miRNAs [220]. miRNAs are 21-23 nt long noncoding RNAs that recognize target mRNAs using 7 bp ‘seed sequences’ and silence them [77]. To identify the means by which KSHV causes tumors, KSHV latency proteins and miRNAs have been studied extensively [220]. Ribonomics approaches to identify targets of KSHV miRNAs have focused exclusively on mRNAs [115, 123].

Recently, lncRNAs have emerged as important regulatory molecules in cancer [221]. LncRNAs play a variety of regulatory roles in both the cytoplasm and nucleus [138, 139]. This group includes all RNA molecules longer than 200 nt with no apparent coding potential, and they
have diverse functions ranging from acting as a scaffold, sponge/decoy or guide aiding in cell-signaling [138, 159]. Owing to their diversity, over 95% of the lncRNAs remain uncharacterized. Disease association is a starting point for identifying and characterizing lncRNAs with important regulatory roles. Using this approach with different cancer types, oncogenic lncRNAs such as MALAT1, ANRIL, UCA1, and tumor suppressor lncRNAs like Gas5 and MEG3 have been functionally characterized [221]. Another important group of disease-relevant lncRNAs includes those involved in the innate immune response following viral or bacterial infections [201]. A few studies have addressed the roles of host lncRNAs during viral infections, for example HULC (Hepatitis-B) and NRON (HIV) [188]. However, the question of whether viruses manipulate specific host lncRNAs to their advantage remains largely unexplored. Understanding deregulation of specific host lncRNAs, especially cancer-related lncRNAs by persistent oncoviruses, such as the γ-herpesviruses, would shed light on how these viruses drive oncogenesis.

Regulatory cross-talk is known to occur between miRNAs and lncRNAs, at multiple levels. LncRNAs like BIC1 and H19 act as precursors for miRNAs[213, 222] and lncRNAs such as HULC and CDR1-AS act as sponges for miRNAs[223, 224]. Conversely, human miRNA miR-9 represses the expression levels of the lncRNA MALAT1[217]. Work from the Steitz laboratory demonstrated that the viral lncRNAs HSUR1 and HSUR2, encoded by Herpesvirus Saimiri, act as sponges for cellular miR-16, miR-142-3p and miR-27 and thereby silence some of these miRNAs in T-lymphocytes, suggesting that γ-herpesviruses can utilize virus lncRNAs to target host miRNAs[225]. Conversely, whether herpesvirus miRNAs can target and downregulate host lncRNAs remains an open question.
In this study, we demonstrate that latent KSHV infection of endothelial cells alters the host lncRNA profile. We provide evidence that KSHV deregulates hundreds of host lncRNAs including many cancer-associated lncRNAs such as UCA1, ANRIL and MEG3 in both a miRNA dependent and independent manner. Furthermore, KSHV appears to manipulate the host lncRNAs to favor proliferation and migration of latently infected endothelial cells.

Results

KSHV Deregulates Host LncRNAs

Previously, we identified the mRNA targetome of viral miRNAs in PEL cells by High Throughput Sequencing-Cross-linking Immunoprecipitation (HITS-CLIP) analysis of the Ago protein[115]. The PEL cell lines we studied were BC-3 and BCBL-1, which are KSHV positive B cell lines. We reanalyzed the HITS-CLIP data for enriched lncRNAs and compared our results with a similar reinvestigation of Ago PAR-CLIP data from lymphoblastoid cell lines infected with Epstein-Barr virus (EBV) [124], a related γ-herpesvirus that causes cancer. We found that approximately 357 and 750 lncRNAs were a part of the KSHV and EBV miRNA targetome, respectively, and 64 lncRNAs were potentially targeted by miRNAs from both viruses.

We aimed to determine the effect of latent KSHV infection on the lncRNA expression profile of endothelial cells and specifically question whether KSHV encoded miRNAs targeted endothelial lncRNAs. To address these questions, we used Telomerase Immortalized Vein Endothelial (TIVE) cells, an in vitro model system to study KS [226]. We performed lncRNA expression profiling on latently infected TIVE cells harboring either the wt-KSHV or Δcluster-KSHV[227, 228], in which a region containing 10 of the 12 miRNA genes is deleted, and used the lncRNA profile of mock-infected TIVE cells as reference. The KSHV latency-associated
region of the wt and mutant bacmid backbones used for this experiment is shown in Figure 2-1A. The profiling analysis revealed that wt-KSHV and Δcluster-KSHV infections deregulate 858 and 2372 host lncRNAs, respectively (Table 2-1), indicating that latent KSHV infection globally affects lncRNA expression. The higher count of deregulated lncRNAs in Δcluster-KSHV infection is likely due to increased spontaneous reactivation rate in the absence of viral miRNAs [228, 229]. We grouped the deregulated lncRNAs into three categories based on a cut-off of fold change ≥ 2.0: upregulated, downregulated and rescued. We defined rescued genes as those that were downregulated in wt-KSHV-infected cells compared to mock, and were upregulated in Δcluster-KSHV-infected cells compared to wt-infected cells. We identified 126 candidates in the rescued category, which are putative direct targets of viral miRNAs (Figure 2-1B).

**Mature Viral MicroRNAs and Ago2 are Present in the Nuclei of KSHV-Infected Cells**

Based on lncRNA localization data from HUVEC cells [230], 9 of the 126 putative lncRNA targets of viral miRNAs we identified are exclusively nuclear localized, and 32 of them are partially nuclear localized. It is important to note that the localization information was available for only 72 out of the 126 rescued lncRNAs. Similarly, the 357 lncRNAs identified from Ago HITS-CLIP of PEL cells include nuclear resident lncRNAs such as ANRIL (CDKN2B-AS1) and MALAT1. Moreover, several of the uncharacterized candidates of these 357 lncRNAs may be nuclear localized. miRNA-mediated regulation of nuclear localized lncRNAs seemed paradoxical at the outset, as mature miRNAs and RISCs including the Ago family proteins are believed to reside and function in the cytoplasm. Recently, several groups showed that Ago2 complexes can be present in the nuclei of different cell types [231, 232]. Moreover, studies in Hodgkin’s lymphoma lines identified that several lncRNAs co-isolate with
Ago protein [233]. To determine whether KSHV miRNAs could regulate nuclear IncRNAs, we investigated the nuclear/cytoplasmic distribution of viral miRNAs and Ago2 in PEL cells.

We fractionated BCBL-1 cells into nucleus and cytoplasm and analyzed the distribution of KSHV miRNAs using stem-loop qRT-PCR, which amplifies mature miRNAs but not their precursors. Mature KSHV miRNAs were found in both the cytoplasmic and nuclear fraction (Figure 2-2A). We probed the fractions for Ago2 using western blotting (Figure 2-2B). Calnexin, an ER resident, was used as a control to ensure that the nuclear preparations were free of endoplasmic reticulum. A significant fraction of Ago2 was localized in the nucleus of BCBL-1 cells. These results were confirmed using immunofluorescence analysis (IFA) of Ago2 in isolated BCBL-1 nuclei by confocal microscopy and 3D-reconstruction. The images in Figure 2-2C show Ago2 in all planes of view (XY, YZ and ZX) with and without DAPI, and it is evident that Ago2 is present inside the BCBL-1 nuclei. We observed similar results with IFA performed on KSHV-infected TIVE cells. Thus, we concluded that Ago2 and viral miRNAs are present in the nuclei of infected cells, and miRNAs could potentially interact via Ago2 with nuclear IncRNAs.

**KSHV MicroRNAs Directly Target Host IncRNAs**

Of the 126 rescued IncRNAs identified based on transcriptional profiling, 98 contained seed sequence matches for at least one KSHV miRNA. Repeated sampling of 126 sequences from randomly generated DNA sequences, controlling for IncRNA length, revealed that the presence of KSHV miRNA seed matches in 98 out of 126 IncRNAs is statistically significant (p-value = 5.79 x 10^-8, one-sided t-test). These data provide genetic evidence for miRNA-dependent deregulation of host IncRNAs during KSHV latency.
In order to validate that KSHV miRNAs can target host lncRNAs in the absence of KSHV infection, we chose four lncRNAs from the 98 containing seed sequences, and transfected pools of corresponding miRNA mimics into uninfected TIVE cells. The pools of mimics transfected were specific to the seed matches that those lncRNAs contained. Their respective mimic pools when compared to control mimic significantly knocked down all four lncRNAs tested, demonstrating that the viral miRNAs target lncRNAs in the absence of KSHV infection (Figure 2-3A).

The miRNA-dependent downregulation of lncRNAs could result from direct targeting of lncRNAs by miRNAs, or from an indirect secondary effect (e.g., through miRNA-mediated downregulation of transcription factors). To investigate direct interaction between KSHV miRNAs and lncRNAs, we performed miRNA pull-down experiments in TIVE-Ex-LTC cells. TIVE-Ex-LTC cells were derived from TIVE cells (see methods), but grow much faster compared to TIVE cells. KSHV negative TIVE-Ex-LTC cells were transfected with biotinylated miRNA mimics for either miR-K12-6-5p, miR-K12-11* or siGLO (lacks biotin) and pull-down experiments were performed 24 h post-transfection. It is important to note that the mimics are dsRNAs that require loading into the RISC in order to bind their targets. Loc541472 has one binding site for miR-K12-6-5p but none for miR-K12-11*, and CD27-AS1 has one binding site for miR-K12-11* but none for miR-K12-6-5p. Biotinylated miR-K12-6-5p mimic pulled down 43.7% of Loc541472 and none of CD27-AS1, and miR-K12-11* mimic pulled down 12.9% of CD27-AS1, but no Loc541472, thus confirming direct miRNA-lncRNA interaction (Figure 2-3B). The fact that we identified putative lncRNA targets of viral miRNAs in PEL and endothelial cells by Ago HITS-CLIP and viral genetics, together with biochemical evidence for direct
miRNA-lncRNA interaction, demonstrated that KSHV deregulates a subset of host lncRNAs in a miRNA-dependent fashion.

**Latent KSHV Deregulates LncRNAs Aberrantly Expressed In Cancer**

To date a very small percentage of all lncRNAs are functionally annotated, making interpretation of lncRNA expression data challenging. As a starting point, we analyzed lncRNAs that were deregulated (upregulated, downregulated and rescued) in response to latent KSHV infection for known or proposed functions in disease processes. Comparison of our dataset to two public databases [187, 234] identified 54 lncRNAs that were previously shown to be aberrantly expressed in various human cancers. These include HOTTIP, DLEU2, HOTAIR1M1, ANRIL, MEG3 and UCA1. Ten of the 54 lncRNAs are listed in Table 2-2, and include oncogenic and tumor suppressor lncRNAs. HOTTIP is upregulated in hepatocellular carcinoma, osteosarcoma, lung, prostate and other cancers [235]; DLEU2 is deleted in lymphocytic leukemia and epigenetically silenced in myeloid leukemia [181, 236]. Knockdown of HOTAIR1M1 has been shown to promote proliferation in promyelocytic leukemia cells [237]. ANRIL is an oncogenic lncRNA that promotes proliferation in numerous cancers including basal cell carcinoma (BCC), glioma, prostate and ovarian cancers [178]. UCA1 is upregulated in multiple cancers, including bladder, endometrial and pancreatic cancer and acts as an oncogenic lncRNA [238]. Loss of MEG3 expression has been reported in a wide spectrum of malignancies ranging from gliomas to colon and liver cancers [239]. To understand the mechanisms by which cancer-related lncRNAs are deregulated by KSHV, and their contribution to pathogenesis, we chose to initially study UCA1, ANRIL and MEG3.
Viral MicroRNAs Downregulate Tumor Suppressor LncRNA MEG3

MEG3 is a tumor suppressor IncRNA which is proposed to act by enhancing transcription from p53-dependent promoters [239]. Studies in HCT116 and U2OS cell lines have identified that MEG3 is a nuclear localized IncRNA [182], which was also confirmed in GM12878 cells by the GENCODE project [240]. According to the microarray data, MEG3 was slightly upregulated during latent KSHV infection. However, when validating MEG3 expression by qRT-PCR, it behaved in a rescued pattern, being suppressed in wt-KSHV infection and restored in Δcluster-KSHV-infected cells, suggesting regulation by KSHV miRNAs (Figure 2-4A). MEG3 contained seed sequence matches for miR-K12-3, K12-5, K12-6-5p, K12-8* and K12-9*. Uninfected TIVE cells were transfected with a pool of three KSHV miRNA mimics (miR-K12-5, K12-6-5p and K12-8*). MEG3 expression was reduced by almost 80% (Figure 2-4B). Furthermore, miRNA pull-down assays using biotinylated miR-K12-6-5p mimic pulled-down 24.5% of MEG3 (Figure 2-4C). miR-K12-11* mimic did not pull down MEG3 lncRNA. These data are consistent with viral miRNAs directly binding to and downregulating MEG3.

Viral MicroRNAs and Latency Proteins Both Target ANRIL

ANRIL is a nuclear localized oncogenic IncRNA that drives proliferation by silencing the INK4 tumor suppressor gene by recruiting PRC2 complexes [178]. The fact that ANRIL was downregulated in KSHV-infected cells from the microarray data suggested that ANRIL does not have a direct role in proliferation; however, ANRIL has recently also been implicated in innate immune responses, albeit in the context of bacterial infection [241]. Analysis of ANRIL expression by qRT-PCR showed a very strong 100-fold downregulation in wt-KSHV-infected cells, and a slightly reduced inhibition in the Δcluster-KSHV-infected TIVE cells (Figure 2-5A).
Such strong repression is not typical of miRNAs. However, the cDNA of ANRIL had a total of 17 6mer seed matches for 12 of 25 mature KSHV miRNAs. To investigate whether the large number of KSHV miRNA seed sequence matches in ANRIL are targeted by KSHV miRNAs, we ectopically overexpressed the shortest isoform (transcript variant 12) of ANRIL from a CMV promoter-driven vector in wt-KSHV-infected and uninfected TIVE-Ex-LTC cells. Since TIVE cells are highly resistant to plasmid transfection, we used TIVE-Ex-LTC cells for this experiment. As shown in Figure 2-5B, the ANRIL expression levels achieved in wt-KSHV-infected cells were 80% less compared to uninfected cells. We note that this expression difference was not due to differences in transfection efficiencies, since a control gene (LSD-1), expressed from the same vector, was expressed at similar levels in both cell lines (Figure 2-5B). Hence, the reduced ANRIL expression levels in infected cells compared to control cells strongly suggested post-transcriptional miRNA-dependent regulation of ANRIL. To test this, we transfected a pool of four miRNA mimics (miR-K12-1*, K12-6-5p, K12-2* and K12-11*) which led to a strong knock-down of ANRIL expression in uninfected TIVE cells compared to the control mimic (Figure 2-5C). Additionally, pull-down experiments in TIVE cells using biotinylated miR-K12-6-5p and miR-K12-11* mimics, for which ANRIL contains two seed matches each, significantly pulled-down 12.7% and 22.7% of ANRIL transcripts, respectively (Figure 2-5D).

Together these data show that ANRIL is targeted by multiple viral miRNAs. Since ANRIL also contained miRNA seed sequence matches for miR-K12-10 and K12-12, which are still present in Δcluster-KSHV (Figure 2-1A), we wanted to test ANRIL expression in the absence of all viral miRNAs. To this end we analyzed ANRIL expression in TIVE cells by infecting with a virus lacking all 12 miRNA genes (Δall-KSHV). Surprisingly we did not
observe significantly altered ANRIL expression compared to wt-KSHV-infected cells (Figure 2-5E). These data suggested that ANRIL may also be negatively regulated by latency associated proteins. To directly address this question, we ectopically expressed the major latency associated proteins of KSHV (LANA, vCyclin, vFLIP and Kaposin) and monitored ANRIL expression by qRT-PCR. Since TIVE-Ex-LTC cells do not express detectable levels of ANRIL, this experiment was performed in HeLa cells, which are known to robustly express ANRIL [242]. vFLIP and vCyclin downregulated ANRIL expression by almost 75% and 53%, respectively (Figure 2-5F). LANA and Kaposin did not have significant effects. The observation that ANRIL is negatively regulated by both miRNAs and latency associated proteins is in congruence with other host genes that are targeted by multiple viral mechanisms [114].

MicroRNA-Independent Deregulation of Host LncRNA UCA1 Promotes Proliferation and Migration

Urothelial Cancer Associated 1 (UCA1) is a lncRNA which was identified as highly upregulated in bladder cancer and has since been implicated in other cancers like colorectal, ovarian and renal carcinomas [238]. UCA1 is partially localized in both the nucleus and the cytoplasm and plays distinct roles in different sub-cellular compartments [243, 244]. Recently, it was shown that UCA1 transcription is induced by HIF-1α, to enhance hypoxic proliferation, migration and invasion of bladder cancer cells[238]. UCA1 was upregulated by approximately 90-fold during wt-KSHV infection and approx. 300-fold during Δcluster-KSHV infection (Figure 2-6A). Since UCA1 was upregulated under both infection conditions, UCA1 is presumably not regulated by miRNA-dependent gene silencing.

To determine which of the four major latency-associated proteins (LANA, vCyclin, vFLIP and Kaposin) upregulates UCA1, we transfected TIVE-Ex-LTC cells with expression
vectors either alone or in combination. Ectopic expression of vCyclin and Kaposin led to a 3.9 and 5.7-fold upregulation of UCA1 as monitored by qRT-PCR, respectively. Furthermore, co-transfection of vCyclin and Kaposin increased UCA1 to almost 15-fold compared to empty vector suggesting synergy (Figure 2-6B). LANA and vFLIP had no effect. The fact that the up-regulation observed in transfected cells is much less than in the context of infection could be a consequence of altered stoichiometry of latency proteins, or mean that other viral genes might contribute to UCA1 upregulation.

To address whether UCA1 directly contributes to KS-associated phenotypes, we knocked-down UCA1 expression using siRNAs in KSHV-infected TIVE cells. At 24, 48, 72 and 96 h post-transfection we observed 60-85% knockdown of UCA1 expression (Figure 2-6C). Next, we assayed for proliferation using the MTS assay. We measured proliferation at 24, 48, 72 and 96 h post-transfection and observed a statistically significant and dose-dependent decrease in proliferation of cells treated with siUCA1 as compared to scrambled control (Scr). Upon treatment with 10 nM siUCA1, the proliferation rate dropped to 72% by day 1 and then progressively to 52% by day 4 (Figure 2-6D). The migration assay (wound healing) involves introduction of a scratch in a monolayer of cells and measuring the percentage of the clear area that gets covered by migration at 12 hours post introduction of the scratch under serum-free conditions (Figure 2-6E). siUCA1-treated cells were consistently slower in migration from day 1 through day 4, as they recovered only between 12-15% of the scratch area, while Scr-treated cells recovered between 26-35% of the area (Figure 2-6F). We observed no significant susceptibility of siUCA1-treated cells to staurosporine-induced apoptosis in comparison with Scr-treated cells (data not shown). These data demonstrate that the induction of UCA1 by the
KSHV latency-associated proteins Kaposin and vCyclin promotes proliferation and migration, and likely contributes to KSHV pathogenesis and tumorigenesis.

**Discussion**

Here we show that latent KSHV infection significantly alters the IncRNA expression profile of endothelial cells. Deregulation of IncRNAs has implications in diseases such as diabetes, neurological disorders, viral infections and cancer [245, 246]. Our study establishes that KSHV employs its latency proteins and miRNAs, either alone or in combination, to target specific IncRNAs and potentially contribute to sarcomagenesis.

Post-transcriptional regulation of IncRNA expression by miRNAs is a novel phenomenon. Yoon et al showed let-7 loaded RISCs targeted lincRNA-p21 in a HuR-dependent manner in cervical carcinoma cells, eventually destabilizing and degrading lincRNA-p21 [169]. In bladder cancer, UCA1 and miR-1 expressions were inversely correlated, and overexpression of miR-1 phenocopied the knockdown of UCA1[218]. Further, MALAT1, a nuclear IncRNA, was reported to be targeted by miR-9 in an Ago2-dependent manner in the nuclei of Hodgkin’s lymphoma and glioblastoma cell lines[217].

We identified 126 IncRNAs as potential targets of viral miRNAs in endothelial cells, and we verified direct miRNA/IncRNA interactions by pull-down experiments with biotinylated KSHV miRNA mimics targeting Loc541472, CD27-AS1, ANRIL and MEG3. Results from the Ago HITS-CLIP experiment further suggest that this regulation proceeds in an Ago and hence RISC-dependent manner. As per our current understanding, RISC-mediated silencing of mRNAs proceeds via translation repression and induction of mRNA turnover [87, 247]. RNA destabilization followed by degradation is perhaps the mechanism relevant to silencing of IncRNAs. However, the details of the mechanism, especially for IncRNAs lacking a cap and/or a poly-A tail, remain to be uncovered.
An alternative and not mutually exclusive mechanism that involves direct engagement of miRNAs and lncRNAs is miRNA sponging by lncRNAs[154]. LincRNA-RoR sponges miR-145-5p thereby increasing the expression of pluripotent stem cell factors Oct4, Nanog and Sox2 [248]. The Steitz lab showed that lncRNAs encoded by Herpesvirus Saimiri, called HSURs, sequester host miRNAs in infected T-lymphocytes[225]. It is plausible that some host lncRNAs could sponge KSHV miRNAs, thereby derepressing downstream targets instead of being targeted by miRNAs themselves.

We demonstrated that viral latency proteins vCyclin and Kaposin synergistically upregulate UCA1 while vFLIP and vCyclin downregulate ANRIL. Thus, aside from miRNAs, the latency proteins play a pronounced role in perturbing lncRNA expression. This is not surprising given we identified 858 differentially expressed lncRNAs during wt-KSHV infection and only 126 were potential miRNA targets. vCyclin, an ortholog of cellular Cyclin D, upregulates expression of cell cycle regulatory genes[31]. Moreover, Kaposin stabilizes cytokine mRNAs thereby increases their turnover time [42]. vCyclin and Kaposin may act cooperatively by augmenting transcription and simultaneously preventing turnover of UCA1. We also showed that ectopically expressed vFLIP strongly downregulates ANRIL. STAT1-mediated activation of the ANRIL locus in vascular endothelial cells has been reported based on GWAS studies [249]. Studies using a mutant virus that lacks vFLIP in HUVEC cells showed activation of STAT1 in a NF-κB-dependent manner, suggesting that vFLIP probably inhibits STAT1 to downregulate ANRIL expression [250]. A recent study in endothelial cells demonstrated that ANRIL expression is induced by pro-inflammatory molecules, especially NF-κB and TNF-α, and silencing of ANRIL expression led to a reduction in IL6/IL8 response [251]. This further
underlines the role of ANRIL in immunity and supports the notion that KSHV may downregulate ANRIL to evade innate immune responses.

KSHV drives latently infected cells towards proliferation by a variety of mechanisms such as encoding orthologs for cell cycle proteins like vCyclin, or interfering with the p53 pathway through LANA[7], encoding miR-K12-11, an ortholog of oncomir-155[107], and the induction of the oncogenic host miRNA cluster miR-17/92[114]. Here we demonstrate that KSHV also upregulates UCA1 to drive proliferation and migration in endothelial cells. UCA1 has also been shown to promote the Warburg effect[252], an effect that has been shown to be required for maintenance of latent KSHV in endothelial cells[253]. We found that 53 additional lncRNAs previously shown to be aberrantly expressed in various malignancies are deregulated by KSHV, suggesting that UCA1 exemplifies how KSHV could similarly exploit lncRNAs that contribute to phenotypes such as proliferation and migration in the context of tumorigenesis.

Given that the majority of lncRNAs we catalogued in this study remain uncharacterized, the repertoire of cancer-relevant lncRNAs regulated by KSHV may be much larger. Although cancer is the pathological consequence of KSHV infection, KSHV could target lncRNAs of biological significance in other cellular processes, for example, lncRNAs involved in inflammation and innate immunity[201]. KSHV continually evades the innate immune response using several approaches, like suppressing TGF-β signaling [114], activation of NF-κB response genes [254] and encoding trace amounts of v-IL-6, a truncated version of human IL-6, during latent infection [255]. Loc541472, which we show here is targeted directly by KSHV miRNAs, is antisense to the hIL-6 promoter, suggesting that targeting of this lncRNA contributes to regulation of IL-6 expression. Indeed, preliminary experiments suggest a correlation between Loc541472 and hIL-6 expression (data not shown) and mechanistic studies are currently ongoing.
We identified a novel paradigm by which KSHV, an oncogenic herpesvirus, regulates cellular gene expression by targeting host IncRNAs with viral miRNAs and latency proteins. Studying IncRNAs deregulated by KSHV may yield novel mechanisms by which viruses evade the host immune response and in the case of EBV and KSHV contribute to tumorigenesis, as exemplified by our data on UCA1 which modulates migration and proliferation. Finally, studies on aberrantly expressed IncRNAs in KSHV-infected cancer cells may aid the functional characterization of cellular IncRNAs and at the same time identify novel virus-specific therapeutic targets for KS.

Material and Methods

Virus and plasmid constructs. The viruses used in this study, wt-KSHV, Δcluster-KSHV and Δall-KSHV, have the viral genome cloned into a Bac-16 backbone, as described in Brulois et al.[227] and Jain et al.[228]. Transcript variant 12 (RefSeq ID: NR_047542.1) of ANRIL was expressed from a pcDNA3.1 vector [256]. LANA, vCyclin, vFLIP and Kaposin were expressed from pcDNA3.2 vectors [257].

Cell culture. Telomerase immortalized vein endothelial cells (TIVE) and long-term cultured KSHV infected cells (TIVE-LTC) were generated by immortalizing passage 2 HUVEC cells (kindly provided by Dr. Keith McCrae, Case Western Reserve University) in our laboratory as described [226]. All uninfected and infected TIVE cells were grown in complete Medium-199 (1% Pen-Strep, 20% FBS), supplemented with Endothelial cell growth supplement (Sigma). TIVE-Ex-LTC cells were obtained by culturing TIVE-LTC cells as single cell dilutions without antibiotic selection, and have lost all copies of viral episomes. TIVE-Ex-LTC cells grow faster and are more transfectable compared to TIVE cells. All uninfected and infected TIVE-Ex-LTC cells were grown in complete DMEM (1% Pen-Strep, 10% FBS). Latently infected TIVE and
TIVE-Ex-LTC cells were maintained under hygromycin (10 µg/mL) to prevent episome loss. Body-cavity-based lymphoma (BCBL-1) cell line was derived from KSHV positive primary effusion lymphoma (PEL) and was kindly provided by Dr. Don Ganem at UCSF [258]. BCBL-1 cells were grown in complete RPMI (2% Pen-Strep, 10% FBS). HeLa cells were grown in complete DMEM (1% Pen-Strep, 10% FBS).

**Bioinformatics analysis.** *Reanalysis of CLIP data:* The BED files generated as a part of the analysis of Ago HITS-CLIP data from PEL cells [115] were compared with GENCODE V19 [259] to obtain a comprehensive list of putative lncRNA targets. These lncRNAs were compared with published tables available from EBV PAR-CLIP [124]. *Statistical test for enrichment of KSHV miRNA seed matches in lncRNAs:* R version 3.3.0 was used for this analysis. 100,000 random DNA sequences of length 1189 nt were generated. This number was obtained by calculating the mean length of the 126 rescued lncRNAs. KSHV miRNA seed matches were counted using repeated sampling (10,000 times) of 126 random DNA sequences. One-sided t-test was performed to compare the average number of seed matches in random sequences to that of rescued lncRNAs.

**Fractionation of PEL cells.** The method for isolating nuclear and cytoplasmic fractions was adapted from [260]. Briefly, 1 x 10^7 BCBL-1 cells were pelleted and washed twice with ice cold PBS. Cells were resuspended smoothly by gentle pipetting in sucrose buffer I (SB-I: 0.32 M Sucrose, 3 mM CaCl₂, 2 mM Mg(Ac)₂, 0.1 mM EDTA, 10 mM Tris-HCl (pH 8), 1 mM DTT, 0.5 mM PMSF and 0.5% NP-40) using 100 µL buffer per 1 x 10^7 cells. Lysis was at room temperature for 60 – 90 s. The nuclei were pelleted at 800 x g, 4 °C for 5 min and the supernatant (cytoplasmic fraction) was frozen immediately and stored at -80°C. The pellet was resuspended smoothly by gentle pipetting in 50 µL of SB-I and allowed to sit for 30 s at RT. The nuclei were
pelleted again at 800 x g, 4 °C for 5 minutes. The supernatant was discarded and the pellet (now whiter) was washed twice in 1 mL ice cold PBS. The resuspension was smooth and easy indicating no nuclear rupture. 10 µL of the 1 mL suspension from the second wash was trypan blue stained and checked by microscopy to verify the purity and integrity of the isolated nuclei. The nuclear fraction was frozen immediately and stored at -80 ºC.

**Immunofluorescence assays.** TIVE cells were grown overnight on coverslips at a dilution of 1 x 10^4 cells per well in a 6-well plate. Nuclei isolated from PEL cells were prepared as described [261], and fixed with a 1:1 ratio of methanol and acetone for 10 min in a humid chamber at 4 ºC. The samples were blocked in PBS with 3% BSA for 1 h at room temperature, and then incubated overnight at 4 ºC with either primary anti-Ago2 antibody or blocking solution (control). After washing, the samples were incubated with Alexa-468 anti-rat secondary antibody for 1 hour at room temperature. The slides were then stored at -20 ºC and imaged using a LEICA TCS SP2 AOBS Spectral Confocal microscope. The images were analyzed and figures were generated using the freeware Vaa3D[262].

**Western blots.** SDS-PAGE and Western blotting were performed using whole cell lysates, or cytoplasmic or nuclear fractions prepared from 100,000 cells/well. The following antibodies were used to probe the membrane: Ago2 (11A9, [263]), β-tubulin (Millipore, CP06-100UG), Sm antigen (Dr. Joan Steitz’s lab, Yale University), Lamin A/C (Active Motif, 39288), Calnexin (ENZO Lifesciences, ADI-SPA-865-D).

**RNA isolation and microarray analysis.** Total RNA was isolated with RNA-Bee (Tel-Test Inc.) using the protocol provided by the manufacturer. Total RNA (5-10 µg) was treated with DNase I (NEB) according to the manufacturer’s instructions and ethanol precipitated overnight. Genome-wide IncRNA microarray analysis was performed with ArrayStar using
Human LncRNA Array v3.0 (8 x 60K, Arraystar). A fold change cut-off of 2.0 was applied to filter lncRNAs into different categories (upregulated, downregulated and rescued) for further analyses. Three technical replicates for each of the three samples were analyzed.

**qRT-PCR of miRNAs.** Total RNA preparations from PEL cell fractions were reverse transcribed using the TaqMan® MicroRNA Reverse Transcription Kit (ThermoScientific). Stem-loop qPCR was performed using the TaqMan® Gene Expression Master Mix and appropriate miRNA assays from Applied Biosystems.

**qRT-PCR of mRNA or lncRNAs.** Total RNA (2 µg) was reverse transcribed using SuperScript III (Life Technologies) using random hexamers according to the manufacturer’s instructions. cDNA corresponding to 50-100 ng RNA was used per 10 µL of qPCR reaction. Instruments used for real-time PCR included ABI StepOne Plus™ (Applied Biosystems) and LightCycler96 (Roche). qPCR primer sequences are listed in Table 2-3.

**miRNA mimic transfections.** TIVE cells were seeded in 48-well plates (50,000 cells/well) and transfected with pools of miRNA mimics (in equimolar ratios and a final concentration of 5 nM) purchased from Qiagen. Comparable transfection efficiencies were ensured by co-transfecting a fluorescent non-targeting small RNA (siGLO, Dharmacon). At 48 h post transfection, the lncRNA expression levels were measured using the Power SYBR® Green Cells-to-CT™ Kit (ThermoFisher). In the cases of ANRIL and MEG3, 10 cm plates were seeded to 70% confluency and qRT-PCR analysis was performed using the conventional approach described above.

**Biotinylated miRNA pull-down.** Biotinylated miRNA mimics (miR-K12-6-5p and miR-K12-11*) were purchased from Exiqon. Pull-down was performed from TIVE and TIVE-Ex-LTC cells according to the previously published protocol[264] with minor changes. Each
replicate started with $6 \times 10^6$ cells for TIVE-Ex-LTCs (instead of $4 \times 10^6$) and $8 \times 10^6$ cells for TIVE cells. Input RNAs saved for analysis were 5% and 20% for TIVE-Ex-LTC and TIVE cells, respectively.

**Ectopic expression of latency genes from plasmids.** TIVE-Ex-LTC cells were reverse transfected in 6-well plates (300,000 cells/well) with 2 µg of plasmid DNA using FuGENE® HD according to the manufacturer’s protocol. HeLa cells were seeded in 6-well plates (150,000 cells/well) and were transfected 24 h later with 2 µg plasmid DNA using Lipofectamine 3000 according to the manufacturer’s protocol. DMEM (10% FBS) was used for transfection of both cell types. Comparable transfection efficiencies were ensured by co-transfecting pmaxGFP. Total RNA was harvested from transfected cells at 72 h post-transfection.

**siRNA Knockdown.** wt-KSHV-TIVE cells were plated in 96-well plates (20,000 cells/well for MTS assay) and 48-well plates (250,000 cells/well for wound healing assay). siRNAs (5nM or 10 nM) against UCA1 (Qiagen) were transfected using Lipofectamine® RNAiMAX reagent (ThermoFisher) according to the manufacturer’s protocol. ON-TARGETplus Non-targeting Control siRNA (Dharmacon) was used as the scrambled negative control. At 4 h post-transfection, the serum free medium was replaced by complete Medium-199.

**Cell proliferation and migration assays.** *MTS assay:* At 24, 48, 72 and 96 h post-transfection of siRNAs, the MTS assay was performed using the CellTiter 96® AQueous Non-Radioactive Cell Proliferation Assay kit (Promega) according to the manufacturer’s instructions. The absorbance of the samples was measured at 490 nm. *Wound-healing assay:* At 24, 48, 72 and 96 h post-transfection with siRNAs, the confluent wells were scratched using a 200 µL pipette tip along the diameter of the well. Images of the scratch were recorded at 0 and 12 h, and analyzed using the freeware TScratch[265].
**Statistics.** Statistical analyses on experimental measurements were done using two-tailed student’s t-test assuming unequal variances.

**Data availability.** Raw data files from the microarray experiment were deposited to the Gene Expression Omnibus under the accession number GSE89114.
Figure 2-1: Expression profiling of wt-KSHV and Δcluster-KSHV infected endothelial cells. (A) Latency associated region of wt-KSHV in a Bac16 backbone. The region deleted in the Δcluster-KSHV virus is highlighted. (B) Heatmap of unsupervised hierarchical clustering of the microarray samples in the ‘rescued’ category of genes (n = 3 technical replicates).
Figure 2-2: KSHV miRNAs and Ago2 are partially localized in the nuclei of latently infected cells. (A) qRT-PCR analysis of mature KSHV miRNA distribution in the cytoplasmic and nuclear fractions of PEL cells. Percentage distribution was calculated by normalizing to expression in whole PEL cells, assuming no loss during fractionation. RNU48 was used as a nuclear control for fractionation. The bar graphs show the mean values (n = 3) ± SEM. p-values: * < 0.05; ** < 0.01; *** < 0.005. (B) Subcellular distribution of Ago2 proteins in PEL cells analyzed using Western blotting. Tubulin was probed as positive control for cytoplasm, Sm and Lamin A/C are positive controls for nuclei and Calnexin is the negative control for endoplasmic reticulum. (C) Localization of Ago2 in PEL nuclei analyzed using IFA and confocal microscopy. Ago2 is shown in green and DAPI in blue. DAPI is shown at half the original intensity.
Figure 2-3: KSHV miRNAs directly bind to and downregulate host lncRNAs. (A) Uninfected TIVE cells were transfected with 5 nM final concentration of miRNA mimic pools (Loc541472: miR-K12-1, K12-6-5p; CD27-AS1: miR-K12-1*, K12-11*; RP11-438-N16.1: miR-K12-1*, K12-8*, K12-11*; Linc00607: miR-K12-2*, K12-11*). Relative expression levels of target lncRNAs were analyzed 48 h post-transfection using qRT-PCR. The bar graphs show the mean values ± SEM after normalization to GAPDH (n = 3). (B) Biotinylated miRNA mimics of miR-K12-6-5p and miR-K12-11* were transfected into uninfected TIVE-ExLTC cells (5 nM final concentration) and were pulled down 24 h later. Target lncRNAs were analyzed using qRT-PCR. siGLO pull-down was used a negative control. The bar graphs show the mean values ± SEM after normalization to input (n = 3). p-values: * < 0.01; ** < 0.005; *** < 0.0005; and **** < 0.0001.
Figure 2-4: Tumor suppressor IncRNA MEG3 is targeted by KSHV miRNAs. All bar graphs show the mean values ± SEM after normalization to GAPDH (n = 3), unless specified otherwise. (A) MEG3 expression in Uninfected, wt-KSHV-infected and Δcluster-KSHV-infected cells measured by qRT-PCR. (B) Uninfected TIVE cells were transfected with 5 nM final concentration of miRNA mimic pool (miR-K12-5, K12-6-5p and K12-8*). Relative expression level of MEG3 was analyzed 48 h post-transfection using qRT-PCR. (C) Biotinylated miRNA mimic of miR-K12-6-5p was transfected into uninfected TIVE cells (5 nM final concentration) and was pulled down 24 h later. MEG3 expression was analyzed using qRT-PCR. siGLO pull-down was used as a negative control. The data were normalized to input. p-values: * < 0.05; ** < 0.01; *** < 0.005; **** < 0.0005; and ***** < 10^-4.
Figure 2-5: LncRNA ANRIL is targeted by both KSHV miRNAs and latency proteins. All bar graphs show the mean values ± SEM after normalization to GAPDH (n = 3), unless specified otherwise. (A) ANRIL expression in Uninfected, wt-KSHV-infected and Δcluster-KSHV-infected cells measured by qRT-PCR. Inset shows the comparison between wt-KSHV-infected and Δcluster-KSHV-infected cells. (B) Uninfected and wt-KSHV-infected TIVE cells were transfected with pcDNA3.1-ANRIL and relative over-expression of ANRIL was measured using qRT-PCR. LSD-1 was used a control to verify comparable transfection efficiencies of uninfected and infected cells. (C) Uninfected TIVE cells were transfected with 5 nM final concentration of miRNA mimic pool (miR-K12-1*, K12-6-5p, K12-2* and K12-11*). Relative expression level of ANRIL was analyzed 48 h post-transfection using qRT-PCR. (D) Biotinylated miRNA mimics of miR-K12-6-5p and miR-K12-11* were transfected into uninfected TIVE cells (5 nM final concentration) and were pulled down 24 h later. ANRIL expression was analyzed using qRT-PCR. siGLO pull-down was used as a negative control. The data were normalized to input. (E) ANRIL expression in Uninfected, wt-KSHV-infected and Δall-KSHV-infected cells measured by qRT-PCR (n = 2). (F) HeLa cells were transfected with latency gene(s) (LANA, vCyclin, vFLIP, Kaposin or vCyclin + Kaposin) expressed from pcDNA3.2 vector. ANRIL expression was analyzed 72 h post-transfection using qRT-PCR. p-values: * < 0.05; ** < 0.01; *** < 0.005; **** < 0.0005; ***** < 10^{-4} and n.s. = not significant.
Figure 2-6: UCA1 is upregulated by KSHV in a miRNA-independent manner. (A) UCA1 expression in Uninfected, wt-KSHV-infected and Δcluster-KSHV-infected cells measured by qRT-PCR. The bar graphs show the mean values ± SEM after normalization to GAPDH (n = 6). (B) Uninfected TIVE cells were transfected with latency gene(s) (LANA, vCyclin, vFLIP, Kaposin or vCyclin + Kaposin) expressed from pcDNA3.2 vector. UCA1 expression was analyzed 72 h post-transfection using qRT-PCR. The bar graphs show the mean values ± SEM after normalization to GAPDH (n = 6 for vCyclin, n = 3 for others). (C) wt-KSHV-infected TIVE cells were transfected with 10 nM concentration of siUCA1 or Scr control. At 24, 48, 72 and 96 h, UCA1 expression was analyzed using qRT-PCR. The bar graphs show the mean values ± SEM after normalization to GAPDH (n = 3). (D) wt-KSHV-infected TIVE cells were transfected with 5 nM or 10 nM concentration of siUCA1 or Scr control. At 24, 48, 72 and 96 h, the samples were subject to MTS assay and absorption was measured at 495 nm wavelength. The bar graphs show the relative absorbance ± SEM (n = 3). (E) wt-KSHV-infected TIVE cells were transfected with 10 nM concentration of siUCA1 or Scr control. At 24, 48, 72 and 96 h, the samples were subject to scratch assay. Plates were imaged at 0 and 12 h and the images were processed using T-Scratch. (E) Representative images of the wound healing assay, scale bar = 100 pixels. (F) The bar graphs show the percentage of scratch area recovered ± SEM (n = 3). p-values: * < 0.05; ** < 0.01; *** < 0.005; and **** < 0.0005.
Table 2-1. Summary of deregulated lncRNAs from the microarray analysis.

<table>
<thead>
<tr>
<th></th>
<th>wt-KSHV vs. Mock</th>
<th>Δcluster-KSHV vs. Mock</th>
<th>Common</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upregulated</td>
<td>325</td>
<td>1107</td>
<td>247</td>
</tr>
<tr>
<td>Downregulated</td>
<td>533</td>
<td>1265</td>
<td>238</td>
</tr>
<tr>
<td>IncRNA</td>
<td>Function</td>
<td>Comparison group(s)</td>
<td>Ref.</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>ANRIL (CDKN2B-AS1)</td>
<td>Epigenetic silencing of tumor suppressor INK4B</td>
<td>KSHV vs. Mock: Down, Δcluster vs. Mock: Down,</td>
<td>[178]</td>
</tr>
<tr>
<td>CRNDE</td>
<td>Upregulates mTOR pathway in gliomas, recently shown to code for a short nuclear peptide</td>
<td>KSHV vs. Mock: Down, Δcluster vs. KSHV: Up,</td>
<td>[266]</td>
</tr>
<tr>
<td>DLEU2</td>
<td>Host gene for tumor suppressor miRNAs miR-15a and miR-16-1</td>
<td>KSHV vs. Mock: Up, Δcluster vs. Mock: Up,</td>
<td>[181, 236]</td>
</tr>
<tr>
<td>HOTAI RM1</td>
<td>Modulates gene expression of cell adhesion molecules</td>
<td>KSHV vs. Mock: Up, Δcluster vs. Mock: Up,</td>
<td>[237]</td>
</tr>
<tr>
<td>HOT TIP</td>
<td>Upregulates transcription of the antisense transcript, HOXA13</td>
<td>KSHV vs. Mock: Up, Δcluster vs. KSHV: Down,</td>
<td>[235]</td>
</tr>
<tr>
<td>PLAC2 (TINCR)</td>
<td>Binds to Stau1 protein and regulates KLF2 mRNA in cells</td>
<td>KSHV vs. Mock: Up, Δcluster vs. Mock: Up</td>
<td>[267]</td>
</tr>
<tr>
<td>PTCSC3</td>
<td>Tumor suppressor IncRNA that acts by downregulating S100A4</td>
<td>KSHV vs. Mock: Down, Δcluster vs. Mock: Down</td>
<td>[268]</td>
</tr>
<tr>
<td>UCA1</td>
<td>Promotes cell cycle progression via PI3K-AKT pathway; also aids pRb1 and SET1A interplay</td>
<td>KSHV vs. Mock: Up, Δcluster vs. Mock: Up,</td>
<td>[238]</td>
</tr>
<tr>
<td>ZEB1-AS1</td>
<td>Promotes EMT by upregulating ZEB1, MMP2, MMP9, N-cadherin, and Integrin-β1</td>
<td>KSHV vs. Mock: Down</td>
<td>[269]</td>
</tr>
</tbody>
</table>

Table 2-2. Examples of oncogenic and tumor-suppressor lncRNAs deregulated by KSHV.
Table 2-3. Oligonucleotides used in this study.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Direction</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>Forward</td>
<td>CCCCTGGCCAAGGATCCATCCA</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>ACACATCGGACAGGCGCATCCA</td>
</tr>
<tr>
<td>CD27-AS1</td>
<td>Forward</td>
<td>ACAGAATGAGTGACAGCGAGGG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>TGGTCTGCTGCTCCCATCCCAC</td>
</tr>
<tr>
<td>Linc00607</td>
<td>Forward</td>
<td>GACGGCTGAGGAAGGATGGGT</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>AGTAATGGAGTGGAGTGAGAAC</td>
</tr>
<tr>
<td>Loc541472</td>
<td>Forward</td>
<td>TGACCTCTGTGGGCAATTTAC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>TCGGTGAAGATGGATGACCT</td>
</tr>
<tr>
<td>RP11-438-N16.1</td>
<td>Forward</td>
<td>TGGAGACCAGCCCGAGAATCTG</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GGAGAAGCTGAGACCGAGGAGG</td>
</tr>
<tr>
<td>MEG3</td>
<td>Forward</td>
<td>TTTTGTGCCAAAGGGCTCTGGGA</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>AGGGACTCAAGGGCCCGCGTAAA</td>
</tr>
<tr>
<td>ANRIL</td>
<td>Forward</td>
<td>TCTGATACAGACAGACAGATCAA</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CAGCACACCTAAGCTGATGC</td>
</tr>
<tr>
<td>LSD1</td>
<td>Forward</td>
<td>CTCTTCTGGAACCTCTATAAGGC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>CATTTCCAGATGATCCTGAGCAA</td>
</tr>
<tr>
<td>UCA1</td>
<td>Forward</td>
<td>CTCTCCATTGGGCTACTAC</td>
</tr>
<tr>
<td></td>
<td>Reverse</td>
<td>GCCGCAGGTCTTAAGAGATGAG</td>
</tr>
</tbody>
</table>

Biotinylated miRNA mimics

<table>
<thead>
<tr>
<th>KSHV miRNA</th>
<th>Direction</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-K12-6p</td>
<td>miRNA</td>
<td>CCAGCGACCCUAUCCCAUGUU/3Bio/</td>
</tr>
<tr>
<td></td>
<td>Passenger</td>
<td>CGAUGGAAUAGGUCGCUAGAG</td>
</tr>
<tr>
<td>miR-K12-11*</td>
<td>miRNA</td>
<td>GGUCACGCUUAAACAUCUCAGAUU/3Bio/</td>
</tr>
<tr>
<td></td>
<td>Passenger</td>
<td>UCUGAAGGUUUAAGCGUGAUCAG</td>
</tr>
</tbody>
</table>

Shown are the RNA sequences of the miRNA strand (the biotinylated strand with miRNA sequence) and the passenger strand (used to make the miRNA double stranded, but does not get incorporated into the RISC). Biotin was added to the 3' end of the miRNA strand.
CHAPTER 3
COMPUTATIONAL ANALYSIS OF RIBONOMICS DATASETS IDENTIFIES LONG NONCODING RNA TARGETS OF GAMMA-HERPESVIRAL MICRORNAS

Summary

Ribonomics experiments involving cross-linking and immunoprecipitation of Ago proteins have expanded the understanding of the miRNA targetome of several organisms. These techniques, collectively referred to as CLIP-seq, have been applied to identifying the mRNA targets of Kaposi’s sarcoma associated herpesvirus (KSHV) and Epstein-Barr virus (EBV) miRNAs. However, these studies focused on identifying only RNA targets of KSHV and EBV miRNAs that are known to encode proteins. Recent studies have suggested that long noncoding RNAs (lncRNAs) are also targeted by miRNAs. In this study, we performed a systematic re-analysis of published HITS-CLIP and PAR-CLIP datasets from KSHV and EBV driven lymphomas and a CLASH dataset from KSHV infected endothelial cells to identify novel lncRNA targets of the viral miRNAs. Here, we catalog the lncRNA targetome of KSHV and EBV miRNAs, and provide some initial insights into lncRNA-miRNA binding interactions based on in silico analysis.

Significance

KSHV and EBV driven cancers are the most prevalent AIDS-associated malignancies. miRNAs encoded by KSHV and EBV have been demonstrated to be important contributors to the development of cancer. Our recent work indicated that KSHV miRNAs can target host lncRNAs, and altered lncRNA expression can contribute to cancer phenotypes. miRNA-mediated targeting of lncRNAs is a newly described route for regulation of gene expression, the mechanism of which remains to be uncovered. Cataloging potential lncRNA targets of KSHV
and EBV miRNAs from available datasets provides an important research tool to understand the exact molecular mechanisms that drive tumorigenesis. Our preliminary results that describe binding interactions between lncRNAs and miRNAs could provide direction to understanding the actual mechanism of lncRNA repression by miRNAs and miRNA sponging by lncRNAs.

Introduction

Kaposi’s sarcoma-associated herpesvirus (KSHV) and Epstein-Barr virus (EBV) are opportunistic human pathogens which belong to the γ-herpesvirus family [8, 270]. These viruses drive cancers in immunocompromised AIDS patients and organ-transplant recipients. Certain strains cause cancers in immunocompetent individuals as well. KSHV causes Kaposi’s sarcoma (KS), primary effusion lymphoma (PEL), and certain forms of multicentric Castleman’s disease (MCD) [220]. EBV causes Hodgkin’s and non-Hodgkin’s lymphoma, Burkitt’s lymphoma and nasopharyngeal carcinomas [270]. Both these dsDNA viruses establish life-long latency in infected individuals during which their genomes remain as circular episomes with restricted gene expression [271]. An important feature of KSHV and EBV is the expression of viral miRNAs during latent infection, and this also extends to some other herpesviruses like HSV and HCMV [272]. Since KSHV and EBV driven cancers harbor latent episomes, several labs have cataloged the mRNA targets of these viral miRNAs to identify mechanisms by which these miRNAs contribute to cancer [102, 115, 117, 121, 123, 124].

Systematic isolation of cross-linked miRISC (miRNA bound to RNA-induced silencing complexes) is often employed to identify simultaneously all mRNA targets of all miRNAs in the cell line of interest [122, 273]. KSHV and EBV miRNA targets have been identified in lymphomas caused by these viruses using ribonomics techniques including Ago HITS-CLIP [115, 121] and Ago PAR-CLIP [123, 124]. Haecker and Renne present a detailed review of
HITS-CLIP and PAR-CLIP for viral miRNAs [125]. A simple graphical outline of Ago HITS-CLIP, PAR-CLIP and CLASH is shown in Figure 1-3. Briefly, Ago HITS-CLIP involves cross-linking of cells or tissue using 254 nm UV irradiation, followed by lysis. The lysate is then subject to immunoprecipitation with antibodies against the Ago protein, which enriches miRNA-Ago-mRNA complexes. The RNAs within isolated complexes are trimmed, size separated, reverse transcribed and then sequenced as two distinct pools: Ago-bound miRNAs and Ago-bound mRNAs [273]. Ago PAR-CLIP is a variant of HITS-CLIP where cells are cultured with non-toxic thiouridine, which is incorporated into nascent RNA in place of uridine, and thus allows for high efficiency cross-linking of RNA to Ago using 365 nm UV irradiation. Since the reverse transcriptase adds a G opposite thiouridine instead of A, use of thiouridine leads to a characteristic T to C mutation in PAR-CLIP reads [122]. The targets identified based on these methods are matched to their targeting miRNAs based on the presence of the miRNA seed sequence in the pool of potential target sequences, using bioinformatics tools [122, 273]. However, this approach ignores targets that undergo non-canonical targeting. Recent studies have reported that a significant proportion of miRNA targeting proceeds via non-canonical binding, i.e., binding independent of classical seed pairing [126, 274, 275]. To address this deficiency in CLIP-based methods directly, the Tollervey laboratory developed Ago Cross-linking and sequencing of hybrids (CLASH), which includes an RNA-RNA ligation step [126, 276]. After isolation of miRNA-Ago-mRNA complexes and trimming, the ligation step enables ligation of miRNAs to their target mRNAs. Each of these chimeric molecules identified via sequencing represents a unique binding event in the cell between a miRNA and its RNA target. This method enables us to identify both canonical and non-canonical miRNA targeting events.
The majority of ribonomics studies to date focused on identifying the miRNA targetome have exclusively investigated mRNA targets [52]. Specifically, most studies have focused on identifying miRNA binding sites in the 3’ UTRs of target mRNAs [52, 277]. However, long noncoding RNAs (lncRNAs) have emerged as an important class of regulatory RNAs, especially in cancer [221]. Several lncRNAs have been implicated in cancer by being aberrantly expressed in multiple cancers, and only a small subset of those are mechanistically well-understood [170, 221, 278]. Recent experimental and computational studies have suggested that miRNAs can also target cellular long noncoding RNAs (lncRNAs) [218, 279]. Recent work from our lab (Chapter 2) confirmed that KSHV miRNAs bind to and downregulate lncRNAs in endothelial cells in a RISC-dependent manner. In this chapter, we catalog the lncRNA targets of KSHV and EBV miRNAs based on previously published Ago HITS-CLIP and Ago PAR-CLIP analyses of EBV and KSHV infected lymphoma cells. We also provide high-confidence lncRNA targets of KSHV miRNAs identified based on CLASH in endothelial cells, representative of KS. Finally, we present some preliminary evidence from in silico study of binding characteristics of lncRNAs and miRNAs.

Results

PIPE-CLIP Analysis Identifies Putative LncRNA Targets of KSHV and EBV MicroRNAs

After the characterization of KSHV and EBV miRNAs, initial studies to identify their targets primarily involved a gene-by-gene approach. Later, seed sequence based prediction algorithms and RNA-immunoprecipitation (followed by microarray) were used for simultaneous identification of multiple targets. However, the establishment of Ago HITS-CLIP and Ago PAR-CLIP allowed for high-throughput identification of viral miRNA targets. Four independent
studies cataloged KSHV and EBV miRNA targets using either HITS-CLIP or PAR-CLIP in lymphoma systems [115, 121, 123, 124]. A summary of these studies is presented in Table 3-1. We chose these four studies for our computational analysis primarily because of the following shared features: (1) They identify miRNA targets of a human γ-herpesvirus, (2) They were all performed on lymphoma cell lines, and (3) They used closely related techniques like HITS-CLIP or PAR-CLIP.

We aimed to identify and catalog the putative lncRNA targets of KSHV and EBV. To do this, we used a recently developed pipeline called PIPE-CLIP [280], which uses a modeling-based approach to identify high confidence miRNA target sites. To uniformly reanalyze the data from these four studies, we downloaded the publicly available raw data from these studies (see methods for file numbers), aligned them to the human genome (hg19) using Bowtie, and fed these alignments to the PIPE-CLIP pipeline. PIPE-CLIP first removes PCR duplicates and then calls for enriched clusters and reliable cross-linking sites from the sequencing alignments. We used a cut-off of FDR < 0.05 to call for enriched clusters. PIPE-CLIP is designed to appropriately handle the differences in HITS-CLIP and PAR-CLIP protocols, such as the T to C mutations found in PAR-CLIP samples. The clusters obtained were then annotated using Bedtools v2.25.0 based on the information from GENCODE V19. It is important to note that we allowed for all possible annotations of any given cluster, that is, if a cluster appears in a genomic region shared by two or more overlapping transcripts, that cluster is annotated multiple times, once for each unique transcript. The number of mRNA clusters identified by PIPE-CLIP was lower than that identified using CLIPZ [281] or PARalyzer [282] in the original studies. This could be due to the differences in algorithms used by these programs. CLIPZ does not perform any statistics on the identified clusters and hence reports all potential clusters, while PIPE-CLIP
uses a zero-truncated negative binomial distribution to model clusters and thus selects clusters with a higher stringency. PARalyzer on the other hand uses a non-parametric kernel density based approach to identify clusters. In addition, both PAR-CLIP based studies [123, 124] used reads longer than 13 nt for alignment, whereas we used reads > 18 nt for alignment using Bowtie.

All the identified clusters were interrogated for the presence of viral miRNA seed sequences. Only the viral miRNAs encoded by the virus in the respective cell types were used to interrogate for seed matches. BC-1 cells harbor both KSHV and EBV episomes and hence miRNAs from both viruses were used for this analysis. We screened for both 7mer (nt 2-8) and 6mer (nt 2-7) seed matches in the cluster sequences [77]. Shown in Table 3-2 and Figure 3-1 are the counts and percentages, respectively, for clusters with viral miRNA seed matches, grouped by RNA class. While mRNAs make up about 60% of all viral miRNA targets, lncRNAs contribute about 35%, thus forming a second important class of miRNA targets. Around 4% of all clusters mapped to regions for which no annotation was available in the GENCODE datasets.

**KSHV and EBV MicroRNAs Target Common LncRNAs**

There are no miRNA orthologs between these two γ-herpesviruses, i.e., none of the KSHV and EBV miRNAs share a seed sequence [100]. However, previous studies have reported an appreciable overlap in the cellular pathways targeted by these two viruses. For example, KSHV encodes miR-K12-11, a miRNA ortholog of cellular oncomir hsa-miR-155 [100, 222]. While EBV does not encode any ortholog of miR-155, it upregulates the expression of the pre-miRNA gene of miR-155, often referred to as BIC [222, 283, 284]. We questioned whether KSHV and EBV miRNAs share a subset of lncRNA targets. To investigate this, we analyzed the overlap in target RNAs between studies (Table 3-3) by pooling all putative targets identified.
from different replicates of each study. We chose to pool the targets since the four studies had different replication structures, with one study using only one biological replicate. From Table 3-3, it is evident that the overlap between HITS-CLIP and PAR-CLIP is less than 50%, for both mRNA and lncRNA targets, suggesting that neither approach is comprehensive. For example, HITS-CLIP analysis of KSHV infected cells recovered 26% and 48% of 7mer and 6mer targets, respectively, which were identified by PAR-CLIP analysis of KSHV positive cells. This is consistent with the observation made by Haecker et al. (HITS-CLIP, [115]) that they recovered 42% of the targets identified by Gottwein et al. (PAR-CLIP, [123]). The minor difference in percentages could be because the original PAR-CLIP paper used both 7mer (nt 2-8) and 7mer1A (nt 2-7 with an A in the mRNA opposite position 1 of miRNA) seed types to identify mRNA targets [123]. In spite of the experiment-dependent differences and the differences in targetomes of KSHV and EBV miRNAs, lncRNAs found in all studies would represent the best set of candidates for further analysis. We found 37 lncRNAs with 7mer miRNA binding sites and 405 lncRNAs with 6mer miRNA binding sites were enriched as viral miRNA targets in all studies (Figure 3-2).

Some LncRNA Targets of KSHV and EBV MicroRNAs are Aberrantly Expressed in Cancer

The majority of the lncRNAs known to date remain functionally uncharacterized. Thus, it is challenging at present to formulate hypotheses based on the identified lncRNA targets to understand the tumorigenic properties of these latent γ-herpesviruses. Aberrant expression of lncRNAs has been reported in several cancers [170]. We used two publicly available databases of lncRNAs aberrantly expressed in cancer to ask if any of the lncRNA targets might be associated with a tumorigenic phenotype [187, 234]. We identified 95 cancer-associated
lncRNAs within the putative viral miRNA targets. Of these, eight were previously implicated in lymphomas, and they include some of the well-characterized lncRNAs such as Gas5, DLEU2 and NEAT1. These eight also included MIR17HG, which is the miR-17-92a-1 cluster host gene, and was recently shown to be induced in KSHV infected endothelial cells, thus inhibiting TGF-β signaling [114]. Interestingly, clusters found on MIR17HG only contained EBV miRNA binding sites, suggestive of alternate mechanisms used by EBV to target similar host pathways to those targeted by KSHV. Given that most lncRNAs are not characterized functionally, the true number of cancer-associated lncRNA targets of KSHV and EBV miRNAs could be much larger.

**Modified CLASH Identifies KSHV MicroRNA Targets in Endothelial Cells**

CLIP-seq based methods have revolutionized the high-throughput identification of miRNA targets. However, the primary caveat with these is the reliance on bioinformatics to identify which miRNA targets which mRNA. Often multiple miRNA seed sequences are found in enriched clusters, thus complicating unambiguous assignment of a miRNA to a target cluster. Hence, these methods allow for identification of putative targets, but further work is required to confirm how many of those are true targets of any given miRNA. In addition, several studies have suggested that miRNA targeting can also happen via non-canonical (seed sequence independent) base-pairing between the miRNA and mRNA [126, 274, 275]. Thus, using the seed sequence as the sole reconstruction criteria in CLIP-seq based methods underestimates the number of true targets for any miRNA. A protocol developed by the Tollervey lab in 2011 overcomes this challenge by enabling direct investigation of RNA-RNA interaction in the context of a particular RNA Binding Protein (RBP) [276]. They called this Cross-linking Ligation and Sequencing of Hybrids (CLASH), an outline of which is presented in Figure 1-3. CLASH is essentially HITS-CLIP with an additional RNA ligation step after the pull down of
Ago-bound RNAs. This allows miRNAs and mRNAs in close proximity, like those in complex with the Ago protein, to ligate and form chimeric molecules. These chimeras, along with other RNAs, are then processed, reverse transcribed and sequenced. A bioinformatics pipeline called Hyb, developed also by the Tollervey lab, can be used to identify these hybrids [127].

Considering the two major cancers caused by KSHV, KS and PEL, viral miRNA targets have been identified only in PEL using HITS-CLIP and PAR-CLIP. However, KS is the most frequent clinical manifestation of KSHV [8]. We performed CLASH analysis to identify KSHV miRNA targets in an endothelial model system of KS, called TIVE-Ex-LTC cells (described in Chapter 2). We adapted the CLASH protocol to minimize procedural losses, by performing most of the steps post-immunoprecipitation on the beads and eliminating the size separation steps presented in the original protocol. Detailed methods and an analysis of mRNA targets identified from this CLASH experiment will be reported elsewhere (unpublished work).

We compared the hybrids from three different samples: uninfected cells, wt-KSHV infected cells, and Δ11-KSHV infected cells. In Δ11-KSHV, the miR-K12-11 sequence is mutated to disrupt the formation of the stem-loop structure characteristic of miRNAs, and hence miR-K12-11 and miR-K12-11* are not expressed [228]. We chose to specifically study miR-K12-11 targets because miR-K12-11 is an ortholog of human oncomir, miR-155, and we hypothesized that miR-K12-11 targets would have direct roles in cancer-related pathways [100]. We performed the CLASH analysis in three biological replicates and the number of hybrids identified is listed in Table 3-4. We found fewer than 15 KSHV miRNA targets in uninfected cells, which represented the low background of sequencing errors. It is important to note that we identified thousands of cellular miRNA-cellular lncRNA hybrids in KSHV infected and
uninfected cells, which suggests that miRNA-lncRNA interaction is a global phenomenon, and is not specific to KSHV miRNAs.

We identified 20 miR-K12-11 specific lncRNA targets by comparing wt-KSHV hybrids with Δ11-KSHV hybrids. To obtain this list we used a stringency of hybrids appearing in at least 2 out of 3 biological replicates of each sample. We interrogated these miR-K12-11 targets for prior implication in cancers by comparing with the two databases discussed above [187, 234]. We found that four lncRNAs targeted by miR-K12-11, namely MIR17HG, MIR155HG, MALAT1 and AFAP1-AS1 have been previously reported to be aberrantly expressed in different cancers. We also analyzed all lncRNA targets of KSHV miRNAs for cancer relevance and identified 35 that are aberrantly expressed in various cancers, a few of which are shown in Table 3-5.

A Subset of LncRNA Targets are Exclusively Nuclear

LncRNAs play important roles in the regulation of gene expression at epigenetic and transcriptional levels [159]. Many important lncRNAs such as MALAT1, ANRIL and NEAT1 have been shown to reside and function in the nucleus of cells [156, 166, 175, 176]. Recent studies have suggested that RNAi proteins, including Dicer and Ago2, are available and functional in the nucleus [231]. In our previous study, we observed that Ago2 and mature viral miRNAs localize to the nuclei of KSHV-infected cells. Other labs have shown that mature cellular miRNAs also partially localize to the nucleus [285-287]. Based on these, we investigated the distribution of the identified lncRNA targets between the nucleus and cytoplasm. To do this, we used the information from an online database called ANGIOGENES, which has a comprehensive list of RNAs from sub-cellular pools of HUVEC (primary endothelial) cells [230]. It is important to know that several RNAs in the database were reported in more than one
of four pools considered: Nuclear, poly A -; Nuclear, poly A +; Cytosolic, poly A –; and Cytosolic, poly A + (Figure 3-3A). We found that a subset of lncRNAs targeted by cellular miRNAs (Figure 3-3B) and a subset targeted by viral miRNAs (Figure 3-3C) were exclusively nuclear, suggesting that miRNAs also target nuclear lncRNAs.

**Comparative Analysis of MicroRNA Interaction with LncRNAs versus MRNAs**

Apart from uncovering high-confidence miRNA targets, CLASH analysis also provides some information on how these RNAs interact in the cell. We mined this information to understand better how miRNAs interact with lncRNAs and how this compares with the well understood miRNA-mRNA interaction. We first considered the orientation of miRNA-target RNA ligation. In the majority of the hybrids identified, the miRNAs were ligated to the 5’ end of the target RNA (Figure 3-4A). Considering exclusively cellular miRNA binding to cellular target RNAs, we identified that mRNA targets had a higher propensity to ligate miRNAs on their 5’ end than lncRNAs (Figure 3-4B). This could reflect the difference in the profile of associated RNA binding proteins (RBPs) (other than Ago) between mRNAs and lncRNAs. KSHV miRNAs also preferentially ligate at the 5’ end of mRNAs and that preference is slightly, yet significantly, lowered for lncRNAs (Figure 3-4B). Next, we questioned whether viral miRNAs had an altered preference for the ligation end when compared to cellular miRNAs. We analyzed this separately for mRNAs and lncRNAs to prevent target-based differences from confounding this analysis. We found that viral miRNAs are slightly less likely to ligate at the 5’ end of mRNAs when compared to cellular miRNAs (Figure 3-4C). However, such differences were not evident for lncRNA targets. Based on the frequency of miRNA ligation at the 5’ end, we conclude that there are significant differences in how miRNAs interact with lncRNAs versus how they interact with
mRNAs. Altered preference for ligation to the 3’ end could reflect that the flexibility and/or steric properties of miRISC bound lncRNAs is different from those of mRNAs.

3’ Ligated MicroRNAs Preferentially Target the 3’ End of LncRNAs

Most target prediction algorithms and even pipelines that analyze ribonomics datasets, like HITS-CLIP or PAR-CLIP, select for mRNA targets based on miRNA binding in their 3’ UTR [79, 80, 281]. However, there are studies that have demonstrated that miRNA binding sites within the coding sequence (CDS) also contribute to miRNA silencing [85, 126]. In fact, it was shown that miRNA binding within CDS contributes to translational inhibition, while miRNA binding within 3’ UTR contributes to mRNA degradation [86]. We interrogated our datasets for the distribution of miRNA binding sites along the length of the target RNA. We found that CDS and 3’ UTRs of mRNAs are equally targeted by all miRNAs, and viral miRNAs tend to target within the CDS more often than 3’UTRs (Figure 3-5A). However, a similar analysis is not possible for lncRNA targets. Therefore, we investigated how the miRNA binding sites are distributed in equally spaced intervals of the lncRNA length, i.e. within the first 20% of the lncRNA, from 21% to 40%, from 41 to 60%, etc. While the viral and cellular miRNAs that ligated to the 5’ end of the hybrid both showed uniform binding along the length of the lncRNA, miRNAs that ligated to the 3’ end of the molecule showed a significant bias for binding towards the 3’ end of the lncRNA (Figure 3-5B). Interestingly, a similar analysis of miRNA binding distribution, along the length of target mRNAs, revealed no such bias (Figure 3-5C), suggesting that the factors that govern where a miRNA binds within the length of the RNA are different for mRNAs and lncRNAs.
LncRNAs are Often Targeted by Non-Canonical Base-Pairing that is Stabilized by Compensatory 3’ Base Interactions

Next, we investigated the seed sequence dependent base-pairing of miRNA to its target RNA. To do this, we used the in silico folding information of each hybrid read generated by the Hyb program. We used the dots and parentheses diagram of RNA folding, which is a diagram where dots and parentheses signify the absence and presence of base-pairing interactions, respectively. We screened for binding patterns fitting four different classes of seed matches: 7mer (nt 2-8), 6mer (nt 2-7), 7mer w/ 1mm (nt 2-8 with 1 mismatch, mismatch not allowed at nt 8), 7mer w/ 2mm (nt 2-8 with 2 mismatches), and other (does not fall in any other category). We identified that > 40% of miRNA targeting events in the case of both mRNAs and lncRNAs were in the ‘other’ category suggesting that there is frequent non-canonical miRNA targeting (Figure 3-6A). This observation is consistent with the data reported in the original CLASH paper [126]. This percentage was > 50% for viral miRNAs, compared with 40% for cellular miRNAs. Moreover, lncRNAs showed higher levels of non-canonical targeting than mRNAs, for both viral and cellular miRNAs (Figure 3-6B). Since a significant percentage of miRNA binding was non-canonical, we searched for compensatory 3’ binding that allows for effective targeting [77]. Compensatory 3’ binding signifies base-pairing at the 3’ end of the miRNA which enhances binding interactions with the target in addition to seed-based base-pairing at the 5’ end of the miRNA. To do this, we checked for binding along the length of the miRNA starting at position (nt) 11 through the end and binned them into four categories: Absent (0 binding events), Weak (1-4 binding events), Moderate (5-7 binding events) and Strong (>7 binding events). For both lncRNA and mRNA targets, as the strength of the seed sequence decreased, the strength of 3’ compensatory base-pairing increased as expected (Figure 3-6C). We observed that in more than 60% of the targets that fell in the ‘other’ category, there was moderate to strong 3’ compensatory
base-pairing. Based on these observations, we conclude that lncRNAs undergo non-canonical
miRNA targeting more frequently than mRNAs, and the base-pairing with miRNA often relies
on increased binding towards the 3’ end of the miRNA.

**MicroRNAs Have Unique Binding Profiles along Their Length**

We questioned whether there are any positional biases for base-pairing along the length
of the miRNA, other than the seed sequence. We plotted binding frequencies along the length of
the miRNA and compared them in three ways: cellular vs. viral miRNAs, 5’ ligated vs. 3’ ligated
miRNAs, and miRNAs targeting mRNAs vs. lncRNAs (Figure 3-7A). The seed sequence was
more pronounced for 5’ cellular miRNAs targeting mRNAs compared with lncRNAs, consistent
with the results in Figure 3-6B. Similarly, we also see increased 3’ compensatory base-pairing in
miRNAs targeting lncRNAs and also all viral miRNAs. In spite of these subtle differences, the
binding profiles for miRNAs look very comparable for mRNA and lncRNA targets. However,
these plots have been averaged over all miRNAs and might not represent individual miRNA
binding profiles. Surprisingly, when we individually plotted the profile for every miRNA, we
saw that each miRNA tends to have a unique binding signature, which is closely comparable
between its mRNA and lncRNA targets. To demonstrate this, we show in Figure 3-7B the plots
for two viral miRNAs, miR-K12-6-5p and miR-K12-3, that represent the extreme cases of strong
seed-dependent and strong non-canonical base-pairing with its target. These suggest that each
miRNA has a preferential binding pattern, and miRNA binding does not drastically change
depending on the nature of the target.
Discussion

In this study, we show that KSHV and EBV miRNAs bind to and interact with cellular lncRNAs in a RISC-dependent manner. We used previously published Ago HITS-CLIP and Ago PAR-CLIP data of KSHV and EBV infected lymphoma cells to identify and catalog the putative lncRNA targets of KSHV and EBV miRNAs in lymphomas caused by these viruses [115, 121, 123, 124]. We found that most miRNA targets that are not mRNAs are indeed lncRNAs. Further, 37 lncRNA targets (using 7mer viral miRNA seed match criteria) were common to all datasets and thus are likely shared targets of KSHV and EBV. We also found eight lncRNAs previously implicated in lymphomas to be targeted by KSHV and EBV, and they probably play important roles in γ-herpesvirus-associated cancers. However, it is important to note that the overlap between different CLIP-seq studies for both mRNA and lncRNA targetomes was often less than 50%, suggesting that these methods are neither comprehensive nor unbiased. Furthermore, based on whether we chose to use a 7mer or 6mer seed sequence criteria to define a target, the number of lncRNA and mRNA targets varied largely, reflecting the subjectivity of defining miRNA targets when using a CLIP-seq based approach.

To overcome these limitations, we performed a CLASH analysis on KSHV infected endothelial cells and identified thousands of lncRNA-miRNA hybrids for cellular and viral miRNAs. Apart from identifying many lncRNA targets of KSHV miRNAs, this analysis also revealed that lncRNAs are widely targeted by cellular miRNAs. This provides strong evidence for global miRNA mediated lncRNA regulation, a regulatory relation that has been shown through very limited examples thus far [218, 279, 288]. Two studies previously showed that the cellular lncRNAs MALAT1 and UCA1 are downregulated by miR-9 and miR-1, respectively.
Our recent work on KSHV infected endothelial cells confirmed that viral miRNAs directly bind to and downregulate some of the cellular IncRNAs (Chapter 2).

In this work, we have used high-throughput methods and identified thousands of IncRNA-miRNA hybrids. However, the mechanism(s) of how miRNAs target IncRNAs and whether it is similar to how miRNAs target mRNAs remain to be worked out. Based on some preliminary information from our CLASH data, miRNA-IncRNA interaction seems remarkably similar to miRNA-mRNA interaction. However, we identified some minor, yet significant differences. First, the propensity of a miRNA to ligate at the 3’ end of a IncRNA during the CLASH procedure is higher than the propensity to ligate at the 3’ end of an mRNA. Second, we observed that miRNAs that ligate to the 3’ end of IncRNAs tend to often target the 3’ end of the IncRNA molecule. However, miRNAs that ligate to the 5’ end showed no such bias. Since not all IncRNAs get polyadenylated, we entertained the question whether the presence of a poly A tail dictates 5’ vs. 3’ ligation of the miRNA. We found no such pattern. Third, we found that miRNAs often target IncRNAs via non-canonical base-pairing rather than seed-dependent base-pairing. While this raises questions about the mechanism of how IncRNAs are loaded into Ago, it is important to consider that almost 40% of mRNAs also undergo non-canonical targeting. We speculate that these differences could be due to the differences in mRNA and IncRNA accessibility due to secondary structures and associated RBP pools including ribosomes, which could significantly influence the accessibility around miRISC complexes. Some studies have shown that a majority of cytosolic IncRNAs are ribosome bound [289, 290], and some of those indeed encode micro-peptides [162]; hence, these would not be true IncRNAs. It would be interesting to investigate whether the IncRNAs that behave similarly to mRNAs in their miRNA binding properties are polyribosome bound. One limitation, particularly in the context of
lncRNAs, is that CLASH identifies miRNA-lncRNA interactions with high confidence but does not provide any insight into the type of regulation that results from this molecular interaction. It is possible that certain lncRNAs serve as sponges for miRNAs, both cellular and viral, thereby derepressing miRNA-targeted mRNAs [154]. To address this question, we are currently performing RNA-seq experiments in KSHV infected cells using viruses that are missing specific miRNAs.

Most lncRNAs interacting with viral miRNAs were present in both the nucleus and the cytoplasm, but we also identified a few that were exclusively nuclear. We have previously shown (Chapter 2) that Ago protein resides in the nucleus of KSHV infected endothelial and B cells, and other studies have detected the entire RNAi machinery is present in the nucleus of HeLa cells [291]. Together, these results suggest that miRNA-lncRNA interactions occur in both the nucleus and the cytoplasm, and more experiments are warranted to understand this mechanistically.

In summary, our computational analysis of deposited HITS-CLIP and PAR-CLIP datasets on EBV and KSHV infected lymphoma cells combined with our recently obtained CLASH data in endothelial cells has identified and cataloged thousands of putative miRNA-lncRNA interactions. Importantly, using CLASH we identified viral and cellular miRNA-lncRNA hybrids. These data provide a resource and a starting point to decipher the biological relevance of such interactions between short and long noncoding RNAs, but at present we have just scratched the surface. Understanding how miRNA targeting of lncRNAs or lncRNA-dependent sponging of miRNA influences gene expression will likely require systems biology approaches. Given the regulatory complexity involved, tackling these questions with viruses
provides a good strategy, since they express only a limited number of viral miRNAs and are easily genetically manipulated.

**Material and Methods**

**Published datasets.** All relevant Sequence Read Archive (SRA) files were downloaded from NCBI. For HITS-CLIP analysis of EBV infected cells (Project: SRP068881), files SRR3122404- SRR3122410 were used. For PAR-CLIP analysis of EBV infected cells (Project: SRP008216), SRR343334- SRR343337 were used. For HITS-CLIP analysis of KSHV infected cells (Project: SRP068881), files SRR580352- SRR580358 were used. Each sequencing file was treated as an independent replicate, although this project had two technical replicates for the third biological replicate of BCBL-1 sample. For PAR-CLIP analysis of KSHV infected cells (Project: SRP016130), files SRR592685- SRR592689 were used.

**Processing of raw reads.** All fastq files used in this study contained single-end sequencing reads. All HITS-CLIP and PAR-CLIP reads were processed using the fastx-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). First, the reads were quality filtered such that at least 80% of the bases in any read had a quality score greater than 20. Then, the appropriate barcodes were removed and reads longer than 18 nt were kept for further analysis. Reads from the CLASH data were trimmed to remove 5’ and 3’ adapters using Trimmomatic 3.0 (http://www.usadellab.org/cms/?page=trimmomatic). All fastq files were quality checked before and after trimming using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

**Alignment to the human genome.** Processed HITS-CLIP and PAR-CLIP reads were aligned to the human genome (hg19) using the Bowtie program (http://bowtie-bio.sourceforge.net/index.shtml). Reads were allowed to have two mismatches in the case of HITS-CLIP and three mismatches in the case of PAR-CLIP. Only the best alignment was
reported for each read that aligned multiple times. The sam files were then converted to bam files and sorted using samtools v1.4 (http://samtools.sourceforge.net/).

**PIPE-CLIP and annotation of BED files.** The publicly available CLIP-seq analysis pipeline called PIPE-CLIP was used to call clusters [280]. PIPE-CLIP handles HITS-CLIP and PAR-CLIP data differently and accounts for the inherent differences in techniques. All sorted bam files were analyzed using PIPE-CLIP, using method 2 for PCR duplicate removal. The output file of enriched clusters for each sample was then annotated using Bedtools v2.25.0 (http://bedtools.readthedocs.io/en/latest/). The reference BED file for this annotation was created from the GENCODE V19 dataset. The gtf file was first converted to a BED file and only ‘transcript’ annotations were retained. The reference file was pre-processed using custom R scripts to eliminate possible duplicates and redundant transcript and exon information.

**Hyb pipeline.** The Hyb pipeline was used to call chimeras (hybrid reads with part miRNA and part lncRNA/mRNA) from the Fastq files of the CLASH analysis [127]. The reference database for the Hyb pipeline was created in-house by downloading cDNA sequences of long-noncoding RNAs from Ensembl biomart. The perl script for this dataset download is available upon request.

**Custom R Scripts.** All analyses of annotated HITS-CLIP and PAR-CLIP files, and chimeras obtained from the Hyb pipeline for CLASH data, were performed using custom R Scripts (available upon request).

**Replicates and statistics.** For all graphs, when averaged over cellular miRNAs, there were N=9 samples (three biological replicates each of uninfected, wt-KSHV infected and Δ11-KSHV infected cells). When averaged over viral miRNAs, there were N=6 samples (three
biological replicates each of wt-KSHV infected and Δ11-KSHV infected cells). All p-values were obtained using student’s t-test assuming unequal variances (two-tailed).
Figure 3-1: Distribution of viral miRNA targets identified from all CLIP datasets. Counts of target RNAs for KSHV/EBV miRNAs from all datasets were summed and their percentage distribution is shown as a pie chart for mRNA, IncRNA, unannotated RNA, and small RNA categories. (A) Results from 7mer seed matched targets. (B) Results from 6mer seed matched targets. Total counts representing 100% are indicated.
Figure 3-2: Venn diagrams of overlapping targets between different CLIP datasets. KSHV and EBV miRNA targets identified from the four CLIP-seq studies were pooled by study and Venn diagrams were generated. The mRNA and lncRNA targets are grouped by the presence of 7mer vs. 6mer seed matches for miRNAs.
Figure 3-3: Sub-cellular distribution of lncRNA targets of cellular and viral miRNAs. (A) Distribution of all RNAs in the ANGIOGENES database between the four sub-cellular fractions nuclear, polyA-; nuclear, polyA+; cytosolic, polyA-; and cytosolic, polyA+ (A, B, C and D, respectively). (B and C) Distribution of lncRNA targets of (B) cellular miRNAs and (C) viral miRNAs between these four sub-cellular fractions.
Figure 3-4: miRNAs ligate more frequently to the 3’ end of target lncRNAs than to the 3’ end of target mRNAs. (A) Distribution of percentage of hybrids with miRNA at either the 5’ or at the 3’ end of the hybrid (B) Percentage distribution of miRNA ligation orientation for mRNA and lncRNA targets grouped by cellular miRNAs or by viral miRNAs (C) Percentage distribution of miRNA ligation orientation for cellular miRNAs and viral miRNAs grouped by mRNA targets or by lncRNAs. Error bars represent mean ± SD. p-values: * < 0.01; ** < 0.005, n.s. = not significant.
Figure 3-5: miRNA binding distribution along the length of the target RNA. Distribution of miRNA binding events along the length of (A) mRNAs, binned by mRNA features such as 5’ UTR, CDS and 3’ UTR (5’UTR-CDS and CDS-3’UTR represent reads overlapping both these features), (B) IncRNAs, binned by % of total IncRNA length and (C) mRNAs, binned by % of total mRNA length. Distributions are grouped by miRNA ligation end (3’ miRNA vs. 5’ miRNA) and miRNA source (cellular vs. viral). Total counts representing 100% are shown in parentheses. Error bars represent mean ± SD.
Figure 3-6: miRNAs often bind lncRNAs via non-canonical base-pairing. Distribution of seed based binding (based on in silico folding of hybrids) events for (A) mRNA targets and (B) lncRNA targets, grouped by miRNA ligation end (3’ miRNA vs. 5’ miRNA) and miRNA source (cellular vs. viral). Total hybrid counts representing 100% are shown in parentheses. Error bars represent mean ± SD. (C) Heatmap representing the frequency of 3’ compensatory base-pairing given every seed match type for lncRNA and mRNA targets. The scale for the heatmaps is percentage. Refer to the text for definition of each category of seed match and 3’ compensatory binding.
Figure 3-7: Percentage binding frequencies along the length of the miRNAs. (A) Percentage of targets having a binding event is plotted for every nucleotide position along the length of the miRNA, averaged over multiple miRNAs. These are grouped using three criteria: target type (mRNA vs. IncRNA), miRNA ligation end (5' miRNA vs. 3' miRNA) and miRNA source (cellular vs. viral). (B) Two representative plots for KSHV miRNAs, miR-K12-6-5p and miR-K12-3, grouped by mRNA and IncRNA targets.
Table 3-1. Summary of CLIP-seq studies of KSHV and/or EBV infected lymphoma cells.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Virus</td>
<td>KSHV</td>
<td>KSHV,EBV</td>
<td>EBV</td>
<td>EBV</td>
</tr>
<tr>
<td>Method</td>
<td>HITS-CLIP</td>
<td>PAR-CLIP</td>
<td>HITS-CLIP</td>
<td>PAR-CLIP</td>
</tr>
<tr>
<td>Cell lines</td>
<td>BC-3, BCBL-1</td>
<td>BC-1, BC-3</td>
<td>Jijoye</td>
<td>LCL</td>
</tr>
<tr>
<td>Virus source</td>
<td>Natural infection</td>
<td>Natural infection</td>
<td>Natural infection</td>
<td>Laboratory strain B95.8</td>
</tr>
<tr>
<td>Biological replicates</td>
<td>3 Bio-reps per cell line</td>
<td>2 Bio-reps per cell line</td>
<td>3 Bio-reps per antibody</td>
<td>1 Bio-rep per virus variant</td>
</tr>
<tr>
<td>Sequencing platform</td>
<td>Illumina GA IIx</td>
<td>Illumina GA IIx</td>
<td>Illumina GA IIx</td>
<td>Illumina GA IIx</td>
</tr>
<tr>
<td>Alignment read length</td>
<td>&gt; 15 nt</td>
<td>&gt; 13 nt</td>
<td>&gt; 18 nt</td>
<td>&gt; 13 nt</td>
</tr>
<tr>
<td>Alignment program</td>
<td>CLIPZ</td>
<td>Bowtie</td>
<td>in-house</td>
<td>Bowtie</td>
</tr>
<tr>
<td>Mismatches allowed</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Cluster identification program</td>
<td>CLIPZ</td>
<td>PARalyzer</td>
<td>in-house</td>
<td>PARalyzer</td>
</tr>
<tr>
<td>Target type reported</td>
<td>genes</td>
<td>3' UTRs</td>
<td>3' UTRs</td>
<td>3' UTRs</td>
</tr>
<tr>
<td>Reference</td>
<td>[115]</td>
<td>[123]</td>
<td>[121]</td>
<td>[124]</td>
</tr>
</tbody>
</table>
Table 3-2. Number of clusters containing 7mer or 6mer seed match for viral miRNAs identified from CLIP-seq datasets.

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Replicate</th>
<th>Virus</th>
<th>mRNA</th>
<th>lncRNA</th>
<th>Unannotated small RNA</th>
<th>Total</th>
<th>mRNA</th>
<th>lncRNA</th>
<th>Unannotated small RNA</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jijoye</td>
<td>1</td>
<td>EBV</td>
<td>4515</td>
<td>2726</td>
<td>270</td>
<td>7525</td>
<td>14400</td>
<td>8607</td>
<td>907</td>
<td>23977</td>
</tr>
<tr>
<td>Jijoye</td>
<td>2</td>
<td>EBV</td>
<td>7865</td>
<td>4827</td>
<td>667</td>
<td>13389</td>
<td>23423</td>
<td>13927</td>
<td>1812</td>
<td>39242</td>
</tr>
<tr>
<td>Jijoye</td>
<td>3</td>
<td>EBV</td>
<td>25561</td>
<td>15326</td>
<td>1416</td>
<td>42410</td>
<td>77896</td>
<td>46517</td>
<td>4508</td>
<td>129163</td>
</tr>
<tr>
<td>Jijoye</td>
<td>4</td>
<td>EBV</td>
<td>4377</td>
<td>2537</td>
<td>256</td>
<td>7217</td>
<td>14003</td>
<td>7881</td>
<td>870</td>
<td>22871</td>
</tr>
<tr>
<td>Jijoye</td>
<td>5</td>
<td>EBV</td>
<td>14429</td>
<td>7829</td>
<td>1019</td>
<td>23353</td>
<td>51133</td>
<td>27505</td>
<td>4774</td>
<td>83590</td>
</tr>
<tr>
<td>Jijoye</td>
<td>6</td>
<td>EBV</td>
<td>7780</td>
<td>4326</td>
<td>342</td>
<td>12518</td>
<td>26230</td>
<td>14528</td>
<td>1137</td>
<td>42052</td>
</tr>
<tr>
<td>EF3DAGO2</td>
<td>1</td>
<td>EBV</td>
<td>528</td>
<td>262</td>
<td>79</td>
<td>884</td>
<td>1516</td>
<td>665</td>
<td>220</td>
<td>2424</td>
</tr>
<tr>
<td>LCL35</td>
<td>1</td>
<td>EBV</td>
<td>427</td>
<td>240</td>
<td>76</td>
<td>757</td>
<td>1051</td>
<td>514</td>
<td>210</td>
<td>1795</td>
</tr>
<tr>
<td>LCLBAC</td>
<td>1</td>
<td>EBV</td>
<td>164</td>
<td>75</td>
<td>17</td>
<td>265</td>
<td>341</td>
<td>170</td>
<td>56</td>
<td>580</td>
</tr>
<tr>
<td>LCLBACD1</td>
<td>1</td>
<td>EBV</td>
<td>156</td>
<td>93</td>
<td>33</td>
<td>295</td>
<td>423</td>
<td>230</td>
<td>99</td>
<td>771</td>
</tr>
<tr>
<td>LCLBACD3</td>
<td>1</td>
<td>EBV</td>
<td>241</td>
<td>150</td>
<td>55</td>
<td>461</td>
<td>743</td>
<td>408</td>
<td>180</td>
<td>1360</td>
</tr>
<tr>
<td>BC-3</td>
<td>1</td>
<td>KSHV</td>
<td>387</td>
<td>251</td>
<td>35</td>
<td>684</td>
<td>1653</td>
<td>930</td>
<td>140</td>
<td>2762</td>
</tr>
<tr>
<td>BC-3</td>
<td>2</td>
<td>KSHV</td>
<td>347</td>
<td>184</td>
<td>25</td>
<td>562</td>
<td>1485</td>
<td>887</td>
<td>93</td>
<td>2496</td>
</tr>
<tr>
<td>BC-3</td>
<td>3</td>
<td>KSHV</td>
<td>379</td>
<td>188</td>
<td>27</td>
<td>604</td>
<td>1844</td>
<td>1032</td>
<td>143</td>
<td>3051</td>
</tr>
<tr>
<td>BCBL-1</td>
<td>1</td>
<td>KSHV</td>
<td>1633</td>
<td>1009</td>
<td>77</td>
<td>2741</td>
<td>6936</td>
<td>3909</td>
<td>402</td>
<td>11373</td>
</tr>
<tr>
<td>BCBL-1</td>
<td>2</td>
<td>KSHV</td>
<td>2349</td>
<td>1389</td>
<td>87</td>
<td>3851</td>
<td>9708</td>
<td>5741</td>
<td>463</td>
<td>16065</td>
</tr>
<tr>
<td>BCBL-1</td>
<td>3</td>
<td>KSHV</td>
<td>670</td>
<td>371</td>
<td>24</td>
<td>1074</td>
<td>2809</td>
<td>1646</td>
<td>137</td>
<td>4667</td>
</tr>
<tr>
<td>BCBL-1</td>
<td>4</td>
<td>KSHV</td>
<td>805</td>
<td>463</td>
<td>29</td>
<td>1308</td>
<td>3046</td>
<td>1767</td>
<td>140</td>
<td>5042</td>
</tr>
<tr>
<td>BC-1</td>
<td>1</td>
<td>EBV</td>
<td>3130</td>
<td>1496</td>
<td>188</td>
<td>4831</td>
<td>9951</td>
<td>4458</td>
<td>524</td>
<td>15003</td>
</tr>
<tr>
<td>BC-1</td>
<td>2</td>
<td>KSHV</td>
<td>908</td>
<td>435</td>
<td>42</td>
<td>1392</td>
<td>4763</td>
<td>2207</td>
<td>242</td>
<td>7235</td>
</tr>
<tr>
<td>BC-1</td>
<td>3</td>
<td>EBV</td>
<td>3291</td>
<td>1511</td>
<td>185</td>
<td>5006</td>
<td>10310</td>
<td>4606</td>
<td>548</td>
<td>15542</td>
</tr>
<tr>
<td>BC-1</td>
<td>2</td>
<td>KSHV</td>
<td>1020</td>
<td>476</td>
<td>49</td>
<td>1551</td>
<td>4964</td>
<td>2284</td>
<td>243</td>
<td>7518</td>
</tr>
<tr>
<td>BC-3</td>
<td>1</td>
<td>KSHV</td>
<td>510</td>
<td>228</td>
<td>40</td>
<td>781</td>
<td>2684</td>
<td>1262</td>
<td>174</td>
<td>4133</td>
</tr>
<tr>
<td>BC-3</td>
<td>2</td>
<td>KSHV</td>
<td>627</td>
<td>265</td>
<td>41</td>
<td>938</td>
<td>2782</td>
<td>1290</td>
<td>170</td>
<td>4257</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>82099</td>
<td>46657</td>
<td>5079</td>
<td>134397</td>
<td>274094</td>
<td>152971</td>
<td>18192</td>
<td>446969</td>
</tr>
</tbody>
</table>


Table 3-3. Percentage overlap between targets identified using CLIP-seq studies of KSHV and/or EBV infected lymphoma cells.

<table>
<thead>
<tr>
<th>mRNA</th>
<th>7mer</th>
<th>6mer</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBV, HITS-CLIP</td>
<td>100</td>
<td>87.67</td>
</tr>
<tr>
<td>EBV, PAR-CLIP</td>
<td>15.75</td>
<td>100</td>
</tr>
<tr>
<td>KSHV, HITS-CLIP</td>
<td>13.35</td>
<td>44.61</td>
</tr>
<tr>
<td>KSHV, PAR-CLIP</td>
<td>5.71</td>
<td>35.49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>lncRNA</th>
<th>7mer</th>
<th>6mer</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBV, HITS-CLIP</td>
<td>100</td>
<td>80.08</td>
</tr>
<tr>
<td>EBV, PAR-CLIP</td>
<td>18.09</td>
<td>100</td>
</tr>
<tr>
<td>KSHV, HITS-CLIP</td>
<td>23.87</td>
<td>36.63</td>
</tr>
<tr>
<td>KSHV, PAR-CLIP</td>
<td>10.88</td>
<td>29.92</td>
</tr>
</tbody>
</table>

Numbers represent the percentage of targets recovered by the study in row label, originally identified by the study listed in the column label.
Table 3-4. Number of miRNA-lncRNA hybrids found in the CLASH experiment.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Replicate</th>
<th>Cellular miRNA</th>
<th>Viral miRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δ11-KSHV</td>
<td>1</td>
<td>31735</td>
<td>1094</td>
</tr>
<tr>
<td>Δ11-KSHV</td>
<td>2</td>
<td>35308</td>
<td>1068</td>
</tr>
<tr>
<td>Δ11-KSHV</td>
<td>3</td>
<td>14269</td>
<td>691</td>
</tr>
<tr>
<td>Uninfected</td>
<td>1</td>
<td>21697</td>
<td>1</td>
</tr>
<tr>
<td>Uninfected</td>
<td>2</td>
<td>36288</td>
<td>3</td>
</tr>
<tr>
<td>Uninfected</td>
<td>3</td>
<td>20109</td>
<td>13</td>
</tr>
<tr>
<td>wt-KSHV</td>
<td>1</td>
<td>20457</td>
<td>1578</td>
</tr>
<tr>
<td>wt-KSHV</td>
<td>2</td>
<td>24869</td>
<td>2281</td>
</tr>
<tr>
<td>wt-KSHV</td>
<td>3</td>
<td>44144</td>
<td>4223</td>
</tr>
<tr>
<td>IncRNA</td>
<td>Function</td>
<td>KSHV-miR-K12-</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>DLEU2</td>
<td>Host gene for tumor suppressor miRNAs miR-15a and miR-16-1</td>
<td>4-3p</td>
<td></td>
</tr>
<tr>
<td>GAS5</td>
<td>Downregulated in multiple different cancers</td>
<td>3, 4-3p, 8*</td>
<td></td>
</tr>
<tr>
<td>H19</td>
<td>Plays a role in tumor initiation, progression and metastasis by interacting with the p53 pathway</td>
<td>9*</td>
<td></td>
</tr>
<tr>
<td>HOTAIRM1</td>
<td>Modulates gene expression of cell adhesion molecules</td>
<td>3, 6-3p</td>
<td></td>
</tr>
<tr>
<td>KCNQ1OT1</td>
<td>Known to play a role in breast, kidney and colorectal cancers</td>
<td>4-3p, 8, 10a</td>
<td></td>
</tr>
<tr>
<td>MALAT1</td>
<td>Upregulated in several cancers; associated with increased proliferation and metastasis</td>
<td>4-3p, 7, 8, 8*, 9*, 10a, 10b, 11, 12*</td>
<td></td>
</tr>
<tr>
<td>PTENP1</td>
<td>Acts as a tumor suppressor since it is a pseudogenes for the tumor suppressor PTEN</td>
<td>3, 4-3p, 11</td>
<td></td>
</tr>
<tr>
<td>TUG1</td>
<td>Downregulated in NSCLC and regulates CELF1 by binding to PRC2 complexes</td>
<td>7, 10b</td>
<td></td>
</tr>
<tr>
<td>UCA1</td>
<td>Promotes cell cycle progression via PI3K-AKT pathway; also aids pRb1 and SET1A interplay</td>
<td>6-3p</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 4
CONCLUSIONS AND FUTURE DIRECTIONS

The work presented in this dissertation expands our current understanding of KSHV biology. We have demonstrated that cellular lncRNAs contribute to an additional layer of regulatory complexity. The pathways involved are likely to play roles in γ-herpesvirus driven cancers. In this chapter, we highlight some of the major conclusions from this work and provide directions for further investigation. The immediate relevance of the results from this study and their broader impacts on the field are also discussed in this chapter.

**Major Conclusions**

**KSHV miRNAs and latency proteins target host cellular lncRNAs in endothelial cells.** Our work in KSHV infected endothelial cells showed that KSHV miRNAs can downregulate expression of a subset of cellular lncRNAs. We identified 126 putative lncRNA targets of viral miRNAs in TIVE cells using microarray technology. Our validation experiments, using miRNA mimics in a virus-free context, confirmed such targeting. We also discovered that KSHV-encoded latency proteins vCyclin and vFLIP downregulate ANRIL, in addition to ANRIL being targeted by KSHV miRNAs. Further, we showed that vCyclin and Kaposin upregulate UCA1 synergistically in endothelial cells transfected with plasmids encoding these proteins. Thus, we present strong evidence that KSHV-encoded latency factors (miRNAs and latency-associated proteins) contribute to alterations in the lncRNA profile of the infected cell, and they also deregulate specific lncRNAs in a targeted manner.

**miRNA-mediated lncRNA targeting proceeds via direct interaction between miRNAs and IncRNAs.** Altered expression levels of lncRNAs observed using microarray technology do not immediately confirm a direct interaction between miRNAs and lncRNAs. Our
pull-down assays using transfected miRNA mimics demonstrated that viral miRNAs indeed directly bind to and interact with host cellular lncRNAs. We used double-stranded miRNA mimics which require a RISC-mediated unwinding and loading to form miRISC. These results, along with our qPCR experiments confirming the knockdown of expression of specific lncRNAs by these miRNAs, confirmed that miRNAs bind to and downregulate lncRNA expression in a RISC-dependent manner. Further, we analyzed CLASH datasets from endothelial cells for miRNA-lncRNA interactions. Since Ago CLASH is based on cross-linking and ligation of RNAs in close proximity to each other followed by pull-down of Ago-bound RNAs, we can conclude that these miRNA-lncRNA interactions were indeed direct and RISC-dependent [126, 276].

**KSHV and EBV miRNAs target host lncRNAs in lymphoma cells.** Reanalysis of published CLIP-seq datasets from lymphoma cells infected with KSHV and/or EBV identified hundreds of putative lncRNA targets for both KSHV and EBV miRNAs [115, 121, 123, 124]. Although these viruses do not share any miRNA orthologs, we found that a subset of lncRNAs were targeted by both viruses [100]. This concurs with previous reports showing that KSHV and EBV target similar host pathways using different approaches related to miRNAs [100, 283, 284]. Moreover, these results suggest that targeting host lncRNAs is a shared feature of γ-herpesvirus miRNAs.

**The viral miRNA targetome includes nuclear resident lncRNAs.** A subset of the putative and confirmed lncRNA targets of KSHV and EBV miRNAs are localized to the nucleus. We determined this by comparing our lists, from the microarray, CLIP-seq and CLASH analyses, to the literature and an online database called ANGIOGENES [230]. miRNA maturation and function have been classically limited to the cytoplasm by virtue of the miRNA biogenesis pathway (reviewed in [61]). However, we based our investigation on recent reports
that have demonstrated the presence of mature miRNAs and RISC components in the nucleus [291]. In fact, one study showed that a lncRNA named MALAT1 is downregulated by human miRNA, miR-9, in the nuclei of Hodgkin’s lymphoma and glioblastoma cells [279]. In KSHV-infected lymphoma cells, we could detect both mature viral miRNAs and Ago proteins in the nucleus, suggesting that nuclear lncRNAs could indeed be targets of KSHV miRNAs. We were also able to detect Ago in the nucleus of endothelial cells. The extension of this finding to EBV miRNAs remains to be investigated.

**lncRNA targets of KSHV and EBV might be direct contributors to virus-driven cancers.** All our attempts to identify lncRNA targets of KSHV and EBV which might be of importance in the development of cancer were severely restricted by the limited number of functionally characterized lncRNAs. Despite this challenge, we used currently available information to catalog which lncRNA targets of these viruses have relevance to cancer phenotypes [187, 234]. We have demonstrated the relevance of UCA1, a lncRNA strongly upregulated by KSHV, to development of sarcomas by showing that knockdown of UCA1 in KSHV-infected cells inhibits their proliferation and migration. UCA1 exemplifies the importance of lncRNA deregulation for pathogenesis of γ-herpesviruses.

**miRNA-mRNA and miRNA-lncRNA interactions are comparable, yet have notable differences.** Analysis of CLASH data allowed us to investigate whether miRNA-lncRNA interactions behave essentially the same as miRNA-mRNA interactions. We found that these two types of interactions were very similar in most of the features we investigated. However, lncRNAs seemed to be a little more flexible in most features, for example, ligation of the miRNA to the 5’ end of a CLASH hybrid was more strictly adhered to by mRNAs than lncRNAs. Also, miRNAs tend to bind lncRNAs via non-canonical base-pairing more frequently
than mRNAs. These observations are merely suggestive at this point and require experimental investigation to delineate how miRNAs bind to IncRNAs, and what the functional consequences are.

**Future Directions**

**Confirmation of cataloged putative targets.** The work in this dissertation has focused on a relatively unexplored area of research, miRNA-mediated IncRNA targeting. We have demonstrated this novel paradigm of regulation in the context of KSHV infection, and have strived to identify putative IncRNA targets of different viruses in different cell types using high-throughput methods. While we confirmed a few of the IncRNA targets identified in KSHV infected endothelial cells, every candidate target needs to be validated using biochemical and genetic experiments before further hypotheses are formulated. Further, we have shown that none of these widely-employed methods are comprehensive or unbiased. Importantly, different methods address different questions, for example, microarray experiments interrogate the expression levels of IncRNAs, but CLASH identifies miRNA-IncRNA binding events. Thus, careful interpretation of these results and experimental validation of viable IncRNA candidates are essential for future work on determining mechanisms.

**Characterization of novel targets.** One of the major challenges in finding immediate applications for the results reported in this study is that most of the putative IncRNA targets identified by us require functional characterization. New techniques like Chromatin Isolation by RNA Purification (ChIRP) are continuously being invented to understand and characterize IncRNAs [292, 293]. However, narrowing down candidates to pursue is a challenge by itself. In this work, we have strived to find overlaps between datasets with the idea that IncRNAs targeted by multiple viruses or identified in more than one cell line might be of higher importance. For
coding genes, phylogenetic conservation has been classically used to stratify functional importance, but how well this applies to lncRNAs remains debatable [294]. In fact, \(\gamma\)-herpesviruses are extremely species-specific, thus lncRNA conservation might contribute little to finding high-confidence lncRNA targets of these viruses. Some other reasonable starting points might include identifying the closest protein coding gene to the lncRNA of interest, or investigating co-expression networks with coding genes [139]. We performed preliminary experiments on two of the lncRNA targets identified based on the microarray, Linc000607 and Loc541472, and observed that changes in their expression level were correlated with expression changes in the nearest protein-coding genes, FN1 and hIL-6, respectively (Figure 4-1). Co-expression based studies are often used to explore lncRNA functions [295, 296]. However, it is important to distinguish correlation from causation when using such approaches.

Delineating mechanistic links between lncRNA targets and cancer phenotypes. For those lncRNAs targets whose functions are known or become characterized in the future, it is important to delineate how these lncRNAs contribute to viral pathogenesis and/or tumorigenesis. For example, our work shows that inhibition of UCA1 in infected cells inhibits proliferation and migration of these cells. It has been shown that HIF1-\(\alpha\) upregulates UCA1, which helps survival in a hypoxic environment [297]. In addition, UCA1 also promotes the Warburg effect [252]. HIF1-\(\alpha\) upregulation and the Warburg effect have both been reported in KSHV-infected tumor cells, suggesting UCA1 might serve as the missing piece of the puzzle [253, 298]. Similar hypotheses need to be addressed for UCA1 and other lncRNA targets to further our understanding of how KSHV drives tumorigenesis.

Mechanistic insights into miRNA-mediated lncRNA regulation. miRNA-mediated lncRNA regulation is a novel regulatory paradigm. Our initial observations from CLASH
experiments suggest that this regulation shares several features with miRNA-mediated mRNA regulation. Nevertheless, we observed significant differences as well. Systematic investigation of miRNA-lncRNA interaction and lncRNA loading into Ago complexes is imperative to completely understand this regulatory mechanism. Based on our observations, it might be worthwhile to investigate the differences in secondary structures and the bound RBP pools between Ago-bound lncRNAs vs. Ago-bound mRNAs. Association with polyribosomes might also contribute to differences in how miRNA-mediated lncRNA regulation occurs. miRNA-lncRNA interactions that occur in the nucleus might be an added layer of complexity that needs to be addressed through experiments that distinguish them from cytosolic miRNA-lncRNA interactions.

Other alternative interpretations of miRNA-lncRNA interactions: sponges. While we have focused on identifying lncRNAs that are inhibited by miRNAs, this does not preclude other possible regulatory interactions between these two classes of RNAs. One of the popular interpretations of miRNA-lncRNA interaction involves lncRNAs acting as miRNA sponges thus inhibiting miRNA function [154]. In other words, this results in de-repression of mRNA targets of the miRNA. Several pseudogenes and circular RNAs have been shown to function as miRNA sponges [299, 300]. It would be interesting to investigate how many of the lncRNA targets identified from CLASH datasets act as viral miRNA sponges. These two regulatory interactions, miRNA-mediated lncRNA repression and lncRNAs acting as miRNA sponges, are not mutually exclusive. Thus, using biochemical experiments, such as CLASH, might not help distinguish these two possibilities. Expression differences in lncRNAs measured by RNA-seq of cells infected with viral miRNA mutants could help narrow down lncRNAs targeted by miRNAs. However, most lncRNAs are often bound by more than one viral miRNA, so RNA-seq of cells
transfected with pools of miRNA mimics could be a better approach. To specifically address the question of sponges, CRISPR/Cas9 based lncRNA mutagenesis screening could be used. Simultaneous measurement of expression levels of viral miRNAs and mRNAs using RNA-seq in cells with a specific lncRNA knocked out could identify lncRNA sponges of KSHV miRNAs.

**Immediate Relevance**

KSHV was identified in 1994 by Chang and Moore and was characterized shortly afterwards [2, 301]. After two decades of research on KSHV driven cancers, we still do not completely understand how the viral latency factors cause cancer [7]. Thus far, all research has exclusively focused on how KSHV deregulates proteins and protein coding mRNAs of the cell. Given the importance of lncRNAs in cancers, understanding how KSHV encoded latency proteins and miRNAs affect cellular lncRNAs would significantly advance our progress towards identifying tumorigenic determinants of the virus. Excitingly, lncRNA targets also seem to be promising therapeutic targets, especially given the recent advances in RNA therapeutics [302].

**Broader Impact**

Latent replication is a characteristic feature of most herpesviruses. Interestingly, in the past decade several herpesviruses have been shown to encode miRNAs, often in their latency associated regions, thus expressing miRNAs during latency (reviewed in detail here [272]). In this dissertation, using microarray and CLASH studies, we have shown that KSHV miRNAs target lncRNAs in endothelial cells. We further extended this to EBV by analyzing putative lncRNA targets of EBV miRNAs from CLIP-seq studies of EBV infected lymphoma cells. It is possible that other herpesviral miRNAs also target lncRNAs in different cell types they infect, such as trigeminal ganglion for herpes simplex virus (HSV), T-lymphocytes for herpesvirus
saimiri (HVS), etc. Identifying the lncRNA targets of herpesviral miRNAs could help understand the variety of diseases caused by these viruses. Further, this could also lead us to identify relevant lncRNAs common to herpesviral infections.

Systems biology approaches have identified thousands of lncRNAs through projects such as ENCODE, FANTOM, etc., a large number of which remain to be characterized [240, 303]. Meanwhile, whether the pervasive transcription of the human genome contributes to functional RNAs or if the transcripts are junk RNA has been a topic of hot debate [304]. LncRNAs deregulated in diseases are probably the best candidates for functional characterization, as their aberrant expression is known to be associated with a disease phenotype. LncRNAs deregulated by KSHV or EBV infection could have functional relevance for either viral infection or progression of the tumorigenic phenotype. Characterization of the identified lncRNA targets could find important players in cellular pathways such as the cell cycle, inflammation, apoptosis, etc.

From the analysis of CLASH datasets and CLIP-seq datasets, we identified thousands of putative lncRNA targets for cellular miRNAs. This suggests that miRNA-mediated lncRNA regulation is a global phenomenon that happens in cells irrespective of viral infection. To further study this novel phenomenon, using viral infection as a tool could prove useful. Landmark discoveries in eukaryotic molecular biology have come from the study of viruses, including splicing and methylation of the 5’ cap during mRNA processing [305-307]. Viruses encode a limited number of miRNAs and allow for easy genetic manipulation, making them useful tools to further understand this novel layer of regulation of gene expression.

LncRNAs continue to emerge as an important class of regulatory RNAs in the cell. The functions performed by this broad class of RNAs span regulation at the epigenetic, transcriptional, translational and post-translational levels. Given the diversity of lncRNAs, their
roles in the extremely complex gene regulatory networks of cells seem to have infinite possibilities at our current level of understanding, and it is safe to claim that we do not yet comprehend all the ramifications of aberrant lncRNA expression. Nevertheless, with novel techniques involving large-scale genome-editing, and Single-Cell sequencing, along with huge improvements in high-throughput sequencing depths and platforms, we are prepared to uncover the layers of complexity in lncRNA biology. Novel techniques and technological advances, combined with current data science and deep learning based computation, will enable us to make great strides towards understanding lncRNAs in an expanded picture of gene regulation.
Figure 4-1: Examples of co-regulation of lncRNAs and neighboring mRNAs. All bar graphs show the mean values ± SEM after normalization to GAPDH (n = 3), unless specified otherwise. (A) Loc541472 and hIL-6 expression in Uninfected, wt-KSHV-infected and Δcluster-KSHV-infected cells measured by qRT-PCR. (B) Linc000607 and FN1 expression in Uninfected, wt-KSHV-infected and Δcluster-KSHV-infected cells measured by qRT-PCR. The genome (blue) arrangement of the lncRNA (green) in reference to the mRNA (maroon) is shown in the cartoon below. p-values: ** < 0.01; *** < 0.005; and **** < 0.0005.
LIST OF REFERENCES


BIOGRAPHICAL SKETCH

Sunantha Sethuraman was born in Chennai, a metropolitan city in India, in the year of 1991. She was born in a family of commerce and economics majors, with one older brother who is employed as an auditor in Chennai, India. She finished her schooling from Modern Senior Secondary School, Chennai, in 2008, ranking first in the country in the Central Board of Secondary Education (CBSE) exams. She graduated with a Bachelor of Technology (B.Tech.) in Industrial Biotechnology from Anna University, Chennai, in 2012. She was awarded a gold medal for finishing first in the class of 2012. During her undergraduate years, she was awarded the prestigious DAAD-WISE fellowship from the German Academic Exchange Service to do a two-month summer internship in the lab of Dr. Thomas Magin, University of Leipzig, Germany. She then moved to the United States in 2012 to join the Interdisciplinary Program in Biomedical Sciences (IDP-BMS) program at the University of Florida (UF) in Gainesville, Florida. She pursued her doctoral research on deregulation of long noncoding RNAs by latent Kaposi’s sarcoma associated herpesvirus in the laboratory of Dr. Rolf Renne. She has contributed to two first-author publications and several co-author publications during her time in the Renne laboratory. She received her Ph.D. in Medical Sciences, with concentration in genetics and a minor in applied statistics from the University of Florida in the summer of 2017.