BHLF1, A LYtic-CYCLE GENE ENCODing A LONG NON-CODING RNA
CONTRIBUTES TO EPSTEIN-BARR VIRUS LATENCY

A Dissertation in
Microbiology and Immunology

by

Kristen Dominique Yetming

© 2017 Kristen D. Yetming

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2017
The dissertation of Kristen Dominique Yetming was reviewed and approved* by the following:

Jeffery T. Sample  
Professor of Microbiology and Immunology  
Dissertation Advisor  
Chair of Committee

Clare E. Sample  
Professor of Microbiology and Immunology

Todd D. Schell  
Professor of Microbiology and Immunology

David J. Spector  
Professor Emeritus of Microbiology and Immunology

Gregory S. Yochum  
Associate Professor of Biochemistry and Molecular Biology and Surgery

Jianming Hu  
Professor of Microbiology and Immunology  
Director of Graduate Program in Microbiology and Immunology

*Signatures are on file in the Graduate School
Abstract

Epstein-Barr virus (EBV) is a lymphotropic, gammaherpesvirus which efficiently establishes a persistent, lifelong infection in humans. As is common with other herpesviruses, EBV has both replicative (lytic) and latent cycles of infection. EBV latency, in which there is no virus production, can be subdivided into several distinct latency programs which are characterized by differential expression patterns of the viral latency-associated proteins: EBNA1, -2, -3A, -3B, -3C, -LP and LMP1, -2A, -2B. The major long-term reservoir of EBV is the memory B cell, and as latently-infected B cells progress from an initial state of EBV-driven cell proliferation to a state of long-term viral latency within the memory B-cell pool, there is a restriction in the expression of the viral latency proteins, due to the epigenetic silencing of the EBNA promoters Wp and Cp. This restriction in latency is essential for the persistence of EBV infection as cytotoxic T lymphocytes can remove infected B cells that continue to express several of the latency-associated proteins.

Historically, it has been thought that latency-associated EBV gene products only contribute to the latent cycle, and lytic-cycle genes only function during lytic infection. However, it has recently been found that there is a subset of “lytic” genes that are expressed upon the initiation of latency, one of which is BHLF1. This, along with several other lines of evidence, suggests that BHLF1 may have a latency-associated function. Thus, the goal of the work presented in this dissertation was to elucidate whether BHLF1 has a role in the establishment or maintenance of EBV latency. Using recombinant EBV (rEBV), we observed that infection of an EBV-negative cell line, BL2, with a wild-type (WT) rEBV is capable of sustaining latency III, whereas infection with mutant rEBVs, in
which BHLF1 has been deleted, results in a transition from latency III to latency I within 3 months post-infection at both the protein and mRNA levels. Disruption of BHLF1 did not significantly influence the expression of other genes near the locus; thus, the phenotype observed is likely a direct consequence of the loss of BHLF1 function. In addition to the mutant phenotype in BL2 cells, we also observed a decrease in the efficiency of immortalization upon infection of primary B cells with the mutant rEBVs.

Although BHLF1 contains a predicted translational open reading frame (ORF), attempts to transiently express a protein from this ORF failed unless we co-expressed the EBV SM protein, whose expression is normally restricted to the lytic cycle. We therefore hypothesize that during latent infection, BHLF1 functions as a long non-coding RNA (lncRNA). The likelihood that BHLF1 primarily functions as an lncRNA is further supported by our recent observation and that of others that the ORF is not conserved among all EBV isolates. Furthermore, RT-PCR analysis of the 5’ ends of the EBNA cDNAs indicated a higher frequency of intron retention, possibly resulting in nonsense-mediated decay. Overall, these data suggest that BHLF1 may play a role in the maintenance of latency III, possibly through the regulation of splicing at the 5’ ends of the EBNA RNAs.
# Table of Contents

List of Figures.................................................................................................................. vii
List of Tables ....................................................................................................................... ix
List of Abbreviations.......................................................................................................... x
Acknowledgements............................................................................................................ xiv
Chapter 1: Introduction ....................................................................................................... 1

Chapter 2: Literature Review ............................................................................................ 6

2.1 Discovery and biology of EBV ................................................................................... 6
2.2 EBV life cycle .............................................................................................................. 11
2.3 EBV latency ................................................................................................................. 14
2.4 Latency-associated proteins ...................................................................................... 18
2.5 Non-coding RNAs ..................................................................................................... 30
2.6 BamHI-H leftward frame 1 (BHLF1) ......................................................................... 33
2.7 Long non-coding RNAs (lncRNAs) ........................................................................ 47
2.8 Viral lncRNAs ........................................................................................................... 52
2.9 BHLF1 as an lncRNA during viral latency ............................................................... 57

Chapter 3: Materials and Methods .................................................................................. 59

3.1 Cell lines .................................................................................................................... 59
3.2 Generation of rEBV ............................................................................................... 60
3.3 Virus production and infection of BL2 cells ......................................................... 65
3.4 Isolation and infection of primary B lymphocytes ................................................. 66
3.5 Induction of EBV lytic replication ......................................................................... 67
3.6 Plasmids and transfection of Louckes cells ......................................................... 67
3.7 Immunoblot analysis .............................................................................................. 68
3.8 RNA isolation and RT-PCR .................................................................................... 68
3.9 RT-qPCR ................................................................................................................ 69
3.10 Isolation and RT-qPCR of miRNA ................................................................. 69
3.11 Subcellular fractionation .................................................................................. 72
3.12 Cloning of BHLF1 into retroviral vector ......................................................... 72
3.13 Production of retrovirus and transduction of BL2 cells ................................... 75

Chapter 4: Results ...................................................................................................... 77

4.1 Contribution of BHLF1 to EBV latency ............................................................. 77
   BHLF1 deletion mutants are unable to sustain latency III ................................. 77
   Deletion of BHLF1 does not negatively affect BHRF1 mRNA and miRNA
   expression ........................................................................................................... 79
   Contribution of BHLF1 to EBV-mediated B-cell immortalization .............. 80
4.2 BHLF1 may be functioning as an IncRNA ......................................................... 112
   Expression of BHLF1 protein is dependent on expression of SM protein .... 112
   The BHLF1 ORF is not conserved among EBV isolates .................................. 112
4.3 Expression and localization of the BHLF1 RNA in different latency
   programs ................................................................................................................. 116
   Expression of BHLF1 RNA in different cell lines ............................................. 116
   Expression of BHLF1 RNA within nuclear and cytoplasmic fractions ............ 116
   Expression of BHLF1 RNA in BL2 cells infected with rEBV ......................... 117

Chapter 5: Discussion ................................................................................................ 121

5.1 Summary of findings in thesis ......................................................................... 121
5.2 Potential mechanisms by which BHLF1 may contribute to EBV latency ....... 125
5.3 Pitfalls and limitations ....................................................................................... 131
5.4 Future directions ............................................................................................... 135
5.5 Concluding remarks ........................................................................................ 139

Bibliography ............................................................................................................. 140

Appendix A: Recombineering of S1 aptamer into BHLF1 ..................................... 175

Appendix B: Cloning of BHLF1 into the retroviral vector, pLHCX .................... 181
List of Figures

Figure 2.1 The EBV genome. ................................................................. 7
Figure 2.2 Structure and coding content of the EBV genome .................. 9
Figure 2.3 EBV infection in healthy carriers. ........................................ 12
Figure 2.4 Expression patterns of EBV latency-associated proteins during the establishment of long-term viral latency. .......................................................... 15
Figure 2.5 EBV promoter usage in different latency programs .................. 16
Figure 2.6 Structure of the $BHLF1$ gene. .............................................. 34
Figure 2.7 EBV isolates with known deletions that include $BHLF1$ .......... 36
Figure 2.8 Different patterns of EBV latent gene expression in BLs and LCLs .... 39
Figure 2.9 Structure of the $LF3$ gene. .................................................. 43
Figure 2.10 Mechanisms of lncRNA function. ........................................ 48
Figure 3.1 Approach used to extend left homology arm of GalK targeting construct into the non-repeat region of $BHLF1$ for the generation of the $BHLF1$-S1 rEBV. ................................................................. 63
Figure 4.1 The $BHLF1$ locus is required to sustain latency III in BL2 cells. .... 82
Figure 4.2 rEBV lacking the BHLF1 ORF alone is unable to sustain latency III in BL2 cells .............................................................. 88
Figure 4.3 Further confirmation that the rEBV lacking the BHLF1 ORF alone is unable to sustain latency III in BL2 cells. ................................. 91
Figure 4.4 Deletion of the $BHLF1$ locus does not negatively impact expression of the $BHRF1$ miRNAs .................................................. 94
Figure 4.5 Deletion of the $BHLF1$ locus does not negatively impact expression of the mRNA encoding $BHRF1$. ............................................... 97
Figure 4.6 $BHLF1$ contributes to EBV-mediated B cell immortalization. .... 100
Figure 4.7 Primary B lymphocytes immortalized by WT, $\Delta B$-S, or $\Delta BHLF1$ rEBV establish latency III. .......................................................... 104
Figure 4.8 Growth curves of transformed primary B-cell lines .................. 110
Figure 4.9 Expression of BHLF1 protein is dependent on SM. ........................................ 114

Figure 4.10 BHLF1 ORF is not conserved between Akata and B95-8/Raji genomes. .................................................................................................................. 115

Figure 4.11 Expression of BHLF1 RNA in different cell lines. ........................................ 118

Figure 4.12 Expression of BHLF1 RNA within nuclear and cytoplasmic fractions. ... 119

Figure 4.13 Expression of BHLF1 RNA in BL2 cells infected with WT rEBV. .......... 120

Figure 5.1 Deletion of the BHLF1 locus increases the frequency of intron retention in the 5′ end of Cp-initiated EBNA mRNAs. ......................................................... 128

Figure A.1 RNA structure prediction of a region of the BHLF1 RNA and the S1 aptamer. ..................................................................................................................... 177
List of Tables

Table 2.1 EBV-associated malignancies................................................................. 17
Table 2.2 Amino acid composition of the hypothetical proteins coded by the regions carrying the NotI and PstI repeats................................................................. 41
Table 3.1 Primers used in recombineering. ............................................................. 62
Table 3.2 Primers used for PCR and RT-PCR......................................................... 70
Table 3.3 Primers and TaqMan probes for RT-qPCR analysis. .............................. 71
Table 3.4 Primers used in repairing pSG5-FLAG-BHLF1 and cloning BHLF1 into retroviral vector pLHCX. ................................................................. 73
Table A.1 Summary of problems encountered during the cloning of BHLF1 into the retroviral vector, pLHCX, and solutions implemented to address them. ..... 183
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-DOG</td>
<td>2-deoxy-galactose</td>
</tr>
<tr>
<td>3C</td>
<td>Chromatin conformation capture</td>
</tr>
<tr>
<td>4C</td>
<td>Circularized chromosome confirmation capture</td>
</tr>
<tr>
<td>BART</td>
<td>BamHI-A rightward transcript</td>
</tr>
<tr>
<td>BCR</td>
<td>B cell receptor</td>
</tr>
<tr>
<td>BHLF1</td>
<td>BamHI-H leftward reading frame 1</td>
</tr>
<tr>
<td>BL</td>
<td>Burkitt lymphoma</td>
</tr>
<tr>
<td>CD40L</td>
<td>CD40 ligand</td>
</tr>
<tr>
<td>ChIP-seq</td>
<td>Chromatin immunoprecipitation-sequencing</td>
</tr>
<tr>
<td>CTCF</td>
<td>CCCTC-binding factor</td>
</tr>
<tr>
<td>CTL</td>
<td>Cytotoxic T lymphocyte</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s modified Eagle medium</td>
</tr>
<tr>
<td>DNMT3a</td>
<td>DNA methyltransferase 3a</td>
</tr>
<tr>
<td>DS</td>
<td>Dyad symmetry</td>
</tr>
<tr>
<td>DS&lt;sub&gt;L&lt;/sub&gt;</td>
<td>Duplicated sequence left</td>
</tr>
<tr>
<td>DS&lt;sub&gt;R&lt;/sub&gt;</td>
<td>Duplicated sequence right</td>
</tr>
<tr>
<td>EBER</td>
<td>Epstein Barr virus-encoded small RNA</td>
</tr>
<tr>
<td>EBNA</td>
<td>Epstein Barr virus-encoded nuclear antigen</td>
</tr>
<tr>
<td>EBNA-LP</td>
<td>Epstein Barr virus-encoded nuclear antigen leader protein</td>
</tr>
<tr>
<td>EBV</td>
<td>Epstein-Barr virus</td>
</tr>
<tr>
<td>EMEM</td>
<td>Eagle’s minimal essential medium</td>
</tr>
<tr>
<td>ESE</td>
<td>Exon splicing enhancer</td>
</tr>
<tr>
<td>ESS</td>
<td>Exon splicing silencer</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>EZH2</td>
<td>Enhancer of zeste homolog 2</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal bovine serum</td>
</tr>
<tr>
<td>Flp</td>
<td>Flippase</td>
</tr>
<tr>
<td>FR</td>
<td>Family of repeats</td>
</tr>
<tr>
<td>FRT</td>
<td>Flippase recombinase recognition target</td>
</tr>
<tr>
<td>G+C</td>
<td>Guanine-plus-cytosine</td>
</tr>
<tr>
<td>Gas5</td>
<td>Growth arrest-specific 5</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>GR</td>
<td>Glucocorticoid receptor</td>
</tr>
<tr>
<td>GRU</td>
<td>Green Raji units</td>
</tr>
<tr>
<td>GRIM-19</td>
<td>Genes associated with retinoid/IFN-induced mortality-19</td>
</tr>
<tr>
<td>GSP</td>
<td>Gene-specific primer</td>
</tr>
<tr>
<td>HCMV</td>
<td>Human cytomegalovirus</td>
</tr>
<tr>
<td>HCV</td>
<td>Hepatitis C virus</td>
</tr>
<tr>
<td>HDAC-1</td>
<td>Histone deacetylase 1</td>
</tr>
<tr>
<td>HHV-4</td>
<td>Human herpesvirus 4</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>hnRNP</td>
<td>heterogeneous nuclear ribonucleoprotein</td>
</tr>
<tr>
<td>HSV-1</td>
<td>Herpes simplex virus 1</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IRF4</td>
<td>Interferon regulatory factor 4</td>
</tr>
<tr>
<td>IRF7</td>
<td>Interferon regulatory factor 7</td>
</tr>
<tr>
<td>ISE</td>
<td>Intron splicing enhancer</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ISS</td>
<td>Intron splicing silencer</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>LCV</td>
<td>Lymphocryptovirus</td>
</tr>
<tr>
<td>LF3</td>
<td>Leftward reading frame 3</td>
</tr>
<tr>
<td>lincRNA</td>
<td>Long intergenic non-coding RNA</td>
</tr>
<tr>
<td>LMP</td>
<td>Latent membrane protein</td>
</tr>
<tr>
<td>IncRNA</td>
<td>Long non-coding RNA</td>
</tr>
<tr>
<td>LTR</td>
<td>Long terminal repeat</td>
</tr>
<tr>
<td>LSM</td>
<td>Lymphocyte Separation Medium</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>miRNA</td>
<td>MicroRNA</td>
</tr>
<tr>
<td>MOI</td>
<td>Multiplicity of infection</td>
</tr>
<tr>
<td>ncRNA</td>
<td>Non-coding RNA</td>
</tr>
<tr>
<td>NPC</td>
<td>Nasopharyngeal carcinoma</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>oriLyt</td>
<td>Origin of lytic-cycle DNA replication</td>
</tr>
<tr>
<td>oriP</td>
<td>Origin of plasmid DNA replication</td>
</tr>
<tr>
<td>PAN RNA</td>
<td>Polyadenylated nuclear RNA</td>
</tr>
<tr>
<td>PBMC</td>
<td>Peripheral blood mononuclear cell</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate-buffered saline</td>
</tr>
<tr>
<td>PI3K</td>
<td>Phosphatidylinositol 3-kinase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PRC2</td>
<td>Polycomb repressive complex 2</td>
</tr>
<tr>
<td>Rb</td>
<td>Retinoblastoma</td>
</tr>
<tr>
<td>RBP-Jk</td>
<td>Jk recombination signal-binding protein</td>
</tr>
<tr>
<td>rEBV</td>
<td>Recombinant Epstein-Barr virus</td>
</tr>
<tr>
<td>RIG-I</td>
<td>Retinoic acid-inducible gene I protein</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RNA-seq</td>
<td>RNA-sequencing</td>
</tr>
<tr>
<td>RNP</td>
<td>Ribonucleoprotein</td>
</tr>
<tr>
<td>SCID</td>
<td>Severe combined immunodeficiency</td>
</tr>
<tr>
<td>sfRNA</td>
<td>Subgenomic flavivirus RNA</td>
</tr>
<tr>
<td>sisRNA</td>
<td>Stable intronic sequence RNA</td>
</tr>
<tr>
<td>snoRNA</td>
<td>Small nucleolar RNA</td>
</tr>
<tr>
<td>SR</td>
<td>Serine/arginine</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single-stranded DNA</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll-like receptor</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumor necrosis factor</td>
</tr>
<tr>
<td>TPA</td>
<td>12-O-tetradecanoylphorbol-13-acetate</td>
</tr>
<tr>
<td>TRAF</td>
<td>Tumor necrosis factor receptor-associated factor</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start site</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-type</td>
</tr>
<tr>
<td>xrRNA</td>
<td>Xrn1-resistant RNA</td>
</tr>
<tr>
<td>ZRE</td>
<td>Zta response element</td>
</tr>
</tbody>
</table>
Acknowledgements

The last seven years have been some of the most challenging and rewarding years of my life. There are many people to whom I am indebted for their unending support. First and foremost, I would like to thank my mentor and advisor, Jeff Sample, for the many years of guidance and encouragement. Not only did he trust me with an incredibly challenging project, but he has also patiently counseled me throughout the numerous obstacles and frustrations that I encountered throughout my time working on it. He has always shown understanding and support, and for that, I am deeply grateful.

I would also like to thank the members of my committee – Clare, Todd, David, and Greg – who have provided an immense amount of support over the years. They have consistently pushed me to be a better scientist, and have often assured me that I am always welcome to ask their advice, even after having successfully defended.

I am also thankful for the members of the Microbiology and Immunology department, and especially the members of both Sample labs, past and present. I can always count on them for words of encouragement (or commiseration!) and consider myself extremely lucky to have been surrounded by such warm and supportive people throughout my time here.

Last but certainly not least, I would like to thank my friends and family. Graduate school has been an incredible journey with many ups and downs (sometimes more downs than ups!). My family and friends have always been my much-needed cheerleaders and made a somewhat painful journey a great deal more enjoyable. Their support and motivation have been invaluable, and without them, I would not be where I am today. Thank you all!
Chapter 1: Introduction

Epstein-Barr virus (EBV) is a gammaherpesvirus which establishes a persistent, lifelong infection in humans and, as is common with other herpesviruses, has both lytic (replicative) and latent cycles of infection. The establishment of latency is critical for EBV’s oncogenic potential and can be subdivided into a number of latency programs, characterized by different expression patterns of the latency-associated proteins as well as by a difference in promoter usage. The major long-term reservoir of EBV is the memory B cell, and as latently infected B cells progress from an initial state of EBV-driven cell proliferation to eventually enter the memory B-cell pool, there is an ordered downregulation in the expression of the latency-associated proteins due to the epigenetic silencing of the latency-specific EBV-encoded nuclear antigen (EBNA) promoters Wp and Cp. This restriction in latency allows for immune evasion and limits the oncogenic potential of EBV, which encodes an oncoprotein and several other proteins whose functions are consistent with a role in EBV-associated lymphoproliferation and tumors.

Generally, EBV genes are expressed either during latency or the lytic cycle; however, there is a subset of “lytic” genes that are expressed upon the initiation of latency, one of which is BHLF1. There is increasing evidence that BHLF1, previously believed to express a protein exclusively during productive infection, in fact encodes a long non-coding RNA (IncRNA) during both the latent and replicative phases of EBV infection. Although the BHLF1 gene encodes a highly expressed, polyadenylated
transcript upon activation of the lytic cycle (1–3), it has been found to be one of the most highly transcribed genes in latently infected B-cell lines (4–6) and RNA mapping studies suggest that BHLF1 has upstream promoters that are active during latent infection (7).

Early sequencing of EBV DNA identified a long open reading frame (ORF) within the DNA sequence encoding the BHLF1 RNA. However, recent sequencing of additional EBV genomes revealed that the BHLF1 ORF is not conserved either among all EBV isolates (8) or between EBV and the highly homologous lymphocryptoviruses of non-human primates (9, 10); this lack of conservation is highly suggestive of a non-protein-coding function. Although it has been demonstrated that protein can be expressed from the BHLF1 ORF (11–14), the only known function for BHLF1 is as an lncRNA during the replicative cycle, where it contributes to oriLyt function via the formation of an RNA:DNA hybrid at the site of transcription (15). The role of the protein or lncRNA encoded by BHLF1 during EBV latency is unknown.

Furthermore, one of the hallmark properties of EBV is its ability to transform primary B lymphocytes; in naturally-occurring non-transforming EBV isolates, all or a portion of the BHLF1 gene is deleted. Though the loss of this hallmark property has largely been attributed to the deletion of the adjacent ORF encoding the latency-associated EBNA2 protein, rescue experiments demonstrating a requirement for EBNA2 in B-cell immortalization did not exclude a potential contribution by BHLF1. Consequently, the work presented in this dissertation sought to address the role of BHLF1 in EBV latency.
**Aim 1: Define the contribution of BHLF1 to EBV latency.**

Initially, we hypothesized that BHLF1 acts to direct silencing of the Cp/Wp locus *in cis* and was therefore critical for the transition from latency III to latency I. Experiments to test this hypothesis showed that this is unlikely to be the case (16). However, one of the experiments performed in the course of those studies suggested the opposite of our original hypothesis: that BHLF1 may instead be essential for the maintenance of latency III.

Upon infection of an EBV-negative cell line, BL2, with either wild-type (WT) recombinant EBV (rEBV) or mutant rEBVs that lack BHLF1, we observed the establishment of latency III. However, by 2 months post-infection, the mutant-virus infections had transitioned to latency I at both the protein and mRNA levels. This suggests that although EBV latency is established in the absence of BHLF1, the locus may in fact be necessary for the maintenance of latency III in BL2 cells. Consistent with these findings, deletion of BHLF1 appears to reduce the transformation efficiency of EBV in primary B cells but does not affect the mRNA and miRNA expression of BHRF1, an adjacent EBV gene that encodes a viral Bcl-2 homolog. Furthermore, RT-PCR analysis of the 5’ ends of EBNA cDNAs indicated a higher frequency of intron retention in mutant-virus infections compared to WT rEBV infections. The analysis of the 5’ ends gives a potential explanation for the lack of a detectable latency III pattern of EBNA protein expression in the mutant-virus infections. Therefore, we hypothesize that BHLF1 may be contributing to viral latency via the regulation of splicing at the 5’ ends of EBNA mRNAs.
Aim 2: Determine if the effect on latency is mediated by the *BHLF1* protein, RNA, or genomic locus.

Although it is clear that the loss of *BHLF1* has an effect on latency, we must determine which form of *BHLF1* (protein, RNA, or the genomic locus itself) mediates this effect. Interestingly, we were able to express the BHLF1 protein in transient transfection experiments only in the presence of SM, a lytic-cycle protein. This suggests that the BHLF1 protein is only expressed during the lytic cycle and thus, we feel it is unlikely to be responsible for the effect that we observe during latency. Additionally, we and others have found that the BHLF1 ORF and/or amino acid sequence is not conserved evolutionarily, either amongst EBV-related lymphocryptoviruses of non-human New World and Old World primates or even among EBV isolates (8–10). In fact, a single base pair deletion within the BHLF1 ORF of the Akata and Mutu I isolates results in a downstream stop codon that is not present in the B95-8 strain (8; our unpublished data). Based on these observations, we hypothesize that *BHLF1* may be functioning at least in part as an IncRNA during latency.

Aim 3: Determine the pattern of expression and subcellular localization of the *BHLF1* RNA in different EBV latency programs.

Previous observations by other groups indicate that the *BHLF1* transcript is the most abundant polyadenylated transcript in latency I, and is highly transcribed during latency III (but is under-represented in the polyadenylated RNA fraction in the cytoplasm) (4, 5). To further elucidate the function of *BHLF1*, we wanted to determine the subcellular localization and pattern of expression of the RNA. As many IncRNAs are localized in the nucleus, and based on our observations that *BHLF1* may play a role in
regulating splicing, we hypothesize that the BHLF1 RNA localizes exclusively or primarily to the nucleus during latency.

**Overall significance**

Although most people are infected with EBV at a young age and exhibit little to no symptoms, healthy adults typically present with infectious mononucleosis which can last weeks to months. Furthermore, EBV is a potentially oncogenic virus and latent EBV infection is associated with several types of cancer, particularly in individuals who are immunosuppressed (such as transplant patients) or immunocompromised. Thus, it is imperative to further our understanding of EBV latency so that novel therapeutics can be developed. Previously, lytic-cycle genes were not thought to play a role during latency; however, with the recent discovery that several of these genes are in fact expressed early in latency, it has become essential to elucidate their potential contributions to latent infection. One of these genes, BHLF1, has not been well-characterized, despite having been identified several decades ago. In fact, a function for BHLF1 during the lytic cycle has only been described fairly recently. Thus, the work presented in this dissertation aims to fill a gap in our knowledge of BHLF1, thereby expanding our understanding of EBV latency.
Chapter 2: Literature Review

2.1 Discovery and biology of EBV

EBV, also known as human herpesvirus 4 (HHV-4), is a lymphotropic, gammaherpesvirus which establishes a persistent, lifelong infection in humans. It was first discovered in 1964 when it was identified in cultured lymphoblasts from Burkitt lymphoma (BL; 17). A major causative agent of infectious mononucleosis, EBV is also a potentially oncogenic virus and has been associated with several tumor types in addition to BL, most commonly Hodgkin lymphoma, gastric carcinoma, and nasopharyngeal carcinoma (NPC). The virus has a linear, double-stranded DNA genome of approximately 172 kbp, which circularizes upon infection and is maintained as an episome in latently infected B cells (Figure 2.1; 18). The genome encodes for about 90 proteins, typically classified as lytic- or latency-associated, as well as a number of small non-coding RNAs (ncRNAs; Figure 2.2).
Figure 2.1 The EBV genome.
**Figure 2.1 The EBV genome**

**Top:** The linear form of the viral genome is flanked by terminal repeats (TR). Repeat elements (IR1-IR4), unique sequence domains (U1-U5), the origin of plasmid replication (OriP), and the origins of lytic replication (OriLyt) are indicated.

**Bottom:** Shortly after infection, the TR elements fuse to generate the episomal form of the viral genome that is maintained in latently infected cells. The latency-associated mRNAs and *EBERs* are indicated. The common *EBNA* promoters Cp and Wp, and the *EBNA1*-exclusive promoter Qp are depicted by bent arrows. Exons that contain ORFs are bracketed.

Figure adapted from **Sample JT, Sample CE.** 2008. Epstein-Barr Virus: Molecular Biology, p. 157-167. *In* Encyclopedia of Virology (Third Edition).
Figure 2.2 Structure and coding content of the EBV genome.
Figure 2.2 Structure and coding content of the EBV genome.

ORFs and orientation of the viral genes are depicted. Genes are classified as latency-associated, early lytic, late lytic, or uncharacterized. Latency-associated genes are annotated above their respective ORFs with commonly used nomenclature. The viral non-coding RNA regions are also indicated (EBERs, BHRF1 miRNAs, and BART miRNA clusters).
2.2 EBV life cycle

Like other herpesviruses, EBV has both lytic (replicative) and latent cycles of infection. Transmission between human hosts occurs primarily via saliva and primary infection is thought to occur in the oropharynx where incoming virus infects epithelial cells. A recent study showed that EBV infection of primary epithelial cells \textit{in vitro} results in the efficient production of virus (19), which then spreads to the B-cell compartment by mechanisms that are still unknown, though it may gain access to and infect B lymphocytes directly through microabrasions in the epithelium. Viral replication in epithelial cells may occur primarily upon exit from the host, as virus produced in epithelial cells is programmed to preferentially infect B cells (20), and readily infects these lymphocytes and establishes life-long latency within them. Infected B cells initially proliferate via the growth program (latency III), in which the full complement of latency-associated viral proteins are expressed (discussed below), but eventually downregulate their expression of the viral latency proteins, enabling them to escape immune detection. Viral replication in memory B cells latently infected with EBV may be reactivated by their differentiation into plasma cells (21), thereby producing virus which can then efficiently infect epithelial cells in the oropharynx, possibly leading to low-level shedding of the virus in this compartment and subsequent transmission of the virus between hosts (Figure 2.3).
Figure 2.3 EBV infection in healthy carriers.
Figure 2.3 EBV infection in healthy carriers.

Primary EBV infection begins in the oral cavity. EBV uses different glycoproteins to infect epithelial cells and naïve B cells. Viral entry results in transport of the EBV genome into the B-cell nucleus, where replication by cellular and viral DNA polymerases begin. EBV gene products activate the B-cell growth program, resulting in the proliferation of blasting B cells. Priming of naïve T cells by antigen-presenting cells occurs in parallel. Normally, these blasting B cells are destroyed by cytotoxic T lymphocytes. Once in the circulation, previously activated memory B cells may continue to undergo lytic replication or, if EBV shuts down most of its protein-encoding genes, latency occurs. At a later time, as cells recirculate between the oral and peripheral compartments, resting B cells may be activated, resulting in viral reactivation and shedding.

Reprinted with permission from the American Society for Microbiology:

2.3 EBV latency

The establishment of latency, i.e., a state in which there is no virus production, is critical for EBV’s oncogenic potential and is subdivided into several distinct latency programs that are characterized by specific expression patterns of the viral latency-associated proteins (Figure 2.4) and promoter usage for the EBNA mRNAs (Figure 2.5). Table 2.1 provides an overview of EBV-associated malignancies and the latency program typically found in these diseases.

The major long-term reservoir of EBV is the memory B cell, and as latently infected B cells progress from an initial state of EBV-driven cell proliferation to eventually enter the memory pool, there is an ordered downregulation in the expression of the latency-associated proteins due to epigenetic silencing of most of the EBV latency-specific promoters, most notably those that drive transcription of the EBNA promoters: Wp and Cp. Upon infection of the B cell, Wp is activated by cellular transcription factors to initiate expression of the EBNA proteins, primarily EBNA2 and EBNA-LP (22–28). This is followed by EBNA2- and EBNA-LP-mediated activation of Cp (23, 29–32), located approximately 3 kbp upstream of Wp, and of the promoters that drive transcription of the latent membrane protein (LMP) 1 and 2A/2B genes (33, 34), thus establishing the latency III program in which the full complement of latency-associated proteins are expressed, i.e., EBNA1, 2, 3A, 3B, 3C and –LP, and LMPs 1, 2A and 2B. Thereafter, Wp activity is downregulated via mechanisms such as DNA methylation (35–38) and potentially other mechanisms. The latency-associated EBNA transcripts continue to initiate from Cp, until it too is epigenetically silenced and the virus transitions to subsequent latency programs in which there is further restriction of EBV protein expression (i.e., latency II and I). In latency II, only the LMPs and EBNA1 are expressed.
EBV has several distinct latency programs, characterized by specific expression patterns of the viral latency-associated proteins. Upon infection of the primary B cells, the virus establishes latency III in which the full complement of latency-associated proteins is expressed (EBNAs 1, 2, 3A, 3B, 3C, -LP, and LMPs 1, 2A, 2B). The virus undergoes gradual restriction of the expression of these latency-associated proteins due to epigenetic silencing, resulting in latency II (EBNA1, and LMPs 1, 2A, 2B) and latency I (EBNA1 only). In latency 0, there is no expression of the latency-associated proteins. When infected memory B cells periodically divide, latency I is reactivated to allow for expression of EBNA1 to maintain the viral episome.
Figure 2.5 EBV promoter usage in different latency programs.

Structure and promoter usage of the EBV *EBNA* mRNAs expressed during the different latency programs. The virus establishes latency III upon primary infection and utilizes Wp and Cp. As the virus transitions to a more restricted, long-term latency (latency I/II), Wp and Cp are silenced and the *EBNA1*-exclusive Qp is active. Exons are in black; genome is shown in its linear form. Not to scale.
<table>
<thead>
<tr>
<th>Malignancy</th>
<th>Latency program</th>
<th>Viral proteins expressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burkitt’s lymphoma (BL)</td>
<td>I</td>
<td>EBNA1</td>
</tr>
<tr>
<td>B-lymphoproliferative disease (BLPD)</td>
<td>III</td>
<td>EBNA1, EBNA2, LMP-1, LMP2A, LMP2B</td>
</tr>
<tr>
<td>Hodgkin’s lymphoma</td>
<td>II</td>
<td>EBNA1, LMP-1, LMP-2A, LMP2B</td>
</tr>
<tr>
<td>T-cell lymphoma</td>
<td>II</td>
<td>EBNA1, LMP-1, LMP-2A, LMP2B</td>
</tr>
<tr>
<td>Nasopharyngeal carcinoma (NPC)</td>
<td>II</td>
<td>EBNA1, LMP-1, LMP-2A, LMP2B</td>
</tr>
<tr>
<td>Gastric carcinoma</td>
<td>II</td>
<td>EBNA1, LMP-1, LMP-2A, LMP2B</td>
</tr>
</tbody>
</table>
while in latency I, EBNA1 is the only detectable latency-associated protein. Ultimately, all known EBV protein expression is silenced as the B cells enter the memory pool and cease to divide, a state sometimes referred to as latency 0. When infected memory B cells do periodically divide, latency I is reactivated so that EBNA1, the viral genome-maintenance protein, can be expressed via the EBNA1-exclusive promoter, Qp, to ensure against loss of the viral episome during successive rounds of cell division (39, 40). This restriction in latency allows for immune evasion and limits the oncogenic potential of EBV, which encodes an oncoprotein (LMP1) and several other proteins whose functions are consistent with a role in EBV-associated lymphoproliferation and tumorigenesis.

2.4 Latency-associated proteins

As mentioned above, EBV encodes nine latency-associated proteins that are expressed differentially during the latency programs (with the exception of latency 0) as well as numerous small ncRNAs. These are described below.

_EBNA1_

EBNA1 is expressed in all EBV latency programs with the exception of latency 0 (in which there is believed to be a complete absence of EBV protein expression), and plays multiple roles during EBV infection. As the only viral protein required for maintenance of the EBV genome via the origin of plasmid DNA replication, _oriP_ (41), EBNA1 is the only viral protein expressed during latency I. _OriP_ contains two functional
domains (42): the dyad symmetry (DS) element which is the origin of DNA replication during latent infection and that contains four EBNA1 binding sites (43), and the family of repeats (FR; ~20 repeats, each of which contains an EBNA1 binding site) which is required for the mitotic segregation and transcriptional activation functions of EBNA1. The mitotic segregation function of EBNA1 is necessary for stable maintenance of the EBV episomes in proliferating cells and involves the tethering of the EBV episomes to the cellular mitotic chromosomes (44–51). Transcriptional activation by EBNA1 occurs when the protein is bound to the FR element and activates expression from the Cp and LMP promoters, which lie within approximately 3 and 10 kbp of oriP in the EBV episome, respectively (52, 53).

In addition to the FR and DS elements of oriP, EBNA1 also binds to the EBV genome immediately downstream of the transcription start site of Qp; binding by EBNA1 here negatively autoregulates its expression from Qp, which is thought to limit its expression to prevent immune detection of cells supporting the latency I and II programs (54–56). Interestingly, EBNA1 has the lowest affinity for these binding sites among the three EBNA1-binding domains within the genome (56, 57), consistent with the notion that this feedback mechanism ensures that there is ample EBNA1 to mediate genome maintenance through oriP within proliferating cells. EBNA1 has also been shown to be important, though probably not directly, for efficient immortalization of primary B cells by EBV (58, 59).

EBNA1 has also been proposed to play a role in cell growth and survival. A study using EμEBNA-1 transgenic mice showed that these mice developed lymphomas in which EBNA1 was detectable, suggesting that EBNA1 has oncogenic potential in transgenic mice (60). However, subsequent studies using EBNA1-transgenic mice of two
other genetic backgrounds were unable to repeat these results, indicating that EBNA1 does not significantly increase lymphoma prevalence in these mice (61, 62). Furthermore, EBNA1 may promote genomic instability (e.g., chromosomal aberrations, DNA double-strand breaks, engagement of the DNA damage response, and telomere dysfunction) via the induction of oxidative stress (63–65), which may be a potential mechanism by which EBNA1 promotes tumorigenesis.

**EBNA2**

EBNA2 and EBNA-LP are the first latency proteins detected after infection of primary B cells (24, 66). While they can be expressed from either Wp or Cp, transcription of the EBNA genes initiates from Wp but switches to Cp later in infection; this correlates with detection of the EBNA2 protein, which with its coactivator EBNA-LP is a strong transactivator of not only Cp (29, 32) but also the LMP promoters (67–70). EBNA2 transactivation of Cp leads to the expression of EBNA1, the EBNA3s, and the LMPs. Neither EBNA2 nor EBNA-LP bind to DNA directly, but are targeted to their responsive promoters by cellular transcription factors with which they interact, notably RBP-Jk and PU.1; these interactions are essential for transactivation of Cp and the LMP1 promoter, respectively (71–73).

In addition to activation of EBV gene transcription, EBNA2 also transactivates the transcription of a number of cellular genes, including *c-myc*, *CD23*, and *CD21* (24, 74, 75). Transactivation of these genes promotes the conversion of resting B lymphocytes to lymphoblastoid cell lines (LCLs) upon infection with EBV. In particular, EBNA2 transactivation of *c-myc* leads to continuous proliferation of the infected B cells, while transactivation of *LMP1* leads to activation of NF-κB and MAP kinase, which upregulates
protein expression of the anti-apoptotic Bcl-2 family. These events establish a transcription program in B cells that mimics that seen in B cells responding to antigen. Furthermore, EBV drives the proliferation of resting B cells by activation of cell growth genes via EBNA2 targeting of H3K4me1-modified, nucleosome-depleted, non-promoter sites (76).

EBNA2 is expressed during latency III but its expression is downregulated as the virus transitions to latency II and I, in which it is no longer expressed.

**EBNA-LP**

As indicated above, EBNA-LP is one of the first latency proteins expressed after infection of primary B cells (24) from either Wp or Cp. EBNA-LP is expressed only during latency III, and its primary if not only function is believed to be as a strong coactivator of EBNA2-induced transcription, although the mechanism(s) by which it fulfills this role remain to be elucidated (33, 34). However, studies have shown that EBNA-LP is not a global coactivator of EBNA2 targets, but instead preferentially coactivates transcription from Cp and a subset of EBNA2-responsive genes, such as LMP1 and LMP2B, and the cellular gene Sp100A (77, 78).

Unlike for the other EBNA proteins, the open reading frame (ORF) for EBNA-LP is composed not of a single- or two-exon portion at the 3’ half of its mRNA, but instead is composed primarily of a two-exon repetitive unit comprising the W1 and W2 exons, which also encodes the 5’ or leader portion of all the EBNA mRNAs that originate from either Wp or Cp. Therefore, all EBNA mRNAs from Cp or Wp are potentially bicistronic, encoding both EBNA-LP and the respective EBNA protein, or monocistronic for that EBNA protein. Whether an EBNA mRNA encodes EBNA-LP is determined by an
alternative splice at the 5′ end of the mRNA that creates the AUG initiation codon. This leads to the production of multiple EBNA-LP isoforms, as transcription can initiate from Wp residing in each repetitive unit. EBNA-LP localizes predominantly in the nucleus due to a nuclear localization signal (79); however, it has been shown that the size of the EBNA-LP isoform (influenced by the number of W repeats) can allow for the localization of EBNA-LP in the cytoplasm and may in fact affect its function (80, 81).

The EBNA3 family

The EBNA3 family consists of three large proteins: EBNA3A, EBNA3B, and EBNA3C. Along with EBNA1, EBNA2, and EBNA-LP, they are expressed from a polycistronic gene under the control of Wp or Cp, and whose primary transcript is alternatively spliced to produce the various EBNA transcripts, with EBNA-LP potentially expressed from the 5′ half of the mRNA (see EBNA-LP, above). Although the EBNA3s exhibit limited sequence homology and are arranged tandemly in the viral genome, each encoded by an ORF comprised of a short and long exon (Figure 2.1), they are not functionally interchangeable. All three proteins have been shown to bind to the same domain on RBP-Jκ as EBNA2 (82, 83), suggesting that they act as antagonists of EBNA2-mediated transactivation. However, while EBNA3A and EBNA3C have been shown to be essential for immortalization of primary B cells (84), EBNA3B is not (85). Thus, this family of proteins appears to have arisen by gene duplication, and while they share at least one mechanism of action (via their interaction with RBP-Jκ), each appears to have acquired a unique function(s). Interestingly, the earliest known EBV-related virus, that of a New World primate (common marmoset), contains only a single ORF in the location of the genome occupied by the three EBNA3 genes in EBV and its highly-
related counterparts of Old World primates (9, 86), consistent with the belief that the EBNA3 genes arose through duplication of a single ancestral gene that has undergone significant evolution over the past 35 million years.

In addition to the role they are believed to play in co-regulating EBV gene transcription, work in the past several years has revealed that EBNA3A and EBNA3C may play roles in cell-cycle regulation and possibly epigenetic modification. EBNA3C was suggested to disrupt the cyclin/CDK-pRb-E2F pathway, which regulates cell-cycle progression in the G1 phase of the cell cycle, and was subsequently found to inhibit the accumulation of p27KIP1 (87) while EBNA3C inactivation led to the accumulation of p16INK4A, a decrease in hyperphosphorylated retinoblastoma (Rb) protein, and a decrease in the proportion of cells in S or G2/M phase (88). EBNA3C was also shown to stabilize and enhance the kinase activity of the cyclin D1/CDK complex, thereby facilitating G1-S transition and leading to deregulation of the cell cycle (89). Similarly, EBNA3A was found to regulate p21WAF1/CIP1 protein expression which maintains Rb in its inactive, hyperphosphorylated form, thereby promoting cell-cycle progression (90). Additionally, the cellular pro-apoptotic protein Bim is cooperatively repressed and epigenetically regulated by EBNA3A and EBNA3C (91–93), which may contribute to EBV's transforming and oncogenic potential.

The EBNA3s are the most immunogenic of all the latency-associated viral proteins; they have been shown to be the primary targets of CD8+ T cells (cytotoxic T lymphocytes, CTLs) in EBV-positive donors (94–96). Studies by two different groups tested CTL preparations from EBV-immune donors and found that the dominant response of all the donors was directed towards one of the EBNA3 proteins, in particular EBNA3A and EBNA3C (94, 95). Similarly, memory CTL responses in long-term EBV
carriers were focused preferentially on the EBNA3 proteins (96). Therefore, it is not surprising that the EBNA3 proteins are only expressed during latency III and that their expression downregulated during the other latency programs.

**LMP1**

LMP1 is the primary oncogene of EBV and is therefore essential for the efficient transformation of primary B cells. LMP1 expression in NIH 3T3 and Rat-1 cells resulted in altered cell morphology and an enhancement in the ability of the cells to grow in low-serum medium, while Rat-1 cells expressing LMP1 were tumorigenic in nude mice (97). The LMP1 gene has also been shown to transform murine Balb/3T3 cells (98). Studies in which the LMP1 ORF was disrupted with nonsense linker insertions showed that these mutants were only able to maintain growth transformation of primary B lymphocytes when wild-type (WT) LMP1 was provided in trans by co-infection with the transformation-defective EBV strain P3HR-1; LMP1 was therefore proposed to be essential for EBV-induced growth transformation of primary B lymphocytes (99). However, a subsequent study investigating a panel of LMP1 mutants demonstrated that while LMP1 was critical for the process of B-cell proliferation, it was not essential under certain conditions, i.e., high virus titers and co-cultivation with human fibroblast feeder cells (100). This study also showed that severe combined immunodeficiency (SCID) mice injected with WT-transformed proliferating B cells developed tumors while those injected with ΔLMP1-mutant-transformed proliferating B cells did not, thereby demonstrating the importance of LMP1 for in vivo oncogenesis.

As mentioned above, the LMP1 promoter is activated by EBNA2 (68–70). The LMP1 protein consists of six transmembrane domains with a short cytoplasmic amino-
terminus and a carboxy-terminal signaling domain. LMP1 mimics a constitutively active tumor necrosis factor (TNF) receptor, CD40, which allows it to affect cell growth and survival pathways (101–108). The interaction of CD40 with its ligand, CD40L, promotes the formation of the germinal center, in which B cells undergo immunoglobulin isotype switching and somatic hypermutation, and results in long-lived plasma cells and memory B cells (109). LMP1 has two essential components within its C-terminal cytoplasmic domain, CTAR1/TES1 and CTAR2/TES2, both of which enable LMP1 to mimic CD40/CD40L signaling. CTAR1/TES1 associates with TNF receptor-associated factors (TRAFs) which activate non-canonical NF-κB signaling (110), while CTAR2/TES2 associates with TNF receptor-associated death domain proteins without generating a death signal (111, 112), thereby leading to the survival of EBV-infected B cells. LMP1 further contributes to cell survival via the upregulation of anti-apoptotic genes, including \( Bcl-2 \) (113–116).

Levels of the LMP1 protein are balanced by autoregulation whereby the LMP1 promoter is activated by LMP1 itself via the JNK signaling pathway and then inhibited via LMP1-activated NF-κB signaling (117). LMP1 also activates the NF-κB or p38 signaling pathways and can induce autophagy or activate the kinase PERK, all of which can upregulate the expression of LMP1 itself (118–121). Furthermore, LMP1 is expressed during both latency III and II, preventing reactivation of the lytic cycle by inhibiting transcription of the viral protein BZLF1, which controls entry into the lytic cycle (122, 123).
**LMP2**

The *LMP2* gene spans the terminal repeats at both ends of the linear EBV genome, and is therefore only expressed when the viral genome circularizes (Figure 2.1). It is comprised of nine exons, the first of which is located at three alternative positions which start at separate promoters. This, along with alternative splicing, results in several isoforms, of which the LMP2A and LMP2B transcripts are the most well-studied. In latency III, LMP2A and LMP2B transcription is induced by EBNA2 from different promoters which are 3 kb apart (124) and which both contain an EBNA2-responsive element as well as RBP-Jκ and PU.1 sites (72, 125–127). In latency II, during which EBNA2 is not expressed, LMP2A autoactivates its expression independently of EBNA2 via the Notch signaling pathway (128). Furthermore, the *LMP2* gene has also been shown to be induced by activation of the viral lytic cycle (129).

Both LMP2A and LMP2B are transmembrane proteins consisting primarily of 12 membrane-spanning domains and which differ by a 119-amino acid N-terminal cytoplasmic domain that is present in LMP2A but not LMP2B (130). LMP2A contributes to the survival of infected B cells by mimicking a constitutively activated B-cell receptor (BCR) by modifying cellular signaling pathways via the recruitment of kinases such as the Src family tyrosine kinase Lyn, as well as other kinases such as Syk, which leads to the constitutive phosphorylation of LMP2A at the multiple tyrosine residues within its N-terminal domain (131–133). The N-terminal domain of LMP2A also contains proline-rich regions which interact with E3 ubiquitin ligases in the Nedd4 family, which can ubiquitinate LMP2A and associated proteins, thereby leading to their degradation (134–138). LMP2A also causes constitutive phosphorylation of phosphatidylinositol 3-kinase...
(PI3K) and the serine/threonine kinase Akt, thereby contributing to the inhibition of apoptosis (139, 140).

Constitutive expression of LMP2A in BL lymphoblasts blocks signal transduction events that follow the cross-linking of surface IgM, CD19, or class II major histocompatibility complex (MHC), similar to the desensitization that follows BCR signal transduction (141–143); e.g., LMP2A blocks the activation of protein tyrosine kinases to prevent activation of the lytic cycle and allow persistence of EBV in latently infected B cells (133, 143). Furthermore, studies have implicated LMP2A, in conjunction with c-myc, in the promotion of tumor development (144, 145). A subsequent study suggested that LMP2A and c-myc play a cooperative role in altering cell-cycle progression by interfering with the function of the cell-cycle inhibitor p27 (146). Thus, LMP2A not only contributes to the survival of EBV-infected B cells, but may also play a role in tumorigenesis.

The LMP2B isoform has not been well-studied, in part due to the lack of LMP2B-only antibodies, and therefore not much is known about its protein function. Unlike LMP2A, LMP2B does not contain the cytoplasmic domain required for tyrosine kinase signaling. It has been shown that LMP2B expression promotes the spread and motility of epithelial cells, similar to LMP2A, and therefore can still modify cellular signaling pathways (147). However, B-cell lines expressing LMP2B showed no effect in BCR signal transduction, in contrast to LMP2A-expressing cells (148). Importantly, LMP2B appears to negatively regulate the function of LMP2A in preventing the switch from latency to lytic replication as studies have shown that overexpression of LMP2B leads to an increase in the switch to lytic replication while silencing of LMP2B reduces activation of lytic replication (149, 150).
**BHRF1 and other early/pre-latency proteins**

Generally, EBV genes are expressed either during latency or the lytic cycle; however, there is a subset of so-called lytic genes that are expressed shortly after the infection of primary B cells (151). These genes are expressed transiently and prior to the establishment of latency, a period sometimes referred to as the prelatent phase. Although the expression of these lytic genes does not lead to the production of virus, they may be initially required for B-cell transformation. This subset of lytic genes includes *BHRF1, BALF1, BZLF1, BCRF1/vIL-10, and BNLF2a* (152–156). *BHLF1* is also expressed early after infection (157) and is further discussed in section 2.6 below.

*BHRF1 and BALF1* are viral homologs of the cellular anti-apoptotic Bcl-2 protein. They are highly expressed in the first 24 hours following infection; however, their expression decreases to almost undetectable levels during latency (154, 158). It has been shown that the expression of *BHRF1* and *BALF1* is required for both the initial evasion of apoptosis as well as the initial cellular transformation of infected cells, but not for the maintenance of latency (154).

*BZLF1* is an EBV transcription factor that is essential and sufficient for the mediation of the switch from latency to lytic replication in EBV-infected cells. Although it has been shown to be expressed early after infection of primary B cells, it does not initiate the lytic phase of the virus due to the lack of CpG-methylated promoters, to which it preferentially binds, in the viral genome at this stage (155, 159).

*BCRF1* is a viral homolog of cellular IL-10 (vIL-10) that is expressed early after infection of primary B cells and has been shown to be important for B-cell transformation by EBV (152, 153, 156). *BNLF2a* is also expressed during the early stages of infection.
and, along with BCRF1, appears to contribute to immune evasion during the prelatent phase of EBV infection (156).

Summary of the establishment of long-term latency

The viral latency-associated proteins have been studied extensively for the last 30 years; as such, we now have a good understanding of the establishment of long-term latency in EBV-infected cells and the roles that these proteins play in this process. As mentioned above, upon EBV infection of a B cell, Wp is activated by cellular transcription factors to initiate expression of the EBNA2 and EBNA-LP proteins. EBNA-LP acts as a coactivator with EBNA2, and together they transactivate Cp and the LMP promoters, as well as several host genes. The transactivation of Cp leads to the expression of EBNA1, the EBNA3s, and the LMPs, thereby establishing the latency III program in which the full complement of latency-associated proteins is expressed. In addition to its transactivation of viral promoters, EBNA2 also contributes to the continued proliferation of the infected B cells via its transactivation of cellular genes, including c-myc. The expression of LMP1 and LMP2A provides a constitutively active TNF receptor signal and BCR, respectively, both of which contribute to the survival of EBV-infected B cells. LMP1 also upregulates anti-apoptotic genes while LMP2A modifies signaling pathways to further promote cell survival. LMP2A also prevents the switch from EBV latency to lytic replication, while LMP2B modulates this function. The EBNA3s co-regulate EBV gene transcription and contribute to cell-cycle regulation and possibly epigenetic modulation. Wp activity is eventually downregulated via epigenetic mechanisms and the EBNA transcripts continue to initiate from Cp until it too is epigenetically silenced. This leads to a restriction in the expression of the latency-associated proteins until only EBNA1 is expressed from the
EBNA1-exclusive promoter Qp, which ensures against loss of the viral episome when infected memory B cells periodically divide.

2.5 Non-coding RNAs

In addition to the EBV proteins expressed during latency, the virus encodes a number of ncRNAs, i.e., that have functions independent of coding for protein. The roles that most of these ncRNAs contribute to EBV latency are poorly understood. EBV expresses several classes of small (<200 nucleotides) ncRNAs during latency, including the EBV-encoded small RNAs (EBERs), two families of microRNAs (miRNAs), and a small nucleolar RNA (snoRNA). The EBV miRNAs are expressed from two clusters, BHRF1 and BamHI-A rightward transcript (BART). Although the role of EBV proteins in latency is reasonably well-understood, the contribution of these viral ncRNAs is less clear.

EBERs

The EBERs (EBER1 and EBER2) are highly abundant, non-polyadenylated, small (<200 nucleotides) ncRNAs expressed in all latently infected cells. In fact, they are the most abundant EBV RNAs in latently infected cells, with EBER1 being about 10-fold more abundant than EBER2 due to the latter’s shorter half-life (160, 161). EBER1 and EBER2 are 167 and 172 nucleotides, respectively, and localize within the nuclei of infected cells (162). They are highly structured and have been shown to interact with cellular proteins such as La and L22, forming ribonucleoprotein complexes (163–165).
The EBERs also confer resistance to Fas-mediated and interferon (IFN)-α-induced apoptosis (166–168). They are recognized by retinoic acid-inducible gene I protein (RIG-I), thereby activating antiviral innate immune signaling to induce type I IFN and interleukin (IL)-10 and also proinflammatory cytokines via toll-like receptor 3 (TLR3) signaling (169–171). Although not essential, the EBERs significantly contribute to the efficient transformation of primary B lymphocytes in vitro (172, 173).

**BHRF1 miRNAs**

The BHRF1 miRNAs (miR-BHRF1-1, miR-BHRF1-2, and miR-BHRF1-3) were first identified in 2004 (174) and shown to be expressed at high levels during latency III when Cp/Wp are active, but not during latency I or II (175, 176). Interestingly, they are also highly expressed in Wp-restricted cell lines which have a transcriptionally active Wp but silent Cp (176). Primary B lymphocytes infected with a mutant EBV that was functionally deleted for the BHRF1 miRNAs appeared to not only have decreased proliferation early after infection, but were also protected from apoptosis (177). Furthermore, the BHRF1 miRNAs appear to promote cell-cycle progression as mutant-virus-infected cell cultures contained fewer cells in S phase compared to control EBV-infected cell cultures (177). Therefore, the BHRF1 miRNAs promote cell-cycle progression and inhibit apoptosis early after EBV infection. They have also been shown to contribute to the B-cell-transforming capacity of the virus (178, 179).

**BARTs and BART miRNAs**

Aside from the small ncRNAs, EBV also expresses the BARTs which are alternatively spliced, polyadenylated transcripts originating from the BamHI-A region of
the viral genome. Although the BARTs are highly expressed during latency I and latency II (e.g., in NPC), they are expressed only at low levels during latency III, including in LCLs (180–182). As suggested by their low levels in LCLs, the BARTs do not appear to be necessary for the transformation of primary B lymphocytes, as the region encoding several of them is deleted in the transformation-competent B95-8 EBV strain. The BARTs contain several short ORFs that have the potential to code for proteins, but thus far, it is unknown whether any proteins are in fact translated from these ORFs (183). The introns of the BARTs serve as precursors for the 22 BART miRNAs, while the fully processed RNAs appear to be nuclear ncRNAs (174, 184). A recent study suggested that these fully processed BART RNAs function as long ncRNAs (IncRNAs; >200 nucleotides), independent of their role in miRNA formation (185).

The functions of the BART miRNAs have not yet been fully elucidated, though some of them appear to contribute to transformation both in vitro and in vivo (186, 187). They have also been shown to target cellular and viral genes, including pro-apoptotic genes such as Bim to promote cell survival, as well as the cellular signaling molecule MAP3K2, and the viral genes BZLF1 and BRLF1 to facilitate the maintenance of latency in infected cells (188–191).

v-snoRNA1

The EBV snoRNA, v-snoRNA1, was first identified in 2009 by Hutzinger et al. and is located within the BART region (192). This study found v-snoRNA1 to be expressed in a number of latently infected cell lines, including LCLs established with viruses unable to initiate lytic replication, indicating that it is indeed a latent transcript. The authors proposed that v-snoRNA1 is processed into a smaller RNA with miRNA-like
activity; however, although it was shown to co-immunoprecipitate with canonical snoRNP proteins (fibrillarin, NOP56, and NOP58), a definitive function for v-snoRNA1 has yet to be determined.

2.6 BamHI-H leftward frame 1 (BHLF1)

As mentioned above, several EBV lytic genes are expressed shortly after infection of primary B cells, including BHLF1 (157). BHLF1 is the first leftward reading frame commencing in the BamHI-H restriction fragment of the EBV genome. The BHLF1 gene is depicted in Figure 2.6 and further described below.

**Discovery and description of the BHLF1 gene**

The BHLF1 transcript was first identified in the early 1980’s in a study by Hummel and Kieff, in which they identified viral DNAs encoding abundant cytoplasmic polyadenylated RNAs in permissively infected B95-8 cells that had been treated with 12-O-tetradecanoylphorbol-13-acetate (TPA) to enter the replicative cycle (1). One of the most abundant RNAs detected in that study was a 2.5-kb RNA encoded by BamHI-H which was characterized as an early RNA as its expression was not inhibited by phosphonoacetic acid treatment, which inhibits viral DNA synthesis and late gene expression. Along with Hummel and Kieff, several other groups also found that BHLF1 encodes a highly expressed, polyadenylated transcript upon activation of the replicative cycle either by TPA induction (2, 3) or by superinfection of latently infected Raji cells with the non-transforming P3HR-1 EBV strain (193). Due to its high expression levels, the
Figure 2.6 Structure of the *BHLF1* gene.

The *BHLF1* locus in relation to the linear viral genome is depicted. Bent arrows indicate mapped transcription start sites; P3 (dashed arrow) was localized by RT-PCR. Transcripts (horizontal arrows) are shown below ORF (colored). Vertical lines indicate the 125-bp *NotI* repeats which are located within the IR2 repeat region.
presence of $BHLF1$ is often used as a marker of lytic infection and viral early gene expression in clinical samples (14, 194–201).

$BHLF1$ consists primarily of a repeat region called IR2 which is comprised of 125-bp tandem direct repeats characterized by single sites for the NotI restriction enzyme (2, 202); these repeats are therefore referred to as the NotI repeats. It has been suggested that different numbers of these repeat units within the IR2 region account for the variation in size of the BamHI-H fragment between strains (3, 203, 204). Jeang and Hayward mapped the deleted region in P3HR-1, a non-transforming EBV isolate that has a 6.8-kb deletion in the BamHI-W, -Y, and -H regions of the viral genome, which includes the $BHLF1$ gene (2). This mapping revealed that the NotI repeat cluster contains a nonintegral number of repeat units and forms part of the template for the $BHLF1$ mRNA, while sequencing of the NotI repeat unit revealed an unusually high guanine-plus-cytosine (G+C) content of 84%. The NotI repeat unit was also sequenced by several other groups and the obtained sequences were either identical to each other or differed by a single nucleotide (205–207). There are approximately 500-600 bp between the NotI repeat cluster and the left duplicated sequence (DL) which flanks the 5′ end of the $BHLF1$ RNA (205, 207).

In addition to P3HR-1, there are several other EBV isolates with known deletions that include $BHLF1$ (Figure 2.7). The EBV genome in the BL cell line Daudi, contains a 7.4-kb deletion that, similar to the deletion in P3HR-1, removes both BHLF1 and the EBNA2 coding sequence or ORF (208). A subset of BLs have also been identified which carry both the WT genome as well as a transformation-defective genome in which the EBNA2 ORF has been deleted (209). These lines (e.g., Sal, Oku, and Ava) maintain what is referred to as Wp-restricted latency in which Wp remains transcriptionally active.
Figure 2.7 EBV isolates with known deletions that include *BHLF1*.
Deletions in Wp-restricted genomes (Sal, Oku, and Ava) relative to those described in P3HR-1 and Daudi. Shown above is a BamHI restriction map of the EBV B95-8 strain virus genome, with an expanded view of the BamHI W, BamHI Y, and BamHI H fragments (arrows indicate restriction sites). Open boxes identify the BamHI W1, W2 and BamHI Y1, Y2, and Y3 exons together with the ORFs encoding EBNA2, BHLF1, and BHRF1. Shown below are the positions of the deletions (solid black bars) and their coordinates relative to the B95-8 nucleotide sequence.

Reprinted with permission from Nature Publishing Group:

and Cp is silent. In these lines, EBNA1, the EBNA3s, and a truncated EBNA-LP are expressed exclusively from Wp, while EBNA2 and the LMPs are not expressed (Figure 2.8).

**Amino acid composition and putative protein**

Early sequencing identified a long ORF within the DNA sequence encoding the *BHLF1* RNA (207). As the transcript is polyadenylated, it was proposed that it would encode a protein which plays a role during the viral lytic cycle. However, the hypothetical 660-amino acid BHLF1 protein would have an unusual amino acid composition, high in proline and arginine (Table 2.2). Nonetheless, it has been demonstrated by several groups that protein can indeed be expressed from the B95-8 BHLF1 ORF (11–14). Lieberman et al. constructed a plasmid that contained the B95-8 BHLF1 ORF under the control of the SV40 early promoter and used monospecific antiserum raised against a 15-mer synthetic peptide from the predicted BHLF1 protein sequence to stain NIH 3T3 cells transfected with the BHLF1 plasmid. They were able to detect by indirect immunofluorescence discrete foci that appeared to colocalize with the cell nucleoli, thereby concluding that the *BHLF1* gene does indeed encode a protein product (11). Marchini et al. utilized this same anti-BHLF1 serum to stain for the BHLF1 protein in lytically reactivated primary B lymphocytes transformed with an EBV strain containing either WT BHLF1 or an interrupted BHLF1 in which the hygromycin phosphotransferase gene had been inserted. They detected nuclear staining in the WT-transformed LCLs, similar to that observed by Lieberman et al.; however, interruption of the *BHLF1* gene eliminated protein expression (13). Using a polyclonal monospecific rabbit antiserum generated against a fusion protein containing two repeat units of the M-ABA EBV strain,
Figure 2.8 Different patterns of EBV latent gene expression in BLs and LCLs.
Figure 2.8 Different patterns of EBV latent gene expression in BLs and LCLs.

Latency I BLs express EBNA1 exclusively from Qp. Wp-restricted BLs express EBNA1, the EBNA3s, and a truncated EBNA-LP from Wp, which remains transcriptionally active while Cp is silent. Latency III LCLs express EBNA1, the EBNA3s, and EBNA-LP from Cp, and the LMPs from separate promoters.

Reprinted with permission from the American Society for Microbiology:

Table 2.2 Amino acid composition of the hypothetical proteins coded by the regions carrying the *Not*I and *Pst*I repeats.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Composition (%) of the region carrying the following repeats:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Not</em>I</td>
</tr>
<tr>
<td>Alanine</td>
<td>15.7</td>
</tr>
<tr>
<td>Arginine</td>
<td>14.1</td>
</tr>
<tr>
<td>Asparagine</td>
<td>1.0</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>2.2</td>
</tr>
<tr>
<td>Cysteine</td>
<td>2.6</td>
</tr>
<tr>
<td>Glutamine</td>
<td>5.5</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>2.3</td>
</tr>
<tr>
<td>Glycine</td>
<td>15.1</td>
</tr>
<tr>
<td>Histidine</td>
<td>2.3</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>0</td>
</tr>
<tr>
<td>Leucine</td>
<td>3.8</td>
</tr>
<tr>
<td>Lysine</td>
<td>0</td>
</tr>
<tr>
<td>Methionine</td>
<td>0.3</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>0.1</td>
</tr>
<tr>
<td>Proline</td>
<td>21.3</td>
</tr>
<tr>
<td>Serine</td>
<td>5.7</td>
</tr>
<tr>
<td>Threonine</td>
<td>5.2</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>1.0</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>0.4</td>
</tr>
<tr>
<td>Valine</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Reprinted with permission from the American Society for Microbiology:

Nuebling and Mueller-Lantzsch were able to detect the BHLF1 protein in EBV-positive cell extracts via immunoblotting. They also observed an increase in the protein expression of BHLF1 after induction with TPA and found that the BHLF1 protein exhibits single-stranded DNA (ssDNA)-binding activity (12). Consistent with the previous studies, Xue et al. also detected the BHLF1 protein which appeared to colocalize with the nucleoli in BL cells (14).

However, recent sequencing of additional EBV genomes revealed that the BHLF1 ORF is not conserved among other EBV isolates, specifically Akata and Mutu (8). The sequencing analysis of the Akata genome revealed a stop codon shortly downstream from the corresponding methionine start codon, while the Mutu genome only has an upstream start codon with an in-frame stop codon soon after. These differences would truncate the BHLF1 protein from 660 amino acids to a mere 24 amino acids in the Akata strain and 35 amino acids in the Mutu strain. In addition to the lack of conservation between EBV isolates, the BHLF1 ORF is also not conserved between EBV and the highly homologous lymphocryptoviruses (LCVs) of Old and New World non-human primates (9, 10). The BHLF1 ORF in the Rhesus LCV has only 63.6% amino acid similarity with the EBV BHLF1 ORF (10), while the C4 gene in the marmoset LCV is a positional homolog of EBV BHLF1 but the two share no sequence homology (9).

Leftward reading frame 3 (LF3)

LF3 is a paralog of BHLF1 that encodes a 2.8-kb mRNA and is associated with the second viral origin of lytic-cycle DNA replication, oriLyt_{right} (Figure 2.9). Like BHLF1, LF3 has a leftward orientation and is comprised primarily of a repeat domain (IR4) which contains approximately 25 102-bp tandem direct repeats characterized by single sites for
Figure 2.9 Structure of the LF3 gene.

The LF3 locus in relation to the linear viral genome is depicted. Bent arrows indicate mapped transcription start sites; P3 (dashed arrow) was localized by RT-PCR. Transcripts (horizontal arrows) are shown below ORF (colored). Vertical lines indicate the 101-bp PstI repeats which are located within the IR4 repeat region.
the restriction enzyme \textit{PstI} and which has an average G+C content of 84.3\% (207). The 5’ end of \textit{LF3} is flanked by the right duplicated sequence (DS\textsubscript{R}) that is almost perfectly homologous to DS\textsubscript{L}; however, the \textit{NotI} and \textit{PstI} repeat clusters themselves show only partial homology. Unlike the \textit{NotI} repeat cluster, the \textit{PstI} repeat cluster is separated from DS\textsubscript{R} only by the dinucleotide AT (207). Similar to \textit{BHLF1}, \textit{LF3} encodes a highly expressed, polyadenylated transcript upon TPA induction of the replicative cycle (3, 210). Early sequencing also identified a long ORF within the \textit{LF3} DNA sequence, which would encode for an 896-amino acid hypothetical protein also of unusual amino acid composition, with high proline and arginine content (Table 2.2, 207). Despite the unusual predicted amino acid composition, it has been demonstrated that protein can indeed be expressed from the \textit{LF3} ORF; however, unlike \textit{BHLF1}, the \textit{LF3} protein does not appear to bind to ssDNA (211, 212). It should be noted that the widely used B95-8 EBV strain is deleted for \textit{LF3}; however, the B95-8/Raji reference sequence that is commonly used includes the B95-8 deletion as determined from the Raji genome (213).

\textit{Function of BHLF1 during the lytic cycle}

The EBV \textit{BZLF1} gene encodes an immediate-early protein called Zta, which is a basic leucine zipper protein that governs lytic reactivation from viral latency. The 5’ promoter region of the \textit{BHLF1} gene overlaps the sequence for ori\textsubscript{Lyt\textsubscript{Left}} which was found to contain binding sites for the Zta protein, i.e., Zta response elements (ZREs). Binding of full-length, bacterially expressed Zta to these ZREs activates transcription from \textit{BHLF1} (214), consistent with the observations that transcription of \textit{BHLF1} greatly increases upon TPA induction of the lytic cycle. It has also been shown that the BMRF1
protein, which is required for viral polymerase processivity, also activates the oriLyt
*BHLF1* promoter, and further activates it in combination with Zta (215).

The *BHLF1* promoter was first mapped by Jeang and Hayward and was found to be retained and functional *in vitro* in the P3HR-1 strain (2). Laux et al. also mapped the *BHLF1* and *LF3* promoters in the M-ABA strain (207). A subsequent study by Xue and Griffin using RNase protection assays identified a lytic promoter (designated P1) in both *BHLF1* and *LF3* that was sensitive to TPA induction (7). Thus far, the only known function for *BHLF1* is as an IncRNA during the replicative cycle, where it contributes *in cis* to oriLyt<sub>Left</sub> function via the formation of an RNA:DNA hybrid at the site of transcription (15). In this study, the authors showed that it is *BHLF1* RNA transcription, rather than the DNA sequence, that is essential for oriLyt<sub>Left</sub> function. Furthermore, transcription alone is not sufficient to promote oriLyt<sub>Left</sub> activation, as the production of G-rich RNA (in this case, *BHLF1*) was required.

*Evidence for function of BHLF1 during latency*

Expression of *BHLF1* during latency was first suggested by the detection of *BHLF1* transcripts or transcription early upon EBV infection of primary B lymphocytes *in vitro* and within EBV-immortalized cells (4, 157, 160, 216). Human tonsillar lymphocytes were infected with B95-8 EBV and treated with cycloheximide, which inhibits protein synthesis. Northern blot analyses of poly(A)+ RNAs from cycloheximide-treated and untreated infected cells showed that while *BHLF1* is expressed early upon infection of primary B cells, cycloheximide treatment had no effect on the transcription of the *BHLF1* mRNA, suggesting that *BHLF1* is an immediate-early viral gene (157). In B95-8 cells, *BHLF1* and the *EBERs* were found via nuclear run-on assays to be the most frequently
transcribed viral genes (160). Similarly, there was a high rate of transcription of \textit{BHLF1} in IB-4 and B95-8 cells (both latency III cell lines) as determined by nuclear run-on assays but little detectable poly(A)+ RNA in IB-4 cells as determined by northern blot (4). \textit{LF3} was not assessed in these lines due to the deletion in these EBV genomes. RNA-sequencing (RNA-seq) analyses of the EBV transcriptome in latency I BL cell lines (Mutu I and Akata) also revealed unexpectedly that the \textit{BHLF1} and \textit{LF3} transcripts are highly abundant, even more so than 98% of cellular polyadenylated transcripts (5). Additionally, very high read counts covering the \textit{BHLF1} transcript were detected via RNA-seq of LCLs generated with B95-8 EBV, which was unexpected as \textit{BHLF1} is typically associated with lytic replication (6).

As mentioned above, lytic promoters (P1) for \textit{BHLF1} and \textit{LF3} were identified which were sensitive to TPA induction. In addition to these, RNA mapping studies also suggested that both \textit{BHLF1} and \textit{LF3} have upstream promoters that are active during latent infection (7). Multiple transcription start sites 5′ to the P1 promoters were identified by RNase protection assays in the presence or absence of chemical inducing agents. These were designated P2 and P3′ for \textit{BHLF1}; transcripts initiating from P3′ were identified but the transcription start site itself was not definitively mapped. The P2 and P3′ promoters are not sensitive to chemical induction but are functionally active, suggesting that they are alternative promoters utilized during latency.

As discussed above, cell lines that maintain Wp-restricted latency have both a transformation-defective viral genome as well as a WT viral genome. It appears that the WT genome in these lines is transcriptionally silent, suggesting that the Wp-restricted lines are able to silence the Cp/Wp locus (209). Based on this, our lab previously investigated the potential contribution \textit{in cis} of \textit{BHLF1} to Cp/Wp silencing (i.e., restricted
latency or latency I), as the \textit{BHLF1} gene is deleted from the transcriptionally active EBV genomes in the Wp-restricted cell lines (16). For this, Kem I cells were utilized, which maintain the endogenous EBV genomes in latency I and support the transition from latency III to I by superinfecting EBV genomes. Superinfection of Kem I cells with either WT or a mutant recombinant EBV (rEBV), in which there is a 3,264-bp deletion that removes the entire \textit{BHLF1} locus, resulted in an eventual transition from latency III to latency I, demonstrating that neither \textit{BHLF1} nor the other elements contained within the introduced deletion are required for the establishment of latency I, at least \textit{in cis} as was originally hypothesized.

2.7 Long non-coding RNAs (IncRNAs)

In 1988, almost three decades after the idea of mRNA was established, the first mammalian IncRNA, \textit{H19}, was described in mice (217). Subsequent analysis of the human \textit{H19} gene revealed it has no conserved ORF and that despite being transcribed by RNA polymerase II, spliced, and polyadenylated, \textit{H19} is not a classical mRNA. Therefore, it was proposed that the functional product of \textit{H19} was not a protein but rather an RNA molecule (218).

Since then, IncRNAs have been defined as a class of ncRNAs that are greater than 200 nucleotides in length. Like mRNAs, IncRNAs are capped and polyadenylated; however, they typically do not encode functional proteins. To date, the majority of IncRNAs that have been identified and characterized have been mammalian IncRNAs. In addition to the lack of functional protein-coding capacity (though they may contain short
Figure 2.10 Mechanisms of IncRNA function.
There are a range of mechanisms by which lncRNAs regulate their targets; many seem to depend on specific features of primary sequence, secondary structure and genomic positioning of lncRNA effector transcripts. (1) Several lncRNAs act as RNA decoys, titrating transcription factors away from their DNA targets by directly binding to them as target mimics. (2) Others work at the post-transcriptional level as microRNA target site decoys, titrating microRNA effector complexes away from their mRNA targets. lncRNAs, the microRNA target sites of which lack the structural sequence features needed for transcript degradation, have the overall effect of ‘sponging’ their microRNA regulators. (3) Many lncRNAs seem to bind to specific combinations of regulatory proteins, potentially acting as scaffold elements within ribonucleoprotein complexes. (4) Recruitment of chromatin-modifying complexes to their DNA targets in cis has also emerged as a well-characterized function for several mammalian lncRNAs. Recruitment in trans is not depicted. A few lncRNAs seem to modulate direct processing of their mRNA targets, including translation (5), splicing (6) and degradation (7).

lncRNA, long non-coding RNA; mRNA, messenger RNA; RNP, ribonucleoprotein.

Reprinted with permission from John Wiley & Sons:

ORFs), IncRNAs are also not evolutionarily conserved. LncRNAs have a number of functions, primarily involved in the regulation of gene expression, which they can fulfil via RNA-protein, RNA-RNA, and RNA-DNA interactions. Several mechanisms of IncRNA function are illustrated in Figure 2.10 and further discussed below. Some mammalian IncRNAs are referred to as long intergenic non-coding RNAs (lincRNAs) as they are transcribed from non-coding regions between protein-coding genes.

**Mechanisms of IncRNA function**

**RNA decoy:** IncRNAs can bind to transcription factors and act as RNA decoys to prevent these proteins from binding to their DNA targets (219, 220); e.g., the IncRNA **growth arrest-specific 5 (Gas5)** has been shown to bind to the DNA-binding domain of the glucocorticoid receptor (GR), thereby competing with other glucocorticoid response elements for binding. This leads to the suppression of GR-induced transcriptional activity of glucocorticoid-responsive genes, thereby affecting cell survival and metabolic activities during starvation (219).

**miRNA sponge:** Similarly, IncRNAs can also act as miRNA sponges where they bind to miRNAs and inhibit them from interacting with their target mRNAs, thereby indirectly upregulating protein expression from those mRNAs. One such example is a muscle-specific IncRNA, **linc-MD1**, which “sponges” miR-133 and miR-135 (221). These miRNAs typically regulate the expression of transcription factors MAML1 and MEF2C, respectively, which activate muscle-specific gene expression. **Linc-MD1** was found to be a target for miR-133 and miR-135 and therefore acts as a natural decoy for these miRNAs. Depletion of **linc-MD1** led to a repression in the levels of MAML1 and MEF2C.
while its overexpression led to increased levels of these proteins, suggesting a role for \textit{linc-MD1} in the modulation of gene expression and myogenesis.

**Ribonucleoprotein (RNP) component and recruitment of chromatin modifiers:** IncRNAs can bind to cellular proteins, such as chromatin modifiers, to recruit them to their target genes. For example, the 5’ domain of the lincRNA \textit{HOTAIR} has been shown to bind polycomb repressive complex 2 (PRC2), while the 3’ domain binds the LSD1/CoREST/REST complex (222). \textit{HOTAIR} thus functions as a bridge and recruits these complexes to their target genes to regulate gene expression via histone modifications such as demethylation (223).

**Modulation of post-transcriptional events:** IncRNAs can also play a role in translation inhibition, splicing, and degradation of their mRNA targets. For example, the mammalian IncRNA \textit{MALAT1} interacts with serine/arginine (SR) splicing factors, which regulate alternative splicing, affecting their phosphorylation and therefore the levels of the active forms of these proteins (224). Other mammalian IncRNAs have also been found to play a role in the post-transcriptional inhibition of translation (225) as well as the mediation of mRNA decay (226).

\textit{Subcellular localization of IncRNAs}

The function of IncRNAs is associated with their subcellular localization. Although many IncRNAs are localized to the nucleus, some must be exported to the cytoplasm to perform their functions. Therefore, determining the subcellular localization of an IncRNA can provide essential clues to its potential functions.

Nuclear IncRNAs are typically involved in the modulation of nuclear processes such as chromatin modification, transcription regulation, and maintenance of subnuclear
structures (227). For example, as described above, the HOTAIR IncRNA regulates gene expression via the binding of its 5’ and 3’ domains to PRC2 and the LSD1/CoREST/REST complex, respectively, thereby recruiting these complexes to their target genes and facilitating chromatin modification (222, 223). MALAT1 is involved in alternative splicing and localizes to nuclear speckles which are enriched in splicing factors (224), while NEAT1 localizes to paraspeckles in the nucleus and is required for their integrity (228, 229).

On the other hand, cytoplasmic IncRNAs can affect post-translational modifications or influence gene regulation by acting as decoys for miRNAs and proteins (221, 225, 226, 230); e.g., as described above, the cytoplasmic IncRNA linc-MD1 sponges miR-133 and miR-135, preventing them from interacting with their target mRNAs and thereby affecting expression of the proteins translated from these mRNAs (221). Similarly, NORAD is a cytoplasmic IncRNA that maintains genomic stability via the sequestration of the cellular PUMILIO proteins, which are negative regulators of gene expression, thereby abrogating their ability to repress their target mRNAs (230).

### 2.8 Viral IncRNAs

Although much work has been done on mammalian IncRNAs, several viruses have been proposed or shown to also produce IncRNAs; some of these are further discussed below.
EBV

As discussed above in section 2.5, EBV expresses the BARTs which are alternatively spliced, polyadenylated transcripts that contain several short ORFs with the potential to code for polypeptides, but which have not been shown to produce detectable protein in natural infections (183, 184). A recent study suggested that fully processed BART RNAs function as nuclear IncRNAs, independent of their role in miRNA formation, possibly by affecting the regulation of cellular gene expression in EBV-infected cells (185).

Although introns are typically unstable and are degraded while still in the nucleus, a novel class of IncRNAs, called stable intronic sequence RNA (sisRNA), have recently been described (231). Recent RNA-seq analyses of EBV identified two sisRNAs, ebv-sisRNA-1 and ebv-sisRNA-2 that arise from the BamHI-W repeat region of the EBV genome which encodes EBNA-LP (Figure 2.1, 232). Each W repeat contains two coding exons, W1 and W2, separated by a short intron (81 nucleotides) and a long intron (2791 nucleotides) which were identified as ebv-sisRNA-1 and ebv-sisRNA-2, respectively. However, very little is known about these sisRNAs and thus what their function may be, if any.

*BHLF1* has been shown to function as an IncRNA during the viral replicative cycle, where it contributes *in cis* to ori*Lyt*Left function via the formation of an RNA:DNA hybrid at the site of transcription (15). However, *BHLF1* may also play a role in latency as an IncRNA; this is further discussed in sections 2.6 and 2.9.
Kaposi’s sarcoma-associated herpesvirus (KSHV)

KSHV produces a highly abundant, early IncRNA designated polyadenylated nuclear (PAN) RNA (233, 234). The PAN RNA is a ~1.1-kb transcript that is believed to lack protein-coding potential due to: 1) suboptimal ORFs that would only code for very small proteins; and 2) its nuclear localization. PAN is the most abundant viral transcript expressed during the lytic phase, estimated to account for >80% of the poly(A)+ RNA fraction (234), and has been shown to interact with several viral and cellular factors. PAN interacts with host proteins such as histones H1 and H2A, IFN regulatory factor 4 (IRF4), demethylases UTX and JMJD3, histone methyltransferase MLL2, and components of PRC2, indicating that PAN may play roles in the mediation of chromatin modification, regulation of gene expression, and immune modulation (235–237). In addition to these cellular factors, PAN has also been shown to bind directly to the KSHV ORF57 protein, a viral post-transcriptional regulator of gene expression homologous to the EBV SM and herpes simplex virus (HSV) ICP27 proteins, which protects PAN from degradation (238). PAN also binds to KSHV latency-associated nuclear antigen (LANA), facilitating the sequestration of LANA away from viral episomes during the early stages of lytic reactivation from latency (239).

Human cytomegalovirus (HCMV)

HCMV produces two IncRNAs: a ~2.7-kb transcript (β2.7) and a 5-kb immediate-early sisRNA. The β2.7 transcript is highly abundant, comprising >20% of the total viral transcriptome (240, 241), and has been shown to inhibit apoptosis via binding of genes associated with retinoid/IFN-induced mortality-19 (GRIM-19), an essential subunit of the mitochondrial enzyme complex I (242). Interestingly, β2.7 lacks a canonical ORF but has
been shown to have protein-coding capacity as it contains multiple short ORFs that are translated into small peptides (243). The 5-kb HCMV sisRNA is a non-polyadenylated, immediate-early RNA that accumulates to high abundance in the nucleus (244). However, a function for the sisRNA has yet to be elucidated, although it has been proposed that it may play a role in viral replication or spread in the host.

*Herpes simplex virus-1 (HSV-1)*

HSV-1 produces a single gene product in high abundance during latency: the latency-associated transcript (LAT) which is capped and polyadenylated. The primary ~8.3-kb LAT is further spliced into an ~6.3-kb exonic product, from which viral miRNAs are produced, and an ~2-kb intronic product which accumulates to high levels (245). The ~2-kb intronic LAT product, due to its stability, has been classified as a sisRNA. It has been shown that expression of the LAT sisRNA promotes survival of infected cells by inhibiting apoptosis (246, 247), although the exact mechanism by which it does so is not yet known. Furthermore, the LAT sisRNA has also been shown to silence lytic gene expression by inducing the formation of facultative heterochromatin on lytic promoters; i.e., high levels of histone modifications characteristic of facultative heterochromatin were found to be associated with viral lytic promoters during latent infection (248, 249).

*Flaviviruses*

Members of the *Flaviviridae* family have been shown to produce an abundant IncRNA, called subgenomic flavivirus RNA (sfRNA) or Xrn1-resistant RNA (xrRNA). The sfRNAs produced by arthropod-borne flaviviruses (e.g., West Nile virus, dengue virus, and Japanese encephalitis virus) are typically 300-500 nucleotides (250, 251) and are
formed when the cellular 5′-3′ exonuclease Xrn1 incompletely degrades the flaviviral genomic RNA, stalling within the 3′ untranslated region (UTR) due to stem-loops which act as signals for Xrn1 resistance; the remaining RNA is the sfRNA (252). The functions of these sfRNAs have yet to be fully elucidated, but there is evidence to suggest that they are involved in modulation of the RNA interference (RNAi) response to infection (253, 254). They have also been implicated in antagonism of the IFN pathway by binding to regulators of the antiviral IFN response (G3BP1, G3BP2, and CAPRIN1), thereby inhibiting their activity (255).

On the other hand, sfRNAs produced by members of the Hepacivirus and Pestivirus genera (e.g., hepatitis C virus [HCV] and bovine viral diarrhea virus, respectively) are produced when Xrn1 stalls within the 5′ UTR of the viral genomic RNA (256), resulting in sfRNAs that are only ~15-130 nucleotides shorter than the full-length ~9.6-12.5-kb genomes. These longer sfRNAs have only recently been discovered and, as such, their functions are not known. However, it has been suggested that the HCV sfRNA contributes to post-transcriptional processes (such as the increased stability of short-lived cellular mRNAs) and the dysregulation of cellular gene expression, thereby promoting cell growth and pathogenesis (256).

*Human immunodeficiency virus (HIV)*

Within the last decade, antisense ncRNAs produced from the HIV genome have been described (257–261). One of these is an IncRNA that originates in the Nef region, close to the 3′ long terminal repeat (LTR), and spans the entire length of the viral genome (260, 261). A recent study found that this IncRNA lacks a polyadenylated tail and therefore may be retained in the nucleus (261). Furthermore, it may be involved in
epigenetic regulation of HIV gene expression, as it appears not only to interact directly with DNA methyltransferase 3a (DNMT3a), but may also recruit other chromatin-modifying proteins such as enhancer of zeste homolog 2 (EZH2; involved in DNA methylation) and histone deacetylase 1 (HDAC-1) to the viral promoter (261).

2.9 *BHLF1* as an IncRNA during viral latency

Although initially classified as a lytic-cycle gene, there are several lines of evidence that suggest a role for *BHLF1* during latency, discussed in further depth above. Briefly, *BHLF1* transcripts or transcription have been detected early upon infection of primary B lymphocytes *in vitro* and within EBV-immortalized cells (4, 157, 160, 216). Furthermore, the *BHLF1* transcript is the most abundant polyadenylated transcript in latency I and is highly transcribed (although under-represented in the cytoplasmic poly(A)+ RNA fraction) during latency III (4–6, 160). In addition to its lytic promoter, *BHLF1* also appears to have latency-specific alternative promoters (7). Although *BHLF1* contains a long ORF (207) and studies have shown that a BHLF1 protein can be detected in infected cells (11–14), the amino acid composition of the BHLF1 protein would be biochemically unusual (Table 2.2). Additionally, the BHLF1 ORF is not conserved evolutionarily – a hallmark of IncRNAs – either among EBV isolates or among EBV-related LCVs of non-human New and Old World primates (8–10). In fact, base differences in the BHLF1 ORF of Akata and Mutu I genomes would result in truncated proteins of 24 and 35 amino acids, respectively (8). This suggests that *BHLF1* lacks protein-coding capacity, another characteristic of many IncRNAs. This, coupled with its
high abundance during latency, points to a potential role for $BHLF1$ during latent EBV infection. The work presented hereafter therefore aimed to investigate what role, if any, $BHLF1$ might be playing during EBV latency.
Chapter 3: Materials and Methods

3.1 Cell lines

Akata (clone 21), Kem I, and Mutu I are EBV-positive latency I BL cell lines. Kem III is a BL cell line derived from same tumor as Kem I but which maintains a latency III program. BL2 and Louckes are EBV-negative BL cell lines. Sal is a BL cell line that maintains a Wp-restricted program of EBV latency gene expression. All BL cell lines were maintained in RPMI 1640 medium (HyClone) supplemented with 10% fetal bovine serum (FBS; HyClone). Akata-LCL (Ak-LCL) and MH LCL are human LCLs generated by infection of primary B lymphocytes in vitro with Ak-GFP-WT rEBV or B95-8 EBV, respectively. Primary B lymphocyte infections and LCLs were maintained in RPMI 1640 medium supplemented with 15% FBS, 50 μg/mL gentamicin sulfate (Lonza), and 1× L-glutamine (HyClone). GP2-293 (Clontech) is a HEK293-based packaging cell line that stably expresses the viral gag and pol genes, used for the production of retroviruses. HT-1080 (ATCC) is a fibrosarcoma cell line used for titering retroviruses. HEK293 and GP2-293 cells were maintained in Dulbecco’s modified Eagle medium (DMEM; Lonza) supplemented with 10% FBS. HT-1080 cells were maintained in Eagle’s minimal essential medium (EMEM; Lonza) supplemented with 10% FBS. All cell lines were maintained at 37°C in 5% CO₂.
3.2 Generation of rEBV

All rEBV used in this study was derived from Ak-GFP-BAC (clone 12-15) (262) and produced from HEK293 cells stably transfected with Ak-GFP-BAC or its derivative. Ak-GFP-BAC DNAs, which contain a chloramphenicol resistance gene, were maintained in *Escherichia coli* strain SW105 under chloramphenicol selection. The ΔB-S rEBV was generated using recombineering as previously described (16). Briefly, SW105 cells carrying Ak-GFP-BAC (Ak-GFP-BAC SW105) were cultured and induced to express recombination proteins. These cells were then transformed by electroporation with a targeting construct containing a tetracycline resistance gene flanked by flippase (Flp) recombinase recognition target (FRT) sites and 250-bp sequences homologous to the regions immediately upstream and downstream of the locus to be deleted. Tetracycline-resistant clones were then selected and expression of the Flp recombinase induced. Chloramphenicol-resistant, tetracycline-sensitive colonies were selected and verified to be deleted for the targeted locus by sequence analysis.

The ΔBHLF1 rEBV was generated using the *galK* positive/negative selection method of recombineering in SW105 as described in Warming et al. 2005. Briefly, the GalK cassette flanked by 50-bp sequences homologous to the regions immediately upstream and downstream ("homology arms") of the BHLF1 ORF was amplified by PCR and used to transform Ak-GFP-BAC SW105 by electroporation, thereby removing the BHLF1 ORF by homologous recombination. GalK-positive clones were selected and a DNA fragment comprising of the desired deletion flanked by the 50-bp homology arms was then used to remove the GalK cassette. GalK-negative clones, which would carry the ΔBHLF1-Ak-GFP-BAC were counter-selected for the ability to metabolize galactose.
using 2-deoxy-galactose (2-DOG) minimal media plates and verified by sequence analysis.

Generation of rEBV containing either a FLAG-tagged BHLF1 ORF (FLAG-BHLF1 rEBV) or the S1 aptamer sequence (BHLF1-S1 rEBV) was also done using the galK positive/negative selection approach. First, the GalK cassette flanked by 50-bp homology arms was amplified by PCR from a plasmid containing the GalK cassette, using so-called GalK primers (Table 3.1) which were designed as follows:

**Forward:** 5′----- 50-bp homology ----- CCTGTTGACAAATATTACATCGGCA-3′
**Reverse:** 5′----- 50-bp homology (complementary strand) ----- TCAGCACTGTCCTGCTCCTT-3′

Ak-GFP-BAC SW105 cells were cultured, transformed by electroporation with the BHLF1-flanked GalK cassette, and plated on minimal media plates. Resulting colonies were struck for isolation on MacConkey plates. GalK+ clones were selected and verified by sequence analysis to contain the GalK cassette in the correct location. Due to the presence of an unexpected premature stop codon in the BHLF1 ORF, generation of the FLAG-BHLF1 rEBV was not continued.

As the GalK cassette for the BHLF1-S1 rEBV seemed to be preferentially inserted into the *LF3* gene, the left homology arm was further extended an additional 300 bp into the non-repeat region of *BHLF1* by annealing overlapping PCR fragments (Figure 3.1) to decrease the likelihood of the construct being inserted into *LF3*. This GalK cassette, flanked by a 350-bp left homology arm and a 50-bp right homology arm, was used as above to generate the GalK intermediate for the BHLF1-S1 rEBV. Next,
Table 3.1 Primers used in recombineering.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5'-3')</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>For insertion of FLAG tag immediately after ATG; primers designed using B95-8 sequence:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLAG-BHLF1-GalK Forward</td>
<td>GGGGCCGGCGTGGCCGCCGGCGCTGGCCCTGGGACCCCTAGGCCTG</td>
<td>GalK primers to amplify the GalK cassette flanked by 50-bp homology arms.</td>
</tr>
<tr>
<td>FLAG-BHLF1-GalK Reverse</td>
<td>GGGTCCCTACGACGTCTGCTGCTCCTT</td>
<td>Oligos with overlapping FLAG sequence.</td>
</tr>
<tr>
<td><strong>For insertion of S1 aptamer and flexible linker in P1; designed using Akata sequence:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3'BHLF1-GalK</td>
<td>GCCCGGGCTGAGGTGATGCACCCGGCAGCTGGAGCGACGC</td>
<td>GalK primers to amplify the GalK cassette flanked by 50-bp homology arms.</td>
</tr>
<tr>
<td>5'BHLF1-GalK</td>
<td>ACCCTAAGGTATGGCACAGAGCCCAACTGAGAATAGAAAAGACGGAGCAGGC</td>
<td>Oligos with overlapping S1 aptamer sequence.</td>
</tr>
<tr>
<td>3'BHLF1-S1</td>
<td>GCCCGGGCTGAGGTGATGCACCCGGCAGCTGGAGCGACGC</td>
<td>To amplify annealed and filled-in BHLF1-S1 oligos.</td>
</tr>
<tr>
<td>5'BHLF1-S1</td>
<td>ACCCTAAGGTATGGCACAGAGCCCAACTGAGAATAGAAAAGACGGAGCAGGC</td>
<td>To extend the left homology arm into non-repeat region</td>
</tr>
<tr>
<td>3'BHLF1-S1fill</td>
<td>GCCCGGGCTGAGGTGATGCAC</td>
<td></td>
</tr>
<tr>
<td>5'BHLF1-S1fill</td>
<td>ACCCTAAGGTATGGCAC</td>
<td></td>
</tr>
<tr>
<td>3'BHLF1arm-ext</td>
<td>CACCCGGCTGGGTGGCACCGGCTGGGCTGG</td>
<td></td>
</tr>
<tr>
<td>5'BHLF1arm-ext</td>
<td>CCTGCCGCGCTGGCAC</td>
<td></td>
</tr>
</tbody>
</table>

GalK sequence in red.

Insertion sequence (FLAG tag or S1 aptamer) in green.
Figure 3.1 Approach used to extend left homology arm of GalK targeting construct into the non-repeat region of BHLF1 for the generation of the BHLF1-S1 rEBV.

**Not to scale.**
Figure 3.1 Approach used to extend left homology arm of GalK targeting construct into the non-repeat region of *BHLF1* for the generation of the BHLF1-S1 rEBV.

The PCR-amplified GalK cassette flanked by 50-bp homology arms (light red and yellow) was being preferentially inserted into the *LF3* gene rather than *BHLF1*. To increase the likelihood of insertion into *BHLF1*, the left homology arm (light red) was extended an additional 300 bp (dark red) into the non-repeat region of *BHLF1*. This was done by PCR amplifying a 350-bp arm extension that overlapped the 50-bp left homology arm. The PCR fragments (arm extension and GalK targeting cassette) were denatured, annealed, and filled in to generate the new GalK targeting cassette that contained a 350-bp left homology arm and a 50-bp right homology arm. The same method was used for extending the left homology arm of the S1 aptamer insert used to replace the GalK cassette.
overlapping oligos (Table 3.1) were annealed and filled in to create an insertion construct containing the S1 aptamer flanked by the 50-bp homology arms. Again, the left homology arm was extended into the non-repeat region of BHLF1 by annealing overlapping PCR fragments as before. The GalK+ intermediate clones were cultured, transformed by electroporation with the S1 aptamer insertion construct, and plated on 2-DOG minimal media plates for negative selection. Resulting colonies were struck for isolation on MacConkey plates. GalK- clones were selected and verified by sequence analysis to contain the S1 aptamer in the correct location. To further verify insertion of the S1 aptamer into BHLF1, 2.5 μg of DNA was digested with BamHI for 6 h at 37°C and electrophoresed on a 0.8% agarose gel with ethidium bromide for 2 days at 40V.

3.3 Virus production and infection of BL2 cells

To produce WT and mutant rEBV, HEK293 cells were transfected with 2 μg of either Ak-GFP-BAC, ΔB-S-Ak-GFP-BAC, or ΔBHLF1-Ak-GFP-BAC using TransIT-293 transfection reagents (Mirus), and individual clones were selected based on green fluorescent protein (GFP) expression and resistance to Geneticin (500 μg/mL). To induce EBV replication, HEK293 clones were transiently transfected with 15.6 μg of pCMV-BALF4 and pCMV-HAZ, and at 24 hours post-transfection, sodium butyrate and TPA were added to the culture medium at concentrations of 20 ng/mL and 4 mM, respectively. After 3 hours, the cells were washed and then incubated in fresh RPMI growth medium for 4 days, after which the culture medium was clarified by centrifugation and passed through a 0.45-μm filter. The filtered viral supernatant was then concentrated by tangential flow filtration using a MidiKros hollow fiber filter module with a
pore rating of 500 kD (Spectrum Labs) and Bio-Rad Model EP-1 Econo Pump at a rate of 10 mL/min. To determine virus titer, 5 × 10^5 Raji cells were mixed with 1 mL of increasing dilutions of concentrated viral supernatant, plated in 6-well plates, and centrifuged at 200 × g for 1 hour at 4°C. After 24 hours of incubation, 2 mL of fresh RPMI culture medium was added to each well. At 3 days post-infection, the cells were analyzed for GFP expression by flow cytometry using the BD FACSCalibur (BD Biosciences). Flow cytometry data was analyzed using FlowJo software and the viral titer expressed as green Raji units (GRU)/mL. WT or mutant rEBV was used to infect 5 × 10^5 EBV-negative BL2 cells in 6-well plates in the same manner as the Raji cells. At 24 h post-infection, most of the medium was aspirated and 3 mL of fresh RPMI growth medium plus 1 mL conditioned medium was added to each well. At 5-7 days post-infection, the infected BL2 cells were placed under G418 selection (500 μg/mL) and subsequently expanded for further analysis.

3.4 Isolation and infection of primary B lymphocytes

Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood using a Lymphocyte Separation Medium (LSM; MP Biomedicals) gradient. CD19^+ B lymphocytes were then isolated by positive selection using human CD19 MicroBeads (Miltenyi Biotec), following the manufacturer’s instructions. The primary B lymphocytes were plated in 96-well plates at a density of 5 × 10^4 cells/well. WT or mutant rEBV was added at a starting multiplicity of infection (MOI) of 0.0675 and two-fold serial dilutions were done. The plates were then centrifuged at 1000 rpm for 1 h at 13°C. The culture medium was changed each week post-infection. At 6 weeks post-infection, the infected
cells were scored microscopically for transformation, indicated by yellow medium, clumping cells, and expression of GFP. Transformed lines were subsequently expanded for further analysis.

3.5 Induction of EBV lytic replication

Akata (clone 21) cells were plated in 6-well plates at a density of $3 \times 10^6$ cells/well. To induce viral lytic replication, goat F(ab')2 fragment to human IgG (Cappel, MP Biomedicals) was added to each well at a concentration of 100 μg/mL. After 48 h of incubation, the induced cells were harvested for subsequent analysis.

3.6 Plasmids and transfection of Louckes cells

To express the BHLF1 protein, a pSG5 vector containing the BHLF1 ORF tagged at the N-terminus with a FLAG epitope (pSG5-FLAG-BHLF1) was used. To express SM, a pcDNA3 vector containing the SM ORF was used (pcDNA3-SM). A pSG5 vector containing FLAG-tagged insulin-degrading enzyme (pSG5-IDE) was used as a positive control for transfection. Louckes cells ($5 \times 10^6$ cells/transfection) were transfected with 10-20 μg of plasmid DNA using Amaxa Nucleofector (Lonza) according to the manufacturer's instructions. Transfected cells were plated in 6-well plates and after 24 hours of incubation, 1 mL RPMI growth medium was added to each well. At 48 hours post-transfection, the transfected cells were harvested for subsequent analysis.
3.7 Immunoblot analysis

Cells were harvested by centrifugation, washed with phosphate-buffed saline (PBS), resuspended in 2× SDS-PAGE buffer with 5% β-mercaptoethanol at 10^6 cells/80 μL and sonicated. Samples were boiled for 5 minutes prior to separation by SDS-PAGE. The separated proteins were then transferred onto PDVF membranes. Cellular and EBV proteins were detected by standard immunoblotting techniques using the following primary antibodies: EBNA1, rabbit antiserum (gift of J. Herring); EBNA2, monoclonal antibody (MAb) PE2; LMP1, MAb S12; EBNA3A, -3B, -3C, sheep antiserum (Exalpha Biologicals, Inc.); β-actin, MAb JLA20 (Calbiochem); β-tubulin, H-234 (Santa Cruz Biotechnology); and FLAG M2 (Sigma-Aldrich).

3.8 RNA isolation and RT-PCR

Total cellular RNA was extracted using RNA-Bee (Tel-Test) according to the manufacturer’s instructions and treated with either RQ1-DNase (Promega) or TURBO DNase (Thermo Fisher Scientific) to remove residual DNA. cDNA was generated from 2 μg total RNA in 19-μL reaction mixtures with 200 U SuperScript III reverse transcriptase (+RT; Invitrogen) according to the manufacturer’s instructions, using either 0.1 pmol gene-specific primer (GSP) or 2.5 μM random decamers (Thermo Fisher Scientific). Corresponding negative-control reactions lacked reverse transcriptase (–RT). End-point PCR reactions were performed using 2 μL of +RT or –RT cDNA as template in a 25-μL reaction volume containing 0.5 μM of each primer, 0.25 mM dNTPs, 1× PCR buffer without Mg, 1.5 mM MgCl₂, and 4 U Platinum Taq DNA polymerase (Invitrogen) using the following cycling conditions: 95°C for 3 min, followed by 30-35 cycles of 95°C for 15
sec, annealing temperature (see Table 3.2) for 30 sec, 72°C for 1 min, and then a final extension at 72°C for 10 minutes.

3.9 RT-qPCR

Expression of EBV RNAs was quantified using cDNA primed as described above with GSPs for BHRF1, BHLF1, and SM. Latent BHRF1 transcripts were detected using a reverse primer and FAM-labelled probe with a forward primer either in the Y2 exon or in the W2 exon (158). Primer and probe sequences used for RT-qPCR analysis are listed in Table 3.3. Reactions were performed using 2 μL of cDNA as template in a 20-μL volume containing 1x TaqMan Universal Master Mix II (Applied Biosystems), 900 nM of each primer, and 250 nM of probe. A commercial GAPDH TaqMan assay (Applied Biosystems) was used as an endogenous control and samples were normalized to an EBV reference line. Reactions were carried out using the Applied Biosystems StepOnePlus Real-Time PCR System.

3.10 Isolation and RT-qPCR of miRNA

miRNA was isolated from cells using the Ambion mirVana miRNA Isolation Kit (Life Technologies). Reverse transcription of 10 ng of miRNA was performed using the TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems). Commercial TaqMan miRNA assays for miR-BHRF1-1, -2, and -3 (Applied Biosystems) were used for reverse transcription and RT-qPCR according to the manufacturer’s instructions. U6 snRNA was used as an endogenous control and samples normalized to an EBV
Table 3.2 Primers used for PCR and RT-PCR.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Annealing site</th>
<th>Sequence (5'-3')</th>
<th>Notes [PCR annealing temp, °C]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qp Fwd</td>
<td>EBNA1 5' (Q)</td>
<td>AAGGCAGGATAGC</td>
<td>Forward primer to detect Qp-derived EBNA1 mRNA [55]</td>
</tr>
<tr>
<td>K exon Rv</td>
<td>EBNA1 3' coding</td>
<td>CTCTATGTCTTGCCCT</td>
<td>Reverse primer to detect Qp- and Cp/Wp-derived EBNA1 mRNA [55]</td>
</tr>
<tr>
<td></td>
<td>exon (K)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y2 exon</td>
<td>Y2 exon</td>
<td>GAGGATGAAGACTAAGTCACAGGCTTA</td>
<td>Forward primer to detect Cp/Wp-derived EBNA1 mRNA, EBNA2 mRNA, EBNA3C mRNA [55]</td>
</tr>
<tr>
<td>EBNA2-Rv</td>
<td>Y1 exon</td>
<td>GAGAGTGACGGTTCCAAGa</td>
<td>Reverse primer to detect EBNA2 mRNA [55]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GAGAGTGACGGTTCCAACb</td>
<td></td>
</tr>
<tr>
<td>EBNA3C-Rv</td>
<td>EBNA3C BERF4</td>
<td>GGAGATGTTAGAAGCCAATGTC</td>
<td>Reverse primer to detect EBNA3C mRNA [55]</td>
</tr>
<tr>
<td></td>
<td>exon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-BHLF1</td>
<td>BHLF1</td>
<td>GATGCTGCATCGCTAGTCC</td>
<td>Forward primer to detect BHLF1 mRNA [60]</td>
</tr>
<tr>
<td>3'-BHLF1</td>
<td>BHLF1</td>
<td>GACCACGCCCTCCTTTAC</td>
<td>Reverse primer to detect BHLF1 mRNA [60]</td>
</tr>
</tbody>
</table>

\(a\) Based on complete EBV genome sequence (type 1 strain), accession number NC_007605.

\(b\) Based on complete Akata EBV genome sequence, accession number KC_207813.
<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5′-3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BHRF1 cDNA primer</td>
<td>TTCTCTTGCTGCTAGCT</td>
</tr>
<tr>
<td>BHRF1 W2 forward primer</td>
<td>TGGTAAGCGGTTCACCTTCAG</td>
</tr>
<tr>
<td>BHRF1 Y2 forward primer</td>
<td>GAGGATGAAGACTAAGTCACAGGTCTTA</td>
</tr>
<tr>
<td>BHRF1 reverse primer</td>
<td>TCCCCGTATACACAGGGCTAACAGT</td>
</tr>
<tr>
<td>BHRF1 TaqMan probe</td>
<td>AATAGGCCATCTTGCTCTACAAGATCTGGCA</td>
</tr>
<tr>
<td>BHLF1 cDNA primer</td>
<td>TCTGGGGGTCGCTGCAT</td>
</tr>
<tr>
<td>BHLF1 forward primer</td>
<td>GTACGCCTGGATTGCG</td>
</tr>
<tr>
<td>BHLF1 reverse primer</td>
<td>AGGTCGGACGACTGAGGATG</td>
</tr>
<tr>
<td>BHLF1 TaqMan probe</td>
<td>CTTGCCTGGTGCTGGAGCTCATC</td>
</tr>
<tr>
<td>SM cDNA primer</td>
<td>ACCGCCAGCATCGACTGT</td>
</tr>
<tr>
<td>SM forward primer</td>
<td>GGGCAAGGGTGACAATGTAATC</td>
</tr>
<tr>
<td>SM reverse primer</td>
<td>AAGAACACAGCCAGAGG</td>
</tr>
<tr>
<td>SM TaqMan probe</td>
<td>ACCGTGGTTTGACATGAGTTG</td>
</tr>
</tbody>
</table>
Reactions were carried out using the Applied Biosystems StepOnePlus Real-Time PCR System.

3.11 Subcellular fractionation

Nuclear/cytoplasmic RNA fractionation was performed on BL cells and LCLs as described in Chen and Carmichael, 2009. Briefly, cells were centrifuged for 10 min at 1000 rpm then rinsed with ice-cold PBS. The cell pellet was resuspended in 200 μL of lysis buffer A (10 mM Tris [pH 8.0], 140 mM NaCl, 1.5 mM MgCl₂, 0.1% Igepal CA-630 [Sigma-Aldrich]), incubated on ice for 5 min, and centrifuged at 1000 rpm for 10 min at 4°C. The cytoplasmic fraction was transferred to a new microcentrifuge tube and further centrifuged twice at maximum speed, transferring the supernatant to a new tube each time. The nuclear pellet was washed twice with lysis buffer A. RNA was isolated from the fractions using RNA-Bee according to the manufacturer’s instructions.

3.12 Cloning of BHLF1 into retroviral vector

To express the BHLF1 RNA in cells, two constructs that span the BHLF1 locus from either the P1 (lytic, BHLF1-P1) or P2 (latent, BHLF1-P2) transcription start sites (TSS) up to the poly(A) signal sequence were cloned into pLHCX, a retroviral vector for the transient or stable constitutive expression of a gene. The pSG5-FLAG-BHLF1 plasmid, which contains the FLAG-tagged BHLF1 ORF, was “repaired” at both the 5’ and 3’ ends to obtain the desired constructs. The primers used to amplify the desired fragments are listed in Table 3.4. First, the non-repeat region at the 5’ end of BHLF1, flanked by an added MfeI restriction site and an existing BstBI site, was PCR amplified
Table 3.4. Primers used in repairing pSG5-FLAG-BHLF1 and cloning *BHLF1* into retroviral vector pLHCX.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’-3’)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF1-P1-MfeI</td>
<td>GATC CAATTGAGGCCAGCTGGCCCGGCTCG</td>
<td>Forward primer to amplify 5’ non-repeat region P1 fragment flanked by MfeI and BstBI</td>
</tr>
<tr>
<td>HF1-P2-MfeI</td>
<td>GATC CAATTGCCACCTGGTATAGGGGC</td>
<td>Forward primer to amplify 5’ non-repeat region P2 fragment flanked by MfeI and BstBI</td>
</tr>
<tr>
<td>BHLF1-BstBI-Rev</td>
<td>CCGGTAGGGTTCGAATGGGCCGTTGTC</td>
<td>Reverse primer to amplify 5’ non-repeat region P1 and P2 fragments flanked by MfeI and BstBI</td>
</tr>
<tr>
<td>BHLF1-3’-EcoRI</td>
<td>GATCGAATTCCTGCTGCCATGGAATGCTC</td>
<td>Forward primer to amplify 3’ end of <em>BHLF1</em> flanked by EcoRI and BamHI</td>
</tr>
<tr>
<td>HF1-Rev-BamHI</td>
<td>GATCGGATCCGATACATGGCTGTTTAGE</td>
<td>Reverse primer to amplify 3’ end of <em>BHLF1</em> flanked by EcoRI and BamHI</td>
</tr>
</tbody>
</table>

Restriction sites in red.
using either Ak-GFP-BAC DNA or pDK286 (plasmid containing the B95-8 BamHI H fragment) as template. These P1- and P2-MfeI/BstBI fragments and pSG5-FLAG-BHLF1 were digested with MfeI and BstBI and ligated together overnight at 16°C in an insert:vector ratio of 2:1 to remove the FLAG tag and insert the P1 and P2 promoter regions at the 5’ end of the BHLF1 insert. MAX Efficiency DH10B Competent Cells (Invitrogen) were transformed with 2 μL of ligation reaction according to the manufacturer’s instructions. Colonies were screened by miniprep (Qiagen QIAprep Spin Miniprep Kit) and MfeI/BstBI digestion. Positive clones were confirmed by sequence analysis.

Next, the 3’ end of BHLF1, flanked by EcoRI and BamHI restriction sites, was PCR amplified using Ak-GFP-BAC DNA as template. The intermediate 5’-repaired pSG5-BHLF1 plasmids and 3’BHLF1-EcoRI/BamHI fragment were digested with EcoRI and BamHI and ligated together overnight at 16°C in an insert:vector ratio of 2:1. MAX Efficiency DH10B Competent Cells were transformed with 2 μL of ligation reaction as above. Colonies were screened by miniprep and BamHI/MfeI digestion. Positive clones were confirmed by sequence analysis.

To move the inserts from pSG5 into pLHCX, the fully repaired pSG5-BHLF1 plasmids (pSG5-BHLF1-P1, pSG5-BHLF1-P2, and pSG5-BHLF1-P1-B95-8) were digested with BamHI and MfeI and the BHLF1 inserts gel-purified using Qiagen MinElute Gel Extraction Kit. These inserts were ligated into Hpal-digested pLHCX overnight at 16°C at an insert:vector ratio of 2:1. One Shot Stbl3 Chemically Competent E. coli (Invitrogen) were transformed with 5 μL of ligation reaction according to manufacturer’s instructions. Colonies were screened by miniprep and HindIII/ClaI digestion for insertion
of BHLF1, and further screened by HindIII/BstBI or Clal/BstBI digestion to determine the orientation of the insert. Positive clones were confirmed by sequence analysis.

3.13 Production of retrovirus and transduction of BL2 cells

DNA from clones carrying the pLHCX-BHLF1-P1, pLHCX-BHLF1-P2, and pLHCX-BHLF1-P1-B95-8 plasmids was isolated by alkaline lysis and purified by cesium chloride gradient. GP2-293 cells were seeded in 100-mm tissue culture dishes at a density of $4 \times 10^6$ cells/plate and incubated for 24 h. After 24 h, the culture medium was changed and the cells transfected with 13 μg of a pLHCX-BHLF1 plasmid and VSV-G using TransIT-293 transfection reagents (Mirus). At 24 h post-transfection, the culture medium on the plates was changed. At 3 d post-transfection, the culture medium was clarified by centrifugation and passed through a 0.45-μm cellulose acetate filter. For viral titering, HT-1080 cells were seeded in 6-well plates at a density of $2 \times 10^5$ cells/well. After 24 h, 10-fold serial dilutions of the filtered viral supernatant was added to the cells with 4 μg/mL Polybrene (Sigma-Aldrich) and centrifuged at 1200 × g for 90 min at 32°C. At 1 d post-infection, the culture medium was changed and the cells placed under hygromycin selection (250 μg/mL). At 3 d and 7 d post-infection, the selection medium was changed. At 8 d post-infection, the cells were washed with fresh medium and then 1 mL of crystal violet in 10% EtOH solution (Electron Microscopy Sciences) was added. The cells were incubated with the stain for 2 min before being washed with sterile H$_2$O. The H$_2$O was aspirated and the plates sealed with parafilm. The number of colonies for each dilution was counted and averaged. The viral titer was calculated by multiplying the number of colonies for the least concentrated dilution by the dilution factor.
BL2 cells were seeded in 6-well plates at a density of $2 \times 10^5$ cells/well and transduced with retrovirus in the same manner as the HT-1080 cells but using 400 μg/mL hygromycin for selection. Transduced BL2 cells were expanded for subsequent analysis.
Chapter 4: Results

4.1 Contribution of BHLF1 to EBV latency

*BHLF1 deletion mutants are unable to sustain latency III.*

To determine the potential contribution of *BHLF1* to EBV latency, we began by infecting an EBV-negative BL cell line, BL2, with either WT rEBV or a mutant rEBV (ΔB-S) in which the entire BHLF1 ORF and 5′ promoter region are deleted. This 3,264-bp deletion removes *oriLyt* left and extends to the right boundary of the deletion in the EBV genome within the BL cell line Sal, which maintains Wp-restricted latency (209). Six independently-derived cell lines from infection with either WT or ΔB-S rEBV were analyzed. We observed the establishment of latency III in both WT and ΔB-S infections as indicated by the expression of the latency-associated EBNA2, LMP1, and EBNA3 proteins at 1 month post-infection (Figure 4.1A, left panel). However, by approximately 2 months post-infection, the ΔB-S infections had transitioned to latency I, evidenced by the downregulation of expression of the latency-associated proteins other than EBNA1 (Figure 4.1A, right panel). The EBNA1 protein was still detectable in these infections, indicating the presence of the EBV genome.

To assess promoter usage for *EBNA1* mRNA expression, end-point RT-PCR was performed on RNA isolated from the infected BL2 cells. The results showed that at 1 month post-infection, the *EBNA1* transcript in both the WT and ΔB-S infections was derived from Cp/Wp which are active during latency III, while by 2 months post-infection,
the transcript was derived exclusively from Qp in the ΔB-S infections, consistent with latency I (Figure 4.1B). This was in contrast to the WT infections (latency III) in which Cp/Wp remained active at 2 months post-infection. Furthermore, RT-PCR analysis of the 3’ ends of EBNA2 and EBNA3C mRNAs showed that the levels of these transcripts were lower at 2 months post-infection (Figure 4.1C), further demonstrating the transition to latency I. Interestingly, at an intermediate time point (41 days post-infection), the EBNA2 and EBNA3C transcripts were still detectable (Figure 4.1C), although the corresponding proteins were not evident by immunoblotting at this time (Figure 4.1D). End-point RT-PCR analysis confirmed that the BL2+ΔB-S cell lines did not express the BHLF1 transcript, as expected (Figure 4.1E).

As the deletion in the ΔB-S rEBV also removes the region encoding the BHRF1-1 miRNA, we generated another mutant rEBV (ΔBHLF1) in which only the BHLF1 ORF is deleted to ensure that loss of the BHRF1-1 miRNA is not responsible for the observed phenotype. Infection of BL2 cells with ΔBHLF1 rEBV also resulted in the establishment of latency III, followed by a transition to latency I within a similar timeframe as the ΔB-S rEBV infections, at both the protein (Figure 4.2A) and mRNA levels (Figure 4.2B). We also confirmed by end-point RT-PCR that the two BL2+ΔBHLF1 cell lines did not express the BHLF1 transcript (Figure 4.2C). To confirm our observations in these two BL2+ΔBHLF1 lines, we repeated the infection of BL2 cells with the ΔBHLF1 rEBV and obtained an additional six independently-derived cell lines. Consistent with the previous two lines, the repeated ΔBHLF1 infections established latency III and transitioned to latency I within a similar timeframe (by 3 months post-infection) as determined by immunoblot analysis of the latency-associated proteins (Figure 4.3A). RT-qPCR analysis
of the WT and ΔBHLF1 infections at 1 month post-infection confirmed that the ΔBHLF1 infections did not express the \textit{BHLF1} transcript (Figure 4.3B).

**Deletion of \textit{BHLF1} does not negatively affect \textit{BHRF1} mRNA and miRNA expression.**

The deletion in the ΔB-S rEBV not only removes the \textit{BHLF1} ORF and 5’ promoter region, but also the first miRNA encoded by the adjacent \textit{BHRF1} gene, miR-BHRF1-1. As the deletion in the ΔBHLF1 rEBV is smaller, the miR-BHRF1-1 sequence is retained in this mutant. To confirm that: 1) the deletion of miR-BHRF1-1 is not responsible for the phenotype that we have observed; and 2) neither deletion affects the expression of the other \textit{BHRF1} miRNAs, we examined the expression levels of the three \textit{BHRF1} miRNAs in the WT- and mutant-infected BL2 cells. As expected, BL2 cells infected with ΔB-S rEBV did not express \textit{miR-BHRF1-1} (Figure 4.4A). However, they expressed similar or higher levels of \textit{miR-BHRF1-2} and \textit{miR-BHRF1-3} compared to WT rEBV-infected BL2 cells. ΔBHLF1 rEBV-infected BL2 cells expressed all three \textit{BHRF1} miRNAs at levels similar to WT rEBV-infected BL2 cells (Figure 4.4B).

We next assessed whether the deletions in the \textit{BHLF1}-mutant rEBVs had affected the expression of the \textit{BHRF1} mRNA. To do this, we compared \textit{BHRF1} mRNA expression levels in the six independently derived WT and mutant rEBV-infected BL2 cell lines by RT-qPCR. We used two different forward primers to amplify the alternatively spliced latent forms of \textit{BHRF1} transcripts that originate from Cp/Wp, W2-BHRF1 and Y2-BHRF1. The W2 and Y2 forward primers anneal within the W2 and Y2 exons, respectively; however, the Y2-containing \textit{BHRF1} transcripts would be undetectable in
Wp-restricted lines in which the Y2 exon is deleted. BL2 cells infected with ΔB-S rEBV expressed higher levels of both W2-BHRF1 and Y2-BHRF1 compared to the WT rEBV-infected BL2 cells (Figure 4.5A). However, ΔBHLF1 rEBV-infected BL2 cells expressed similar levels of W2-BHRF1 and slightly higher levels of Y2-BHRF1 compared to the WT-infection controls (Figure 4.5B). As expected, Sal cells (which maintain Wp-restricted latency) did not express Y2-BHRF1, but did express high levels of W2-BHRF1.

**Contribution of BHLF1 to EBV-mediated B-cell immortalization.**

Thus far, the data suggested that BHLF1 is needed in the context of BL2 cells for the long-term maintenance of latency III. To examine the role of BHLF1 in a more biologically relevant environment, primary B cells isolated from four different donors were infected with either of the two BHLF1-mutant rEBVs to determine if BHLF1 is required for the immortalization of primary B cells, as would be expected if latency III could not be sustained. Infection with the WT rEBV as a control resulted in efficient immortalization of primary B cells isolated from the first donor (Figures 4.6A, 4.7A). However, ΔB-S rEBV was unable to immortalize the primary B cells from the same donor (Figure 4.6A), while infection of these cells with ΔBHLF1 rEBV resulted in inefficient immortalization (Figure 4.6B). Surprisingly, when primary B cells from a second donor were infected with ΔBHLF1 rEBV, they were immortalized as efficiently as with WT rEBV (Figure 4.6C). However, these cells in general appeared to also be more efficiently immortalized by WT rEBV than the cells from the first donor (Figure 4.6D), suggesting that perhaps the cells from the second donor were more susceptible to immortalization. Due to this striking difference in results between the two donors, we
infected primary B cells from two additional donors, obtained from Stemcell Technologies. Unfortunately, cells from the third donor (ST-1) were not efficiently immortalized, even by WT rEBV (Figure 4.6E). However, the WT and ΔBHLF1 lines that proliferated upon expansion all established latency III, as determined by immunoblot analysis (Figure 4.7D).

WT rEBV efficiently immortalized the cells from the fourth donor (ST-2; Figures 4.6F, 4.7E), while infection with ΔBHLF1 rEBV resulted in limited immortalization, similar to what had been observed in B cells from donor 1 with this mutant virus. Interestingly and unexplainably, several wells from the infections done with intermediate MOIs grew out, while wells from the highest MOIs did not (Figure 4.6F).

As the mutant rEBV appeared to transform B cells in some instances, we asked whether the resultant lines were indeed comparable to WT rEBV-infected lines. To do this, we compared the growth of WT and ΔBHLF1 rEBV-infected primary B cells from the fourth donor. Cells were seeded in 6-well plates at a density of 10^5 cells per well, and viable cells were counted daily for approximately two weeks. The three WT rEBV-infected lines proliferated at nearly identical rates and to the same saturation density. Two of the ΔBHLF1 rEBV-infected lines, by contrast, exhibited little to no ability to proliferate beyond the initial seeding density (Figure 4.8B), while the other two ΔBHLF1 rEBV-infected lines proliferated more slowly and to a lower density than the WT rEBV-infected lines (Figure 4.8). Thus, while we did observe some immortalization by the BHLF1-mutant rEBV, the resulting lines appeared to be less vigorous in their growth, even though they appeared to sustain latency III.
Figure 4.1 The *BHLF1* locus is required to sustain latency III in BL2 cells.
Figure 4.1 The *BHLF1* locus is required to sustain latency III in BL2 cells.
Figure 4.1 The *BHLF1* locus is required to sustain latency III in BL2 cells.
Figure 4.1 The *B*HLF1 locus is required to sustain latency III in BL2 cells.
Figure 4.1 The *BHLF1* locus is required to sustain latency III in BL2 cells.
Figure 4.1 The *BHLF1* locus is required to sustain latency III in BL2 cells.

(A) Immunoblot detection of EBV latency-associated proteins expressed in BL2 cells at 30 and 68 days post-infection with either WT or ΔB-S rEBV. WT infections establish and maintain a latency III-specific pattern of protein expression even at 68 days post-infection, while ΔB-S infections establish latency III early but later transition to a latency I-specific pattern of protein expression. All cell lines were derived from independent infections by the respective rEBV. Kem III and BL2 served as positive and negative controls, respectively. Actin and β-tubulin served as loading controls. (B) End-point RT-PCR analysis of *EBNA1* promoter usage at 28, 41, and 68 days post-infection indicated a shift from Cp/Wp (latency III) to Qp (latency I) in BL2 cells infected with ΔB-S rEBV, corresponding to the loss of latency III-specific protein expression in these infections. Kem I, Kem III, and BL2 cells served as latency I, latency III, and EBV-negative controls, respectively. (C) Similar decreases in *EBNA2* and *EBNA3C* mRNA expression were detected in BL2 cells infected with ΔB-S but not WT rEBV. (D) Immunoblot detection of EBV latency-associated proteins in the same infected BL2 cells at 41 days post-infection revealed that the latency III-specific proteins had already mostly been downregulated. (E) End-point RT-PCR to detect the presence of *BHLF1* confirmed that BL2 cells infected with ΔB-S rEBV did not express the *BHLF1* transcript. Kem I and Kem III served as positive controls for the presence of the *BHLF1* gene while uninfected BL2 cells served as a negative control.
Figure 4.2 rEBV lacking the BHLF1 ORF alone is unable to sustain latency III in BL2 cells.
Figure 4.2 rEBV lacking the BHLF1 ORF alone is unable to sustain latency III in BL2 cells.
Figure 4.2 rEBV lacking the BHLF1 ORF alone is unable to sustain latency III in BL2 cells.

(A) Immunoblot detection of EBV latency-associated proteins in BL2 cells infected with ΔBHLF rEBV at 18-386 days post-infection revealed a shift from a latency III to latency I-specific pattern of EBV protein expression by late time post-infection. Kem III and BL2 served as positive and negative controls, respectively. Actin and β-tubulin served as loading controls. (B) End-point RT-PCR analysis of EBNA1 promoter usage at 23, 60, and 92 days post-infection indicated a shift from Cp/Wp (latency III) to Qp (latency I) in BL2 cells infected with ΔBHLF1 rEBV. Kem I, Kem III, and BL2 cells served as latency I, latency III, and EBV-negative controls, respectively. A similar decrease in EBNA3C mRNA expression was also detected; sequencing of the smaller bands in infection #2 at 60 and 92 days post-infection revealed that they were alternatively spliced transcripts of EBNA3C. (C) End-point RT-PCR to detect the presence of BHLF1 confirmed that BL2 cells infected with ΔBHLF1 rEBV did not express the BHLF1 transcript. Kem I served as a positive control for the presence of the BHLF1 gene.
Figure 4.3 Further confirmation that the rEBV lacking the BHLF1 ORF alone is unable to sustain latency III in BL2 cells.
Figure 4.3 Further confirmation that the rEBV lacking the BHLF1 ORF alone is unable to sustain latency III in BL2 cells.
Figure 4.3 Further confirmation that the rEBV lacking the BHLF1 ORF alone is unable to sustain latency III in BL2 cells.

(A) Immunoblot analysis at 1, 2, and 3 months post-infection of an additional six independently-derived BL2 cell lines infected with ΔBHLF1 rEBV confirmed the shift from a latency III to latency I-specific pattern of EBV latency-associated protein expression. Kem III and BL2 served as positive and negative controls, respectively. Actin and β-tubulin served as loading controls. (B) RT-qPCR of BHLF1 expression in triplicate confirmed that BL2 cells infected with WT rEBV express BHLF1 while those infected with ΔBHLF1 rEBV do not. Sal and uninfected BL2 cells served as negative controls. Expression values are relative to the BHLF1 RNA level in the reference line Ak-LCL. Error bars indicate the standard error of the means.
Figure 4.4 Deletion of the *BHLF1* locus does not negatively impact expression of the *BHRF1* miRNAs.
Figure 4.4 Deletion of the *BHLF1* locus does not negatively impact expression of the *BHRF1* miRNAs.
Figure 4.4 Deletion of the BHLF1 locus does not negatively impact expression of the BHRF1 miRNAs.

Expression of miR-BHRF1-1, miR-BHRF1-2, and miR-BHRF1-3 in BL2 cells was analyzed by RT-qPCR at 1 month post-infection in BL2 cells infected with either (A) WT or ΔB-S rEBV; or (B) WT or ΔBHLF1 rEBV. Each bar represents the mean relative level of expression determined, in triplicate, for each of the three BHRF1 miRNAs in six independently infected BL2 cell lines. The Sal cell line has a naturally occurring deletion which removes miR-BHRF1-1 but not the other two miRNAs. Expression values are relative to the respective BHRF1 miRNA level in the reference line Ak-LCL. Error bars indicate the standard error of the means.
Figure 4.5 Deletion of the BHLF1 locus does not negatively impact expression of the mRNA encoding BHRF1.
Figure 4.5 Deletion of the BHLF1 locus does not negatively impact expression of the mRNA encoding BHRF1.
**Figure 4.5 Deletion of the *BHLF1* locus does not negatively impact expression of the mRNA encoding *BHRF1*.**

Expression of the latency-associated *BHRF1* mRNA was analyzed by RT-qPCR at 1 month post-infection in BL2 cells infected with either (A) WT or ΔB-S rEBV; or (B) WT or ΔBHLF1 rEBV. Latent *BHRF1* transcripts were detected using a forward primer either in the Y2 or W2 exon. The Sal cell line has a naturally occurring deletion which removes the Y2 exon. Each bar represents the mean relative level of expression determined, in triplicate, for the *BHRF1* mRNA in six independently infected BL2 cell lines. Expression values are relative to the respective *BHRF1* mRNA level in the reference line Ak-LCL. Error bars indicate the standard error of the means.
Figure 4.6 BHLF1 contributes to EBV-mediated B cell immortalization.
Figure 4.6 BHLF1 contributes to EBV-mediated B cell immortalization.
Figure 4.6 BHLF1 contributes to EBV-mediated B cell immortalization.
Figure 4.6 *BHLF1* contributes to EBV-mediated B cell immortalization.

Primary B lymphocytes from four adult donors were plated in 96-well plates at a density of $5 \times 10^4$ cells/well and infected *in vitro* with either WT, ΔB-S (donor 1 only), or ΔBHLF1 rEBV at the indicated MOIs (two-fold serial dilutions with a starting MOI of 0.0675). Immortalization was scored at 6 weeks post-infection and is presented as percent of wells that contained transformed cells. (A) Donor 1 infected with WT or ΔB-S rEBV. (B) Donor 2 infected with WT or ΔBHLF1 rEBV. (C) Donor 2 infected with WT or ΔBHLF1 rEBV. (D) Combined graphs of Donors 1 and 2 infected with WT or ΔBHLF1 rEBV. Red line and arrow show the increase in efficient immortalization of donor 2 cells compared to donor 1 cells. (E) Donor 3 and (F) Donor 4 infected with WT or ΔBHLF1 rEBV.
Figure 4.7 Primary B lymphocytes immortalized by WT, ΔB-S, or ΔBHLF1 rEBV establish latency III.
Figure 4.7 Primary B lymphocytes immortalized by WT, ΔB-S, or ΔBHLF1 rEBV establish latency III.
Figure 4.7 Primary B lymphocytes immortalized by WT, ΔB-S, or ΔBHLF1 rEBV establish latency III.
Figure 4.7 Primary B lymphocytes immortalized by WT, ΔB-S, or ΔBHLF1 rEBV establish latency III.
Figure 4.7 Primary B lymphocytes immortalized by WT, ΔB-S, or ΔBHLF1 rEBV establish latency III.
Figure 4.7 Primary B lymphocytes immortalized by WT, ΔB-S, or ΔBHLF1 rEBV establish latency III.

Immunoblot detection of EBV latency-associated proteins in infected primary B lymphocytes after successful expansion confirmed that these cell lines establish a latency III-specific pattern of protein expression. (A) Donor 1 infected with WT rEBV. No ΔB-S lines obtained. (B) Donor 2 infected with WT or ΔBHLF1 rEBV. (C) Donor 2 infected with WT or ΔBHLF1 rEBV. (D) Donor 3 and (E) Donor 4 infected with WT or ΔBHLF1 rEBV. Kem III and BL2 cells served as positive and negative controls, respectively. Actin and β-tubulin served as loading controls.
Figure 4.8 Growth curves of transformed primary B-cell lines.
EBV-transformed cell lines that were successfully expanded after infection of primary B lymphocytes from donor 4 with (A) WT or (B) ΔBHLF1 rEBV were plated in triplicate at a starting density of 10^5 cells/well and monitored daily for growth over a two-week period. Three independent WT lines and four independent ΔBHLF1 lines were examined. (C) Combined graph showing both WT and ΔBHLF1 growth curves.
4.2 *BHLF1* may be functioning as an IncRNA

**Expression of BHLF1 protein is dependent on expression of SM protein.**

Having observed the deficiency in sustaining latency III upon infection with the *BHLF1*-mutant rEBVs, we originally had to consider a role for the BHLF1 protein. To begin to assess its potential role, but without an antibody to detect its expression, we first cloned the BHLF1 ORF from the DNA of the B95-8 isolate, fused in-frame at the 5′ end to the sequence encoding the FLAG epitope, into the expression vector pSG5. Transfection of EBV-negative Louckes BL cells with only the BHLF1 expression vector resulted in little to no detectable expression of the FLAG-tagged BHLF1 protein (Figure 4.9). We therefore reasoned that the presence of another viral protein, particularly a lytic-cycle one, may be critical for expression of the BHLF1 protein. SM is an early lytic-cycle RNA-binding protein that promotes the expression of a number of viral lytic-cycle mRNAs. Upon co-transfection with an expression vector containing the SM ORF, we did indeed observe a dose-dependent increase in FLAG-BHLF1 expression. This suggested that the BHLF1 protein is efficiently expressed only in the presence of SM, i.e. during the lytic cycle, and is unlikely to be responsible for the inability of *BHLF1*-mutant rEBVs to sustain latency III.

**The BHLF1 ORF is not conserved among EBV isolates.**

To explore the expression of the BHLF1 protein in the context of viral infection, we next sought to use recombineering to generate an rEBV with a FLAG epitope in the N-terminal region of the BHLF1 ORF, immediately downstream of the ATG start codon.
Surprisingly, during the course of this work, our sequence analysis of the Ak-BAC BHLF1 across the insertion site for the FLAG-encoding element revealed a single base-pair deletion in the Akata EBV genome that resulted in a frameshift and subsequent stop codon. This would truncate the 660-amino acid BHLF1 protein after 24 amino acids. In addition to this single base pair deletion, there were also other base pair substitutions within the ORF and promoter region, compared to the published consensus B95-8/Raji sequence (Figure 4.10). We confirmed these base differences via sequence analysis of DNA isolated from A21 cells derived from the Akata BL, which contain the Akata EBV genome. Furthermore, the base differences we observed were consistent with a subsequent report of the complete sequence of the Akata EBV genome by Lin et al. (8). Sequencing analysis of the Mutu EBV genome by this group also revealed base differences in this isolate that would result in a truncated 35-amino acid BHLF1 protein (8). Thus, the BHLF1 ORF is not conserved among all EBV isolates, providing further evidence of a non-coding function of the gene and its RNA. Due to the presence of the premature stop codon in the Ak-BAC BHLF1, we did not proceed with the generation of the rEBV that would encode a FLAG-tagged BHLF1 ORF.
Figure 4.9 Expression of BHLF1 protein is dependent on SM.

EBV-negative Louckes cells were transfected using Amaxa Nucleofector with a pSG5 vector containing the BHLF1 ORF tagged at the N-terminus with a FLAG epitope with or without increasing amounts of a pcDNA3 vector containing the ORF of the EBV early lytic-cycle protein SM. Transfected cells were plated in 6-well plates and fresh growth medium was added at 24 hours post-transfection. The cells were harvested at 48 hours post-transfection and the FLAG-tagged BHLF1 protein detected via immunoblot. Expression of the BHLF1 protein in transiently transfected EBV-negative Louckes cells requires the co-transfection of a vector containing SM. Insulin-dependent enzyme (IDE) served as a positive control for detection of the FLAG tag. Actin served as a loading control.
Figure 4.10 BHLF1 ORF is not conserved between Akata and B95-8/Raji genomes.

Sequencing of the Ak-BAC *BHLF1* and DNA isolated from A21 cells, which contain the Akata genome, revealed base pair differences both upstream and downstream of the ATG start codon (arrow) compared to the B95-8/Raji reference genome; these differences are highlighted in yellow. The Akata genome contains a single base pair deletion (red box) downstream of the start codon which truncates the BHLF1 protein from 660 to 24 amino acids in the Akata strain.
4.3 Expression and localization of the \textit{BHLF1} RNA in different latency programs

\textbf{Expression of \textit{BHLF1} RNA in different cell lines.}

As it appears that BHLF1 protein is unlikely to be expressed during latency, we examined \textit{BHLF1} RNA expression levels in different cell lines using RT-qPCR. To get a better picture of when the RNA is expressed, we examined B-cell lines that maintain either latency I (Kem I, Mutu I, and A21), latency III (Kem III, MH LCL, and Ak-LCL), or Wp-restricted latency (Sal), as well as A21 cells in which EBV replication had been induced by IgG cross-linking of the B-cell receptor. We also assessed expression of the lytic-cycle gene \textit{SM} as an indicator of the extent to which any given cell line may contain cells that have spontaneously entered the lytic cycle. As expected, induced A21 cells had the highest expression levels of both \textit{BHLF1} and \textit{SM} RNA, while uninduced A21 cells expressed very low levels of both genes (Figure 4.11). Kem I, Mutu I, Kem III, MH LCL, and Ak-LCL cell lines all expressed similar levels of \textit{BHLF1}, while expectedly, Sal cells did not express \textit{BHLF1}. Surprisingly, Kem III cells expressed high levels of \textit{SM}, higher than even the induced A21 cells, while the other uninduced cell lines expressed relatively low levels of \textit{SM}. Thus, there is not always a strict correlation between \textit{SM} and \textit{BHLF1}. Furthermore, these results show that the \textit{BHLF1} RNA is indeed expressed in B-cell lines that maintain both latency I and III.

\textbf{Expression of \textit{BHLF1} RNA within nuclear and cytoplasmic fractions.}

To determine the subcellular localization of the \textit{BHLF1} RNA, we extracted RNA from the nuclear and cytoplasmic fractions of latently infected B-cell lines (Kem I, Kem
Ill, A21, and WT rEBV-infected BL2 cells) and lytically reactivated A21 cells, and quantified BHLF1 RNA expression by RT-qPCR. BHLF1 expression was higher in the nuclear fractions of the Kem I, Kem III, A21, and WT rEBV-infected BL2 cells compared to the cytoplasmic fractions (Figure 4.12). As expected, BHLF1 expression in induced A21 cells was increased in both the nuclear and cytoplasmic fractions; however, the nuclear:cytoplasmic ratio of BHLF1 was slightly higher in the induced A21 cells compared to the uninduced A21 cells. These results suggest that the BHLF1 RNA localizes predominantly to the nucleus.

**Expression of BHLF1 RNA in BL2 cells infected with rEBV.**

We next used RT-qPCR to assess the expression of the BHLF1 RNA in BL2 cells infected with the WT rEBV, including at 3 months post-infection, by which time the BL2+ΔBHLF1 cell lines (which do not contain or express BHLF1) had already transitioned from latency III to I. Although we found that, overall, BHLF1 expression levels were decreased at 3 months post-infection compared to those at 1 month post-infection, these BL2+WT lines did not express high levels of BHLF1 even at 1 month post-infection (Figure 4.13). These results could be significant if BHLF1 is indeed functioning as an IncRNA, particularly via post-transcriptional or epigenetic mechanisms, where only low amounts of the IncRNA are required to effect changes.
Quantitative determination of levels of *BHLF1* RNA expression was done in triplicate by RT-qPCR in B-cell lines maintaining either latency I (Kem I, Mutu I, and A21), latency III (Kem III, MH LCL, Ak-LCL), or Wp-restricted latency (Sal) as well as A21 cells in which viral lytic replication had been induced by addition of IgG (induced A21). Quantification of the RNA encoded by the EBV early-gene SM served as an indicator of the relative level of lytic-cycle gene expression. Uninfected BL2 cells served as a negative control. Expression values are relative to the *BHLF1* RNA level in the reference line Ak-LCL. Error bars indicate the standard error of the means.
Figure 4.12 Expression of *BHLF1* RNA within nuclear and cytoplasmic fractions.

Nuclear/cytoplasmic fractionation was performed on various cell lines and the levels of *BHLF1* RNA expression quantitatively determined in triplicate by RT-qPCR. Quantification of the RNA encoded by the EBV early-gene SM served as an indicator of the relative level of lytic-cycle gene expression. Uninfected BL2 cells served as a negative control. Expression values are relative to the *BHLF1* RNA level in the reference line Ak-LCL. Error bars indicate the standard error of the means.
Figure 4.13 Expression of *BHLF1* RNA in BL2 cells infected with WT rEBV.

Levels of *BHLF1* RNA expression at 1 and 3 months post-infection in BL2 cells infected with WT rEBV were quantitatively determined in triplicate by RT-qPCR. (A) Shown here are the data for each of the six BL2+WT cell lines. Sal and uninfected BL2 cells served as negative controls. (B) All values for the six cell lines were grouped for 1 and 3 months post-infection. Expression values are relative to the *BHLF1* RNA level in the reference line Ak-LCL. Error bars indicate the standard error of the means.
Chapter 5: Discussion

5.1 Summary of findings in thesis

Although classified as an immediate-early lytic gene, BHLF1 is highly transcribed in infected B-cell lines that maintain either latency III or the more restricted latency I program (4, 5). It has been shown to function as an IncRNA during lytic-cycle DNA replication via the formation of an RNA:DNA hybrid at the site of viral transcription, adjacent to oriLytLeft (15); however, a function for BHLF1 during EBV latency has not been described. Thus, in this work, we investigated the contribution of BHLF1 to viral latency and began to elucidate its potential latency-associated functions.

Contribution of BHLF1 to EBV latency

Initially, we hypothesized that BHLF1 acts to direct silencing of the Cp/Wp locus in cis and was therefore critical to the transition from latency III to latency I. This was based on observations that the WT genome in cells maintaining Wp-restricted latency is transcriptionally silent, suggesting that the Wp-restricted lines (which lack BHLF1 in their transcriptionally active but transformation-defective viral genomes) are able to silence the Cp/Wp locus (209). Our experiments to test this hypothesis showed that this is unlikely to be the case as superinfection of Kem I cells with the ΔB-S rEBV resulted in an eventual transition from latency III to latency I, demonstrating that BHLF1 is not required in cis for the establishment of latency I as we originally hypothesized (16).
We have shown here instead that \textit{BHLF1} does indeed appear to contribute to the maintenance, though not the establishment, of EBV latency in EBV-negative BL2 cells. Infection of BL2 cells with WT rEBV or the mutant ΔB-S rEBV both resulted in the establishment of latency III, as expected. However, the mutant-virus infections eventually transitioned to latency I by \sim 2\ months post-infection, both at the protein and mRNA (and presumably transcriptional) levels. Interestingly, the \textit{EBNA2} and \textit{EBNA3C} transcripts in the ΔB-S infections were still present at 2 months post-infection, whereas the respective proteins were no longer detectable. This is consistent with observations made by a previous postdoctoral scholar in our lab, David Hughes, in which the deletion of \textit{BHLF1} resulted in an increase in intron retention at the 5′ end of \textit{EBNA} transcripts originating from Cp/Wp (discussed in further detail below). To rule out the possibility that removal of the \textit{BHRF1-1} miRNA contributes to the mutant phenotype, we also utilized a second mutant rEBV, ΔBHLF1, in which only the BHLF1 ORF is deleted, leaving the coding region for the \textit{BHRF1-1} miRNA intact. Narrowing the deleted region again resulted in the establishment of latency III in the infected cells followed by a transition to latency I by 3 months post-infection.

In the context of a more biologically relevant environment, our findings show that deletion of the \textit{BHLF1} locus (ΔB-S rEBV) does not result in immortalization of primary B cells, although it is important to note that infection with the ΔB-S rEBV was only carried out using primary B cells isolated from a single donor. Interestingly, infection with the rEBV in which the BHLF1 ORF alone is deleted (ΔBHLF1 rEBV) does not always result in a lack of immortalization of primary B cells. However, whether the primary B cells are immortalized or not may be dependent on the donor from which they are isolated. For example, infection of primary B cells from the first donor with either ΔB-S or ΔBHLF1
rEBV resulted in either no or inefficient immortalization, respectively. By contrast, B cells from a second donor appeared to be more susceptible to immortalization with both WT and ΔBHLF1 rEBV, with no apparent difference in immortalization efficiency between the two rEBVs. Conversely, cells from a third donor were also not efficiently immortalized by the ΔBHLF1 rEBV, though we were able to expand several immortalized lines from this donor. These findings suggest that there may be cellular factors that balance the effect of BHLF1 in host cells, and that the levels of these factors may differ between hosts, thereby modulating the effect of BHLF1 in cells from different donors.

BHLF1 as an IncRNA

Although it is clear that the loss of BHLF1 has an effect on latency, we wanted to determine which form of BHLF1 (protein, RNA, or genomic locus) mediates this effect. Previous studies by several groups demonstrated detectable BHLF1 protein (11–14); however, there is evidence to suggest that BHLF1 would not in fact encode a functional protein. Not only would the amino acid content of the hypothetical BHLF1 protein be highly unusual (207), but the ORF and/or amino acid sequence of BHLF1 is not conserved among EBV isolates (8); in fact, translation of the BHLF1 ORFs in the Akata and Mutu I EBV genomes would result in truncated proteins of 24 and 35 amino acids, respectively, instead of the full-length 660 amino acids. Our sequencing analysis of EBV DNA isolated from A21 cells, which carry the Akata EBV genome, confirm the nucleotide differences (relative to the prototypic EBV isolate B95-8) identified by Lin et al. which would result in a truncated BHLF1 protein (8). Further, our findings suggest that the BHLF1 protein would only be efficiently expressed during the lytic cycle, when the lytic-cycle protein SM, which our data indicate is required for BHLF1 expression, is
expressed. Based on these observations, we conclude that the BHLF1 protein is unlikely to be responsible for the effect that we observe during latency following infection with BHLF1-mutant rEBVs.

**Pattern of expression and localization of the BHLF1 RNA**

Previous observations by other groups indicated that the BHLF1 transcript is the most abundant polyadenylated transcript in latency I, and is highly transcribed during latency III (but is under-represented in the polyadenylated RNA fraction in the cytoplasm) (4, 5). As discussed in Chapter 2.7, the function of an lncRNA can, to some extent, be inferred based on its subcellular location in the cell. Since many lncRNAs are localized in the nucleus, and based on our observations that suggest that BHLF1 may play a role in regulating splicing (discussed below), we hypothesized that the BHLF1 RNA localizes exclusively or primarily to the nucleus during latency. Thus, to further elucidate the function of BHLF1, we attempted to determine the subcellular locations and pattern of expression of the RNA. Our findings suggest that the BHLF1 RNA does indeed localize to the nucleus; however, both the nuclear and cytoplasmic fractions were normalized to GAPDH, which may not accurately reflect the nuclear:cytoplasmic ratio of BHLF1.

**BHRF1 mRNA and miRNA expression**

In Wp-restricted BL cell lines, BHRF1 expression is increased and contributes to the apoptosis resistance of these cells (209). In our ΔB-S rEBV infections, which have a deletion that extends to the right boundary of the Sal deletion, BHRF1 mRNA levels are increased. This increase is possibly due to the removal of alternative splice donor sites.
within the deleted region. However, the ΔBHLF1 infections express similar levels of
* BHFRF1* mRNA as WT infections. Deletion of * mir-BHFRF1-1* is not expected to significantly
affect the transformation efficiency of the mutant virus but may increase *BHFRF1* mRNA
levels (179). Our *BHFRF1* mRNA RT-qPCR results are consistent with this.

### 5.2 Potential mechanisms by which *BHLF1* may contribute to EBV latency

Although it is clear that the loss of *BHLF1* has an effect on latency, we were not
able in these experiments (due to time constraints and other limitations, discussed
below) to establish a mechanism by which it may be doing so. One of the main
questions that remains to be answered is what form of *BHLF1* mediates this effect: the
protein, RNA, or genomic locus itself. Based on our observations, which suggest that
any BHLF1 protein would only be expressed during the lytic cycle, we feel it is unlikely
that a functional protein is produced during latent infection and therefore would not be
responsible for the phenotype that we have observed. Additionally, we and others have
found that the BHLF1 ORF and/or amino acid sequence is not conserved evolutionarily
(8). Thus, we hypothesize that *BHLF1* may be functioning at least in part as an IncRNA
during latency. Our results strongly support a role for *BHLF1* as an IncRNA during
latency; however, we do not yet have definitive evidence showing that it is indeed the
RNA and not the locus *per se*.

**BHLF1 as an IncRNA**

Having ruled out the BHLF1 protein as being responsible for the effect that we
see during latent infection, we hypothesize that *BHLF1* may be functioning as an IncRNA
for several reasons, which have already been expanded upon elsewhere: 1) the gene is transcribed at high levels during latency; 2) the BHLF1 RNA appears to localize to the nucleus (many IncRNAs are nuclear); 3) lack of evolutionary conservation of the BHLF1 ORF (characteristic of IncRNAs); and 4) the only known function of BHLF1 is as an IncRNA during the lytic cycle (15).

As IncRNAs often function via RNA-protein interactions, the identification of any proteins that may be associating with the BHLF1 RNA would provide important and much-needed clues as to the function of BHLF1. The approach we chose to use involved the S1 aptamer, which is a structured RNA motif/tag that is bound by streptavidin with high affinity and can be eluted from streptavidin beads with biotin (263, 264). We planned to insert the S1 aptamer within the BHLF1 RNA and then either engineer the S1-tagged sequence into WT virus and use it to infect cells, or clone it into an expression vector to create cell lines. In both cases, the associated proteins could be pulled down with streptavidin beads and subsequently identified by mass spectrometry. Importantly, the pulldown assays would have been performed with and without cross-linking. Performing the pulldown assays in combination with cross-linking would have allowed us to capture transient protein interactions and increase the efficiency of protein recovery. Thus far, we have generated three rEBV DNA clones that were confirmed to contain the S1 aptamer sequence in the desired location within BHLF1 (described in Appendix A). However, before proceeding, we would like to further confirm that the S1 aptamer sequence was not inserted elsewhere in the viral genome, e.g., in LF3.
Potential RNA-binding protein partners of the BHLF1 RNA

As mentioned above, potential functions can be assigned to the BHLF1 RNA based on the proteins that bind to it. Unfortunately, as we have yet to move forward with the pulldown assays due to the prolonged difficulty in generating the S1-aptamer-containing rEBV, we do not yet know what, if any, proteins bind to the BHLF1 RNA. However, the BHLF1 RNA appears to localize to the nucleus which, along with the aforementioned lines of evidence, suggests that BHLF1 contributes to EBV latency as a nuclear lncRNA.

Nuclear lncRNAs fulfil their functions via several different mechanisms, one of which is the regulation of splicing. Furthermore, preliminary data from a previous post-doctoral scholar in our lab, David Hughes, suggest that the deletion of the BHLF1 locus increases the frequency of intron retention in the 5’ end of Cp-initiated EBNA mRNAs. Specifically, sequencing of the 5’ ends of EBNA cDNAs from BL2 cells infected with WT or ΔB-S rEBV revealed that at early time post-infection, more of the ΔB-S infections had alternative splice structures which retain an 81-bp intron between the W1/W01 and W2 exons, and at late time post-infection, none of the ΔB-S infections had the normal/expected exon structure (Figure 5.1). Although he also observed alternative splice structures in the WT infections, these were the only structures seen in the ΔB-S infections. Interestingly, one of the alternative splice structures (Figure 5.1, indicated by red arrow) is predicted to activate nonsense-mediated decay as it results in a premature stop codon just after the exon-intron junction. These alternative splice structures correlate with the loss of protein that we observe in the ΔB-S infections, but may or may not be responsible for it. Therefore, these results suggest that BHLF1 ensures proper splicing within the 5’ termini of EBNA mRNAs.
Figure 5.1 Deletion of the *BHLF1* locus increases the frequency of intron retention in the 5′ end of Cp-initiated *EBNA* mRNAs.

Splice structure of the 5′ termini of Cp-initiated *EBNA* transcripts in BL2 cells infected with WT or ΔB-S rEBV as determined by cDNA sequence analysis. Normal or expected structures are shown above the dashed line, while alternative splice structures are shown below the dashed line. The red arrow indicates an alternative splice structure that is predicted to activate nonsense-mediated decay. ND, not detected.
Based on these preliminary results, we would expect to recover splice regulatory factors using our pulldown approach with the S1 aptamer. There are two major types of cis-acting RNA sequence elements known as exon and intron splicing enhancers (ESEs and ISEs, respectively) and silencers (ESSs and ISSs, respectively). Splicing enhancers are sites to which splicing activator proteins, such as serine/arginine (SR) proteins, bind to promote the use of a splice site. On the other hand, splicing silencers are sites to which splicing repressor proteins, such as heterogeneous nuclear ribonucleoproteins (hnRNPs), bind to repress the use of a splice site. As deletion of BHLF1 results in increased frequency of intron retention, we might expect to recover hnRNPs that might normally act to repress the splice donor and/or acceptor on either side of the retained intron. The BHLF1 IncRNA may therefore be acting as a decoy or sponge for such hnRNPs. In the absence of BHLF1, these hnRNPs would be unavailable in sufficient amounts to bind to their target silencer sites, disrupting the proper splicing of the EBNA transcripts by increasing the frequency of intron retention.

Alternatively, we may instead recover SR proteins. If the BHLF1 IncRNA acts as a scaffold/guide rather than a decoy, it may be binding SR proteins to help direct them to splicing enhancer sites, especially suboptimal sites within primary EBV transcripts, to promote proper splicing. In the absence of BHLF1, therefore, critical SR proteins might not be available, thereby decreasing the frequency of splicing, leading to intron retention within the EBNA transcripts. MALAT1 is an example of a mammalian IncRNA which is involved in the regulation of alternative splicing. MALAT1 localizes to nuclear speckles, which are subnuclear compartments enriched in pre-mRNA splicing and processing factors, and has been shown to modulate the distribution and levels of active SR proteins (224). When MALAT1 is depleted, levels of dephosphorylated (i.e., inactive) SR
proteins increase and exhibit a more homogenous nuclear distribution, resulting in changes in alternative splicing of pre-mRNA. Perhaps *BHLF1* plays a similar role in EBV-infected cells by affecting the phosphorylation and distribution of active SR proteins such that the splicing of *EBNA* mRNAs is affected.

Nuclear lncRNAs are also involved in the regulation of epigenetic mechanisms. As discussed in the literature review (Chapter 2), epigenetic changes to the EBV genome play a crucial role in the establishment and maintenance of the different latency programs. Thus, it is possible that *BHLF1* acts as a guide to direct chromatin-modifying proteins to the viral genome. If so, we would expect to recover histone- and DNA-modifying proteins such as histone deacetylases (HDACs) or DNA methyltransferases in our pulldown experiments.

**BHLF1 as a control locus**

Although we hypothesize that *BHLF1* is acting as an lncRNA during latency, we have not ruled out a possible role for it as a control locus. The latency I and latency III programs exhibit different chromatin conformations wherein the origin of plasmid DNA replication, *oriP*, is in close proximity to Qp in latency I, and to Cp in latency III (265). The DNA-binding protein CCCTC-binding factor (CTCF) binds to several areas in the EBV genome, including the promoter regions upstream of both Cp and Qp (266). CTCF has since been implicated in the mediation of chromatin looping between *oriP* and Qp (265). Mutating the CTCF binding site at Qp resulted not only in a failure to form a loop between *oriP* and Qp, but also in the formation of an alternative loop between *oriP* and Cp, suggesting that the failure to form a latency I loop conformation resulted in the formation of a latency III conformation. CTCF depletion with siRNAs led to a complete
loss of loop formation at both Qp and Cp, suggesting that CTCF is essential for loop formation at both loci. However, work done by our lab suggests that CTCF contributes to but is not essential for the establishment of restricted latency, i.e., latency I (267). Therefore, factors other than CTCF and chromatin looping appear to also contribute to the differential regulation of latency programs.

As shown by our results, deletion of the BHLF1 locus leads to the transition from latency III to latency I. Furthermore, CTCF chromatin immunoprecipitation-sequencing (ChIP-seq) analysis of the EBV genome revealed a strong CTCF binding motif near BHLF1 (268). Thus, it is possible that BHLF1 contributes to chromatin looping in such a way that favors the latency III conformation, while in its absence, the chromatin ultimately adopts a latency I loop conformation.

5.3 Pitfalls and limitations

Although the work described here makes important first steps in characterizing the role of BHLF1 during latency, we were unable within the proposed timeframe to determine a mechanism of action for the phenotype that we observed. This was primarily due to the difficulty in manipulating the BHLF1 gene. As discussed in the literature review (Chapter 2), BHLF1 is a leftward-reading gene that encodes a 2.5-kb transcript containing twelve 125-bp repeats and is 82% GC-rich. This means that a large portion of the transcript is comprised of this repeat region which, as one may imagine, has significant secondary structure. Furthermore, the non-repeat region at the 5′ end of BHLF1 has almost perfect sequence homology with its viral paralog, LF3, leaving a roughly 500-bp unique stretch of sequence with which to work. These factors made
Manipulation of the gene particularly challenging and time-consuming, as extra confirmatory steps must be taken when inserting or deleting sequences to ensure that we do not inadvertently affect either LF3 or the secondary structure of BHLF1 itself, as well as when we are trying to detect BHLF1 to ensure that we are not instead detecting LF3.

Cloning of BHLF1 into the pLHCX vector

To confirm a role for the BHLF1 IncRNA in the maintenance of latency III, we wanted to rescue the mutant phenotype by expressing BHLF1 in trans in cells infected with the mutant rEBV. As described in Appendix B, we cloned the BHLF1 gene into a retroviral vector, pLHCX, for expression of the BHLF1 RNA in primary B cells and BL2 cells infected with the ΔBHLF1 rEBV. The two constructs span the BHLF1 locus from either the P1 (lytic) or P2 (latent) TSS up to the poly(A) signal sequence. Although seemingly straightforward, we encountered a number of obstacles related to the amplification and cloning of the inserts (further detailed in Appendix B and summarized in Table A.1), primarily due to the large repeat region and high GC content of the BHLF1 gene, as well as its high degree of homology to LF3. Since our mutant viruses, ΔB-S and ΔBHLF1, were engineered using a BAC containing the Akata EBV genome, we would have ideally liked to use the BHLF1 sequence from the same genome. However, we were ultimately unable to amplify the P1 and P2 inserts from Akata DNA, despite our best efforts. We instead amplified the inserts from pDK286, a plasmid containing the B95-8 BamHI-H fragment; however, we then encountered difficulties in cloning them into pLHCX. After trying several different approaches, we were finally able to obtain a single pLHCX clone containing the P1 insert but unfortunately, based on restriction digest, it
appeared that roughly four of the repeats were missing. As repair of the repeat region in the pLHCX-BHLF1-P1 clone would have been difficult (and since we were unable to obtain a positive P2 clone), we opted to instead modify the pSG5-BHLF1-FLAG vector, which contains the BHLF1 ORF FLAG-tagged at the 5’ end, to obtain our desired inserts. This entailed rebuilding the 5’ and 3’ ends of the current pSG5-BHLF1 insert to remove the FLAG tag and include the additional desired sequences outside of the ORF which we were able to amplify from Akata DNA. We were able to successfully append the additional sequences for both P1 and P2 and subsequently moved the rebuilt BHLF1 inserts from pSG5 into pLHCX. However, we found via sequencing that the appended 5’ end of the rebuilt P2 insert actually contained the Akata LF3 sequence rather than the BHLF1 sequence. Therefore, we re-amplified the 5’ end of P2 from pDK286 and used this to rebuild the 5’ end of the P2 insert. As a result, we had three final retroviral clones with which we produced retrovirus. However, the titers of the virus obtained were very low, and transduction of BL2 cells did not result in the expression of the BHLF1 RNA.

RNA-protein pulldowns

As mentioned in the previous section, the recovery of RNA-binding protein partners of the BHLF1 RNA would provide important clues about the role that BHLF1 may be playing during latency. We originally planned to use two complementary approaches to identify any proteins that may be interacting with the BHLF1 RNA. The first approach involved synthesizing the BHLF1 RNA in vitro with biotin-labeled uridine triphosphate (UTP); the biotinylated RNA would then have been incubated with cell extracts and any associated proteins pulled down with streptavidin beads and identified using mass spectrometry. However, the random incorporation of biotin into the BHLF1
RNA may disrupt the secondary structure of the RNA, which is presumably essential to its function as an IncRNA. Therefore, we opted not to use this approach and instead focused on using a different approach for the identification of RNA-binding proteins.

The second approach involved inserting the S1 aptamer within the $BHLF1$ RNA, and then either engineering the S1-tagged sequence into WT virus and using it to infect cells, or cloning it into an expression vector to create cell lines. The associated proteins could then be pulled down, with or without cross-linking, with streptavidin beads and identified using mass spectrometry. Performing the pulldown assays in combination with cross-linking would have allowed us to capture transient protein interactions and increase the efficiency of protein recovery. As the maintenance of secondary structure is important for both the aptamer and (most likely) the target RNA, we used RNA structure prediction software to identify an unstructured region of $BHLF1$ to insert the S1 aptamer plus a flexible linker region reported to stabilize the aptamer structure. The location that we chose is in the 5′ untranslated region of $BHLF1$ and in silico allows for the correct folding of both the aptamer and the $BHLF1$ RNA. To engineer the S1-tagged sequence into WT virus, we used a recombineering method that utilizes a galK cassette for positive and negative selection. Unfortunately, due to the perfect homology of the chosen region of $BHLF1$ with $LF3$, initial attempts to insert the galK cassette (the first step in the recombineering procedure) resulted in the insertion of the cassette into $LF3$ rather than $BHLF1$. To remedy this, we extended the left homology arm into the unique region of $BHLF1$, thereby theoretically targeting the galK cassette (and subsequently, the sequence for the S1 aptamer and linker region) exclusively to $BHLF1$. Thus far, we have completed the recombineering itself and generated three clones that were confirmed via sequencing to contain the S1 aptamer sequence in the desired location within $BHLF1$. 
Subcellular localization of BHLF1

As mentioned above, our RT-qPCR results suggest that the BHLF1 RNA may be localizing to the nucleus; however, it is important to note that BHLF1 transcripts can initiate from either lytic or latent promoters: P1 (lytic), or P2 and P3 (latent) (7). The BHLF1 probe and primers that were used in our studies would detect all BHLF1 transcripts, not just those initiating from the latency-specific promoters. At any given time, a small percentage of EBV-positive cells in latently infected cell lines may be undergoing spontaneous reactivation of the lytic cycle; therefore, it is possible that some of the BHLF1 RNA that we detect in latently infected cell lines may be initiating from P1, the lytic promoter. However, there is currently no data published on the ratio of expression of these transcripts in infected cell lines, or whether all are expressed during latency. In an attempt to address the contribution of lytic transcripts to the overall expression levels of BHLF1, we also examined the expression levels of the lytic-cycle gene SM as an indicator of lytic reactivation. However, there did not appear to be a consistent direct correlation between the levels of SM and BHLF1 in the cell lines we tested. Future RT-qPCR experiments being conducted in the lab will utilize different primers and probes to distinguish between BHLF1 transcripts initiating from P1, P2, and P3.

5.4 Future directions

Rescue of mutant phenotype

Thus far, our data suggests that BHLF1 is needed in BL2 cells for the maintenance of latency III and may contribute to the efficiency of the virus to immortalize
primary B cells. To further confirm this, we would like to rescue the mutant phenotype in both BL2 cells and primary B cells by expressing *BHLF1 in trans* in cells infected with the mutant rEBV. For the rescue of the mutant phenotype, we cloned the *BHLF1* gene into a retroviral vector, pLHCX (described above). Although we were able to produce virus, the titers were very low and did not result in *BHLF1* expression when used to infect BL2 cells. Thus, our next step would be to transfect the cells with the retrovirus DNA constructs themselves using Amaza nucleofection. If we are able to get expression of the *BHLF1* RNA via nucleofection, our next step would be to infect these cells with our mutant rEBVs to determine if we are able to rescue the mutant phenotype, in which case the *BHLF1*-mutant virus infections would maintain latency III long-term rather than transitioning to latency I. In primary B cells, we would expect an increase in the efficiency of immortalization.

*RNA-protein pulldowns using the S1 aptamer*

As described above, we have completed the recombineering to insert the S1 aptamer and flexible linker region within *BHLF1* and generated three clones that were confirmed via sequencing to contain the S1 aptamer sequence in the desired location within *BHLF1*. *BamHI* digest of the clones also suggested that the S1 aptamer was indeed inserted within *BHLF1*. Ideally, we would like to confirm that the S1 aptamer sequence was not also inserted elsewhere in the genome, e.g., *LF3*. To do this, we would perform a Southern blot on the *BamHI*-digested clones to determine if an appropriate probe hybridizes to the fragment containing *BHLF1* but not the one containing *LF3*, indicating that the S1 aptamer sequence was inserted correctly. Once we have verified correct insertion of the S1 aptamer sequence within only *BHLF1*, we
can transfect HEK293 cells to make a producer cell line, which can then be induced to produce virus. This virus can then be used to infect cells to determine if the function of the BHLF1 RNA during latency has been disrupted due to the location of aptamer insertion.

Once we have recovered and identified proteins via this method, we would choose potentially relevant proteins to continue studying and perform reciprocal pulldown assays to see if we can recover the BHLF1 RNA. This would confirm that the chosen protein(s) and the BHLF1 RNA do indeed interact, rather than the protein binding to the S1 aptamer. Thereafter, we would conduct knockdown experiments in which the identified protein is depleted to determine if there is any effect on viral latency similar to what we observe with our mutant viruses. Depending on the types of proteins that are recovered, we would pursue the respective avenue (e.g., splicing, chromatin modification) to further elucidate how BHLF1 contributes to viral latency.

Tracking of the BHLF1 RNA to confirm subcellular localization

Another approach that our lab is working on is the insertion of RNA aptamers (Spinach2 and Baby Spinach) at the 5' and 3' ends of the BHLF1 P1 and P2 inserts to enable live-cell imaging of the BHLF1 RNA. These aptamers when bound by the compound DFHBI-1T are mimics of GFP. This approach should not only allow us to determine the subcellular localization of the BHLF1 RNA, but also has the added benefit of allowing us to distinguish between the lytic (P1) and latent (P2) transcripts. Thus far, other members of the lab have cloned the aptamers on a tRNA scaffold and generated the RNA by in vitro transcription, and demonstrated that both aptamers fluoresce when
bound to DFHBI-1T \textit{in vitro}. Both are being cloned into the aforementioned P1 and P2 retroviral constructs.

An alternative approach to identify the subcellular localization of \textit{BHLF1} is RNA fluorescence \textit{in situ} hybridization (FISH). An example of this is the ViewRNA ISH Cell Assay manufactured by Affymetrix that is capable of detecting single RNA transcripts. This assay utilizes a target-specific probe set that typically contains approximately 20 probes that hybridize to the target RNA. Signal amplification is achieved via the use of amplification molecules which bind to the hybridized probes. However, we would have to design the probes such that they target the unique region of \textit{BHLF1} (to avoid detecting \textit{LF3}), which means that we would be unable to distinguish between the P1 and P2 transcripts. Although not an ideal approach, this RNA FISH assay would be an acceptable alternative should we not be able to get the Spinach2 and Baby Spinach aptamers to work.

\textit{Chromatin conformation capture assays to investigate DNA looping}

To investigate the possibility that the \textit{BHLF1} locus contributes to chromatin looping, we would utilize a chromatin conformation capture (3C) approach. The 3C assay involves cross-linking DNA followed by restriction digest and ligation, and the reversal of cross-linking. The resulting ligated products can then be amplified by PCR and analyzed by agarose gel detection and/or RT-qPCR. This approach would allow us to determine if the \textit{BHLF1} locus is in close proximity to either the Qp or Cp CTCF binding sites during the different latency programs.

Additionally, we can also employ the use of circularized chromosome conformation capture (4C), which would allow us to detect the interaction between the
BHLF1 locus and unknown DNA regions. The procedure is similar to 3C; however, after the cross-link reversal, the fragments are digested with a second restriction enzyme to create sticky ends and allow for self-circularization of the fragments which are then amplified by inverse PCR. The resulting library can then be analyzed via DNA sequencing or microarray. This would allow us to determine if the BHLF1 locus interacts with other regions of the EBV genome.

5.5 Concluding remarks

Although BHLF1 is classified as a lytic-cycle gene, numerous lines of evidence have suggested that it plays a role during EBV latency. Deletion of the BHLF1 locus resulted in a transition from latency III to I upon infection of EBV-negative BL2 cells with BHLF1-mutant rEBVs, suggesting that BHLF1 is critical for the maintenance of EBV latency in these cells. We also observed a decrease in the efficiency of immortalization upon infection of primary B cells with the BHLF1-mutant rEBVs. Furthermore, expression of BHLF1 protein may be restricted to the lytic cycle as it appears to require the presence of SM, another lytic-cycle protein, for efficient expression. Therefore, we hypothesize that BHLF1 is acting as an IncRNA during latent infection. It is important, however, to note that the results presented here do not rule out a role for the BHLF1 genomic locus itself. Further studies will aim not only to determine whether it is the RNA or the genomic locus (or both) that contributes to EBV latency, but also the mechanism by which it does so.
Bibliography


Definition of the Sequence Requirements for Binding of the EBNA-1 Protein to Its Palindromic Target Sites in Epstein-Barr Virus DNA. J Virol 64:2369–2379.


Carcinoma Nuclear Proteome Induced by the EBNA1 Protein of Epstein-Barr Virus Reveal Potential Roles for EBNA1 in Metastasis and Oxidative Stress Responses. J Virol 86:382–94.


RBP-J kappa interact with an Epstein-Barr virus nuclear antigen 2 responsive cis-element. EMBO J 13:5624–32.


79. **Peng R, Tan J, Ling PD.** 2000. Conserved Regions in the Epstein-Barr Virus Leader Protein Define Distinct Domains Required for Nuclear Localization and


but both it and activated CD40 can prolong their survival. Embo J 15:7070–7078.


160. Metzenberg S. 1989. Relative Rates of RNA Synthesis across the Genome of


168. **Ruf IK, Lackey KA, Warudkar S, Sample JT.** 2005. Protection from Interferon-Induced Apoptosis by Epstein-Barr Virus Small RNAs Is Not Mediated by


208. Jones MD, Foster L, Sheedy T, Griffin BE. 1984. The EB virus genome in Daudi Burkitt’s lymphoma cells has a deletion similar to that observed in a non-
transforming strain (P3HR-1) of the virus. EMBO J 3:813–21.


255. Bidet K, Dadlani D, Garcia-Blanco MA. 2014. G3BP1, G3BP2 and CAPRIN1 Are Required for Translation of Interferon Stimulated mRNAs and Are Targeted by


Appendix A: Recombineering of S1 aptamer into \textit{BHLF1}

Based on our observations, it is unlikely that the BHLF1 protein is responsible for the effect that we see during latency. Since the \textit{BHLF1} gene is transcribed at high levels during latency, it may be functioning as an IncRNA. As IncRNAs often function via RNA-protein interactions, we wanted to identify any proteins that may be associating with the \textit{BHLF1} RNA. For this, we planned to utilize the S1 aptamer, which is a structured RNA motif/tag that is bound by streptavidin with high affinity and can be eluted from streptavidin beads with biotin. The S1 aptamer was to be inserted into the \textit{BHLF1} RNA and the S1-tagged sequence either engineered into WT virus and used to infect cells or cloned into an expression vector to create cell lines. The associated proteins could then be pulled down with streptavidin beads and identified using mass spectrometry.

As the maintenance of secondary structure is important for both the aptamer and the target RNA, we used RNA structure prediction software to determine a suitable location at which to place the aptamer such that: 1) it assumes the correct structure; and 2) does not disrupt the secondary structure of the \textit{BHLF1} target RNA. One of the best spots for aptamer insertion is at the end of a pre-existing stem loop (Figure A.1A). It also appears that for the aptamer to fold correctly, we needed to include a flexible linker region to stabilize it (Figure A.1B). Using the structure prediction software, the aptamer plus the linker region was predicted to fold correctly, while also maintaining the secondary structure of the \textit{BHLF1} 5' untranslated region within which it was being inserted (Figure A.1C). We then wanted to insert the sequence for the aptamer and linker region into the \textit{BHLF1} gene by recombineering. However, due to the perfect
homology of this region with $LF_3$, initial attempts to insert the galK cassette (the first step in the recombineering procedure) resulted in the insertion of the cassette into $LF_3$ rather than $BHLF_1$. To remedy this, we extended the left homology arm into the unique region of $BHLF_1$ (Figure 3.1), thereby theoretically targeting the galK cassette (and subsequently, the sequence for the aptamer and linker region) exclusively to $BHLF_1$. Thus far, the recombineering has been completed and three clones were generated that were confirmed via sequencing to contain the S1 aptamer sequence in the desired location within $BHLF_1$. *Bam*HI digest of these clones also suggest that the S1 aptamer was inserted within $BHLF_1$. However, we would like to further confirm that the S1 aptamer sequence was not also inserted elsewhere in the genome, e.g. $LF_3$. 
Figure A.1 RNA structure prediction of a region of the *BHLF1* RNA and the S1 aptamer.
Figure A.1 RNA structure prediction of a region of the *BHLF1* RNA and the S1 aptamer.
Figure A.1 RNA structure prediction of a region of the *BHLF1* RNA and the S1 aptamer.
Figure A.1 RNA structure prediction of a region of the *BHLF1* RNA and the S1 aptamer.

(A) *BHLF1* sequence from P1, one of the major TSS. As one of the best spots for aptamer insertion is at the end of a pre-existing stem loop, we deleted the 9 nucleotides that make up the end of the indicated stem loop (red circle). (B) A flexible linker region was included with the S1 aptamer to stabilize it and ensure that it folds correctly. This additional sequence also included an *NdeI* restriction site. (C) The S1 aptamer and flexible linker region were inserted at the end of the desired stem loop. The aptamer sequence appears to fold correctly while maintaining the secondary structure of the surrounding *BHLF1* sequence.
Appendix B: Cloning of BHLF1 into the retroviral vector, pLHCX

For the rescue of the mutant phenotype, we wanted to clone the BHLF1 gene into a retroviral vector, pLHCX, for expression of the BHLF1 RNA in primary B cells and BL2 cells infected with the ΔBHLF1 rEBV. The two constructs span the BHLF1 locus from either the P1 (lytic) or P2 (latent) TSS up to the poly(A) signal sequence. Unfortunately, we encountered a number of obstacles related to the amplification and cloning of the inserts (detailed below and summarized in Table A.1), primarily due to the large repeat region and high GC content of the BHLF1 gene, and to its high degree of homology to its paralog, LF3.

Since our mutant viruses, ΔB-S and ΔBHLF1, were engineered using a BAC containing the Akata EBV genome, we would have ideally liked to use the BHLF1 sequence from the same genome. However, we were unable to amplify the P1 and P2 inserts from Akata DNA, even with a GC-rich-compatible DNA polymerase, and thus had to instead use a plasmid containing the B95-8 BamHI H fragment (pDK286). Using pDK286, we were able to amplify the inserts, but encountered difficulty in cloning them even when using varying insert:vector ratios for ligation. Therefore, we tried using a different bacterial strain that is supposed to be more efficient at transforming unstable inserts such as lentiviral DNA containing direct repeats. After screening upwards of 300 colonies from the transformations with these varying conditions, we obtained one clone that we confirmed, via sequencing, contained the BHLF1 P1 insert (we never obtained any P2 clones). However, based on restriction digest, it appeared that roughly four of the repeats were missing from this clone.
As repair of the repeat region in the pLHCX-BHLF1-P1 clone may be difficult (and since we were unable to obtain a positive P2 clone), we opted to modify the BHLF1 sequence in the pSG5-BHLF1-FLAG vector, which contains the BHLF1 ORF FLAG-tagged at the 5’ end, to obtain our desired inserts. This entailed rebuilding the 5’ and 3’ ends of the current pSG5-BHLF1 insert to include the additional sequences outside of the ORF, which we were able to amplify from Akata DNA, by digestion of pSG5-BHLF1 and ligation of the PCR-amplified additional sequences. We were able to successfully append the additional sequences for both P1 and P2 and subsequently moved the rebuilt BHLF1 inserts from pSG5 into the pLHCX vector. Although we obtained the correct pLHCX-BHLF1-P1 clone, we found via sequencing that the appended 5’ end of the rebuilt P2 insert actually contained the Akata LF3 sequence rather than BHLF1 sequence. Therefore, we re-amplified the 5’ end of P2 using pDK286 as the template and used this to rebuild the 5’ end of the P2 insert. As a result, we now have three final retroviral clones (pLHCX-BHLF1-P1, pLHCX-BHLF1-P2, and pLHCX-BHLF1-P2-B95-8) with which we produced retrovirus. Although the retroviral titers were fairly low, we are attempting to transduce BL2 cells to create stable lines and to verify that the BHLF1 RNA is indeed being expressed.
Table A.1 Summary of problems encountered during the cloning of BHLF1 into the retroviral vector, pLHCX, and solutions implemented to address them.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unable to amplify desired sequences from Akata DNA</td>
<td>• Amplified sequences from pDK286 which contains B95-8 BamHI H fragment.</td>
</tr>
<tr>
<td></td>
<td>• Used a GC-rich-compatible DNA polymerase.</td>
</tr>
<tr>
<td>Unable to clone P1 and P2 inserts into retroviral vector</td>
<td>• Used chemically competent cells which are recommended for cloning unstable inserts, e.g. DNA containing direct repeats.</td>
</tr>
<tr>
<td>Some of the P1 repeats were deleted upon cloning of insert into retroviral vector</td>
<td>• Rebuilt 5’ and 3’ ends of BHLF1 sequence in pSG5-BHLF1-FLAG, which originally contained only the FLAG-tagged BHLF1 ORF.</td>
</tr>
<tr>
<td>Unable to clone P2 into retroviral vector</td>
<td></td>
</tr>
<tr>
<td>P2 5’ end amplified from Akata LF3 instead of BHLF1</td>
<td>• Amplified P2 5’ end using pDK286.</td>
</tr>
</tbody>
</table>
VITA
Kristen Dominique Yetming

EDUCATION
The Pennsylvania State University College of Medicine 2009-2017
Hershey, PA
Ph.D., Microbiology and Immunology

The University of Michigan 2005-2009
Ann Arbor, MI
B.S., Microbiology and Asian Studies

HONORS/AWARDS
Pasteur Prize in Microbiology and Immunology September 2013
The Pennsylvania State University College of Medicine

PUBLICATIONS


PRESENTATIONS