Global transcriptome analysis reveals novel EBV encoded lncRNA structures: antisense, circular and extensive alternative spliced transcripts at the LMP2 locus

A Dissertation
Submitted On The 27th Day Of March 2015
To The Graduate Program In Biomedical Sciences
In Partial Fulfillment Of The Requirements
Of The School Of Medicine
Of Tulane University
For The Degree
Of
Doctor Of Philosophy Of Science
By

____________________
Monica Concha

Approved:
____________________
Erik Flemington, Ph.D.
Director

____________________
Matthew Burow, Ph.D.

____________________
Astrid Engel, Ph.D.

____________________
Cindy Morris, Ph.D.

____________________
Haitao Zhang, Ph.D.
Global transcriptome analysis reveals novel EBV encoded lncRNA structures: antisense, circular and extensive alternative spliced transcripts at the LMP2 locus

A Dissertation
Submitted On The 27th Day Of March 2015
To The Graduate Program In Biomedical Sciences
In Partial Fulfillment Of The Requirements
Of The School Of Medicine
Of Tulane University
For The Degree
Of
Doctor Of Philosophy Of Science
By

________________________________________
Monica Concha

Approved: ____________________________________________
Erik Flemington, Ph.D.
Director

______________________________________________
Matthew Burow, Ph.D.

______________________________________________
Astrid Engel, Ph.D.

______________________________________________
Cindy Morris, Ph.D.

______________________________________________
Haitao Zhang, Ph.D.
This dissertation is dedicated to my mother Monica Concha-Seminario, my father Ernesto Concha and my beloved dog Dexter Morgan.
ABSTRACT

Using an enhanced RNA-Seq pipeline to analyze the Epstein-Barr virus (EBV) transcriptome, we investigated viral and cellular gene expression in the Akata cell line following B-cell-receptor-mediated reactivation. Robust induction of EBV gene expression was observed, with most viral genes induced more than 200-fold and with EBV transcripts accounting for 7% of all mapped reads within the cell. After induction, a novel antisense transcript, a circular transcript, and multiple of candidate viral splicing events were detected in the LMP2 locus using splice junction aligner, TopHat. A more detailed analysis of lytic LMP2 transcripts and LMP2A transcripts expressed during type III latency revealed alternative splicing events and novel transcripts during reactivation. Characterization of novel antisense and circular RNA transcripts has lead to identifying them as virus long non-coding RNA (vlncRNA) transcripts. Pathway analysis of viral and cellular gene expression changes following knockdown of the antisense LMP2 transcript has revealed that the antisense LMP2 transcript increases the magnitude of EBV lytic activation following BCR reactivation.
ACKNOWLEDGMENTS

I would like to thank each member on my committee for your guidance in the in the process of developing my thesis. Thank you to my committee chair, Dr. Erik Flemington for the opportunity to work in his laboratory. I will always appreciate the patience and encouragement. I would also like to thank past and current members in the Flemington lab: Wang, Yi, Tom, Subing, Tina, Mike, Claire, Melody, Zhen, Hani, Jennifer, and Adrianne. Thank you for your friendships and patience.

I would also like to thank to my aunts Kari, Perla, and Anita, and my uncles Jorge, Juan, Pacho, and Luis for the love throughout my life. A special thanks to my cousin Roberto, who I regard to as a brother, for his love and support. To my beloved dog, Dexter, who kept me company for hours of writing, listened to my presentations, and kept me smiling throughout my journey.

Finally, I would like to thank my parents Monica and Ernesto Concha for never setting limits to my dreams and supporting me every step of my life. Everything I do in life is to make you proud.
# Table Of Contents

ACKNOWLEDGEMENTS ........................................................................ii

LIST OF FIGURES ..............................................................................iv

INTRODUCTION ......................................................................................1

1. The latent and lytic stages of the Epstein-Barr virus .........................1

2. Next-Generation Sequencing ..........................................................2

CHAPTER 1. Identification of new viral genes and transcript isoforms during Epstein-Barr virus reactivation using RNA-Seq .................................................................4

1. Abstract .........................................................................................4

2. Introduction ...................................................................................5

3. Materials and Methods .................................................................6

4. Results ..........................................................................................9

5. Discussion ....................................................................................15

CHAPTER 2. Identification and Characterization of two novel Epstein-Barr virus long non-coding RNA in the LMP2 locus ......................................................32

1. Abstract .......................................................................................32

2. Introduction ..................................................................................32

3. Materials and Methods ...............................................................37

4. Results .........................................................................................47

5. Discussion ...................................................................................53

LIST OF REFERENCES ........................................................................73
LIST OF FIGURES

Figure 1.1  Exonic, intronic, and intergenic read count following BCR activation in Akata cell………………………………………………………………………………..21

Figure 1.2  Fold induction of EBV genes in cytoplasmic Akata fractions………………..23

Figure 1.3  BZLF1 and BLLF1 read coverage from uninduced and induced Akata cells………………………………………………………………………………..24

Figure 1.4  LMP1 read coverage from uninduced and induced Akata cells.................26

Figure 1.5  Read coverage and splicing across LMP2A……………………………………..27

Figure 1.6  Nuclear retention of latency genes following B-cell receptor activation in Akata cells……………………………………………………………………..28

Figure 1.7  Coverage across novel transcripts BCLT1 and BCRT2………………..29

Figure 1.8  Quantitative RT-PCR validation of selected cellular gene changes following BCR activation………………………………………………………….30

Figure 1.9  Higher nuclear retention of BCRT2 and BCLT1 than the average of lytic genes following B-cell receptor activation in Akata cells………………..31

Figure 1.10 Illustration of EBV genome coverage gap where the genome has been split for subsequent EBV RNA-Seq analysis……………………………………..31

Figure 2.1  Strand-specific coverage of LMP2 locus in JY cell-line and Akata cells 24-hours post-induction…………………………………………………………..55

Figure 2.2  Time-course strand-specific read coverage across the LMP2 locus following B-cell receptor activation in Akata cells…………………………………56

Figure 2.3  PCR evidence of alternative splicing between LMP2 exon1-exon6 and LMP2 exon1-exon7 24-hours post induction in Akata cells…………………..57
Figure 2.4  Cytoplasmic to nuclear ratio of LMP2 transcripts..........................57
Figure 2.5  Coding Potential Calculator assess the protein-coding potential of the
LMP2 transcripts.................................................................58
Figure 2.6  Illustration of LMP2 1-6 and 1-7 truncated protein.........................59
Figure 2.7  Quantitative RT-PCR of LMP2 transcripts 24-hours post-BCR activation
during viral replication inhibition.............................................59
Figure 2.8  Immunofluorescence of transient transfection of FLAG-tagged LMP2A
cDNA, LMP2A 1-6, and LMP2A 1-7 vector constructs in HeLa cells…60
Figure 2.9  Strand specific coverage of the aLMP2 lncRNA 24-hours post-induction
in Akata cells.................................................................61
Figure 2.10  Time-course qRT-PCR of aLMP2, cLMP2, and selected immediate early,
early, late, and latent EBV genes............................................62
Figure 2.11  Localization of the aLMP2 transcript............................................62
Figure 2.12  Fold change of aLMP2 expression during viral replication inhibition…..63
Figure 2.13  Antisense LMP2 GapmeR-mediated knockdown: EBV genes.............64
Figure 2.14  Antisense LMP2 GapmeR-mediated knockdown: LMP2 locus............65
Figure 2.15  Antisense LMP2 GapmeR-mediated knockdown: Cellular genes........66
Figure 2.16  Discovery of circular LMP2 in induced Akata cells.........................67
Figure 2.17  Quantitative RT-PCR of cLMP2 expression across multiple EBV-positive
cell lines.................................................................68
Figure 2.18  Illustration of cLMP2 and cLMP2 SA mutant vector constructs………69

Figure 2.19  DG75 retroviral stable cell line expressing control and cLMP2………..70

Figure 2.20  Nucleofection in Mutu I of control, cLMP2, and cLMP2 SA mutant vectors……………………………………………………………………71
INTRODUCTION

1. The latent and lytic stages of the Epstein-Barr virus

Epstein-Barr virus (EBV) is a member of the gammaherpesvirus family that is carried by more than 90% of the human population (13). EBV initially infects resting B-lymphocytes and persists for the life of the host in a primarily latent state. In latency, EBV has oncogenic potential as evidenced, in part, through its association with numerous types of cancers including Burkitt's lymphoma (BL), immunoblastic lymphoma, nasopharyngeal carcinoma, and Hodgkin's disease. EBV uses different programs of gene expression to establish latency in B-cells that variably include the 6 nuclear antigens (EBNAs) and the latent membrane proteins, LMP1 and LMP2. The EBNA (-1, -2, -3a, -3b, -3c, -LP) proteins have various nuclear functions including the regulation of gene expression. In latency I, the program displayed in Burkitt’s lymphoma, the only viral protein expressed is EBNA1 (44,45,46). EBNA1 ensures replication of the EBV genome during cell division, therefore, EBNA-1 is expressed in all programs of latency. Epstein-Barr virus encoded RNAs (EBERs) and the BamHI A rightward transcripts (BARTs) are also expressed. Latency II is the program associated with nasopharyngeal carcinoma and Hodgkin’s lymphoma where expression of LMP1 and LMP2 along with EBNA1, EBERs, and BARTs is observed (47,48). The LMP1 and LMP2A proteins mimic the cellular CD40 and B-cell receptors, respectively, to help facilitate the viral infection cascade. Lastly, the Latency III program expresses all the latent genes, which include EBNA (-1, -2, -3a, -3b, -3c, -LP), LMP1, LMP2, EBERs, and BARTs.
Stimulation of the B-cell receptor (BCR) by an antigen in EBV infected B-cells induces EBV lytic replication. In an effort to maintain latency, the EBV LMP2A protein can prevent BCR activation (1). Normally, BCR signaling stimulates expression of the two EBV IE (immediate early) genes, BZLF1 and BRLF1. Expression of BZLF1 and BRLF1 begins the cascade of EBV replication followed by viral packaging and release. BZLF1 and BRLF1 initiate expression of the early EBV genes which are involved in viral DNA replication (5,6,7). The six major EBV genes involved in viral replication are BALF5 (DNA polymerase), BMRF1 (DNA polymerase processivity factor), BBLF4 (helicase), BALF2 (single-stranded DNA binding protein), BSLF1 (primase), and BBLF2/3 (primase-associated protein) (41). Viral replication requires the binding of BZLF1 and BRLF1 to oriLyt, which allows the formation of a triple helix DNA structure (42,43). EBV replication utilizes a rolling-circle mechanism that begins at the lytic origin of replication (oriLyt). At the end of replication the viral genome is cut at the terminal repeat region. The EBV late genes are expressed once replication has been initiated and encode proteins involved in viral packaging. Capsids are assembled and the mature virion exits the nucleus through budding. The virions are processed in the trans-Golgi pathway and released through exocytosis.

2. Next-Generation Sequencing

Next-generation sequencing (NGS) has a number of applications including DNA-sequencing, chromatin immunoprecipitation (ChIP)-sequencing, and RNA-sequencing (transcriptome profiling). RNA-Seq technology allows the simultaneous interrogation of gene expression and transcript structure at a high level of accuracy and at a single-
nucleotide resolution. The objective is to identify mRNAs, non-coding RNAs, small RNAs, 5’ and 3’ ends, splicing events, and isoform patterns in response to biological stimuli. In principle, the RNA can be poly(A)-selected or ribo-depleted (used to detect some noncoding-RNAs) before fragmentation and cDNA library generation. For Illumina sequencing, the cDNA library is constructed by ligating adaptors on both ends of the RNA. These adaptors can contain uniquely tagged sequences to identify 5’ and 3’ ends for directional sequencing. A sample-specific index may also be ligated to each fragment to allow multiple samples to be sequenced together in a single sequencing lane (multiplexing). Once the cDNA library is complete, the fragments can be selected by size and amplified by PCR. The cDNA fragments are delivered onto a flow cell where their adapters anneal to their complementary partners in the flowcell. The annealed fragment is then subjected to bridge amplification by DNA replication followed by denaturation. This three-step process (hybridization, replication, denaturation) is repeated to amplify several million clusters in each flow cell. Next, a fluorescently-labeled nucleotide is added to the flow cell a nucleotide at a time. After the fluorescently-labeled nucleotide is added the image is recorded. The fluorescently-labeled nucleotide emits a specific color indicating each nucleotide (A, T, G, C). The process is repeated until the entire fragment is sequenced.
CHAPTER 1. Identification of new viral genes and transcript isoforms during Epstein-Barr virus reactivation using RNA-Seq

1.1 Abstract

Using an enhanced RNA-Seq pipeline to analyze Epstein-Barr virus (EBV) transcriptomes, we investigated viral and cellular gene expression in the Akata cell line following B-cell-receptor-mediated reactivation. Robust induction of EBV gene expression was observed, with most viral genes induced >200-fold and with EBV transcripts accounting for 7% of all mapped reads within the cell. After induction, hundreds of candidate splicing events were detected using the junction mapper TopHat, including a novel nonproductive splicing event at the gp350/gp220 locus and several alternative splicing events at the LMP2 locus. A more detailed analysis of lytic LMP2 transcripts showed an overall lack of the prototypical type III latency splicing events. Analysis of nuclear versus cytoplasmic RNA-Seq data showed that the lytic forms of LMP2, EBNA-2, EBNA-LP, and EBNA-3A, -3B, and -3C have higher nuclear-to-cytoplasmic accumulation ratios than most lytic genes, including classic late genes. These data raise the possibility that at least some lytic transcripts derived from these latency gene loci may have unique, noncoding nuclear functions during reactivation. Our analysis also identified two previously unknown genes, BCLT1 and BCRT2, that map to the BamHI C-region of the EBV genome. Pathway analysis of cellular gene expression changes following B-cell receptor activation identified an inflammatory response as the top predicted function and ILK and TREM1 as the top predicted canonical pathways.
1.2 Introduction

Epstein-Barr virus (EBV) is a human pathogen that causes malignancies including Burkitt's lymphoma, Hodgkin's disease, and nasopharyngeal carcinoma (13). EBV has a complex infection cycle involving a number of different viral gene expression programs. These individual programs facilitate distinct tasks that are required for specific infection stages. Like all herpesviruses, EBV utilizes both latent gene expression programs, in which only limited numbers of viral genes are expressed, and a replicative gene expression program, in which the bulk of EBV genes are expressed to produce infectious virus.

Efficient and synchronous virus reactivation can be modeled by activating the B-cell receptor (BCR) in the EBV-positive Burkitt's lymphoma cell line Akata (15). In this system, reactivation leads to an ordered induction of immediate-early (e.g., BZLF1 and BRLF1), early (e.g., BMRF1), and late genes, with immediate-early genes peaking at approximately 2 to 6 h and late genes peaking at approximately 6 to 24 h postinduction (6, 15, 20). Interestingly, Yuan et al. (20) found that EBV latency genes were induced following reactivation, suggesting a role for these genes in the lytic cycle (20).

Second-generation RNA-Seq technology allows the simultaneous interrogation of gene expression and transcript structure at a high level of accuracy and at a single-nucleotide resolution. We have recently shown the application of RNA-Seq to the interrogation of EBV transcriptomes in two type I latency cell lines, Akata and Mutu I (11). Here we have improved our RNA-Seq pipeline and have applied it to the analysis of the EBV transcriptome during viral reactivation in the synchronous Akata BCR-mediated reactivation system.
1.3 Materials and Methods

Cell culture. All cells were grown in RPMI 1640 (Thermo Scientific, catalog no. SH30027) plus 10% fetal bovine serum (FBS; Invitrogen-Gibco, catalog no. 16000-069) with 0.5% penicillin and streptomycin (pen/strep; Invitrogen-Gibco, catalog no. 15070). Cells were grown at 37°C in a humidified, 5% CO2 incubator.

Lytic cycle induction. Akata and Mutu I cells were grown to near saturation, at which time an equal volume of fresh RPMI (plus 10% FBS and 0.5% pen/strep) was added. The following day, cells were spun down and resuspended in an equal volume of freshly warmed RPMI (plus 10% FBS and 0.5% pen/strep) with or without 10 µg of anti-IgG (Akata) or anti-IgM (Mutu I)/ml. The cells were harvested 24 h after treatment and subjected to RNA extraction or protein isolation (for Western blot analysis).

RNA extraction. Total RNA was prepared using an RNeasy minikit (Qiagen, catalog no. 74104) according to the vendor's protocol. Cytoplasmic and nuclear RNA were extracted with a Norgen Biotek cytoplasmic and nuclear RNA purification kit (catalog no. 21000) according to the vendor's protocol.

Western blot analysis. After a single 1× phosphate-buffered saline (PBS) wash, the cells were immediately suspended in five pellet volumes of sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer (125 mM Tris [pH 6.80], 10% glycerol, 2% SDS, 5% 2-mercaptoethanol, 0.05% bromophenol blue) and boiled for 20 min to shear the genomic DNA. Protein concentrations were measured using a Bio-Rad protein assay kit (catalog no. 500-0006) according to the manufacturer's instructions. Equal weights of cell lysates were subjected to SDS-PAGE electrophoresis and
transferred to nitrocellulose membranes. The blots were blocked for 30 min in a blocking buffer (0.05 M Tris, 0.138 M NaCl, 0.0027 M KCl pH 8.0, and 0.1% Tween 20 at pH 8.0) containing 5% Bio-Rad blotting grade blocker nonfat dry milk (catalog no. 170-6404) and 1% FBS and then incubated with the primary antibody (in blocking buffer) overnight at 4°C. The blots were washed four times with PBS, incubated with the secondary antibody for 1 h, washed four more times with PBS, and then analyzed using a Li-Cor Odyssey infrared imaging system. Primary antibodies used were mouse anti-Zta (Argene, catalog no. 11-007), mouse anti-Rta (Angene, catalog no. 11-008), mouse anti-BMRF1 (Capricorn, catalog no. EBV-018-48180), and goat anti-Actin (Santa Cruz Biotechnology, Inc., catalog no. sc-1615). The secondary antibodies used were goat anti-mouse (Li-Cor/Odyssey, catalog no. 926-32220) and donkey anti-goat (Li-Cor/Odyssey, catalog no. 926-32224).

**RNA sequencing.** RNA samples were poly(A) selected, and libraries were prepared using the Illumina TruSeq RNA sample preparation protocol (catalog no. RS-930-2001). Poly(A) selection was carried out on cytoplasmic RNA fractions prior to sequencing, while nuclear RNA fractions did not undergo poly(A) selection prior to sequencing. Two × 100-base paired-end sequencing was performed using an Illumina HiSeq instrument.

**Real-time RT-PCR analysis.** First-strand synthesis was carried out with RNA samples using the SuperScript III first-strand synthesis system (Invitrogen, catalog no. 18080-051) using oligo(dT) primers. All real-time reverse transcription-PCR (RT-PCR) analyses were performed using an SsoFast EvaGreen Supermix (Bio-Rad, catalog no. 172) on a Bio-Rad CFX96 machine. PCRs were carried out using the following
conditions: 95°C for 30 s, followed by 40 cycles of 95°C for 3 s and 60°C for 3 s. The expression of EBV transcripts was determined by the 2\(\Delta\Delta\)CT method, in which CT is the threshold cycle. All primer sequences are shown in Table S1 in the supplemental material.

**5’ and 3’ RACE.** 5’ and 3’ RACE (rapid amplification of cDNA ends) was performed using a SMARTer RACE cDNA amplification kit from Clontech (catalog no. 634923) according to the manufacturer's protocol. Two primers were designed corresponding to different positions within each of the BCLT1 and BCRT2 loci (the primers are listed in Table S1 in the supplemental material). PCR products from reactions were cloned into the pCR4-TOPO vector and the inserts were Sanger sequenced to identify the start sites of BCLT1 and BCRT2.

**Data analysis.** RNA-Seq reads were aligned to a human (hg19 assembly) plus EBV genome [B95-8-Raji (National Center for Biotechnology Information [NCBI] accession number NC_007605) index using Novoalign). Splice junctions were identified using TopHat (17) run on a human plus EBV genome Bowtie index. RPKM calculations (i.e., reads per kilobase of exon per million mapped reads) and genome coverage files (wiggle files) were generated using SAMMate (19). The Integrative Genomics Viewer (IGV) (14) was used to visualize sequence alignments, genomic annotations, and splice junctions. In the cytoplasmic fractions, the percentage of EBV or human reads were calculated by dividing the total number of EBV or human mapped reads by the sum of EBV and human mapped reads and then multiplying that value by 100. The numbers of reads mapping to exonic, intronic, and intragenic were determined using SAMMate with annotation files containing the coordinates of all human exons, all human introns, and all human intragenic blocks. The percentage of reads in each category was determined by
dividing exonic, intronic, or intragenic regions mapped read numbers by the total number of mapped reads and multiplying by 100). Ingenuity pathway analysis (IPA; Ingenuity Systems, Inc.) was performed by inputting all cellular genes showing a >4-fold increase or decrease in expression following BCR activation. Input values were the log2 of relative expression.

**Sequencing data accession number.** RNA-Seq data are available from the NCBI Short Read Archive under accession no. SRA047981.2.

1.4 Results

We have previously reported a pipeline for EBV transcriptome analysis in the context of the human transcriptome using RNA-Seq (11). The pipeline was applied to whole-cell RNA preps from the type I latency cell lines Mutu I and Akata to illustrate the quantitative and qualitative value of RNA-Seq in assessing viral transcription. For the work described here, we have enhanced our approach to gain greater clarity in EBV transcriptome analysis. First, we have used 100-base sequencing reads (versus 50-base reads), and the sequencing was carried out on an Illumina HiSeq instrument (versus an Illumina GA2X) to increase the overall sequencing depth. Second, we have now adapted the pipeline to the EBV type I strain B95-8/Raji genome assembly (2) (versus the type II strain AG876 [3]) to achieve a more accurate read alignment for cells harboring this more common strain of virus. Third, for alignment purposes we have split the circular EBV genome between the BBLF2/3 and the BGLF3.5 lytic genes rather than at the terminal repeats so that LMP2 splicing events spanning the terminal repeats can be captured.
Lastly, we have carried out nuclear and cytoplasmic fractionation to specifically isolate and analyze compartmentalized RNAs.

Nuclear and cytoplasmic RNAs were isolated from parallel cultures of Akata cells in which one culture was treated with an anti-human-IgG antibody (24 h) to induce BCR-mediated viral reactivation. Nuclear and cytoplasmic enrichment was considered good in all cases since we observed an ∼30-fold enrichment of the nuclear U2sn transcripts and the cytoplasmic actin transcripts by real-time RT-PCR (data not shown). Mapping of the sequencing reads from uninduced or induced Akata cells to the human genome showed that 81% and 76% of the nuclear reads mapped to introns and that 7% and 6% of the reads mapped to exonic regions, respectively (Fig. 1.1A). This is consistent with an abundance of unprocessed coding transcripts in the nucleus (note that for any single cellular gene, the total length of intron sequences is typically much greater than the total length of exon sequences). In contrast, 70% and 80% of the uninduced and induced cytoplasmic reads mapped to exons (Fig. 1.1A), indicating an abundance of completely processed transcripts. Figure 1B illustrates this difference in coverage across an individual cellular gene, ZFR.

Prior to sequencing, we had determined that the level of induction was reasonable as assessed by Western blot analysis of the immediate-early and early Zta, Rta, and BMRF1 proteins (Fig. 1.2B). Quantitative analysis of the cytoplasmic RNA-Seq data showed robust induction of most viral genes, with the bulk of viral genes being induced 200- to 700-fold (Fig. 1.2). Consistent with previous microarray studies (20), we also observed the induction of latency genes, although the level of induction tended to be somewhat less than for most lytic genes (red-boxed genes in Fig. 1.2). Despite the high
level of viral gene induction, we were still surprised to find that viral transcripts account for 7% of all polyadenylated cytoplasmic transcripts following induction (Fig. 1.2C). This is especially significant in light of the fact that the EBV genome is 1/25,000 the size of the human genome. These results are a testament to the significant redirection of the cellular transcriptional machinery to viral genes during reactivation.

**Alternative splicing analysis of BZLF1 and BLLF1 lytic genes.** Following BCR activation, the immediate-early gene BZLF1 was induced 196-fold (Fig. 1.3). This level of induction is reflected in the coverage data showing nearly 2000 reads spanning some regions of the BZLF1 gene locus under induced conditions (Fig. 1.3A). Eight hundred eighteen and three hundred thirty-nine junction reads were found to span introns 1 and 2 of the BZLF1 gene, respectively. Eighty-one reads mapped to the exon 1-exon 3 junction, reflecting splicing of the BZLF1 inhibitory variant RAZ (7). Based on the relative numbers of junction spanning reads, we estimate that RAZ is expressed at ca. 7% of wild-type BZLF1 under these induced conditions.

The lytic transcript BLLF1 encodes the glycoprotein 350/220, which binds to CR2/CD21 on the surface of B cells to initiate infection (9, 12, 16). After BCR activation, BLLF1 transcripts are increased 269-fold (Fig. 1.2), with coverage data showing nearly 9,000 reads across some regions of the BLLF1 locus following induction (Fig. 1.3B). Junction data showed ample evidence of the splice variant that gives rise to the gp220 isoform with 576 junction-spanning reads (Fig. 1.3B). High expression of this splice variant was also detected by quantitative RT-PCR following BCR-mediated reactivation in both Akata and Mutu I cells (Fig. 1.3C). In addition to the previously annotated splice variant, we also detected evidence of a new splice junction within the BLLF1 locus.
which is represented by 101 junction reads (Fig. 1.3B). Quantitative RT-PCR validated
the expression of this new alternative splicing event in induced Akata and Mutu I cells
(Fig. 1.3C). Consistent with this transcript originating from the BLLF1 promoter, splice
donor/acceptor analysis of these junction reads by TopHat predicts this splicing event to
occur in the leftward direction. The resulting transcript is therefore likely to encode a
truncated BLLF1 variant encoding a 126-amino-acid protein product.

**Alternative splicing of latency genes LMP1 and LMP2 following reactivation.**

LMP1 transcript levels were increased 186-fold following induction (Fig. 1.2). Previous
studies have shown the existence of a lytic promoter within intron 1 (8), and the RNA-
Seq data shown in Fig. 1.4 nicely illustrates the existence of this alternative promoter
following induction. Whereas in JY cells (type III latency), there is robust coverage
throughout exon 1, and there are high numbers of junction reads spanning intron 1 and
intron 2, there is only low coverage across exon 1 in induced Akata cells and
substantially fewer junction reads (Fig. 1.4). These Akata data are consistent with
activation of the lytic LMP1 RNA, with transcription initiated in intron 1 and low but
detectable splicing of intron 2 (8) (Fig. 1.4). Notably, this lytic LMP1 transcript is
noncoding in most EBV isolates, including Akata cells (4), and may therefore have a
unique function in the EBV infection cycle.

Following BCR activation in Akata cells, LMP2A transcript levels increase 182-
fold (Fig. 1.2), as represented by the induced cytoplasmic coverage data (Fig. 1.5A).
Nevertheless, there was relatively low evidence of classic splicing events compared to the
type III latency JY cells (Fig. 1.5A). Consistent with this observation, there is substantial
read coverage across the LMP2 introns in Akata compared to JY cells which show read
coverage drop-offs across the introns (Fig. 1.5A). The low sequential splicing values observed in Akata cells is not due solely to a slightly lower read coverage, because quantification of splicing reads/LMP2 coverage reads shows greater values in JY cells than in Akata cells at each splice junction (Fig. 1.5B). These data indicate that while classic LMP2 splicing does occur in induced Akata cells, there is an abundance of incompletely spliced, polyadenylated LMP2 transcripts in the cytoplasmic RNA fraction. We also noted evidence of alternative splicing within the LMP2 locus that was not observed in JY cells (Fig. 1.5A). Splicing from exons 1 to 6 and exons 1 to 7 were validated by quantitative RT-PCR in induced Akata and Mutu I cells, and little evidence of these splicing events was observed in the type III cell lines Jijoye, X50-7, and JY (Fig. 1.5C). This indicates that these alternative splicing events are specific to lytic conditions and that they are not represented well in type III latency.

**Latency transcripts induced during reactivation have lower cytoplasmic-to-nuclear read ratios than most lytic genes.** Considering the possibility that some of the latency transcripts, like LMP2, may have distinct functions during reactivation, we investigated the ratio of cytoplasmic-to-nuclear reads of all lytic and latent EBV genes following reactivation. Strikingly, while LMP1's cytoplasmic-to-nuclear RPKM ratios were similar to the bulk of lytic transcripts, the cytoplasmic-to-nuclear RPKM ratios for the latency transcripts LMP2A, EBNA-3A, EBNA-3B, EBNA3C, EBNA-2, and EBNA-LP were comparatively low (Fig. 1.6). This raises the possibility that some of these transcripts may have nuclear functions during reactivation.

**Identification of the lytic transcripts BCLT1 and BCRT2.** Although most of the EBV genome is annotated with gene structures, there remain a few regions with no
known genes. We have noted that some of these regions are, in fact, transcribed following induction, as evidenced by the presence of reads. For example, we observed an abundance of reads between the end of the LMP2 gene and the beginning of the EBER1 gene (Fig. 1.7A). 5’ RACE identified two start sites in opposite orientations (Fig. 1.7B and C), indicating that there are at least two overlapping divergent transcripts in this region. Real-time RT-PCR analysis detected expression of both of these transcripts in induced Akata and Mutu I cells and not in Jijoye, X50-7, or JY cells (Fig. 7B and C). Analysis of the coding capacity of the leftward transcript shows no reading frames encoding more than 34 amino acids long, suggesting that this transcript is noncoding. The longest predicted reading frame for the rightward transcript is 72 amino acids. Based on this analysis, we are tentatively calling these transcripts BCRT2 (for BamHI C-fragment right transcript 2, thereby distinguishing it from another rightward transcript, BCRF1) and BCLT1 (for BamHI C-fragment left transcript 1).

**Cellular gene expression changes following BCR activation.** Cellular gene expression was analyzed in uninduced and induced Akata cells using cytoplasmic reads. A total of 148 expressed cellular genes were found to increase or decrease by 4-fold or more (for genes with RPKM values of 1 or more in at least one condition). Nine of these were selected for validation by quantitative RT-PCR (Fig. 1.8A). A relatively good correspondence was observed between RNA-Seq data and quantitative RT-PCR, indicating that the RNA-Seq data were generally reliable. It should be noted, however, that quantitative RT-PCR measures levels of only a portion of the gene, whereas RNA-Seq measures an average of all isoforms of a gene which may account for some amplitude differences in the changes observed by the two methods. Notably, many early
response genes, such as Fos and EGR1, are induced, consistent with BCR activation responses. In addition, there are a substantial number of immune regulatory factors whose levels change by >4-fold, including the quantitative RT-PCR-validated membrane signaling proteins CD7, IGSF1, IL2RB, SLAMF7, and GBP2 (Fig. 1.8). Consistent with this finding, IPA of this 148-gene set found inflammatory response to have the highest significance (Fig. 1.8B). The pathways with the greatest predicted significance in this data set were ILK (integrin linked kinase) and TREM1 (triggering receptor expressed on myeloid cells 1) (Fig. 1.8C), indicating the alteration of these signaling pathways in BCR-activated B cells.

1.5 Discussion

For the analysis performed here, we have generated the necessary tools to perform alignments and to visualize the data against the more common strain of EBV, type I. Accessory files required for the analysis of both type I and type II virus sequence data are available at www.flemingtonlab.com (files include fasta files, Bowtie index files, genome viewer annotation files, and annotation files for EBV RPKM analysis [for the analysis of either EBV alone or the analysis of EBV in the context of the human genome]). The development of these tools for the analysis of the type I EBV strain is important since it is a more common strain observed in tissue culture models and in vivo. For most of the EBV genome, reasonable alignment data can be obtained even when using sequencing data from a type I strain and aligning to a type II strain genome (11). Nevertheless, the capture of sequences from less well-conserved regions, such as the EBNA2 locus, requires the use of the appropriate genome strain during the alignment process (11; data not shown). Whether sequence variations within strains pose any alignment difficulties
has not yet been tested, but since most alignment approaches allow for a limited number of mismatches, this is not likely to be a significant problem. Nevertheless, we have ongoing studies to sequence the Mutu I and Akata genomes to determine whether intrastrain differences cause inaccuracies at subregions of the EBV genome.

Nearly all RNA-Seq analysis tools have been developed for linear genomes since they are generally designed for the analysis of eukaryotic cellular organisms. The implementation of these tools to the analysis of circular genomes requires the artificial linearization of the respective genome at an arbitrary genome position. Although the terminal repeats represent a logical breakpoint for representation of EBV genome features in the NCBI database, it poses a problem for the analysis of reads spanning the terminal repeats and, more importantly, the LMP2 splicing events that span this region. For our analysis, we split the genome between the BBLF2/3 and BGLF3.5 genes. From our preliminary studies aligning against a genome that was split at the terminal repeats, we noted a natural depletion of reads between these two genes, and we did not detect any bona fide splicing events across this region in either uninduced or induced Akata cells (Fig. 1.9). We therefore tentatively recommend utilizing this genome configuration in EBV transcriptome analysis pipelines.

Outside of the lytic LMP1 RNA, the structures of lytic transcripts derived from latency gene loci have not been previously probed in detail. Our more detailed analysis of lytic transcripts from the LMP2 loci shows a limited level of classic consecutive splicing events and an abundance of intron-spanning reads compared to type III latency cells. We also observed alternative splicing that was not observed in type III latency cells. Analysis of the exon 1-exon 6 splice and the exon 1-exon 7 splicing events reveals frameshifts at
the splice junctions with termination codons located soon after the splice junctions. These transcripts are predicted to encode 171- and 145-amino-acid peptides, respectively, and we detected a 20 and 25 kdal products in cells transfected with corresponding expression vectors. These two alternative splicing events are also detected in BCR-activated Mutu I cells, indicating that these splicing events are not spurious anomalies specific to Akata cells. It is also interesting that the number of junction reads capturing the exon 1-exon 6 and the exon 1-exon 7 splicing events is slightly greater than the number of junction reads capturing the classic consecutive exon 1-exon 2 splicing event. Further, transcripts bearing these alternative splice junctions are localized in the cytoplasm and are polyadenylated, further indicating a function for these isoforms.

Considering the enriched localization of EBNA-3A, EBNA-3B, EBNA-3C, EBNA-2, EBNA-LP, and LMP2 transcripts in the nucleus (Fig. 1.6), we need to consider the possibility that some of these transcripts may perform noncoding functions in the nucleus. Cellular encoded long noncoding cellular RNAs (lncRNAs) have been shown to have profound influences on cellular reprogramming through their interactions with lineage specific promoters (10, 18). Many late genes do not contain response elements for the immediate-early transactivators Zta or Rta, but they are induced to very high levels through relatively unknown mechanisms. It is not hard to imagine the possible involvement of at least some of these nuclear latency transcripts in helping facilitate the high level of EBV promoter activation that is observed during reactivation. Another possible role for one or more of these nuclear transcripts could be to help facilitate lytic DNA replication and/or linking replication to late gene expression.
It is important to note that our analysis was performed using a 24-h time point. These data therefore represent a snapshot in the lytic cycle. We need to consider the possibility that some of the nuclear transcripts derived from at least some of these latency gene loci are simply late genes that have not yet been fully processed and exported to the cytoplasm. If this is the case, however, their localization is not representative of the bulk of late genes, such as BFRF2/3, BcLF1, BLLF1, BXLF2, and BKRF2, which clearly have high cytoplasmic-to-nuclear transcript ratios (Fig. 1.6). Yuan et al. (20) have shown that EBNA2, EBNA-3A, and EBNA-3C proteins are detected 2 days after BCR activation in Akata cells. This suggests the presence of enough appropriately processed and transported transcripts to be translated due to a high level of these latent transcripts at these time points or that there is a substantial delay in processing, possibly allowing for nuclear functions to occur before they become utilized as protein coding transcripts.

We have identified read coverage across a region between the end of the LMP2 gene and the EBER1 gene in induced Akata cells (Fig. 1.7). Using quantitative RT-PCR, we also detected evidence of transcription across this region in induced Mutu I cells but not in uninduced Akata or Mutu I cells or in the type III latency cell lines Jijoye, X50-7, or JY. Using 5′ RACE, we were able to determine that there are at least two overlapping, inversely oriented RNAs transcribed from this region (Fig. 1.7). We did not detect any significant reading frames for these predicted transcripts, so we tentatively assumed that they are of the long noncoding class of RNA transcripts. These transcripts are polyadenylated since our PCR and sequencing reactions detected them in poly(A)-selected RNA fractions (Fig. 1.7 and data not shown). We have analyzed these transcripts primarily using cytoplasmic RNA to help ensure that our results represent fully matured
transcripts. However, the cytoplasmic-to-nuclear RPKM ratios for these transcripts are significantly lower than the average ratio for coding lytic genes (Fig. 1.10A). On the other hand, the cytoplasmic-to-nuclear ratios for BCRT2 and BCLT1 are in line with or lower than the ratios observed for most of the induced latency genes. Finally, we readily detect both BCRT2 and BCLT1 using oligo(dT)-primed RNA from the nuclear fraction, which is consistent with the presence of processed BCRT2 and BCLT1 in the nucleus (Fig. 1.10B). We therefore hypothesize that these newly identified transcripts play a noncoding role in the nucleus during reactivation.

Analysis of anti-IgG-mediated changes in cellular genes identified a significant number of factors involved in regulating immune cell signaling and phenotype (Fig. 1.8). This is consistent with a role of BCR signaling in facilitating cellular reprogramming. It is noteworthy, however, that although there are a substantial number of cellular genes that change by >4-fold, the level of induction of most of these are significantly below the level of induction of most viral lytic genes. This may partly explain the substantial fraction of reads mapping to EBV compared to the cellular genome following reactivation (Fig. 1.2). We can assume that some, if not many or most, of the induced cellular genes are induced through BCR-initiated signaling events rather than through virally encoded transcriptional activators. This is a testament to the clear bias that the virus has in the effective production of viral factors over the production of cellular factors.

From a practical standpoint, however, this observation is a little surprising given that the immediate-early transcription factor Zta can efficiently bind to cellular AP-1 sites (5). The relatively inefficient induction of cellular genes may be due to the lack of synergistically configured response elements to the second immediate-early transcription
factor, Rta, which may cooperate with Zta in activating viral promoters. On the other hand, the viral chromatin may be modified in a more suitable configuration to accommodate robust induction of lytic genes (1). We suggest the possibility that there are chromatin-associated long noncoding viral RNAs that may help facilitate the recruitment of activating transcription complexes that may bind viral promoters through sequence complementarity. BCRT2 and BCLT1 are possible candidates in such processes and warrant further investigation.
Figure 1.1. (A) Human exonic, intronic, and intragenic read distributions for uninduced and induced nuclear and cytoplasmic sequencing data. (B) Nuclear and cytoplasmic RNA-Seq coverage data at the human ZFR locus. The y axis is the number of reads spanning each genomic coordinate (x axis). Four known isoforms for ZFR are shown at
the bottom of the figure. Thick lines represent coding sequences, medium thickness lines represent untranslated regions, and thin lines with leftward arrows represent introns.
Figure 1.2. (A) Fold induction of EBV genes in cytoplasmic Akata fractions. Red boxes indicate latency genes. (B) Western blot analysis of the immediate-early and early Zta, Rta, and BMRF1 proteins performed as a quality control prior to sequencing of RNAs.
(C) Percentages of human and EBV reads in uninduced and induced cytoplasmic Akata fractions.

Figure 1.3. (A) Read coverage data (Novoalign) from uninduced and induced Akata cells and splicing evidence (from TopHat) from induced Akata cells are shown for the
BZLF1/Zta locus. The y axis is the number of reads spanning each genomic coordinate (x axis). Thick lines represent coding sequences, medium thickness lines represent untranslated regions, and thin lines with leftward arrows represent introns. Evidence for the canonical BZLF1 splicing are represented by 818 and 339 reads spanning intron 1 and intron 2, respectively. Eighty-one reads spanning the exon 1-exon 3 splicing event correspond to the dominant-negative variant RAZ. (B) Coverage data and splicing evidence at the BLLF1 (gp350/gp220) locus show 101 reads spanning a novel splicing event. The peak within the annotated BLLF1 intron probably does not represent a stand-alone exon since no junction-spanning reads were identified at the peak edges. Instead, these reads are likely attributable to the unspliced version of BLLF1. (C) Relative expression of the BLLF1 splice variant and the alternative splice as determined by quantitative RT-PCR. RNAs from uninduced or induced Akata or Mutu I cells and from the type III latency cell lines Jijo, X50-7, and JY are shown.
Figure 1.4. (A) Read coverage across the LMP1 locus illustrates the lytic LMP1 transcript structure in induced Akata cells and the classic type III latency LMP1 transcript structure in JY cells. The y axis is the number of reads spanning each genomic coordinate (x axis). Thick lines represent coding sequences, medium thickness lines represent untranslated regions, and thin lines with leftward arrows represent introns. Red values are
the number of reads that span the indicated junctions. Coverage data were derived from Novoalign, and junction evidence was taken from TopHat. (B) Quantitative RT-PCR was used to determine the ratio of transcripts spanning exon 2 to exon 3 versus exon 1 to 2 in cytoplasmic fractions from uninduced or induced Akata cells and the type III latency cell lines JY, Jijoye, and X50-7.
Figure 1.5. (A) Unique coverage and splicing of LMP2 transcripts in induced Akata cells versus the type III cell line JY. The y axis is the number of reads spanning each genomic coordinate (x axis). Thick lines represent coding sequences, medium thickness lines represent untranslated regions, and thin lines with leftward arrows represent introns. Red values are the number of reads that span the indicated junctions. Coverage data were derived from Novoalign, and junction evidence was taken from TopHat. (B) Ratio of
sequential splicing events to total coverage across LMP2 locus in induced Akata cells and JY cells. Total coverage calculation excluded regions of LMP2 which overlaps with BMRF1. (C) Real-time RT-PCR validation of alternative splicing in induced Akata and induced Mutu I cells.

Figure 1.6. Higher nuclear retention of latency genes (sans EBNA1 and LMP1) than lytic genes following B-cell receptor activation in Akata cells.
Figure 1.7. (A) Coverage across new transcript region in induced Akata cells. Red arrows represent primers used for 5′ RACE. Green arrows represent PCR primers used for quantitative RT-PCR. (B) Real-time RT-PCR analysis (values are relative to the no-RT control in uninduced and induced conditions) of BCLT1 expression (B) and BCRT2 (C) and 5′ RACE identification of start sites. 5′ RACE products were cloned and sequenced. The start site was determined to be identical using either primer.
Figure 1.8. (A) Quantitative RT-PCR validation of selected cellular gene changes following BCR activation. (B) Top cellular functions predicted to be influenced by expressed cellular genes with changes of >4-fold. (C) Top canonical pathways predicted to be influenced by expressed cellular genes with changes of >4-fold.
Figure 1.9. (A) Higher nuclear retention of BCRT2 and BCLT1 than the average of lytic genes following B-cell receptor activation in Akata cells. (B) Real-time RT-PCR validation of cytoplasmic to nuclear BCLT1 and BCRT2 transcript detection relative to actin RNA in induced Akata and Mutu I cells.

Figure 1.10. Illustration of EBV genome coverage gap where the genome has been split for subsequent EBV RNA-Seq analysis.
CHAPTER 2. Identification and Characterization of two novel Epstein-Barr virus long non-coding RNA in the LMP2 locus

2.1. Abstract

Using an enhanced RNA-Seq pipeline to analyze the Epstein-Barr virus (EBV) transcriptome, we investigated viral and cellular gene expression in the Akata cell line following B-cell-receptor-mediated reactivation. Robust induction of EBV gene expression was observed, with most viral genes induced more than 200-fold and with EBV transcripts accounting for 7% of all mapped reads within the cell. After induction, a novel antisense transcript, a circular transcript, and multiple of candidate viral splicing events were detected in the LMP2 locus using splice junction aligner, TopHat. A more detailed analysis of lytic LMP2 transcripts and LMP2A transcripts expressed during type III latency revealed alternative splicing events and novel transcripts during reactivation. Characterization of novel antisense and circular RNA transcripts has lead to identifying them as virus long non-coding RNA (vlncRNA) transcripts. Pathway analysis of viral and cellular gene expression changes following knockdown of the antisense LMP2 transcript has revealed that the antisense LMP2 transcript increases the magnitude of EBV lytic activation following BCR reactivation.

2.2 Introduction

The LMP2 gene and signaling pathways.

The LMP2 gene locus encodes two mRNAs that express the hydrophobic, intermembrane proteins, LMP2A and LMP2B. LMP2A and LMP2B share the last 8 exons encoding 12 membrane-spanning domains but differ in their first 5’ exon. The
LMP2B transcript has a short non-coding 5’ exon while the first exon of LMP2A encodes a cytosolic signaling sequence. The amino-terminal signaling domain of LMP2A contains binding sites for the tyrosine kinases Lyn and Syk and the ubiquitin ligase Nedd4/Itchy (1,3). The Lyn and Syk tyrosine kinases are essential for LMP2A’s B-cell receptor (BCR) signaling function and allow viral latency to persist by blocking normal BCR activation. In uninfected primary B-cells, stimulation of the BCR leads the complex into lipid rafts to interact with the tyrosine kinases. The tyrosine kinase phosphorylates the BCR immunoreceptor tyrosine activation motifs (ITAMs) and initiates a signaling pathway that leads to cell proliferation and differentiation. Normal BCR signaling allows the IgM-negative immature B-cell to transition into an IgM-positive immature B-cell and exit the bone marrow into the periphery. LMP2A expression inhibits BCR from expressing IgM, which usually leads to apoptosis in the cell. However, LMP2A deregulation of MYC protects B-cells from apoptosis (23,24).

MYC is a proto-oncogenic transcription factor that plays a role in cell cycle progression, cellular differentiation, and apoptosis. All forms of Burkitt's lymphoma are characterized by a c-myc translocation, which moves the c-Myc gene from chromosome 8 to either chromosome 2, chromosome 14, or chromosome 22 (24). Translocation of c-myc leads to abnormal expression and eventually induces apoptosis in response to various stresses. The MYC apoptotic pathway can be dependent on a tumor suppressor protein, p53. The p53 signaling pathway is involved in cell growth that can result in the cell exiting the cell cycle or apoptosis. Normally, the p53 protein is negatively controlled by MDM2, which is an ubiquitin ligase that is inhibited by ARF. Interestingly the p53 protein acts as its own negative regulator by upregulating expression of MDM2.
activation induces the p53-MDM2-ARF pathway towards apoptosis. During apoptotic conditions p53 activates BAX and BAK, which bind to the outer mitochondrial membrane to increase permeability. LMP2A has been shown to interfere with Bax by inducing anti-apoptotic Bcl-XL, which inhibits Bax (25,26) thereby helping prevent apoptosis signaling induced from the c-myc translocation.

The LMP2A N-terminal signaling domain also contains ubiquitin ligase binding sites that play a role in exosome secretion as well as LMP2A degradation. When the LMP2A protein is sequestered into cholesterol-rich lipid rafts it is constitutively phosphorylated and ubiquitinated. The phosphorylated and ubiquitinated LMP2A protein can leave the plasma membrane and be transferred into early endosomes. The LMP2A protein progresses into late endosomes where it is dephosphorylated. At this stage the ubiquitinated LMP2A protein can move into lysosomes where it will be degraded, or it can be transferred into exosomes and released into the extracellular space (27). LMP1 has also been shown to be transferred into exosomes and induce ICAM-1 expression in recipient cells (28). EBV exosome secretion may be used as a vehicle for intercellular communication.

Although the LMP2A protein inhibits BCR signaling, its relative, LMP2B, has been shown to induce BCR signaling. As mentioned above, the LMP2B transcript has a short non-coding 5’ exon while the first exon of LMP2A encodes a cytosolic signaling sequence. Since both LMP2A and LMP2B proteins share similar structural homology, they can colocalize within same area of the plasma membrane and interact. Coexpression of LMP2A and LMP2B in B-cells showed that LMP2B does not modify the localization of LMP2A, but blocks phosphorylation of LMP2A residues. LMP2B aggregates with
LMP2A and blocks phosphorylation, which releases the tyrosine kinases, Lyn and Syk. The Lyn and Syk can then be recruited by the BCR, restoring signaling (22,27). Recently our lab has discovered novel isoforms and transcripts in the LMP2A loci suggesting that this locus plays a larger and more varied role in BCR than previously realized.

**Long non-coding RNAs.**

In the past RNA was considered to be only an intermediary between DNA and protein. It has been previously described that approximately two-thirds of genomic DNA is transcribed, however, only ~2% of that is translated into proteins. Now, with advancements in RNA-sequencing technologies and computational biology, long non-coding RNAs (lncRNAs) are rapidly being identified and characterized and being found to have a multitude of non-coding functions in the cell. The number of lncRNAs far exceeds the number of protein-coding genes. LncRNAs are operationally defined as RNA transcripts that are longer than 200 nucleotides and lack an open reading frame (ORF). They are often but not always polyadenylated. One of the most functionally investigated lncRNAs is HOTAIR (Homeobox C transcript antisense RNA). The HOTAIR gene is located on chromosome 12 and encodes a 2.2kb RNA. HOTAIR has been shown to recruit PRC2 (Polycomb Repressive Complex 2) and silence the HOXD (Homeobox D) locus through chromatin remodeling (30). Initial findings supported the idea that lncRNAs regulated chromatin modification, however, innovative techniques have lead to the discovery of vastly diverse regulatory functions and are being considered as viable biomarkers in cancer. Elevated expression of metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) has been identified as a biomarker in pancreatic ductal
adenocarcinoma (PDAC) tissues. Elevated MALAT1 expression is correlated to low survival rates in PDAC (31). The EBV encoded EBERs are routinely used as markers for EBV infection. EBER1 and 2 are viral non-coding RNAs that are expressed throughout all programs of latency and lytic replication. EBERs have the ability to stimulate IL-10, IFN type 1, IGF1, and IL-9 expression in various cell lines (32, 33, 42). Identifying the roles non-coding RNAs play in the cell may lead to new insights on understanding tumor progression.

Circular RNAs.

While the majority of RNAs have been thought to be linear, a new class of RNAs has emerged called circular RNA (circRNA). Previously, circRNA transcripts were only found in a handful of genes, however, with advancements in deep-sequencing technology more than 40,000 circRNAs have now been identified (35). They have been identified in diverse species including mouse, archaea, and humans. CircRNAs can be composed of one or more exons that are generated by back-splicing of protein-coding genes. Unlike their linear protein-coding gene counterparts, circRNAs lack a poly(A) tail and a 5’ CAP structure which prevents the circRNAs from generating a protein. While circRNAs have not been found to generate proteins, they are predominantly found in the cytoplasm, supporting the notion that they are not artifacts. Most of the investigation into possible functions of circRNAs has been focused on their potential role as microRNA (miRNA) sponges. In order to successfully work as a sponge, the circRNA transcript would contain multiple binding sites that recognize a specific miRNA seed region. The majority of
circRNAs that our lab has identified do not contain multiple miRNA binding sites, which leads us to consider other possible functions.

2.3 Materials and Methods

**Cell culture.** All cells were grown in RPMI 1640 (Thermo Scientific, catalog no. SH30027) plus 10% fetal bovine serum (FBS; Invitrogen-Gibco, catalog no. 16000-044) with 0.5% penicillin and streptomycin (pen/strep; Invitrogen-Gibco, catalog no. 15140-122). Cells were grown at 37°C in a humidified, 5% CO2 incubator.

**Lytic cycle induction.** Akata and Mutu I cells were grown to near saturation, at which time an equal volume of fresh RPMI (plus 10% FBS and 0.5% pen/strep) was added. The following day, cells were spun down and resuspended in an equal volume of freshly warmed RPMI (plus 10% FBS and 0.5% pen/strep) with or without 10 µg/ml of anti-IgG (Akata) or anti-IgM (Mutu I). The cells were harvested 24 hours after treatment and subjected to RNA extraction or protein isolation (for Western blot analysis).

**RNA preparation.** Total RNA was isolated using TRIzol® Reagent (Life Technologies, Cat. #15596-018) or the miRNeasy Mini Kit (Qiagen, Cat. #217004) following the respective vendor’s protocols. Nuclear and cytoplasmic RNA was isolated using the Cytoplasmic and Nuclear RNA Purification Kit (Norgen Biotek Corp., Cat. #21000) following the vendor’s protocol. All RNA preparations were subjected to DNase treatment using the DNA-free™ Kit (Life Technologies, Cat. #AM1906).

**RNase R enrichment.** 5 µg of DNase I-treated, 24hr post IgG-treated Akata RNA was incubated with or without 20 U of RNase R (Epicentre Biotechnologies, Cat. #RNR07250) at 37 °C for 1 hour. The RNase R-treated RNA was subsequently purified by the RNA Clean and Concentrator-5 kit (Zymo Research, Cat. #R1015). RNA was
amplified by taq-PCR (New England Biolabs, Cat. #E5000S) following the vendor’s protocol. PCR products were run on a 2% agarose gel at room temperature. PCR products were cut out and purified using the PCR Clean-Up and Gel Extraction Kit (Clontech, Cat. #740609.250). The resulting PCR fragments were cloned into the pCR4-TOPO vector (Invitrogen, Cat. #450071) and the inserts were Sanger sequenced.

PCR primers:
- ActinB forward: 5’- AGGGCATACCCCTCGTAGAT
- ActinB reverse: 5’- GTCTTCCCCTCCATCGTG
- cLMP2 exon 3 forward: 5’- GGATTGAGGACCCACCTTTT
- cLMP2 exon 4 forward: 5’- TTCTGGTGATGCTTGTGCTC
- cLMP2 exon 5 forward: 5’- TCACTGATTTTTGGGCACACTT
- cLMP2 exon 6 forward: 5’- ATCGCTGGTGCCAGTATTTT
- cLMP2 exon 7 forward: 5’- GCTCTCGCACTCTTTGGCT
- cLMP2 exon 2 reverse: 5’- AGTAGGGCGCAACAATTAC

5’ and 3’ RACE. 5’ RACE (rapid amplification of cDNA ends) was performed using a SMARTer RACE cDNA Amplification Kit from Clontech (Cat. #. 634923) according to the manufacturer's protocol. Primers were designed corresponding to different positions within the LMP2 locus. PCR products from these reactions were then cloned into the pCR4-TOPO vector and the inserts were Sanger sequenced to identify the start and end sites of aLMP2.

5’RACE aLMP2 primers:
- aLMP2 5’R #1 5’- TGCCACATGTTTTCTGGACACAGGAC
aLMP2 5'R #2 5' - TCGTGCCTTCATGCACCTGACCAA
aLMP2 5'R #3 5' - GTGTGACACCAACAGGTGTGCTTG
aLMP2 5'R #4 5' - CAGGTGGCCATTGCTATATGCACGCTTC
aLMP2 5'R #5 5' - TTTTATGTGCTCCGCGGCTGCT
3'RACE aLMP2 primers:
  aLMP2 3'R #1 5' - CTCACAAGTGACAACCGCAGTAAGCA
  aLMP2 3'R #2 5' - GTAAAGTTGACGTCATGCCAAGGC
  aLMP2 3'R #3 5' - GAAGATATGCCAGCGACACTACG
  aLMP2 3'R #4 5' - GCGCAACAATTACAGGCAGGCATAC
  aLMP2 3'R #5 5' - CGGTGCCGGTCACAACCGGTACTAAC
  aLMP2 3'R #6 5' - TGGCCACTGCTGCCAAGATGAGAAG
  aLMP2 3'R #7 5' - TCACCGGTGTCAGCAGTTCTTGTG
  aLMP2 3'R #8 5' - TATGCATTGTAATGGTGCGTG

**Standard qRT-PCR and strand-specific qRT-PCR.** Standard reverse transcription was performed using an iScript™ cDNA Synthesis Kit (Bio-Rad, Cat. #170-8891) according to the manufacturer’s protocol. Strand-specific reverse transcription was carried out using modified gene-specific primers. cDNA was synthesized at 65°C for 50 min using ThermoScript Reverse Transcriptase (Life Technologies, Cat. #12236-022) and treated with Exonuclease I (New England Biolabs, Cat. #M0293L) to remove the excess gene-specific primers. qPCR analysis was performed using iQ SYBR Green Supermix (Bio-Rad, Cat. #170-8882) on a Bio-Rad CFX96 instrument as follows: 1 µl cDNA and 1 µl 10 µM primers were mixed with 10 µl of SYBR green supermix and 8 µl nuclease-free H2O to a 20 µl reaction volume. Polymerase was activated and cDNA was
denatured at 95 °C for 5 minutes. cDNA was then amplified for 40 cycles with 15 s
denaturation at 95°C, 60 s annealing/extension and plate reading at 60°C. Melting curve
analysis was performed at temperatures from 60°C to 90°C with 0.5°C increment per 5 s.
Expression fold changes were calculated using the comparative CT method (2-\( \Delta \Delta CT \)).

Modified gene-specific primers (GSP) used for strand-specific RT:

aLMP2 GSP: 5’- CGCAGCTCTAACCTTGGCATCcgggeCAGTGGTTCT(40)
ActinB GSP: 5’- GTACAGGTCTTTGCGGATGTTtAAtAAtACACTTCATG
GAPDH GSP: 5’- GGGTGATTGATGGCAACggegeCCACTTTACC

Primers used for qPCR:

LMP2A Exon 1-2 forward: 5’- CCTACTCTCCACGGGATGAC
LMP2A Exon 1-2 reverse: 5’- CGGTGTCAGCAGTCTTTCTTT
LMP2A Exon 1-6 forward: 5’- CGACCGTCACTCGACTATC
LMP2A Exon 1-6 reverse: 5’- TGGACATGAAGACACGAAG
LMP2A Exon 1-7 forward: 5’- CGACCGTCACTCGGACTATC
LMP2A Exon 1-7 reverse: 5’- GAGTATGCCAGCGACAATCA
LMP-1 Exon 1-2 forward: 5’- AATTCCAGAGAGCGATGAC
LMP-1 Exon 1-2 reverse: 5’- TCCTCTTTGGCGCTACTGTCT
LMP-1 Exon 2-3 forward: 5’- AGTCATCGTGTTGTTGTTCA
LMP-1 Exon 2-3 reverse: 5’- GGCATTGTTCCTTGGAATTG
BLLF1 splice variant Exon 1-2 forward: 5’- TGTGCTGATAGAGGCTGGTG
BLLF1 splice variant Exon 1-2 reverse: 5’- TGCTGATCCCAATACACGAG
BLLF1 Alternative Splice Exon 1-2 forward: 5’- CGAACTGTTGAGACACATG
BLLF1 Alternative Splice Exon 1-2 reverse: 5’- TCTACTGGAGCTCTTTGGTG
aLMP2 forward: 5’- CGCAGCTCTAACTTGGGCATC
aLMP2 reverse: 5’- CTTGGTCAGTGACATGAAGAG

cLMP2 exon 8 forward: 5’- GCCCTCTTGGGTCATTAG
cLMP2 exon 2 reverse: 5’- ATATATGAAGAAGCGGGCAGAG

LMP2A 1-2 forward 5’- CCTACTCTCCACGGGATGAC
LMP2A 1-2 reverse 5’- CGGTGTCAGCAGTTTCTTT
LMP2A 1-6 forward: 5’- CGACCGTCACCTCGGACTATC
LMP2A 1-6 reverse: 5’- TGGACATGAAGAGCAGAAG
LMP2A 1-7 forward: 5’- CGACCGTCACCTCGGACTATC
LMP2A 1-7 reverse: 5’- GAGTATGCCAGGACAATCA

Zta forward: 5’- CACGACGTACAAGGAAACCA
Zta reverse: 5’- GAAGCCACCTCAGGATG

ActinB forward: 5’- CACTCTTCCACCCCTTCTTC
ActinB reverse: 5’- GTACAGGTCTTTGCGGATGT

GAPDH forward: 5’- GTCACCAGGGGCTTGT
GAPDH reverse: 5’- GGGTCATTGATGGCAAC

KCNQ1OT1 forward: 5’- TACCCGGATCCAGGGTTTGCAGTACA
KCNQ1OT1 reverse: 5’- GCTGATAAAGGCAACGGAAGGAAA

**RNA fluorescence in situ hybridization (FISH).** Stellaris FISH probes with CAL Fluor Red 610 fluorophore were designed and purchased from Biosearch Technologies. Antisense and circular LMP2-FISH was performed according to the Biosearch Technologies online protocol for suspension cells. Cells were washed 24 hours later and incubated with DAPI and Alexa Fluor® 488 goat anti-mouse IgG (H+L)
secondary antibody (Life Technologies, Cat. #A31619) for 1 hour in the dark. Images were taken with a 100/1.35 oil objective lens on a Leica DMRXA2 deconvolution upright microscope using Slidebook and Volocity software.

**Plasmid constructs.** The antisense LMP2 plasmid construct contains sequences spanning nucleotides 63342 to 65293 of the Akata genome. The cLMP2 and cLMP2 splice acceptor mutant plasmid constructs contain sequences spanning nucleotides 61658 to 65293 of the Akata genome. These EBV sequences were amplified by PCR and inserted into the pMSCV-puro expression vector. The cLMP2 splice acceptor mutant plasmid was generated using ??? Kit to mutate the splice acceptor near LMP2A exon 2 from AG to TC.

**Amaxa Nucleofection for Mutu I cLMP2.** Nucleofection of Mutu I cells with control, cLMP2, and cLMP2 SA mutant vectors was performed using Solution R (Lonza, Cat. #VCA-1001) and program G-16 on a Nucleofector II device. Mutu I cells were placed in antibiotic-free RPMI media two days before electroporation. For each nucleofection, 5x10^6 Mutu I cells per well with 1ml of RPMI using a 12-well plate. Nucleofection was performed in triplicates using 5µg of each plasmid (control, cLMP2, and cLMP2 SA mutant) combined with 100µl of Solution R. Immediately after electroporation, ~500µl of warm-RPMI media (no antibiotics) was added to cells and transferred to the 12-well plate. The following day (24 hours later) 1.5ml of fresh RPMI-media was added to each well. The cells were harvested 48 hours after nucleofection and RNA was isolated from triplicate infected pools using TRIzol® Reagent. RNA samples were treated with DNase I for 1 hour at 37°C. Quantitative RT-PCR was preformed for each sample for validation of cLMP2 expression. cDNA libraries (Illumina, Cat. # RS-
930-2001) were prepared from poly(A)-selected RNA using the strand-specific Illumina TrueSeq protocol and sequenced on an Illumina HiSeq 2000 instrument. Triplicate samples were multiplexed per sequencing lane.

**Constitutive cLMP2 expression for FISH.** Stable cLMP2-expressing DG-75 cells were generated through retroviral transduction and selection of infected pools. Briefly, 293T cells were seeded on 10 cm plates in 10 ml DME/High Glucose medium (supplemented with 10% FBS and 0.5% pen/strep) one day before transfection. 10µg of each packaging plasmid (pVPACK-Gp-dI and pCI-VSV-G) and 10µg of the cLMP2, cLMP2 SA mutant, or empty vector control plasmid (for a total of 30µg DNA) were mixed and ethanol precipitated at -20°C overnight to sterilize DNAs. On the day of transfection, the media on the 293T plates was replaced with 8 ml of fresh supplemented DMEM. Plasmid DNAs were spun down and washed with 70% ethanol and resuspended with 30 µl nuclease-free H2O. Plasmids were added to 500 µl HBS (0.5% HEPES, 0.8% NaCl, 0.1% dextrose, 0.01% anhydrous Na2HPO4, 0.37% KCl at pH 7.2) for each transfection. 30 µl 2.5 M CaCl2 was added to each tube and mixed by vortex on a low setting.

DNA/HBS/CaCl2 mixtures were incubated for 20 minutes at room temperature, added dropwise to 293T cell cultures, the plates were gently rocked back and forth and then transferred to a 37°C, 5% CO2 incubator. The following day the media was aspirated and replaced with 10 ml fresh medium. Virus-containing supernatant was collected 3 days post-transfection and filtered through a 0.45 µm SFCA filter to eliminate residual cells. DG-75 cells were prepared for infection by spinning them down and resuspending them at 2-4 x 106 cells/ml in RPMI (supplemented with 10% FBS and
0.5% pen/strep) plus 24 µg/ml polybrene. One milliliter of DG-75 cell suspension was added to single wells of a 6-well plate. One milliliter of control or cLMP2 or cLMP2 SA mutant virus preparations was added to duplicate wells for each, plates were spun at 1000g for 1 hour and then incubated in a 37°C, 5% CO2 incubator for 4 hours. Cell suspensions were transferred to 15 ml tubes, spun down, resuspended in 4 ml fresh RPMI (with 10% FBS and 0.5% pen/strep), transferred to 25 cm² flasks and cultured in a 37°C, 5% CO2 incubator. Two days later, 4 ml of fresh RPMI (with 10% FBS and 0.5% pen/strep) was added and puromycin was added to a final concentration of 1 µg/ml to start selection. Cells were maintained in RPMI (with 10% FBS, 0.5% pen/strep and 1 µg/ml puromycin) for approximately two weeks to select for infected pools. Cells were used for FISH assay as stated above.

**Western blot analysis.** After a single 1× phosphate-buffered saline (PBS) wash, the cells were immediately suspended in five pellet volumes of sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer (125 mM Tris [pH 6.80], 10% glycerol, 2% SDS, 5% 2-mercaptoethanol, 0.05% bromophenol blue) and boiled for 20 min to shear the genomic DNA. Protein concentrations were measured using a Bio-Rad protein assay kit (catalog no. 500-0006) according to the manufacturer's instructions. Equal weights of cell lysates were subjected to SDS-PAGE electrophoresis and transferred to nitrocellulose membranes. The blots were blocked for 30 min in a blocking buffer (0.05 M Tris, 0.138 M NaCl, 0.0027 M KCl [pH 8.0], and 0.1% Tween 20 at pH 8.0) containing 5% Bio-Rad blotting grade blocker nonfat dry milk (catalog no. 170-6404) and 1% FBS and then incubated with the primary antibody (in blocking buffer) overnight at 4°C. The blots were washed four times with PBS, incubated with the
secondary antibody for 1 h, washed four more times with PBS, and then analyzed using a Li-Cor Odyssey infrared imaging system. Primary antibodies used were mouse anti-Zta (Argene, catalog no. 11-007), mouse anti-Rta (Argene, catalog no. 11-008), mouse anti-BMRF1 (Capricorn, catalog no. EBV-018-48180), and goat anti-Actin (Santa Cruz Biotechnology, Inc., catalog no. sc-1615). The secondary antibodies used were goat anti-mouse (Li-Cor/Odyssey, catalog no. 926-32220) and donkey anti-goat (Li-Cor/Odyssey, catalog no. 926-32224).

**Akata and JY RNA sequencing.** RNA samples were poly(A)-selected or ribodepleted and libraries were prepared using the Illumina ScriptSeq v2 or TruSeq RNA sample preparation protocols (Cat. #. SSV21106 and RS-122-2101). Poly(A) selection was carried out on whole cell RNA or cytoplasmic RNA fractions prior to sequencing, while nuclear RNA fractions did not undergo poly(A) selection prior to sequencing. Two \( \times 100 \)-base paired-end sequencing was performed using an Illumina HiSeq instrument.

**Akata and JY cell line RNA-seq data analysis.** RNA-Seq reads were aligned to a human (hg19 assembly) plus EBV genome ((B95-8-Raji (National Center for Biotechnology Information [NCBI] accession number NC_007605)) index using Novoalign. Splice junctions were identified using TopHat (17) run on a human plus EBV genome Bowtie index. RPKM (reads per kilobase of exon per million mapped reads) calculations and genome coverage files (wiggle files) were generated using SAMMate (19). The Integrative Genomics Viewer (IGV) (14) was used to visualize sequence alignments, genomic annotations, and splice junctions. In the cytoplasmic fractions, the percentage of EBV or human reads was calculated by dividing the total number of EBV or human mapped reads by the sum of EBV and human mapped reads and then
multiplying that value by 100. The numbers of reads mapping to exonic, intronic, and intragenic regions was determined using SAMMate with annotation files containing the coordinates of all human exons, introns, and intragenic blocks. The percentage of reads in each category was determined by dividing the exonic, intronic, or intragenic region mapped read numbers by the total number of mapped reads and multiplying by 100.

Ingenuity pathway analysis (IPA; Ingenuity Systems, Inc.) was performed by inputting all cellular genes showing a >4-fold increase or decrease in expression following BCR activation. Input values were the log2 of relative expression.

**Antisense LMP2 GapmeR knockdown/RNA-sequencing.** Customized GapmeRs targeting aLMP2 and negative control GapmeRs were purchased from Exiqon. GapmeRs were transfected into Akata cells by Amaxa electroporation using Nucleofector Solution R (Lonza, Cat. #VCA-1001) and program G-16 on a Nucleofector II device. Akata cells were placed in antibiotic-free RPMI media two days before electroporation. For each transfection, 2x10⁶ cells in 100 µl solution R were electroporated with 3 µl negative control or aLMP2 GapmeR #1 or aLMP2 GapmeR #1-3 together (200pmole/µl) and transferred to a 6-well plate with 1.5 ml media per well. Transfections were performed in triplicate for all samples. 24 hours later, 1.5ml fresh RPMI medium and anti-human IgG (final concentration of 10 µg/ml) was added to each well to induce EBV lytic reactivation. Cells were harvested for RNA isolation 24 hours post-induction. cDNA libraries were prepared from poly(A)-selected RNA using the strand-specific Illumina TruSeq protocol and sequenced on an Illumina HiSeq 2000 instrument. Triplicate samples were multiplexed per sequencing lane.
**RNA-Seq data analysis for aLMP2 GapmeR knockdown.** Deposition of RNA-seq data to NCBI Gene Expression Omnibus is in process. FASTQ files were aligned to human reference genome hg19 and the Akata EBV genome (Inverted) using the sequence aligners STAR and TopHat. EBV reads and junctions were loaded on Integrative Genomics Viewer (IGV) for visualization. The cellular and EBV gene expression was quantified by RSEM and EBSeq. Cellular pathway analysis was evaluated on Qiagen’s Ingenuity Pathway Analysis (IPA).

### 2.4 Results

**Isolation and characterization of alternatively spliced LMP2A transcripts.**

Strand-specific time-course RNA-seq junction data demonstrated that alternatively spliced transcripts were expressed at 24-hours post-induction, whereas canonical wt-LMP2A junctions were detected at 4-hours post-induction (Fig. 2.1A and Fig. 2.1B). This timing of alternative splicing was also verified by time-course qRT-PCR for LMP2A 1-6 and 1-7 transcripts (Fig. 2.1B). Using qRT-PCR we found that these alternatively spliced rightward transcripts were localized in the cytoplasm (Fig. 2.4). This is similar to the localization of the canonical wt-LMP2A transcript (Fig. 2.4). The CPC (Coding Potential Calculator (39)) predicted that the LMP2A 1-6 and 1-7 transcripts would each have a large coverage of the predicted ORF, which is a good indicator of good ORF quality (Fig. 2.4). Furthermore, they contain the same 5’ ORF end as the wild-type LMP2A, which suggests that they are very likely to be translated. These alternatively spliced transcripts cause a frameshift in the second exon of each transcript (corresponding to wild type exons 6 or 7) that generates a stop codon, which will generate a truncated protein (Fig.
Through the course of constructing cDNA expression vectors for the exons 1-6 and exons 1-7 splice variants, our PCR amplification revealed 18 unique alternatively spliced LMP2A transcripts (Fig. 2.3). Since only one LMP2A spliced transcript has been previously annotated, this significant level of diversity was surprising.

Similar to late EBV genes, expression of the LMP2A 1-6 and 1-7 isoforms were inhibited by phosphonoacetic acid (PAA) with a 90% and 99% decrease, respectively (Fig. 2.7). PAA inhibits the synthesis of herpes virus DNA through the virus-specific DNA polymerase. This raises the possibility that the alternatively spliced LMP2A transcripts are expressed from the newly synthesized genome after viral replication. Using LMP2A 1-6 and 1-7 expression vectors, the LMP2A 1-6 and 1-7 truncated proteins were detected through Western Blot analysis and were found to migrate at 30 and 25 kDa, respectively, while the wild-type LMP2A was detected at 50kDa (Fig. 2.6B). To investigate where LMP2A 1-6 and 1-7 proteins were localized in the cell, immunofluorescence microscopy was used to detect expression of LMP2A 1-6 and 1-7 in HeLa cells. Both alternatively spliced LMP2A proteins were localized in the cytoplasm (Fig. 2.8). During latency the signaling domain of LMP2A is constitutively active through phosphorylation. The signaling domain of the truncated LMP2A proteins could also be actively phosphorylated, which could suggest the role they play in viral replication.

**Strand-specific RNA-seq analysis reveals an antisense IncRNA in the LMP2 gene locus.** Changes in the transcriptome of EBV during reactivation were analyzed by directional sequencing. Surprisingly, coverage was observed in the antisense strand of the LMP2A gene (aLMP2) following BCR activation in Akata cells (Fig. 2.9A). The
complete sequence of aLMP2 was obtained by 5’ and 3’ RACE (Rapid Amplification of cDNA Ends), which indicated that the transcript was approximately 1.8kb. The RACE/sequencing assay uses a oligo(dT) primer, however, our RNA-sequencing data reveals that the aLMP2 transcript may not be polyadenylated. This suggests that the aLMP2 transcript may be much longer (Fig. 2.9B). Using the Coding Potential Calculator, aLMP2 was classified as a non-coding transcript due to the presence of only a short, poor quality open reading frame (Fig. 2.5) (8). As a result we considered the aLMP2 to be a long non-coding RNA (lncRNA).

We next verified expression of the aLMP2 transcript with a strand-specific qRT-PCR (ss-qRT-PCR) technique (40). In this assay the cDNA primer sequence was modified to detect a change in the post-amplification melting curve, and to distinguish between primer-initiated transcripts and nonspecifically primed cDNA. The aLMP2 expression levels increase 20-fold in Akata cells 24-hours post-induction through ss-qRT-PCR analysis whereas little expression of the aLMP2 transcript was detected in type III latency (JY, Jijoye, and X50-7), EBV-expressing AGS gastric carcinoma, and nasopharyngeal carcinoma (C666-1) cell lines (Fig.2.9C). Time-course directional RNA-sequencing of the LMP2 locus revealed the time of aLMP2 expression to be 24-hours post-induction. The timing of aLMP2 expression was also confirmed by ss-qRT-PCR (Fig. 2.10). This suggests that the aLMP2 transcript is a late lytic transcript.

In contrast to the protein coding wt-LMP2A transcript, which is exported to the cytoplasm, we found that the aLMP2 was localized in the nucleus. Following reactivation, Akata RNAs from both cytoplasmic and nuclear fractions were isolated and the nuclear-to-cytoplasmic ratio of aLMP2 was determined to be >40:1 (Fig. 2.11A). Utilizing the
Stellaris FISH (Fluorescent in situ Hybridization) technology, which uses multiple probes to identify an RNA species, the aLMP2 RNA transcript was detected 24-hours post-induction. The Stellaris FISH probes also demonstrated that the aLMP2 transcript was restricted to the nuclear compartment (Fig. 2.11B). Similar to late EBV genes and the LMP2 alternatively spliced transcripts, expression of the aLMP2 was inhibited 16-fold by PAA (Fig. 2.13). This suggests that aLMP2 may be play role in EBV replication.

**Knockdown of the aLMP2 decreases the magnitude of EBV lytic reactivation.**

The high expression and nuclear confinement of aLMP2 during reactivation indicates that it may play a role in EBV gene regulation and/or replication. We set out to investigate the role aLMP2 IncRNA plays during EBV lytic reactivation by using GapmeRs. GapmeRs are LNA-DNA antisense oligonucleotides that degrade their RNA targets through an RNase H dependent degradation mechanism. To knockdown aLMP2 expression, three aLMP2 GapmeRs (aG1, aG2, and aG3) were designed to target the start, middle, and end of the aLMP2 IncRNA (Fig. 2.15B). Akata cells were transfected in triplicate with a control GapmeR, aG1, and a combination of aG1, aG2 and aG3. 24 hours later transfected cells were treated with an anti-IgG antibody to induce reactivation. Transfected Akata cells were harvested 24 hours post induction.

The expression of the aLMP2 IncRNA following GapmeR-mediated-knockdown was quantified by ss-qRT-PCR. Upon knockdown, the aLMP2 IncRNA decreased 4.5-fold (Fig. 2.14A and Fig. 2.15B). Each aLMP2 knockdown sample along with the controls was sequenced through strand-specific Illumina-sequencing. Each RNA-seq data set for each condition (uninduced-control GapmeR, induced-control GapmeR, induced-aG1, and induced-aG123) were aligned against hg19 and EBV genomes using the STAR
(Spliced Transcripts Alignment to a Reference) aligner (36). The expression was quantified using RSEM (RNA-Seq by Expectation-Maximization) and differential expression between all four RNA-Seq experiments were performed using EBSeq (empirical Bayes hierarchical model) (37, 38). Upon aLMP2 knockdown, 76 EBV genes were significantly reduced. In the absence of aLMP2 the degree of EBV lytic activation following BCR activation is reduced (Fig. 2.14B and Fig. 2.15A). This supports the idea that aLMP2 plays an important role in EBV lytic replication phase.

To further investigate aLMP2A’s role in EBV replication following BCR activation, we examined BCR-associated cellular genes. Twenty-six BCR-associated genes were regulated upon aLMP2 knockdown including JUND, JUNB, LYN, SYK, and SMAD3. Genes associated with c-Myc and p53 pathways were also affected (Fig. 2.16).

**Identification of a novel circular RNA at the LMP2 locus.** Using a circular RNA-finding algorithm we investigated ribo-depleted EBV RNA-seq data and found predicted circular RNA reads in the sense strand of the LMP2 locus (Fig. 2.2). The circular LMP2 (cLMP2) transcript back-splices from LMP2 exon 8 to exon 2. The cLMP2 transcript was verified by RNase R, which degrades linear RNA, digestion of total RNA from Akata cells 24-hours post-induction. Convergent and divergent primers were created for the ActinB and LMP2 loci, respectively and amplified using PCR. Following RNase R digestion, cLMP2 RNAs were still present while the ACTB RNAs were degraded. The cLMP2 PCR products were sequenced to confirm the back-splicing from LMP2 exon 8 to exon 2. Although the majority of introns were spliced out of the cLMP2 PCR sequences, some PCR fragments still contained introns (Fig. 2.17). This
suggests that in some cases, the back-splicing from exon 8 to exon 2 occurs first followed by internal splicing.

**Characterizing the cLMP2 transcript.** Primers spanning from exon 8 to exon 2 were used to distinguish expression of cLMP2 vs wild-type LMP2A. Following BCR reactivation, qRT-PCR showed an increase of cLMP2 expression in Akata and Mutu I of >300 and >30-fold, respectively. No cLMP2 expression was found in EBV-expressing AGS, C666-1, JY, Jijoye, and X50-7 cell lines implying the cLMP2 transcript is expressed during the lytic cycle (Fig. 2.18). The timing of cLMP2 expression was also verified at 24-hours post-induction by ss-qRT-PCR. This suggests that the cLMP2 transcript is a late lytic transcript (Fig. 2.10).

Following reactivation, Akata RNAs from both cytoplasmic and nuclear fractions were isolated and the cytoplasmic-to-nuclear ratio was determined by qRT-PCR. The cytoplasmic-to-nuclear ratio of cLMP2 was >30:1, which indicted that the cLMP2 is exported into cytoplasm (Fig. 2.4). This trend is similar to the majority of cellular circular RNAs, which are also exported into the cytoplasm.

Since the cLMP2A sequences are contained in its entirety in both LMP2A and LMP2B, specifically detecting the cLMP2A RNA transcript in Akata post-induction was somewhat problematic. We created an expression vector using sequence from LMP2 intron 1 to intron 8. We also created a cLMP2 splice acceptor (SA) mutant where the splice acceptor at the intron 1/exon 2 junction was mutated to prevent circularization (Fig. 2.19).
The localization of the cLMP2 transcript was detected using Stellaris FISH probes. Since the sequences of the LMP2A, LMP2B and cLMP2 transcripts are similar we could not use cells expressing these transcript. The wild-type cLMP2 and control vectors were retrovirally transduced into an EBV-negative B-cell line, DG75. Circular LMP2 expression levels in the wild-type transductancts were determined to be reasonable. The cLMP2 FISH probes were detected in the cytoplasm (Fig. 2.20A and Fig. 2.20B).

Both the wild-type and mutant retroviral expression vectors were transduced in the Mutu I EBV+ cell line, and the circularization from exon 8 to exon 2 was assessed by qRT-PCR. The expression of the cLMP2 transcript in cells transfected with the wild-type cLMP2 vector was increased 100-fold, while cells transfected with the SA mutant cLMP2 vector expressed levels similar to the vector control. This suggests that the wild-type cLMP2 vector can successfully circularize, while the SA mutant cLMP2 vector does not. RNA was isolated from the Mutu I cells transduced with each of these vectors (control, wt-cLMP2, and mut-cLMP2) and was then sequenced using RNA-sequencing to examine which potential cellular and EBV genes are regulated by the cLMP2 transcripts (Fig. 2.21A).

2.5 Discussion

Further investigation of the transcriptome of EBV with strand-specific sequencing has revealed that the LMP2 gene is more complicated than previously described. Using bidirectional-sequencing has proven to be a valuable tool in identifying two novel LMP2 lncRNAs, an antisense LMP2 and a circular LMP2. Mammalian lncRNAs have been involved in each stage of protein biosynthesis by regulating gene expression epigenetically, initiating transcription, and modulating translation.
Two known circular RNAs that are localized in the cytoplasm have shown to act as microRNA sponges. Although circLMP2 is exported into the cytoplasm, it is unlikely that it regulates microRNAs. CircLMP2 lacks multiple binding sites that recognize a specific microRNA seed region. Transient transfection of the circLMP2 vector into Mutu I cells lead to an increase of 18 late EBV genes, while the circLMP2 SA mutant vector did not. The EBV late genes are involved in virion assembly which is followed by virion release through exocytosis via the Golgi pathway. CircLMP2 localization and expression as a late gene may indicate its function in facilitating virion assembly and release.

The aLMP2 lncRNA may be implicated in initiating transcription because it is expressed during reactivation and sequestered in the nucleus. An important question to be addressed is how the aLMP2 transcript increases EBV lytic activation. Upon aLMP2 knockdown expression all EBV genes are reduced, most notably is the LMP2B RNA. The LMP2B protein has been shown to increase the magnitude of EBV lytic replication by blocking LMP2A signaling. The role of LMP2A is to maintain latency by inhibiting BCR signaling. Blocking LMP2A signaling is crucial in EBV switching from latency to lytic form. The aLMP2 lncRNA could play a role in promoting LMP2B expression to increase EBV switching.
Figure 2.1. (A) Time-course strand-specific read coverage across the LMP2 locus following B-cell receptor activation in Akata cells. (B) Time-course strand-specific LMP2 sense, antisense, wild-type splicing and alternative splicing reads.
Figure 2.2. (A) Strand-specific coverage of LMP2 locus in JY cell-line and Akata cells 24-hours post-induction. (B) Junction reads of JY cell-line and Akata cells 24-hours post-induction. Circular evidence in LMP2 locus discovered with circFinder.
Figure 2.3. PCR evidence of alternative splicing between LMP2 exon1-exon6 and LMP2 exon1-exon7 24-hours post induction in Akata cells.
Figure 2.4. Cytoplasmic to nuclear ratio of wild-type LMP2 1-2, LMP2A 1-6, LMP2A 1-7, cLMP2, aLMP2, GAPDH, and Kcnq4lotl 24-hour post-induction in Akata cells. Normalized to Kcnq4lotl.

Coded Potential

LMP2A Antisense

<table>
<thead>
<tr>
<th>Feature</th>
<th>Hit Num</th>
<th>Hit Score</th>
<th>Frame Score</th>
<th>Coverage</th>
<th>Log-Odds Score</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOMOLOGY</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>13.27%</td>
<td>26.60</td>
<td>Partial</td>
</tr>
<tr>
<td>ORF FRAMEFINDER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMP2A 1-6</td>
<td>3</td>
<td>70.31325</td>
<td>18806</td>
<td>99.65%</td>
<td>118.22</td>
<td>Full</td>
</tr>
<tr>
<td>LMP2A 1-7</td>
<td>3</td>
<td>56.045518</td>
<td>46229</td>
<td>98.97%</td>
<td>55.56</td>
<td>Partial</td>
</tr>
</tbody>
</table>

Figure 2.5. Coding Potential Calculator assess the protein-coding potential of the aLMP2, circLMP2, LMP2A 1-6, and LMP2A 1-7 transcripts.
Figure 2.6. (A) Illustration of predicted LMP2 1-6 and 1-7 truncated protein. (B) Western blot of endogenic LMP2A expression in JY and Akata treated with or without IgG (24-hours post-induction). Transient transfection of LMP2A pMSCV-puro vectors (control, LMP2A-wt, LMP2A 1-6, and LMP2A 1-7) into HEK293. Wild-type LMP2A was detected at ~50kDa, while the truncated proteins generated from LMP2A 1-6 and 1-7 vectors were detected at 30kDa and 25 kDa, respectively.
Figure 2.7. Quantitative RT-PCR of LMP2 transcripts 24-hours post-BCR activation with or without phosphonoacetic acid.

Figure 2.8. Immunofluorescence of transient transfection of FLAG-tagged LMP2A cDNA, LMP2A 1-6, and LMP2A 1-7 vector constructs in HeLa cells.
Figure 2.9. (A) Strand specific coverage of the aLMP2 lncRNA 24-hours post-induction in Akata cells. (B) 5’ and 3’ RACE identification of start and end sites. RACE products were cloned and sequenced. Blue arrows indicate 5’ RACE primers and green arrows indicate 3’ RACE primers. (C) Strand-specific qRT-PCR of aLMP2 expression in AGS-EBV positive, C666-1, JY, Jijoye, X50-1, Akata uninduced, and Akata 24-hours post-induction cells lines. Normalized to ACTB.
Figure 2.10. Time-course qRT-PCR of aLMP2, cLMP2, and selected immediate early, early, late, and latent EBV genes.
Figure 2.11. (A) Strand-specific qRT-PCR of aLMP2 expression 24-hours post-induction. Normalized to ActinB. (B) IF-FISH assay of aLMP2 RNA and BMRF1 protein 24-hours post-induction in Akata cells.

Figure 2.12. Fold change of aLMP2 expression in the presence of phosphonoacetic acid using strand-specific qRT-PCR.
Figure 2.13. (A) Strand-specific quantitative RT-PCR of aLMP2 expression during GapmeR knockdown with or without IgG. (B) Heatmap of aLMP2 regulation on lytic replication.
Figure 2.14. Antisense LMP2 RNA GapmeR-mediated knockdown of uninduced Akata transfected with control GapmeR, induced Akata (24-hours post-induction) transfected with control GapmeR, and induced Akata (24-hours post-induction) transfected with the
aLMP2 GapmeRs (aG1, aG2, and aG3) in Akata cells. (A) Coverage of whole EBV transcriptome. (B) Coverage of LMP2 locus containing both LMP2A and LMP2B genes.
Figure 2.15. Induced to uninduced fold change heatmap of BCR, MYC, and p53 associated genes during aLMP2 RNA GapmeR-mediated knockdown. Green arrows illustrate genes up-regulated by aLMP2 and red arrows illustrate gene down-regulated by aLMP2.
Figure 2.16. RNA from Akata 24 hours post-induction treated with and without RNase R (A) PCR of the ACTB gene displaying linear degradation when treated with RNase R. Illustration of convergent primers across the ACTB gene in green (B) Presence of a circular transcript in the LMP2 gene when treated with RNase R. Illustration of divergent primers across the LMP2 exon 8 to exon 2 backsplice in orange and green.

Figure 2.17. Quantitative RT-PCR of cLMP2 expression across multiple EBV-positive cell lines. Circular LMP2 expressed 24 hours post-induction.
Figure 2.18. Illustration of cLMP2 and cLMP2 SA mutant vector constructs. The cLMP2 vector construct used sequence from intron 1 to intron 8 and can generate linear and circular LMP2. The cLMP2 SA mutant vector used the same sequences as above but the splice acceptor next to exon 2 was mutated to inhibit circularization.
Figure 2.19. DG75 retroviral stable cell line expressing control and cLMP2. (A) Quantitative RT-PCR validation of cLMP2 expression. (B) FISH staining the cLMP2 RNA in control and cLMP2.
Figure 2.20. Nucleofection in Mutu I of control, cLMP2, and cLMP2 SA mutant vectors in triplicates. (A) Quantitative RT-PCR validation of cLMP2 expression in Mutu I cells.
(B) Heatmap illustrating EBV genes induced during nucleofection of cLMP2 in Mutu I cells. (C) Coverage of LMP2 locus containing both LMP2A and LMP2B.
LIST OF REFERENCES


BIOGRAPHY

Monica Concha was born in Lima, Peru on October 4th, 1986. She received her Bachelor of Science degree in Microbiology at Louisiana State University in August 2008. Monica was enrolled in the Biomedical Sciences Ph.D. program in August 2009 and joined the laboratory of Dr. Erik K. Flemington in pursuit of a Ph.D.