Extensive host shutoff and transcriptome remodeling during Epstein-Barr virus lytic replication

By

Alejandro Casco

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Cancer Biology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2024

Date of final oral examination: 05/09/2024

This dissertation is approved by the following members of the Final Oral Committee:

Elaine T. Alarid, Professor, Oncology
Eric C. Johannsen, Professor, Oncology and Medicine
Donna M. Neuman, Associate Professor, Ophthalmology and Visual Sciences
Nathan M. Sherer, Professor, Oncology
William M. Sugden, Professor, Oncology

Acknowledgements

The completion of this dissertation would not have been possible without the guidance, support, and inspiration of countless individuals. First and foremost, I want to thank my advisor, Eric. Your unwavering support, insightful guidance, and belief in my potential have carried me through. Eric, your analogies – like "a few hours in the library can save you weeks in the lab," and the reminder that "Egyptians built small pyramids first" – will stay with me. You fostered my growth as a scientist by encouraging me to embrace thoughtful planning, to start with smaller, focused experiments, and by providing the right balance of independence and steady support.

To everyone in the Johannsen lab: thank you for everything! You've all made this a truly special experience, and I'll cherish the memories we've built together. Makoto, I couldn't have asked for a better lab mentor. For a while, it was just the two of us, and I'm so glad I didn't drive you crazy with my endless questions, but you were always so patient and helpful (even if it meant putting your own work on hold). You've made me a better scientist – thank you! Quincy, Emilie, and Mariah – where were you guys all my life? Even though we've only known each other for less than a year, it feels like family already. I'm going to miss your energy and humor which have transformed a great work environment into a place I genuinely look forward to every day. I know you all have such bright futures ahead – keep shining!

I want to thank thank those outside the Johannsen lab whose support has positively influenced me as an individual and scientist. Bill Sugden, your wisdom and advice have been invaluable. To everyone in the Sugden lab, thank you; your support made a world of difference. Mitch, your bioinformatics expertise and incredible experimental support were essential. Thejas, our coffee walks, brainstorm sessions, and workouts were a lifeline. Rebecca and Vish, thank you for the salsa nights, the laughs, and the memories. Madilyn, Beniah, and everyone else –

your humor and camaraderie brightened my days. To Dr. Shannon Kenney, thank you for the collaborative lab meetings and stimulating discussions that greatly enriched my research experience. I extend my thanks to the rest of my committee– Dr. Elaine Alarid, Dr. Donna Neumann, and Dr. Nate Sherer – for their constructive feedback, insights, and expertise. Their guidance has significantly strengthened my work and shaped my professional development. I also thank them for their availability and support in pursuing my next career steps.

Amanda, Katie, Heron, and Santina – who knew when we started that I'd find such a kind and supportive group? Thanks for sharing the good times, the bad times, and for always lifting me up. I'm excited to see where we all end up! Amanda, extra thanks – our chats and laughter kept me sane, and I'll miss them so much.

To the McArdle staff and admin – thank you for your patience throughout my time here! I enjoyed getting to know so many of you through my ordering duties – from Randy and Patty's constant support to Brad's daily cheer and Deb's warmth (enjoy your travels!). And Chris, thanks for the bond over soccer – that beer after a particularly tough World Cup match was a lifesaver!

I'm deeply grateful for those outside the McArdle and scientific community who've supported me along the way. My family has been my bedrock throughout this journey. Mom and Dad, thank you for nurturing my curiosity and love of questioning from a young age – it ultimately led me to this research I adore. Mom, your love and persistence in staying connected were vital, even when I got distracted with my work. I'm so glad I always picked up the phone – it kept us close. Dad, your support means everything. Your hard work is an endless inspiration, and much of my success comes from striving to follow your example. To my brother and sister – you two are the best! Thank you for your amazing personalities, the visits, making my time home so special, and the silly humor that constantly brightens my days during this journey.

To Merrisa's family, my deepest thanks. Naomi and Tim, your support has been incredible from the moment Merrisa and I moved across the country. Your visits to Madison, our visits home, and time at your lake house are times I'll always treasure. Kaylyn and James, our times together are a true escape – thank you for the friendship. And Emmett, watching you grow up has been a joy, filled with so much fun and laughter.

To my closest friends – Fate, Felipe, and Truman – I couldn't have done this without you. I'm so grateful we've stayed connected all these years. You guys always keep me grounded and energized. Felipe and Amanda, our adventures are always a highlight. Truman, your energy and zest for life are contagious. Fate, you're a major reason I'm where I am today – from encouraging me to find a path I love, to our college years of hard work and conversations that shaped my thinking. Thank you all for being my anchors throughout this journey.

To my dear partner, Merrisa, my gratitude for you knows no bounds. From day one, your boundless love, encouragement, and support through every challenge has been my greatest source of strength throughout this journey. Our workouts together aren't just about fitness; they're a shared sanctuary. The adventures you plan are my escape – they remind me there's a world beyond work, thanks to you. Whether it's a well-timed joke to lighten my mood, your empathy when I'm down, or just the way you know how to ground me… Every small, joyful moment in my life exists because of you. Thank you for everything.

Abstract

Epstein-Barr virus is a human herpesvirus that is associated with multiple malignancies and autoimmune disease. EBV's ability to establish life-long latent infection in memory B cells and reactivate from this reservoir asymptomatically is crucial to a strategy that allows it to chronically infect about ~95% of adults. A significant barrier to the study of EBV reactivation is that EBV enters lytic replication inefficiently and asynchronously. To address this problem, we developed a dual-fluorescent lytic reporter (DFLR) EBV that expresses a green fluorescent protein during early lytic replication and a red fluorescent protein during late lytic replication. Using lymphoblastoid cell lines transformed by DFLR EBV as a model for EBV reactivation from B cells, we observed extensive remodeling of the host cell transcriptome during lytic cycle progression. This entailed widespread shutoff of host gene expression and disruption of mRNA processing, including splicing defects and transcriptional readthrough beyond typical polyadenylation sites. Unexpectedly, host shutoff remained extensive even in cells infected with DFLR EBV deleted for the BGLF5 nuclease, previously considered the primary driver of this process. These findings implicate BGLF5-independent mechanisms as the primary drivers of host transcriptome remodeling during EBV lytic replication.

DFLR EBV offers an efficient and highly physiologic platform to allow study of cells in early or late phases of EBV replication or to remove spontaneously lytic cells from latently infected cell populations. While further research is needed to determine the mechanisms responsible for BGLF5-independent host shutoff, this work lays the groundwork for their elucidation, ultimately advancing our understanding of how EBV manipulates the host-cell environment to ensure successful replication.

Table of Contents

ACKNOWLEDGEMENTS ..... **I**

ABSTRACT ..... **IV**

TABLE OF CONTENTS ..... **V**

LIST OF FIGURES ..... **VIII**

LIST OF TABLES ..... **X**

LIST OF ABBREVIATIONS ..... **XI**

CHAPTER 1 INTRODUCTION **1**

The Role of EBV Lytic Replication in Tumorigenesis 2
The Life Cycle of EBV 3
Insights into EBV Lytic Replication Virus-Host Interactions 6
Concluding Remarks 6

CHAPTER 2 UNDERSTANDING HOST SHUTOFF STRATEGIES IN EBV **8**

Host Shutoff Mechanisms Are Highly Divergent, Even among Herpesviruses 9
Host Shutoff in EBV Infection 10
Selectivity 13
Host Shutoff Escape 15
Secondary HSF Effects May Augment Shutoff 16
Host Genes Targeted by HSF during Lytic Infection 17
HSF Interactions with Viral Gene Expression 18
Non-HSF Viral Proteins Contribute to Impaired Host Gene Expression 19
Technical Barriers to Defining HSF Effects 20

CHAPTER 3 MATERIALS AND METHODS **25**

KEY RESOURCES TABLE 26
RESOURCE AVAILABILITY 31
Data and Code Availability 31
EXPERIMENTAL MODEL DETAILS 31
Cell lines and reagents 31
METHOD DETAILS 31
Plasmids construction 31
Lentivirus production 32
Creation of HEK293 cell lines stably transduced with Z-HT, mGL, or mSI 32
Construction of recombinant EBV BACmids 33
Generation of WT and ΔBGLF5 DFLR EBV LCLs 34
DNA extraction and PCR 35
Microscopy and flow analysis of gp350-stained DFLR LCLs 36
FACS and sample collection for RNA-seq 37
RNA-sequencing 38
QUANTIFICATION AND STATISTICAL ANALYSIS 38
Reference genome and annotation assembly 38

Read alignment and differential expression analysis 39
Calculation of RNA-seq metrics 42
Visualization Tracks 42
Absolute quantification of mRNA molecules per cell 42
Alternative splicing analysis 43
Identification of DoGs and read-in transcripts 44
Data analysis 44
Lytic Raji Burkitt lymphoma RNA-seq data acquisition and analysis 45

CHAPTER 4 INVESTIGATING THE TRANSCRIPTOMIC LANDSCAPE DURING EBV LYTIC REPLICATION AND BGLF5'S ROLE IN HOST SHUTOFF 46

INTRODUCTION 47
RESULTS 49
Development and characterization of dual-fluorescent lytic reporter (DFLR) EBV transformed LCLs. 49
DFLR LCLs can be efficiently sorted into latent, early lytic, and late lytic subpopulations. 53
Antibody staining for gp350 expression is less sensitive and specific than DFLR LCLs for purifying cells in the late lytic phase of EBV replication. 57
EBV lytic replication is associated with marked decrease in almost all host mRNAs. 59
EBV replication is associated with marked qualitative disruption of the host transcriptome. 65
Host shutoff induced by EBV reactivation is predominantly BGLF5 independent. 70

CHAPTER 5 DISCUSSION AND FUTURE DIRECTIONS 79

DISCUSSION 80
EBV induces widespread host shutoff during lytic replication 81
BGLF5-independent mechanisms are the predominant drivers of host shutoff 81
Many host mRNAs expressed during EBV reactivation are dysfunctional 83
Evidence of host and early viral gene transcriptional suppression during the late phase 85
A proteomic and transcriptomic discrepancy 86
DFLR EBV is a versatile tool 87
FUTURE DIRECTIONS 87

REFERENCES 90

APPENDIX I A MACS-COMPATIBLE EBV MUTANT FOR ISOLATION OF LATE LYTIC CELLS 104

INTRODUCTION 105
RESULTS 105
DISCUSSION 107
METHODS 108
Development of late lytic membrane reporter EBV. 108
Indirect immunofluorescence. 109
Magnetic-activated cell sorting (MACS). 109
Immunoblotting 110
Live-cell direct immunofluorescence. 111
Statistics. 111

REFERENCES 112

APPENDIX II PRIMERS AND OLIGONUCLEOTIDES **113**

APPENDIX III A COMPUTATIONAL METHOD FOR QUANTIFYING HERPESVIRUS GENE EXPRESSION **120**

INTRODUCTION 121
RESULTS 123
Conventional (ORF based) RNA-seq read assignment is inaccurate for viral gene quantification but performs well for host genes. 123
Accounting for viral transcript overlap substantially improves estimates of lytic gene expression from RNA-seq data. 126
Identification of EBV genes and gene clusters most affected by the RNA-seq read assignment method 133
Premature polyadenylation may explain the low abundance of BaRF1 seen by RNA-seq. 137
DISCUSSION 141
MATERIALS AND METHODS 145
CAGE-seq data acquisition and analysis. 145
RNA sequencing and alignment. 146
Northern blotting. 146
Conventional RNA-seq analysis. 147
Implementation of the UCDS method. 148
Implementation of the UTS method. 150
Implementation of effective transcript lengths for the UTS$^+$ method. 151
Availability and implementation of UTS/UTS$^+$ in Python. 151
Visualization tracks. 152
Statistical analysis. 152
Data availability. 152
REFERENCES 153

APPENDIX IV EBV ZTA RECRUITS P-TEFB TO PROMOTE RTA-SYNERGY GENE EXPRESSION **155**

INTRODUCTION 156
RESULTS 157
METHODS 162
Cell lines 162
Construction of recombinant EBV BACmids 163
Generation of B95-8 DFLR HEK293.ZHT cells 163
Cell chemical treatments and transfections 163
FACS, sample collection, and RNA-seq 164
Immunoblotting 165
Data analysis 165

List of Figures

Figure 2.1 Canonical and non-canonical mechanisms of EBV-mediated host shutoff. 12

Figure 2.2 Barriers and potential solutions for accurate quantification host shutoff. 24

Figure 3.1 BCR immunoglobulin isotype composition and host gene expression filtering.. 41

Figure 4.1 Establishment and validation of a dual-fluorescent lytic reporter Epstein-Barr virus (DFLR EBV). 51

Figure 4.2 The F-factor present in short-cultured LCLs is lost in longer cultures. 52

Figure 4.3 Early lytic and late lytic populations can be purified from DFLR LCLs induced for EBV replication. 54

Figure 4.4 Profiling of EBV gene expression during early and late lytic replication. 56

Figure 4.5 GP350 staining is sensitive, but not specific for cells in late phase. 58

Figure 4.6 Extensive host shutoff during EBV replication. 61

Figure 4.7 Relationship of mRNA levels during latency or mRNA stability in LCLs with mRNA downregulation during early or late lytic EBV replication in DFLR LCLs. 62

Figure 4.8 Early lytic replication gene expression changes in DFLR LCLs versus Buschle et al. Raji Burkitt lymphoma cells. 64

Figure 4.9 Comprehensive analysis of EBV lytic replication-induced splicing alterations reveals a systemic shift towards host shutoff programs. 66

Figure 4.10 Global disruption of cellular transcription termination during EBV lytic replication. 69

Figure 4.11 BGLF5-KO schematic and validation. 72

Figure 4.12 Extensive BGLF5-independent host shutoff during EBV lytic replication. 75

Figure 4.13 Global dysregulation of mRNA processing in DFLRΔBGLF5 LCLs. 76

Figure 4.14 Relationship of mRNA levels during latency or mRNA stability in LCLs with mRNA downregulation during early or late lytic EBV replication in DFLRΔBGLF5 LCLs. 77

Appendix Figure I-1 tLNGFR expression in lytic induced EBV BILF2p-tLNGFR HEK293 cells reflects late kinetics. 106

Appendix Figure I-2 Enrichment efficiency of EBV BILF2p-tLNGFR HEK293 cells by MACS. 107

Appendix Figure III-1 Herpesvirus gene clusters, which share a polyadenylation signal, result in overlapping transcripts that confound conventional RNA-seq read assignment to genes. 122

Appendix Figure III-2 Conventional RNA-seq analysis produces inaccurate estimates of viral gene expression relative to the gold standard, CAGE-seq. 125

Appendix Figure III-3 Schematic of the Unique TranScript (UTS) and unique coding sequence (UCDS) methods for abundance estimation of overlapping genes. 127

Appendix Figure III-4 Example of improved EBV transcript quantification by the UTS and UCDS methods. 129

Appendix Figure III-5 Read depth plots as a function of transcript position for RNA-seq data. 130

Appendix Figure III-6 Global comparison of EBV gene transcript quantification of RNA-seq data by various methodologies relative to the gold-standard, CAGE-seq. 132

Appendix Figure III-7 Performance of UTS versus conventional methods relative to CAGE-seq across all EBV genes and by gene cluster. 135

Appendix Figure III-8 RNA-seq read depth and CAGE-seq data for BLRF1-BLRF2 and BOLF1-BPLF1 clusters. 136

Appendix Figure III-9 Example of CAGE-seq discrepancy contributing to observed variance between CAGE-seq and RNA-seq. 139

Appendix Figure III-10 Examination of the discrepancy between RNA-seq read depth and CAGE-seq signal strength for the BaRF1 transcript. 140

Appendix Figure IV-1 P-TEFb regulation during early lytic replication. 159

Appendix Figure IV-2 Impact of BRD4 and 7SK snRNP inhibition of EBV lytic reactivation. 160

Appendix Figure IV-3 Zta promotes activation of Rta^{synergy} genes through P-TEFb. 161

Appendix Figure IV-4 Proposed model of enhanced EBV lytic reactivation through redirection of P-TEFb towards Zta-mediated viral transcription. 162

List of Tables

Table 2.1 Diverse herpesvirus genes mediate host shutoff. 10

Appendix Table II-1 List of primer sequences used in PCR assays. 114

Appendix Table II-2 Gene block sequences. 115

Appendix Table III-1Summary statistics of Conventional, UCDS, UTS, UTS+ and RSEM RNA-seq quantification of RNA-seq data compared to CAGE-seq for overlapping genes. 144

Appendix Table III-2 Summary of UCDS and UTS regions used in this study. 148

List of Abbreviations

4HT	4-hydroxytamoxifen
snRNA	Small nuclear RNA
snRNP	Small ribonucleoprotein
AE	Alkaline exonuclease
AEN	Apoptosis enhancing nuclease
ARTDeco	Automatic readthrough transcription detection
ASE	Alternative splicing event
ATAC-seq	Assay for transposase-accessible chromatin sequencing
BAC	Bacterial artificial chromosome
BAM	Binary alignment map
BCR	B cell receptor
BHI	Brain heart infusion
BL	Burkitt lymphoma
BRD4	Bromodomain-containing protein 4
BRLF1	BamHI rightward open reading frame 1
BSA	Bovine serum albumin
BXLF1	BamHI X fragment leftward open reading frame 1
BZLF1	BamHI Z-fragment leftward open reading frame 1
C19ORF66	Chromosome 19 open reading frame 66
CAGE-seq	Cap analysis gene expression seq sequencing
CASTOR1	Cytosolic arginine sensor mTORC1 subunit 1
CD1d	Cluster of differentiation 1d
CD2	Cluster of differentiation 2

CD21	Cluster of differentiation 21
CD35	Cluster of differentiation 35
CD8	Cluster of differentiation 8
CDK9	Cyclin-dependent kinase 9
CDS	Coding sequence
CPC2	Coding potential calculator 2
CPSF	Cleavage and polyadenylation specificity factor
DESeq2	Differential expression analysis for RNA-seq 2
DEXSeq	Differential exon usage analysis for RNA-seq
DFLR	Dual-fluorescent lytic reporter
DIG	Digoxigenin
Dis3L2	DIS3 Like 3'-5' Exoribonuclease 2
DNase	Deoxyribonuclease
DoG	Downstream of gene
dsRNA	Double-stranded RNA
EBV	Epstein-Barr virus
EF1a	Eukaryotic translation elongation factor 1 alpha
eGFP	Enhanced green fluorescent protein
EphA2	Ephrin type-A receptor 2
ERCC	External RNA controls consortium
ES	Exon skipping
FACS	Fluorescence-activated cell sorting
FastQC	Fast quality control
FDR	False discovery rate

FP	Fluorescent protein
FPKM	Fragments Per Kilobase per Million mapped fragments
FR	Family of repeats
GADD45B	Growth arrest and DNA damage-inducible 45 beta
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
gB	Glycoprotein B
GCV	Ganciclovir
GDC	Genomic data commons
GENCODE	Gene Annotation by the ENCODE Consortium
GFP	Green fluorescent protein
gH	Glycoprotein H
gL	Glycoprotein L
gM	Glycoprotein M
gN	Glycoprotein N
GO	Gene ontology
gp350/220	Glycoprotein 350/220
gp42	Glycoprotein 42
HEXIM1	Hexamethylene bis-acetamide inducible protein 1
HLA	Human leukocyte antigen
HMBA	Hexamethylene bisacetamide
hNSG	humanized NOD-scid IL2Rγnull
HS	Host shutoff
HSF	Host shutoff factor
HSV	Herpes simplex virus

Hyg^r	Hygromycin resistance
ICP27	Infected cell protein 27
IDBR	Intrinsically disordered binding regions
IDR	intrinsically disordered regions
IFN	Interferon
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IL-6	Interleukin-6
ILF3	Interleukin enhancer binding factor 3
iNKT	Invariant natural killer T cell
IR	Intron retention
IRES	Internal ribosome entry site
ISG	Interferon-stimulated gene
IUPRed2A	Intrinsically unstructured protein predicted 2A
KEGG	Kyoto encyclopedia of genes and genomes
KO	Knockout
KSHV	Kaposi's sarcoma-associated herpesvirus
LARP7	La-related protein 7
LCL	Lymphoblastoid cell line
LEC	Little elongation complex
MA	Log ratio (M) and mean average (A)
MACS	Magnetic-activated cell sorting
MCS	Multiple cloning site
MePCE	Methylphosphate capping enzyme

mGL	mGreenLantern
MHV68	Murine herpesvirus 68
MOMP	Mitochondrial outer membrane depolarization
MPC	Molecules per cell
MSCV	Murine stem cell virus
mSI	mScarlet-I
muSOX	Murine shutoff and exonuclease
Neo^r	Neomycin resistance
NGS	Next-generation sequencing
NLS	Nuclear localization signal
NMD	Nonsense-mediated decay
NMHC-IIA	Non-muscle myosin heavy chain IIA
NRP1	Neuropilin 1
ns	Non-significant
nt	Nucleotide
OAS	2'-5'-oligoadenylate synthetase
OIP5	Opa-interacting protein 5
ORF	Open reading frame
ORF37	Open reading frame 37
OriLyt	Origin of lytic replication
PABPC	PolyA binding protein cytoplasmic 1
PBE	PBS supplemented with BSA and EDTA
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction

PEG	Polyethylene glycol
PFA	Paraformaldehyde
Pfam	Protein families
PIC	Pre-initiation complex
PKR	Protein kinase R
PNPT1	Polyribonucleotide nucleotidyltransferase 1
PolyA	Polyadenylation
PSI	Percent spliced in
PTC	Premature termination codon
P-TEFb	Positive transcription elongation factor b
qPCR	Quantitative polymerase chain reaction
RBP	RNA-binding protein
RFP	Red fluorescent protein
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II
RNAPII	RNA polymerase II
RNase	Ribonuclease
RNase L	Ribonuclease L
RNA-seq	RNA sequencing
RNP	Ribonucleoprotein
RPKM	Reads Per Kilobase per Million mapped fragments
RPM	Rotations per minute
RSEM	RNA-seq by expectation maximization
Rta	R transactivator

SAM	Sequence alignment map
SCID	Severe combined immunodeficiency
SDS-PAGE	Sodum dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Standard error of the mean
siRNA	Small interfering RNA
SLAM-seq	Thiol(SH)-linked alkylation for the metabolic sequencing of RNA
SOX	Shutoff and exonuclease
SRE	SOX-resistance element
STAR	Spliced transcripts alignment to a reference
TBC1D10A	TBC1 domain family member 10A
TBST	Tris-buffered saline with tween
TF	Transcription factor
tLNGFR	Truncated low-affinity nerve growth receptor
TLR	Toll-like receptor
TPM	Transcripts per million
tPT2A	Tandem P2A and T2A
P2A	Picornavirus 2A
T2A	Thosea asigna virus 2A
TSS	Transcription start site
UCDS	Unique coding sequence
UTR	Untranslated region
UTS	Unique transcript
vhs	Virion host shutoff
vPIC	Viral pre-initiation complex

WT	Wild-type
Xrn1	5'-3' exoribonuclease 1
ZHT	Zta fusion to 4HT-responsive estrogen hormone binding domain
Zta	Z transactivator
Δ	Delta symbol representing gene deletion
CRIPSR	Clustered regularly interspaced short palindromic repeats
DAP	Diaminopimelate
DAPI	4',6-diamidino-2-phenylindole
DAVID	Database for annotation, visualization and integrated discovery
Cas9	CRISPR-associated protein 9

Chapter 1

Background and Introduction

Epstein-Barr virus (EBV) is a gammaherpesvirus that infects more than 90% of the adult human population and persists for life. This chronic infection is typically asymptomatic, but occasionally results in disease. EBV is associated with over 300,000 annual cancers worldwide and approximately 1.8% of all cancer-related deaths [1,2]. These cancers include multiple types of lymphoma and carcinomas of the nasopharynx and stomach. EBV is also linked to the development of autoimmune diseases, with strong evidence suggesting a causal role in multiple sclerosis [3].

The Role of EBV Lytic Replication in Tumorigenesis

Although the role of EBV in cancer has primarily been attributed to latent viral infection, accumulating evidence supports a role for lytic replication in tumorigenesis [4]. For example, in humanized NOD/LtSz-scid/IL2Rγnull (hNSG) or severe combined immunodeficiency (SCID) mice with EBV mutants deficient for lytic replication were substantially impaired in their ability to form tumors compared to wild-type EBV [5,6]. The contribution of EBV replication to tumorigenesis has been attributed to horizontal transmission, which expands the reservoir of latently infected cells that can undergo transformation as well as paracrine effects in which cells undergoing EBV replication contribute to the growth and/or survival of latently infected tumor cells.

Further interest in EBV replication derives from a promising strategy to target EBV-positive tumors is a treatment called “lytic induction therapy” [7]. This approach utilizes a two-pronged “shock and kill” mechanism. The initial “shock” involves the administration of latency-reversal agents, which induce reactivation of dormant EBV within tumor cells, driving it into lytic replication. The “kill” phase employs antiviral drugs that selectively target and eradicate cells harboring active replication. Lytic induction therapy capitalizes on the differential EBV

infection rates between peripheral blood B cells (1-50 per million B cells) [8] and EBV-positive tumor cells, which are dependent on latent EBV infection for continued growth and/or survival [9]. This allows for precise targeting of malignant cells while minimizing off-target effects on healthy tissue.

While lytic induction therapy holds promise for treating EBV-positive tumors, it faces challenges that hinder clinical success. Incomplete viral reactivation in tumor cells, the off-target toxicity of latency reversal agents, and the potential for antiviral drug resistance all limit its effectiveness. A major contributor to these hurdles is our incomplete understanding of the mechanisms governing EBV reactivation. Deeper knowledge of this process is crucial to developing more targeted latency reversal agents and designing effective combinatorial therapies that could overcome limitations of lytic induction therapy.

The Life Cycle of EBV

The life cycle of EBV infection within the host organism typically initiates with the exchange of saliva in the oral cavity. According to the prevailing model, transcytosis across the mucosal epithelium of oropharyngeal lymphoid tissue allows saliva-borne viral particles to reach and infect resting naïve B cells [10,11]. These infected B cells become activated, proliferate, and migrate into the B cell follicles of tonsillar lymphoid tissue [12,13]. There, they participate in a germinal center-like reaction in which EBV latency-associated proteins mimic signals normally activated by antigen and T-cell help [14]. This process ultimately leads to the generation of EBV-infected memory B cells. These memory B cells exhibit a restricted latency gene expression profile and do not actively proliferate, allowing EBV to evade immune detection and persist within the long-lived memory B cell pool. Circulating in the peripheral blood, EBV-infected memory B cells eventually return to the oropharyngeal lymphoid tissue [15]. Upon

stimulation, these memory B cells can differentiate into plasma cells, reactivating EBV from latency to enter the lytic replication cycle [16]. This process generates virions with an epithelial cell-specific glycoprotein complex (gH/gL), facilitating infection of the mucosal epithelium from both the basal and apical surfaces [17-19]. Epithelial cell infection is believed to amplify virion production and promote viral transmission. Notably, virions shed from epithelium and released back into saliva possess a B-cell-tropic glycoprotein complex (gH/gL/gp42) [17,20], effectively completing the cycle and initiating a new round of EBV transmission.

At the cellular level, EBV entry is orchestrated by a series of interactions between viral and cellular membrane molecules (reviewed in [19]). Several viral glycoproteins are key players: gp350/220, the gH/gL or gH/gL/gp42 complex, gB, gM/gN, and BMRF2/BDLF2. These proteins engage with distinct host cell receptors depending on the target cell type. B cells are infected as a result of gp350/220 interactions with CD21/CD35 and gH/gL/gp42 interactions with HLA-II molecules. By contrast, epithelial cell infection requires interaction between gH/gL and NRP1, NMHC-IIA, and EphA2. These interactions promote attachment and trigger a conformational change in gB which promotes fusion between the enveloped EBV virion and the plasma membrane. The nucleocapsid then traffics to the nuclear pore complex, likely through exploitation of the host cell's cytoskeleton [21]. The linear EBV genome is then released into the nucleus [22] where it circularizes and establishes a latent infection, particularly in B cells [23].

During latency, EBV expresses a limited number of viral genes, helping it to evade immune detection. This latent state is crucial for the virus's long-term persistence and is involved in the development of several EBV-associated diseases. Different latency programs, characterized by unique patterns of viral latent gene expression, can be established, potentially influencing the behavior and fate of the infected cell [24]. While various factors can

subsequently trigger EBV reactivation from latency into lytic replication, cellular differentiation appears to serve as the principal mediator of this process *in vivo* [16,25].

The EBV lytic replication cycle can be subdivided into three temporal stages of viral gene expression. This process begins with a reactivation event mediated by cellular transcription factors that activate expression of Rta and Zta, the master lytic transcription factors (reviewed in [26]). These factors act as molecular switches, initiating a cascade of events. Early genes, responsible for viral DNA replication and subverting host defenses, are expressed first. Viral DNA amplification occurs in replication compartments that are phase separated from and progressively compact the chromatinized portion of the nucleus [27,28]. Late genes are transcribed from the newly replicated viral DNA within replication compartments [29,30]. These late genes encode structural, packaging components, and egress proteins, essential for virion assembly. Notably, the regulatory mechanisms governing viral gene transcription change dynamically throughout the cycle. Early gene expression mirrors cellular gene regulation, relying on the host cell's pre-initiation complex (PIC) and the existing histone code. In stark contrast, late gene transcription exhibits a unique spatial regulation. It occurs within viral replication compartments and utilizes a virus-encoded PIC specifically recognizing atypical "TATT" motifs present in the promoters of late genes on newly replicated, histone-free DNA [29,31-35]. This early-late mode of transcription is a common feature observed not only in EBV but also in other herpesviruses and some nuclear replicating viruses. However, the evolutionary advantage of maintaining this seemingly redundant system, given the limited viral coding capacity, remains an intriguing question in viral biology.

Insights into EBV Lytic Replication Virus-Host Interactions

Studies investigating EBV lytic replication have revealed that significant disruption of host cell gene expression occurs during this phase. Interestingly, some of these alterations may occur even before the onset of early viral gene expression [36]. During the early stages of EBV lytic replication, the host cell undergoes extensive transcriptional reprogramming, likely driven by the virus to create an environment conducive to viral replication [36-41]. This phase is characterized by a prominent virus-induced host shutoff phenomenon that leads to global inhibition of cellular protein synthesis. The late lytic phase is marked by progressive chromatin compaction and margination to the nuclear periphery, suggesting far-reaching effects on chromatin structure and gene expression [28,35]. Despite the advancement in this field, our comprehensive understanding of the impact EBV lytic replication has on host cell processes remains incomplete. Historically, this limitation stems from technical challenges, especially the difficulty of reactivating the virus from latency.

Concluding Remarks

Studying EBV lytic replication is important for understanding the pathogenesis of EBV-associated diseases and developing effective therapies. Current methodological limitations present significant obstacles to progress in this field, demanding the implementation of novel experimental approaches. This dissertation addresses these challenges by the development of a dual-fluorescent lytic reporter EBV and apples it to studying the EBV-induced host shutoff phenomenon, offering refined insights into virus-host interactions.

In Chapter 2 of this dissertation, I explore the current understanding of host shutoff during EBV lytic replication. This chapter highlights two key points: (1) the field's heavy reliance on overexpression techniques, which may not fully reflect the natural infection, and (2)

the significant gaps in our understanding of this process. Chapter 2 also explores the technical challenges associated with studying EBV lytic replication.

In Chapter 4, supplemented by methodology of Chapter 3, I describe the development and application of a new tool: a dual-fluorescent lytic reporter EBV. This innovation allowed me to conduct the first ever comprehensive and detailed analysis of gene expression throughout EBV's lytic cycle at high resolution. The data revealed an extensive reprogramming of the host transcriptome, encompassing a robust host shutoff event and dysregulation of cellular transcriptional processes. Notably, knocking out the BGLF5 gene, previously thought to be solely responsible for host shutoff, did not prevent this process within the context of the whole virus. This observation implicates alternative mechanisms as the predominant drivers of host shutoff during EBV lytic replication. Chapter 5 discusses these results and their implications, exploring potential BGLF5-independent mechanisms responsible for host shutoff. It also outlines future research directions aimed at uncovering the specific mechanisms involved.

Chapter 2

Understanding Host Shutoff Strategies in EBV

This chapter is composed primarily of a review article entitled “EBV Reactivation from Latency is a Degrading Experience for the Host” by Alejandro Casco and Eric Johannsen published in the journal *Viruses* March 2023.

Host Shutoff Mechanisms Are Highly Divergent, Even among Herpesviruses

Although host shutoff is employed by many viruses from different viral families, the effector mechanisms vary widely. Within the herpesvirus family, the term "virion host shutoff (vhs)" was coined to describe the phenomenon of inhibition of cellular protein synthesis and host mRNA degradation upon infection of alphaherpesviruses such as herpes simplex virus (HSV) [42]. As the name implies, the protein mediating host shutoff (HS) in alphaherpesviruses is packaged within the virion and exerts its effects during initial infection. However, no homologs of the alphaherpesvirus vhs protein are encoded by beta or gammaherpesviruses (Table 2.1). Moreover, the betaherpesvirus family does not exhibit host shutoff during its replication [43]. A significant breakthrough in our understanding of host shutoff was the discovery that the DNA alkaline exonuclease (AE) found in all herpesviruses mediates host shutoff in gammaherpesviruses. Subsequent work demonstrated that in the gammaherpesvirus subfamily, this protein also has RNase activity, but unlike vhs, gammaherpes host shutoff factors (HSFs) are not packaged in virions [44-47]. Thus, in gammaherpesvirus infected cells, including EBV, host shutoff strictly occurs during reactivation from latency.

Table 2.1 Diverse herpesvirus genes mediate host shutoff. Virion host shutoff (vhs) protein homologs are found only in alphaherpesviruses. In gammaherpesviruses, host shutoff is mediated by the alkaline exonuclease (AE) proteins. Collectively, herpesvirus proteins that promote degradation of host mRNAs are referred to as host shutoff factors (HSF) and are shaded green. Note that host shutoff is not observed during betaherpesvirus replication.

	virus	VHS	AE
alpha	HSV	UL41	UL12
	VZV	ORF17	ORF48
beta	CMV	–	UL98
	HHV6	–	U70
	HHV7	–	U70
gamma	EBV	–	BGLF5
	KSHV	–	ORF37/SOX

Host Shutoff in EBV Infection

Shortly after its discovery in KSHV, host shutoff was described for EBV by Rowe and colleagues, who used a CD2/eGFP lytic reporter gene to enrich Akata Burkitt lymphoma cells undergoing EBV lytic replication [48]. They observed a near-global inhibition of total de novo protein synthesis in these cells, including the synthesis of both HLA class I and II molecules. This effect occurred even when viral DNA synthesis was blocked, implicating an early lytic gene product in the effect. They further demonstrated that expression of BGLF5, the EBV homolog of KSHV ORF37 (also called SOX for ShutOff and eXonuclease), was sufficient to mediate host shutoff. Interestingly, both ORF37/SOX and BGLF5 appeared to exert their shutoff effects on mRNA stability [48] despite these proteins being AEs without known RNase activity.

Mechanistic studies of the gammaherpes HSFs BGLF5 and SOX initially suggested their shutoff function was independent of exonuclease activity. The activities were genetically separable: random mutagenesis screens produced BGLF5 and SOX mutants that were selectively impaired for either AE or HS activity [49-51]. The activities are also subcellularly separated.

Processing of viral DNA by AE occurs exclusively in the nucleus and the HSV AE, which lacks HS function, localizes exclusively to the nucleus. By contrast, gammaherpesvirus AEs display nucleocytoplasmic localization [48,49,52]. This subcellular localization appears to play a role in regulating BGLF5/SOX function: AE activity in the nucleus and HS in the cytoplasm. Indeed, restricting SOX localization to the cytoplasm by abolishing its nuclear localization signal did not affect its HS activity [49]. Similarly, fusion of the murine gammaherpesvirus 68 SOX homolog (muSOX) to a nuclear retention signal restricted its localization to the nucleus and abolished HS activity [52]. Interestingly, structural studies established that purified BGLF5 and SOX have intrinsic RNase activity that depends on the same catalytic machinery as AE DNA processing [53,54]. Thus, the genetic separability of AE and HS may be due to the involvement of different residues in substrate recognition (and/or subcellular localization) rather than separability of activities per se. In a landmark study, Covarrubias et al. demonstrated that SOX degrades mRNAs via a two-step process: The initial endonucleolytic cleavage is mediated by SOX, making the two fragments susceptible to degradation by cellular exonucleases (Figure 2.1A) [55,56]. This two-step mechanism was subsequently shown to be conserved across other host shutoff factors, including BGLF5 [57]. This mechanism contrasts with basal cellular mRNA decay that initiates at mRNA ends via poly(A) removal and decapping. The internal cleavage of target mRNAs by BGLF5 and homologs is hypothesized to have a two-fold effect: (1) rapid inactivation of transcripts rendering them incompetent for translation and (2) a mechanism for transcript selectivity. However, the extent of selectivity is incompletely defined.

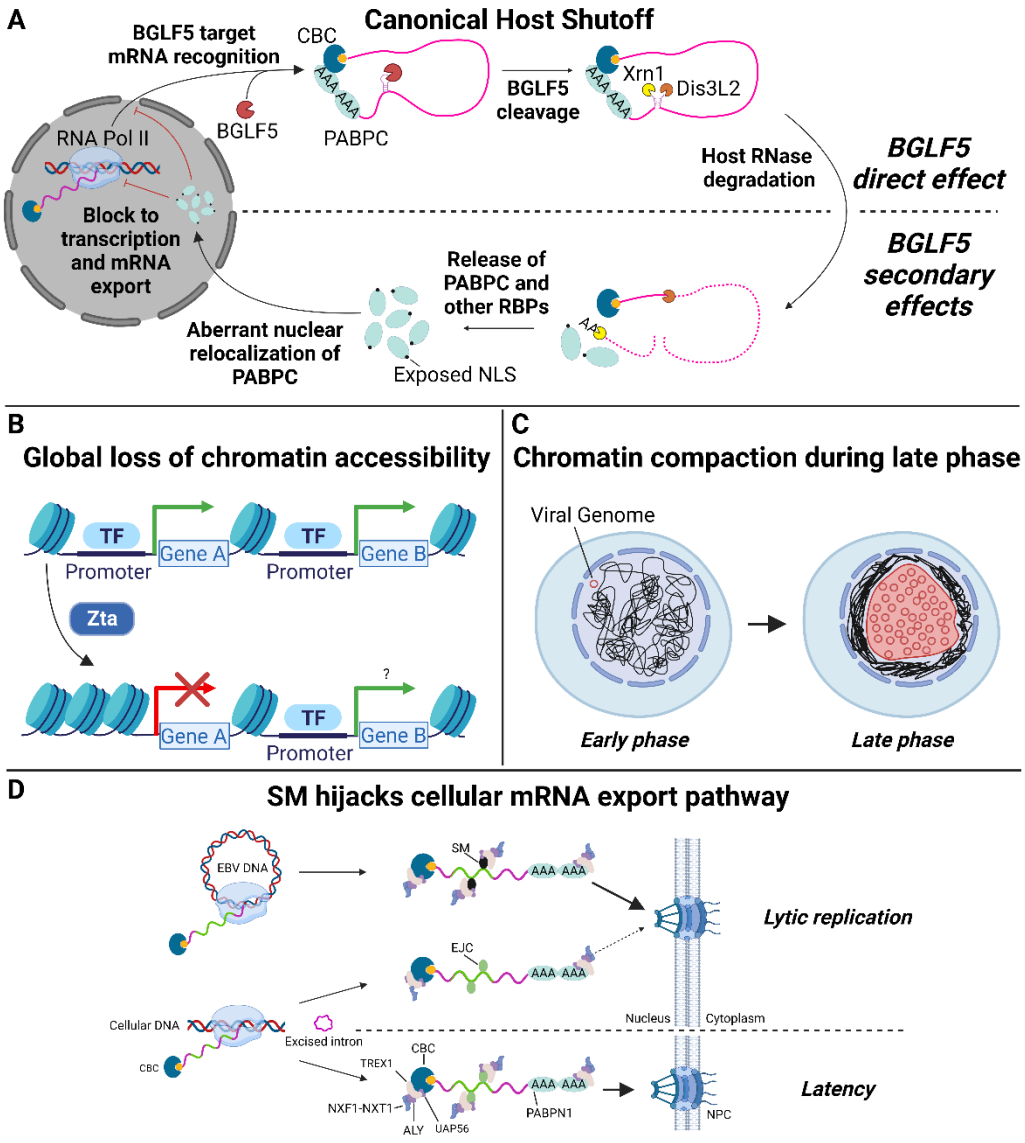


Figure 2.1 Canonical and non-canonical mechanisms of EBV-mediated host shutoff. (A) Canonical EBV host shutoff occurs in the cytoplasm via a two-step process: initial endonucleolytic cleavage by BGLF5 and subsequent degradation by cellular exonucleases (Xrn1, Dis3L2, and likely the exosome complex (not shown)). mRNA degradation releases RNA binding proteins (RBPs), which are recycled into the nucleus. The accumulation of RBPs, in particular the cytoplasmic poly(A)-binding protein (PABPC), induces a state of cellular stress that inhibits host gene expression by inhibiting transcription by RNA polymerase II and possibly by suppressing the nuclear export of mRNAs. (B–D) Non-canonical EBV host shutoff mechanisms include: (B) global loss of host chromatin accessibility induced by the immediate early protein Zta. (C) During the late phase of replication, host chromatin compaction occurs in parallel with the expansion of viral replication compartments that is believed to further disrupt the ability of the host cell to express its gene repertoire. (D) Preferential export of non-spliced viral mRNA mediated by the EBV SM protein. Abbreviations: NLS, nuclear localization signals; CBC, cap-binding complex; TF, transcription factor. Figure created with BioRender.com.

Selectivity

Several early studies indicated that BGLF5 and its homologs have target specificity despite inducing widespread RNA degradation. For example, RNAs transcribed by polymerase II (RNAPII) are degraded, but noncoding RNAs and even untranslatable mRNAs are not. This specificity is thought to arise through targeting of mRNAs associated with the translation machinery [54,56-58]. Indeed, in vitro studies have suggested that BGLF5 and SOX association with RNA is substantially weaker than with DNA and may require additional factors in vivo to efficiently target and degrade RNA [53,54]. Consistent with this suggestion, polysome profiling revealed that SOX co-sediments with the 40S ribosomal subunit [56,57], which would allow preferential targeting of mRNAs and may be important for enhancing binding to RNA and/or the efficiency of cleavage. Future studies examining whether BGLF5 also associates with the 40S subunit and the extent to which BGLF5/SOX mutants selectively defective for HS activity are impaired for 40S association will help clarify the significance of this proposed mechanism for selectivity.

Gammaherpesvirus HSFs also appear to mediate site-specific cleavage of their target transcripts. For example, primary cleavage of reporter mRNAs by HSF (prior to exonuclease degradation) produces fragments of consistent size [56,57]. Furthermore, primary cleavage of different transcripts of similar length results in distinct cleavage intermediates [56]. In addition, cleavage of the same transcript by BGLF5 versus SOX resulted in different intermediates [57]. Thus, despite their shared mechanism, gammaherpesvirus HSFs likely recognize different target sequences in host mRNAs and therefore may preferentially target distinct populations of host mRNAs. Characterization of cleavage intermediates produced by SOX from three reporter mRNAs identified a five-base motif (UGAAG) upstream of the cleavage site that was necessary

but not sufficient for SOX cleavage [56]. In contrast, a larger 200nt sequence containing this motif did confer susceptibility to SOX cleavage. Given that the 200nt sequences from 3 reporter mRNAs were different, this finding supports a requirement for specific mRNA structure(s) in addition to the cis acting motif. A transcriptome-wide analysis of SOX cleavage was also consistent with site-specific cleavage [59]. Although no simple consensus sequence emerged, sites could be predicted by a position-specific weight matrix. RNA structure prediction indicated susceptible sequences are characterized by a secondary RNA stem-loop that is frequently followed by several unpaired adenine residues just upstream of the cleavage site. Mendez et al. demonstrated SOX binds specifically to this polyadenosine tract and that both binding and cleavage require an open loop structure [60]. Further in vitro and in silico studies, including the solved structure of SOX bound to a target RNA, confirmed that cleavage occurs within unpaired nucleotides found within a stem-loop or bulge-loop structure [61]. These authors also found that SOX makes no sequence specific contacts with RNA. Thus, the four adenine residues upstream of the cleavage site, by virtue of their decreased propensity to form RNA duplexes, are promoting loop structures, not sequence specific binding. The SOX residues responsible for binding these RNA-loop structures are only partially conserved in BGLF5. This difference in the binding pocket provides additional evidence that gammaherpesvirus HSFs target different, but potentially partially overlapping, sets of host mRNAs. Collectively, these studies have demonstrated that the gammaherpesvirus host shutoff factors target mRNAs at an early stage during translation and that cleavage occurs at specific sites determined largely by RNA structural motifs. Such motifs have been shown to be widely distributed across host and viral RNAs [59], explaining the ability of these factors to cause widespread degradation while selectivity targeting mRNAs competent for translation.

Host Shutoff Escape

A surprising feature of herpesvirus HSFs is that they also target viral mRNAs for degradation [50,58,61-63]. HSF cleavage activity needs to be sufficiently limiting, at least at physiologic levels of BGLF5 or SOX expression, to ensure adequate viral protein levels are obtained from the surviving viral mRNAs. In addition, some host mRNAs are resistant to HSF cleavage. The first such resistant mRNA identified was IL-6, which accumulates at high levels in KSHV infected cells undergoing lytic replication despite SOX expression [64]. Deep sequencing of mRNAs from HEK293 cells identified multiple other transcripts that were not decreased upon SOX expression. Some, like the apoptosis enhancing nuclease (AEN) transcript, were directly resistant to cleavage, while others were susceptible to cleavage, but did not decrease in abundance, presumably due to other regulatory effects such as increased transcription. The 3′ UTRs of the IL-6, GADD45B, and C19ORF66 transcripts contain elements that are sufficient to mediate resistance to HSF cleavage when transferred to a heterologous reporter mRNA [65-67]. Some of these SOX-resistance elements (SREs) also mediate BGLF5 resistance. In contrast, SREs found in the AEN and GADD45B transcripts failed to confer resistance to muSOX or BGLF5 cleavage [66,68].

The mechanism by which these SREs confer resistance is incompletely understood. Several studies implicated host ribonucleoprotein complexes binding to SREs as being essential for their role in mediating HSF resistance [65,66,69,70]. SREs from different transcripts appear to bind distinct but partially overlapping sets of RNPs. In several cases, siRNA knockdown of components of these SRE-binding RNP complexes restored susceptibility to HSF degradation. Whether differences in the RNP complexes account for the ability of some SREs to confer resistance to cleavage by multiple HSFs and other SREs to confer resistance to a single HSF is

unclear. Despite the uncertainty regarding the exact mechanism by which SREs act, it is clear that differential transcript resistance and susceptibility to specific HSF proteins combine to define the spectrum of host mRNAs targeted for degradation during replication of a given gammaherpesvirus. It is likely that some host mRNAs (as occurs with viral mRNAs—see section entitled 'HSF Interactions with Viral Gene Expression') are expressed at sufficiently high levels to be translated into their protein products despite susceptibility to HSF degradation. Furthermore, because HSF acts at the mRNA level, many host proteins persist despite degradation of their corresponding mRNAs. Therefore, it is predominantly new protein expression that is affected [48,71].

Secondary HSF Effects May Augment Shutoff

Extensive degradation of mRNAs by gammaherpesvirus HSFs may further augment shutoff of host gene expression by dysregulating mRNA stability factors. BGLF5 or SOX expression leads to marked redistribution of RNA binding proteins including the cytoplasmic poly(A)-binding protein (PABPC) into the nucleus as well as hyperadenylation of nascent mRNAs [62,72,73]. PABPC redistribution is downstream of the RNA degradation activity of HSFs, presumably due to release of large amounts of PABPC from degraded host mRNAs [74]. Nuclear accumulation of PABPC may impair mRNA export and be responsible for mRNA hyperadenylation induced by HSFs [72]. Remarkably, this HSF-induced RNA degradation is accompanied by decreased RNAPII recruitment to host promoters by an unknown mechanism, leading to repressed transcription of many host genes [55]. While not formally demonstrated to occur with BGLF5, this degradation-induced transcriptional repression is seen with SOX and HSV vhs, suggesting it is a general feature of herpesvirus host shutoff. Notably, transcription of viral mRNAs does not occur via this mechanism. As these secondary host shutoff mechanisms

only occur as a consequence of HSF-induced mRNA degradation, it is not clear that their contributions are essential for achieving shutoff. However, these secondary effects do provide a means by which HSF can impair expression of mRNAs that resist their endonuclease activity.

Host Genes Targeted by HSF during Lytic Infection

Host shutoff may contribute to viral replication by at least two mechanisms: abrogating protein expression required to mount an anti-viral response and promoting preferential translation of viral mRNAs. There is substantial evidence that EBV host shutoff can impair the innate and adaptive immune responses. EBV lytic reactivation in Burkitt lymphoma cells is associated with downregulation of multiple Toll-like receptors, including TLR1, TLR6, TLR7, TLR9, and TLR10. Overexpression of BGLF5 inhibits TLR2 and TLR9 expression in HEK293 or MelJuSo melanoma cells [75,76]. Both TLR2 and TLR9 are implicated in sensing and inhibiting primary EBV infection and reactivation [77-81]. BGLF5 also targets CD1d, a non-classical HLA that presents lipid antigens to invariant natural killer T (iNKT) cells [75,82]. The iNKT response has been implicated in control of primary EBV infection and may play a role in limiting EBV reactivation during chronic infection [83,84]. In contrast, TLR4 which has not been reported to sense EBV infection, was not downregulated by BGLF5 overexpression, indicating some degree of selectivity of innate immune factors targeted by BGLF5 [75,79].

BGLF5 is capable of targeting genes essential for adaptive immune responses, including HLA class I and II molecules [48,51,62,75] Further, expression of BGLF5 or SOX protects HEK293 cells against HLA-A2-restricted CD8+ T-cell responses compared to control [51]. This protection was conferred by BGLF5 mutants defective for alkaline exonuclease but not BGLF5 lacking host shutoff activity—providing a direct link between host shutoff and immune evasion. Evidence that BGLF5 protects EBV-infected B cells from adaptive immune responses is less

clear. Knockdown of BGLF5 in lymphoblastoid cell lines (LCLs) only produced a slight increase in CD8+ T-cell response to specific lytic antigens compared to knockdown or knockout of other EBV immune evasion genes (BILF1 or BNLF2a) [85]. This discrepancy highlights the potential redundancy of EBV anti-immune factors; however, it is at least noteworthy that BGLF5 is not sufficient to impair the CD8+ T-cell response (i.e., in the BILF1 and BNLF2a knockdown conditions). It should also be noted that these CD8+ T-cell responses were maintained despite decreased lytic protein expression observed during the knockdown of BGLF5. One important caveat of these studies is the BGLF4 and BGLF5 transcripts overlap such that any shRNA that targets BGLF5 will also knockdown BGLF4 transcript and significant functional interactions between BGLF4 and BGLF5 have been reported [63,86,87]. Nonetheless, BGLF5 does abrogate adaptive immune responses, though further studies in relevant cell types are required to define the magnitude and extent of this effect.

HSF Interactions with Viral Gene Expression

It is unlikely that HSF promotes preferential translation of viral mRNAs. Multiple studies have demonstrated that viral mRNAs are susceptible to HSFs, an effect that is presumably overcome by high-level transcription of lytic mRNAs [50,58,61-63]. Thus, HSFs may fine tune expression levels of viral mRNAs and HSF knockout has the potential to disrupt the stoichiometry of viral proteins required for optimal viral replication. Although knockdown of BGLF5 impairs nucleocapsid maturation and slightly impairs DNA replication, the former may be due to loss of AE rather than host shutoff activity [88]. The situation is clearer for MHV68, where a host shutoff-specific muSOX mutant (R443I) produces virions with abnormal morphology and composition [58]. Interestingly, the R443I mutant did not exhibit noticeable replication defects during acute infection in vivo but did impair the establishment of latency as

evidenced by reduced numbers of infected splenocytes and lower levels of viral DNA during chronic infection [50].

Viral genes may in turn regulate HSF activity. This possibility is best established for herpes simplex virus, where its vhs is inactivated at late stages of replication to facilitate accumulation of late mRNAs. It is not known whether this inactivation occurs for gammaherpesvirus HSFs. For EBV, it has been suggested that the BGLF4 protein kinase counters BGLF5 shutoff based on their opposing effects on several viral mRNAs in transcomplementation assays [63]. However, evidence is conflicting regarding whether BGLF4 promotes BGLF5 phosphorylation [89,90]. Furthermore, when the number of cells in the late phase is taken into account, BGLF5-knockout produces subtle increases in late mRNA abundance [63,88,91]. One important limitation of these studies is that they were performed with bulk populations of asynchronously replicating cells. It is therefore difficult to isolate events occurring during the late phase of replication.

Non-HSF Viral Proteins Contribute to Impaired Host Gene Expression

Although canonical host shutoff by gammaherpes HSFs is sufficient to dramatically curtail host gene expression, several other viral proteins create additional “non-canonical shutoff” barriers that thwart expression of host genes (Figure 2.1B–D). For example, Buschle et al. recently demonstrated that the immediate early transcription factor Zta, in addition to initiating the EBV lytic cascade, induces a global restructuring of host chromatin with widespread loss of chromatin accessibility and chromatin–chromatin interactions [41]. They did not observe a decrease in host mRNA levels after 6 h of Zta expression, but this may not have been sufficient time for steady state expression levels to be achieved. Nascent transcription has been examined in Burkitt lymphoma cells induced for EBV replication with sodium butyrate.

Reduction in transcription of multiple host genes was observed, potentially due to Zta effects and/or secondary effects downstream of sodium butyrate [36]. Further, Park et al. reported decreased synthesis of some host proteins in response to Zta expression [73]. Additionally, herpesvirus replication is characterized by a shift from preferential export of spliced mRNAs to preferential export of non-spliced viral mRNAs. In EBV this change is mediated by SM (also known as Mta or EB2) and appears to be important for both export of viral mRNAs and their translation [92-98]. Progression of the lytic cascade is accompanied by such extensive compaction and margination of host chromatin that further host transcription becomes untenable [28,35]. These additional mechanisms may contribute to the ability of herpesviruses to usurp their host's translational machinery. Additionally, any effort to define BGLF5 targets under physiologic conditions must be able to distinguish bona fide host shutoff effects from those due to the potentially confounding host shutoff mechanisms.

Technical Barriers to Defining HSF Effects

Despite remarkable progress in understanding the mechanism(s) by which gammaherpesviruses induce host shutoff, many important questions remain. Much of the existing research has relied on the overexpression of HSFs to define their effects. We know that HSF degradation is surmountable by viral mRNAs and likely by host mRNAs as well. Thus, overexpression-based approaches may exaggerate the true extent of host shutoff and will not capture the regulatory effects of other viral genes on HSF activity.

The reliance on overexpression stems from the inefficient and asynchronous nature of gammaherpesvirus reactivation from latency. This results in heterogeneous cell populations where a small percentage of cells are undergoing lytic reactivation, while the majority remain latent. Consequently, detecting cellular changes specifically triggered by the lytic replication

process becomes challenging. This limitation highlights the complex and multifaceted cellular requirements for reactivation, which remain an active area of investigation. Current research suggests a dependence on the G1/S phase of the cell cycle for EBV lytic reactivation [35]. Additionally, this process is further governed by a dynamic and intricate balance of pro-reactivation and anti-reactivation factors [99].

To overcome these limitations, previous studies have employed diverse strategies, each with its own set of limitations. Traditional approaches often rely on the use of non-physiologic cell lines exhibiting abnormally high reactivation rates, albeit still below 50% [36,37,39-41]. However, to identify biologically relevant host shutoff targets and escapees, it's critical to study lytic replication in physiologically normal cells. Cell-type specific variations in transcripts affected by (or escaping) HSF-mediated degradation have been previously described [65,66,68]. These differences are due in part to expression of lineage specific mRNAs but may also be due to differences in HSF expression levels, and the presence/absence of proteins or mRNA isoforms that promote host shutoff escape. As HSFs are under selective pressure fsor their ability to allow gammaherpesviruses to replicate in normal cells, it is essential that HSF effects be defined in model systems that most closely approximate them. Research using lymphoblastoid cell lines and EBV-infected oral keratinocytes provides a valuable foundation for identifying essential targets for EBV host shutoff.

Currently, no cell model enables reactivation rates sufficient for analysis of lytic replication at high-resolution. This limitation drives the variety of approaches seen in prior studies. Further complicating the issue, asynchronous reactivation obscures the distinct impacts of early and late lytic phases. One technique utilizes plasmid-based reporters expressing markers from early lytic promoters, combined with cell sorting. An example is the use of reporter cell

surface markers in conjunction with magnetic-activated cell sorting (MACS) [39,41,48,76,100]. However, MACS has limitations when enriching lytic reactivated cells, especially in physiologic cell lines where reactivation is rare. Firstly, MACS enrichment efficiency decreases as the target population becomes rarer [101,102]. Secondly, this approach is typically restricted to enriching a single target population, whereas asynchronous lytic reactivation necessitates separate enrichment of early and late lytic sub-populations (Appendix I).

Similar approaches utilize fluorescent protein reporters with fluorescence-activated cell sorting (FACS). However, early lytic promoter-driven plasmid reporters cannot distinguish between replication stages. To circumvent this, some studies have used replication-defective viruses [41] or viral DNA replication inhibitors [103] to block late-phase progression. Unfortunately, these methods hinder our understanding of the late phase and create an artificial accumulation of early viral proteins [87], potentially obscuring relevant cellular changes. Late-phase analysis with plasmid reporters is further complicated by the OriLyt element required for late promoter activation. OriLyt seeds individual replication compartments [104] and can lead to disruptive co-amplification of plasmid and viral genomes, potentially altering late gene expression patterns. Additionally, plasmid reporter systems suffer from issues like transfection efficiency, plasmid loss during stable transfections [105], and leaky expression. Lastly, while antibodies against the viral glycoprotein gp350 have been used for late-phase cell sorting, the reliability of this approach for detecting or enriching late cells remains questionable. Conflicting results between gp350-based sorting and alternative methods, such as plasmid reporters or intracellular Zta/gp350 staining, suggest that this antibody-based technique may have poor efficacy [105-107]. Single-cell RNA-seq approaches could be employed, although the low frequency of lytic infection may make the cost of this approach prohibitive.

Another technical challenge in studying host shutoff lies in the assumptions made by most RNA and protein quantification methods (including RNA-seq and mass spectrometry). These methods make the implicit assumption that the majority of RNAs and proteins remain unchanged across conditions [108,109], an assumption clearly violated by extensive host shutoff-mediated mRNA degradation and the resultant inhibition of protein synthesis. This issue is highlighted by two studies analyzing Akata Burkitt lymphoma lytic transcriptomes: Ramasubramanyan et al. [39], using conventional RNA-seq normalization, observed minimal changes in host gene expression. In contrast, Buschle et al. [41], accounting for transcript abundance differences by use of spike-in RNA standards, demonstrated widespread host shutoff. Furthermore, general RNA-seq analysis might be less informative in the context of host shutoff, as pathway analysis tools may not easily identify meaningful biological processes given the dominance of downregulation. An alternative approach is to focus on unregulated genes (host shutoff escapees), a less conventional but potentially more useful analysis. Tools like DESeq2 provide statistical methods for this purpose [110].

In summary, achieving biologically meaningful insights into EBV-induced host shutoff necessitates addressing (i) limited efficiency of lytic reactivation, (ii) asynchronous nature of reactivation, and (iii) incompatibility of high-throughput methodologies with conventional normalization assumptions (Figure 2.2). Overcoming these obstacles will facilitate the use of physiologically relevant cell models and identification of key host shutoff targets and escapees.

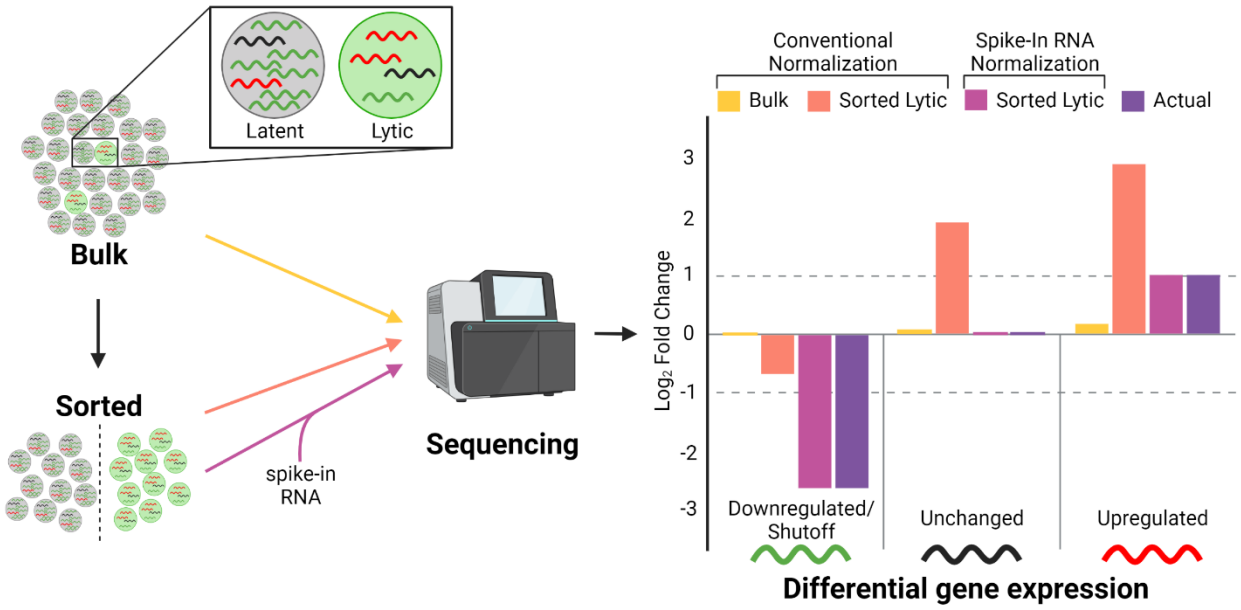


Figure 2.2 Barriers and potential solutions for accurate quantification host shutoff. The inefficient and asynchronous entry of EBV into lytic replication means that bulk populations of EBV infected cells are predominantly latently infected with low levels of early and late lytic infection. As a consequence, detection of differentially expressed genes is diminished, especially mRNAs that are downregulated (or degraded) in lytic cells, such as those subject to host shutoff (compare bulk to actual). This limitation can be addressed via cell sorting approaches; however, given the marked differences in total mRNA in latent cells versus lytic cells, conventional normalization methods fail to accurately measure changes in gene expression. To account for such differences, exogenous spike-in RNAs can be added on a per cell basis. These (or similar) modifications are essential to capture host shutoff effects accurately on the cell gene expression. Figure created with BioRender.com.

Chapter 3

Materials and Methods

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse Anti-Epstein Barr Virus Monoclonal antibody, Unconjugated, Clone 72a1	MilliporeSigma	Cat#MAB10219; RRID:AB_827201
F(ab')2-Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647	Thermo Fisher Scientific	Cat#A-21237; RRID:AB_2535806
AffiniPure F(ab')2 Fragment Goat Anti-Human IgA + IgG + IgM (H+L)	Jackson ImmunoResearch Labs	Cat#109-006-064; RRID:AB_2337548
Normal Goat Serum	Abcam	Cat#ab7481; RRID:AB_2716553
Human Gamma Globulin	Thermo Fisher Scientific	Cat#31879; RRID: AB_2532171
Bacterial and virus strains		
NEB® 5-alpha Competent *E. coli*	New England Biolabs	Cat#C2987
GS1783 *E. coli*	[111]	N/A
BM2710/pGB2Ωinv-hly *E. coli*	[112]	N/A
EBV BAC strain M81	[106]	N/A
DFLR EBV BAC strain M81	This study	N/A
DFLR EBVΔBGLF5 BAC strain M81	This study	N/A
Chemicals, peptides, and recombinant proteins		
Lipofectamine 2000	Thermo Fisher Scientific	Cat#11668019
Polyethylene glycol 8000	MilliporeSigma	Cat#P2139
Polybrene	MilliporeSigma	Cat#TR-1003
2,6-Diaminopimelic acid	MilliporeSigma	Cat#D1377
(Z)-4-Hydroxytamoxifen	MilliporeSigma	Cat#H7904
Cyclosporin A	MilliporeSigma	Cat#C1832
4',6-Diamidino-2-Phenylindole, Dihydrochloride	Thermo Fisher Scientific	Cat#D1306
Ganciclovir	MilliporeSigma	Cat#345700
TRIzol™ LS Reagent	Thermo Fisher Scientific	Cat#10296028
ERCC RNA Spike-In Mix	Thermo Fisher Scientific	Cat#4456740
Gibson Assembly® Master Mix	New England Biolabs	Cat#E2611S
Phusion™ High-Fidelity DNA Polymerase	Thermo Fisher Scientific	Cat#F530S
Critical commercial assays		

ZR BAC DNA Miniprep Kit	Zymo Research	Cat#D4049
Direct-zol RNA Miniprep	Zymo Research	Cat#R2050
DNeasy Blood & Tissue Kit	Qiagen	Cat#69506
Deposited data		
Type 1 EBV reference genome (fasta)	GenBank	NC_007605
Type 1 EBV reference genome (gtf)	[86]	N/A
Inverted Akata EBV reference genome (fasta)	[113]	https://github.com/flemingtonlab/public/blob/master/annotation/chrEBV_Akata_inverted_2.fa
EBV M81 reference genome (fasta)	GenBank	KF373730
ERCC spike-in mix reference genome (fasta and gtf)	Thermo Fisher Scientific	https://assets.thermofisher.com/TFS-Assets/LSG/manuals/ERCC92.zip
GDC GRCh38 reference genome (fasta)	GDC	https://gdc.cancer.gov/about-data/gdc-data-processing/gdc-reference-files
GENCODE GRCh38 primary assembly reference genome v44 (fasta)	GENCODE	https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_44/GRCh38.primary_assembly.genome.fa.gz
GRCh38 comprehensive reference genome v44 (gtf)	GENCODE	https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_44/gencode.v44.primary_assembly.annotation.gtf.gz
EBV lytic Raji Burkitt lymphoma RNA-seq	[41]; ArrayExpress	E-MTAB-7820
LCL mRNA half-lives	[114]	N/A
DFLR LCL RNA-seq	This study	N/A
Experimental models: Cell lines		
HEK293.MSCV-ZHT	This study	N/A
HEK293.MSCV-mGL	This study	N/A
HEK293.MSCV-mSI	This study	N/A
DFLR EBV BAC strain M81 HEK293.MSCV-ZHT	This study	N/A
DFLR EBVΔBGLF5 BAC strain M81 HEK293.MSCV-ZHT	This study	N/A
DFLR EBV BAC strain M81 LCL	This study	N/A

DFLR EBVΔBGLF5 BAC strain M81 LCL	This study	N/A
Mutu LCL	[115]	N/A
Oligonucleotides		
PCR primers were listed in Table S1	IDT	N/A
BAC recombineering gene blocks were listed in Table S2	IDT	N/A
Recombinant DNA		
pCDNA3-ZHT	[46]	N/A
pCDH-MSCV-MCS-EF1a-GFP+Puro	System Biosciences	Cat#CD713B-1
pCDH-MSCV-MCS-EF1a+Puro	This study	N/A
pCDH-MSCV-ZHT-EF1a-puro	This study	N/A
pCDH-MSCV-mGL-EF1a-puro	This study	N/A
pCDH-MSCV-mSI-EF1a-puro	This study	N/A
psPAX2	Addgene	Cat#12260; RRID: Addgene_12260
pCMV-VSV-G	[116]; Addgene	Cat#8454; RRID: Addgene_8454
pRK5-BALF4	[117]	N/A
Software and algorithms		
R version v4.3.0	[118]	https://www.r-project.org/
flowCore v2.14.0	[119]	https://bioconductor.org/packages/release/bioc/html/flowCore.html
flowAI v1.32.0	[120]	https://bioconductor.org/packages/release/bioc/html/flowAI.html
ggcyto v1.30.0	[121]	https://bioconductor.org/packages/release/bioc/html/ggcyto.html
LiftOff v1.6.3	[122]	https://github.com/agshumate/Liftoff
FastQC v0.11.8	[123]	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
STAR v2.7.6a	[124]	https://github.com/alexdobin/STAR
RSEM v1.3.1	[125]	https://github.com/deweylab/RSEM

DESeq2 v1.42.0	[110]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
apeglm	[126]	https://bioconductor.org/packages/release/bioc/html/apeglm.html
Picard CollectRnaSeqMetrics v3.1.0	Broad Institute	http://broadinstitute.github.io/picard/
deepTools bamCoverage v3.5.1	[127]	https://deeptools.readthedocs.io/en/develop/content/tools/bamCoverage.html
SparK v2.6.2	[128]	https://github.com/harbourlab/SparK
SpliceWiz v1.5.2	[129]	https://www.bioconductor.org/packages/release/bioc/html/SpliceWiz.html
IsoformSwitchAnalyzeR v2.2.0	[130,131]	https://bioconductor.org/packages/release/bioc/html/IsoformSwitchAnalyzeR.html
DEXSeq v1.48.0	[132,133]	https://bioconductor.org/packages/release/bioc/html/DEXSeq.html
CPC2	[134]	https://cpc2.gao-lab.org/
Pfam	[135]	https://pfam.xfam.org/
IUPred2A	[136]	https://iupred2a.elte.hu/
SignalP-6.0	[137]	https://services.healthtech.dtu.dk/services/SignalP-6.0/
ARTDeco	[138]	https://github.com/sjroth/ARTDeco
ggplot2 v3.4.4	[139]	https://cran.r-project.org/web/packages/ggplot2/

ggpubr v0.6.0	[140]	https://cran.r-project.org/web/packages/ggpubr/
ggstatsplot v0.12.1	[141]	https://cran.r-project.org/web/packages/ggstatsplot/
eulerr v7.0.0	[142]	https://cran.r-project.org/web/packages/eulerr/
ComplexHeatmap v2.18.0	[143]	https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
Other		
UltraComp eBeads™ Compensation Beads	Thermo Fisher Scientific	Cat#01-2222-42
Sony MA900 Cell Sorter	Sony Biotechnology	N/A

Resource Availability

Data and Code Availability

All RNA-seq data described in this manuscript have been deposited to NCBI SRA under BioProject accession number PRJNA1088113.

Experimental Model Details

Cell lines and reagents

All cells were cultured at 37°C in a humidified atmosphere with 5% CO_2. HEK293FT, HEK293, and derivatives were cultured in Dulbecco's modified Eagle's medium (DMEM, Gebco) supplemented with 10% fetal bovine serum (FBS, HyClone) and 1% penicillin-streptomycin (Gibco). BACmid and lentiviral-infected HEK293 cells were maintained with 600 μg/mL G418 and 1 μg/mL puromycin, respectively. LCLs were cultured in RPMI 1640 supplemented with 10% FBS and 1% penicillin-streptomycin. For BCR stimulation, LCLs were seeded at 300,000 cells/mL in culture media and treated with 20 μg/mL AffiniPure F(ab)2 anti-human pan-Ig antibody (Jackson ImmunoResearch) for 72 hours.

Method Details

Plasmids construction

The base plasmid, pCDH-MSCV-MCS-EF1a+Puro, was derived from the commercial plasmid pCDH-MSCV-MCS-EF1a-GFP+Puro (#CD713B-1, System Biosciences). The GFP fragment was removed through BmtI and BsiWI restriction digestion and replaced with an MCS gene block (Appendix Table II-2) using Gibson assembly. Both pCDH-MSCV-mGL-EF1a+Puro and pCDH-MSCV-mSI-EF1a+Puro were constructed using an EcoRI-digested base plasmid as the backbone. This was combined with Phusion polymerase (#F-530S Thermo Scientific) PCR

amplified fragments of mGreenLantern (mGL) and mScarlet-I (mSI), respectively. The PCR used the FP primers (Appendix Table II-1) with BXLF1-to-mGL and BILF2-tPT2A-mSI gene blocks (Appendix Table II-2) serving as templates. To create pCDH-MSCV-ZHT-EF1a+Puro, the Acc65I/BspEI and BspEI/EcoRI digest fragments from pCDNA3-ZHT were ligated into the BsrGI/EcoRI digested base plasmid. The ZHT gBlock contains Zta fused to a 4HT responsive estrogen hormone binding domain (Z-HT) [46].

Lentivirus production

HEK293FT cells at 70-80% confluency in 10-cm dishes were co-transfected with three plasmids using Lipofectamine 2000: 10 µg of the lentiviral vector, 12 µg of psPAX2 (Addgene #12260), and 4 µg of pCMV-VSV-G (Addgene #8454). The transfection was performed using 30 µL Lipofectamine per 10-cm dish in 4 mL of serum-free DMEM. After 4 hours, the transfection media was replaced with culture media supplemented with 50 mM HEPES. Lentivirus harvests occurred at 24 and 48 hours, with supernatants clarified by centrifugation (2000 x g, 10 min) and filtration (0.8 µm membrane). Supernatants were then concentrated via PEG-8000 precipitation. Briefly, one-part supernatant was mixed with an equal volume of 2x lentivirus concentrator solution (20% PEG-8000, 0.6 M NaCl, pH 7.4), rocked for 60 seconds, and incubated overnight at 4°C with constant rocking at 60 RPM. The supernatant was centrifuged (1600 x g, 60 min, 4°C), and the pellet resuspended with $1/20^{th}$ the original volume of serum-free media. The concentrated lentivirus was snap-frozen and stored at -80°C until use.

Creation of HEK293 cell lines stably transduced with Z-HT, mGL, or mSI

The concentrated lentiviruses pCDH-MSCV-ZHT-EF1a-puro, pCDH-MSCV-mGL-EF1a-puro, and pCDH-MSCV-mSI-EF1a-puro were thawed from -80°C stocks. HEK293 cells at 70-80% confluency in 6-well plates were infected with 1 mL of each lentivirus, supplemented

with 10 μg/mL polybrene, and incubated for 2 hours at 37°C before media replacement with culture media. Following a 24-hour incubation period, cells infected with lentivirus were subjected to puromycin selection (1 μg/mL).

Construction of recombinant EBV BACmids

The dual-fluorescent lytic reporter (DFLR) EBV used in this study was derived from the EBV M81 strain BACmid [106]. To construct this, the recombineering technique *En Passant* mutagenesis was employed in the GS1783 *E. coli* strain [111] using gene blocks (gBlocks) found in Appendix Table II-2. Initially, the F-factor was modified to facilitate selection in hygromycin-resistant cell lines, such as hTERT-immortalized normal oral keratinocytes. This entailed the substitution of the SV40 promoter-driven hygromycin with the neomycin/kanamycin eukaryotic selection marker using the hyg^r-to-neo^r gBlock. To be compatible with a fluorescent reporter system, the CMV-driven eGFP present in the F-factor was deleted using the GFP-KO gBlock. The early expressed and non-essential BXLF1 ORF was then substituted with the GFP-derivative mGreenLantern (mGL) [87,144-146] using the BXLF1-to-mGL gBlock. The RFP-derivative mScarlet-I (mSI) was fused to a tandem P2A and T2A sequence (tPT2A) between the stop codon and the ORF of the late expressed BILF2 gene [87,147,148] using the BILF2-tPT2A-mSI gBlock. This late reporter modification not only enables the bi-cistronic expression of the orphan BILF2 protein with mSI [149], but also preserves that BART promoter region located within the BILF2 ORF [150].

The DFLR EBVΔBGLF5 construct was generated by introducing a deletion within the BGLF5 ORF mirroring a deletion previously described and characterized by Feederle et al. [88] using the BGLF5-KO gBlock. This deletion (aa7 – aa306) encompasses key motifs and residues critical for BGLF5's catalytic activity and nucleic acid binding functions [54]. To preserve the

downstream BBLF1 gene, the deletion terminates upstream of the BBLF1 promoter. To maintain the integrity of the upstream BGLF4 ORF, the first six amino acids of BGLF5, which overlap out-of-frame with the BGLF4 ORF, were retained. Furthermore, a stop codon was introduced immediately following these retained BGLF5 residues.

The integrity of BACs in GS1783 and BM2710 *E. coli* strains was verified through restriction enzyme digestion. Further validation of the BAC sequences in GS1783 *E. coli* was performed by NGS DNA sequencing at the CCIB DNA Core Facility at Massachusetts General Hospital. Throughout the cloning process, regular PCR screening of the full-length family of repeats (FR) within the *oriP* sequence was implemented for all recombinant BACs present in *E. coli* and mammalian cells using primers flanking the FR primers found in Appendix Table II-1. Notably the FR which plays a crucial role in maintaining the EBV genome in latently infected cells [151] , exhibits instability in *E. coli* [152], as observed in the M81 strain.

Generation of WT and ΔBGLF5 DFLR EBV LCLs

The diaminopimelate (DAP) auxotroph and invasive *E. coli* strain BM2710 was employed as a BAC delivery vector for EBV BACmids into HEK293.MSCV-ZHT cells. This strain carries both the invasin-encoding gene (*inv*) from *Yersinia pseudotuberculosis* and the listeriolysin O-encoding gene (*hly*) from *Listeria monocytogenes* within the pGB2Ωinv-hly plasmid. Bactofection was performed by co-culturing HEK293.MSCV-ZHT cells (70-80% confluency in 6-well plate with 2 mL culture media) with BACmid-carrying BM2710 *E. coli* (100 µL overnight cultures grown in BHI media containing 500 µM DAP, 25 µg/mL spectinomycin, and 50 µg/mL kanamycin) for 2 hours at 37°C. Following PBS washes, the bactofected HEK293.MSCV-ZHT cells were incubated with culture media supplemented with

100 μg/mL gentamicin for 24 hours. 48 hours post-bactofection, the cells were exposed to G418 selection (600 μg/mL). Individual G418-resistant colonies were then isolated using cloning rings.

To produce infectious particles for B cell infection, WT and ΔBGLF5 DFLR EBV HEK293.MSCV-ZHT cells were cultured to 70-80% confluency in 10-cm dishes with 10 mL DMEM culture media. Simultaneously, these cells were treated with 200 nM 4-hydroxytamoxifen (4-HT) and transfected with 4 μg of pRK5-BALF4 (gp110) using Lipofectamine 2000 at a 3:1 ratio. Following a 24-hour incubation period, the transfection media was replaced with RPMI culture media supplemented with 200 nM 4-HT. Five days post-inductions, the supernatants were clarified by centrifugation (2000 x g, 10 min, 4°C). These supernatants were then concentrated by PEG-8000 precipitation as described in the 'Lentivirus production' section, except 2x EBV concentrator solution (16% PEG-8000, 2M NaCl, pH 7.4) was used in place of the 2x lentivirus concentrator solution and concentration performed without snap freezing. The concentrated virus was stored at 4°C and used within a week.

Peripheral blood mononuclear cells (PBMCs) obtained from the same donor underwent isolation through Ficoll-Hypaque density centrifugation. Subsequently, these PBMCs were infected with the concentrated virus and sustained in RPMI supplemented with 10% FBS, 1% penicillin-streptomycin, and 0.5 μg/mL cyclosporine A until outgrowth of LCLs. The growing LCLs in each well were selected as distinct clones for further analysis.

DNA extraction and PCR

BAC DNA from *E. coli* was extracted using the ZR BAC DNA Miniprep Kit (Zymo Research). Genomic DNA was isolated from BAC-infected HEK293.MSCV-ZHT and short-term (~2 months) and long-term (~6 months) cultured LCLs using the DNeasy Blood & Tissue Kit (Qiagen). PCR amplification of the F-factor, the modified BXLF1 and BILF2 EBV genes,

and the human GAPDH gene regions were performed using primers found in Appendix Table II-1. The Neo^r and SopA/B primers correspond to the F-factor regions.

Microscopy and flow analysis of gp350-stained DFLR LCLs

For microscopy imaging and flow analysis, $1x10^6$ BCR-stimulated DFLR LCLs were washed with PBS. Cells were resuspended in wash/block buffer (PBS with 5% goat serum and 12 mg/mL human IgG) and blocked for 30 minutes on ice. Blocking with 12 mg/mL human IgG was performed to mitigate non-specific antibody interactions with Fc receptors as previously described [153]. Additionally, cells were blocked with 5% goat serum to reduce undesired binding of the goat-raised secondary antibody. Cells were then stained with 1:50 anti-gp350 72a1 antibody on ice for 30 minutes. Cells were washed three times with wash buffer and incubated with 2 μg/mL of secondary goat anti-mouse Alexa Fluor 647-conjugated antibody on ice for 30 minutes. Cells were washed three times with wash buffer before being resuspended in 200 μL FACS buffer (1x DPBS -Ca^{2+}/-Mg^{2+}; 2% FBS) containing 0.5 μg/mL of the vital dye DAPI. Stained cells were then subjected to microscopy imaging or flow cytometry analysis on the Sony MA900 (UW Flow Cytometry, Madison, WI) with compensation. Unstained and non-reporter Mutu LCLs served as the negative control, while a DAPI-treated dead cell population (heat treated at 72°C for 5 minutes) controlled for DAPI fluorescence. Additionally, HEK293.MSCV-mGL and HEK293.MSCV-mSI cells provided mGreenLantern and mScarlet-I calibrations, respectively. UltraComp eBeads Compensation Beads (ThermoFisher) stained with the secondary Alexa Fluor 647 antibody were used for compensating spillover by this antibody.

Flow data analysis was performed within the R environment (v4.3.0) using the flowCore package v2.14.0 [119]. This involved extracting the spillover data for compensation and applying logical transformation [154] on the fluorescence marker data. Additionally, event

outliers were identified and removed using the flowAI package v1.32.0 [120]. Flow graphs were generated using the ggCyto package v1.30.0 [121].

FACS and sample collection for RNA-seq

72-hour BCR-stimulated DFLR LCLs were harvested by centrifugation at 300 x g for 10 minutes. DFLR LCLs were then resuspended to a final concentration of $10x10^6$ cells/mL in ice-cold FACS buffer containing 0.5 μg/mL of the vital dye DAPI and remained on ice until sorting. Prior to sorting, compensation controls described in the 'Microscopy and flow analysis of gp350-stained DFLR LCLs' section were included to ensure accurate fluorescence interpretation. Additionally, a BCR-stimulated DFLR LCL sample with or without 100 μg/mL ganciclovir was used to establish the gating strategy. Gates were sequentially implemented: (1) LCL population, (2) doublet exclusion, (3) DAPI-based removal of necrotic cells, (4) mGl-/mSI-gate, (5) mGL+/mSI- gate, and (6) mGL+/mSI+ gate. The mSI boundary was set based on the GCV treatment condition and the mGL boundary referenced the negative control.

Sorting was performed using the Sony MA900 cell sorter (UW Flow Cytometry, Madison, WI) at 4°C using a 100 μm nozzle and ultra-pure settings to achieve >99% purity. Samples were collected by sorting 40,000 cells from each sub-population directly into RNase-free tubes containing 1.5 mL TRIzol-LS (Thermo-Fisher) and the equivalent of a 1:125 dilution of 1 μL ERCC Spike-In RNA (Invitrogen, #4456740). For prolonged sorting, samples were vortexed every 10 minutes to ensure cell lysis. The volume of the lysate was then adjusted to 2 mL to maintain a 3:1 TRIzol to sample ratio, and lysates were thoroughly mixed and centrifuged before being stored at -80°C. Each replicate was BCR-stimulated and sorted on different days.

RNA-sequencing

Total RNA was extracted from FAC-sorted samples using the Direct-Zol RNA mini prep kit (Zymo Research). The optional on-column DNase treatment was applied to eliminate genomic DNA contamination. PolyA-enriched strand-specific libraries were generated using the Illumina TruSeq Stranded mRNA kit, followed by 101 bp paired-end sequencing on the Illumina NovaSeq 6000 system. Library preparation and sequencing was performed by the Yale Center for Genome Analysis. All DFLR LCLs sequenced were derived from the same donor. The WT condition comprised two independent clones, totaling five biological replicates. The ΔBGLF5 condition comprised one clone, totaling three biological replicates.

Quantification and Statistical Analysis

Reference genome and annotation assembly

An inverted DFLR M81 genome and its corresponding annotation reference were generated as follows. Frist, the genome was annotated by lifting over genes from the previously published annotation [86] of the type 1 EBV reference genome (based on GenBank accession number NC_007605) using LiftOff v1.6.3 [122]. Manual adjustments were made to mGreenLantern, BILF2-tPT2A-mSI, and neighboring genes. Additionally, the latent and lytic isoforms of the LMP1 gene were divided into individual genes within the updated DFLR M81 annotation. Subsequently, the DFLR M81 genome was inverted to align its origin with the previously published inverted Akata genome [113], followed by a lift-over of the DFLR M81 annotation to the inverted version.

To enable ERCC spike-in normalization, we sourced ERCC sequences and annotation directly from ThermoFisher's database (https://assets.thermofisher.com/TFS-Assets/LSG/manuals/ERCC92.zip). This dataset encompasses the 92 polyadenylated RNAs

designed for ERCC spike-in applications. The final GDC_GRCh38.ERCC.inverted_DFLR_M81 genome was generated by incorporating the ERCC sequences as artificial chromosomes into the GDC GRCh38-based genome and replacing the EBV decoy with the inverted DFLR M81 sequence. Similarly, a gencode.GRCh38.v44.primary_assembly.ERCC.inverted_DFLR_M81 annotation was generated by combining the ERCC and inverted DFLR M81 annotations with the GENCODE GRCh38 v44 primary assembly annotation.

Read alignment and differential expression analysis

Verification of read quality was conducted using FastQC v0.11.8 [123]. For read alignment, STAR v2.7.6a [124] was employed, and gene expression quantification performed using RSEM v.1.3.1 [125]. Both tools utilized the GDC_GRCh38.ERCC.inverted_DFLR_M81 genome and gencode.GRCh38.v44.primary_assembly.ERCC.inverted_DFLR_M81 annotation references. The parameters utilized in this analysis are outlined in the NASA GeneLab RNA-seq pipeline [155].

Raw expected read counts from RSEM were imported into the R environment (v.4.3.0) for differential gene expression analysis performed between pairs of conditions using DESeq2 package v1.42.0 [110]. To mitigate the impact of low-abundance transcripts, genes were filtered out if their FPKM values were less than 1 across the smallest condition's sample size. The analysis focused on transcripts with the GENCODE biotype tag coding (*i.e.*, protein-coding) and non-coding (Mt_rRNA, Mt_tRNA, miRNA, misc_RNA, scRNA, snRNA, snoRNA, ribozyme, sRNA, scaRNA, and lncRNA). Normalization was then performed by using the ERCC spike-in genes as the control genes for estimating the DESeq size factors. The 'apeglm' shrinkage estimator was employed to refine the estimation of gene-wise dispersion parameters, thereby

enhancing the accuracy of differential expression analysis, particularly for genes with low counts [126].

To isolate BGLF5-specific effects, we filtered out potential confounding factors arising from inherent differences in immunoglobulin classes between WT (IgM) and ΔBGLF5 (IgG) LCLs (Figure 3.1). We focused on genes with consistent expression levels in their respective latent fractions by applying the DESeq2 “lessAbs” alternative hypothesis (lfcThreshold = 1, FDR $\leq$ 0.1, and betaPrior = FALSE).

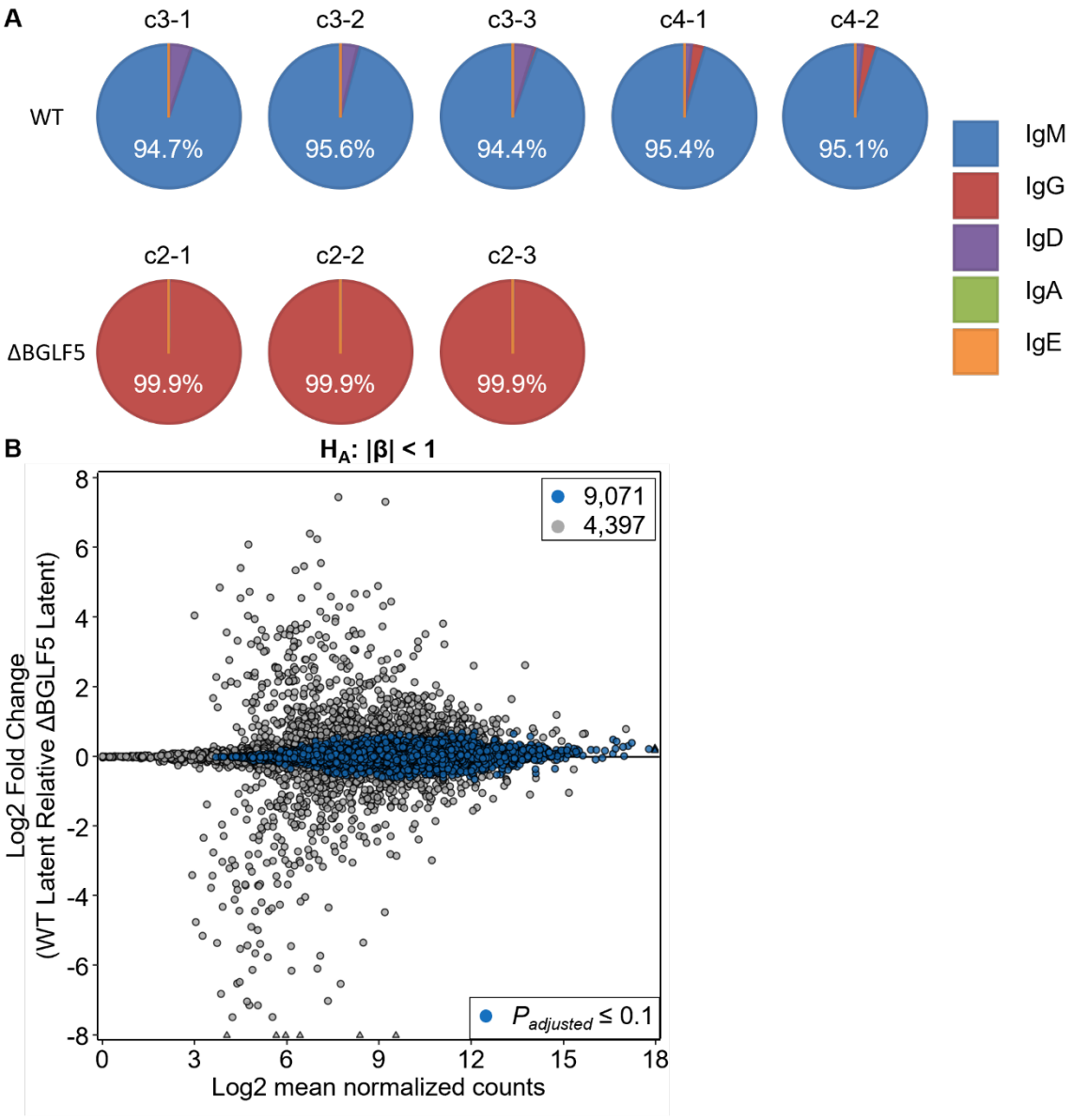


Figure 3.1 BCR immunoglobulin isotype composition and host gene expression filtering between DFLR LCLs and DFLRΔBGLF5 LCLs. (A) Pie charts showing the percentage composition of BCR immunoglobulin isotypes in DFLR LCLs (WT) and DFLRΔBGLF5 LCLs (ΔBGLF5). (B) MA plot showing host gene expression fold changes between latent fractions of DFLR LCLs (WT) and DFLRΔBGLF5 LCLs (ΔBGLF5). Non-differentially expressed genes ($P_{adjusted} \leq 0.1$) are indicated in blue, using the DESeq2 "lessAbs" alternative hypothesis (HA: $|\beta| < 1$).

Calculation of RNA-seq metrics

CollectRnaSeqMetrics v3.1.0 from Picard tools (http://broadinstitute.github.io/picard/) was utilized to estimate metrics from BAM files that contained reads aligned to chr1-22, chrX, and chrY of the GRCh38 genome reference.

Visualization Tracks

To generate NGS coverage tracks, the scaling factors for each sample were initially computed. This involved taking the inverse of the averaged DESeq2 gene normalization factors, which were estimated by using ERCC spike-in genes as control genes during the simultaneous analysis of all samples. Subsequently, coverage from BAM files was calculated and bedgraph files were generated using deeptools bamCoverage v3.5.1 [127] with the specified parameters: -bs 1 -of bedgraph -region <chr:start-end> --filterRNAstrand reverse --scaleFactor <sample scaling factor>. The depth coverage of bedgraph files for each condition were then averaged. Bedgraph files for different regions of interest (*e.g.*, ASEs and DoGs) were split into distinct files. Coverage graphs were plotted with SparK v.2.6.2 [128], using the split bedgraph files for each region of the same sample as individual control groups. The custom scaling option was applied for equal scaling across conditions. The coverage graphs displaying the split regions of interest for each sample were color-coded, overlaid, and then plotted alongside other samples.

Absolute quantification of mRNA molecules per cell

FPKM values obtained from RSEM were imported into the R environment (v4.3.0) for further processing. To ensure a robust analysis, only reliably detected genes retained in the filtered output of DESeq2 were considered. Subsequently, the known absolute abundance of the remaining spike-in genes was utilized to establish a standard curve. This calibration curve related the ERCC FPKM values to their absolute mRNA quantities per cell. The linear regression

equation derived from this standard curve was then applied to the FPKM values of the filtered gene dataset.

Alternative splicing analysis

For alternative splicing analysis, only genes reliably detected in the filtered output of DESeq2 were analyzed. Splicing-based analysis was performed with the SpliceWiz software (v1.5.2) with default filter settings [129] .

To comprehensively analyze isoform switching events and their predicted functional consequences across different conditions, the R package IsoformSwitchAnalyzeR (v2.2.0) was employed [130,131]. This computational tool leverages the DEXSeq (v1.48.0) methodology to assess differential isoform usage between experimental conditions [132,133] and the spliceR algorithm to identify alternative splicing events contributing to these switches [156].

To predict transcript sensitivity to nonsense-mediated decay (NMD), IsoformSwitchAnalyzeR employs a two-step process based on transcript model coordinates [157]. Initially, it extracts the nucleotide sequence for each transcript from the provided reference genome. Using this sequence, the software identifies the most probable open reading frame (ORF) by prioritizing the longest in-frame region. Subsequently, NMD sensitivity is evaluated by searching for premature termination codons (PTCs) within the predicted ORF. Two criteria are applied: firstly, the stop codon must not reside within the final exon, and secondly, the distance between the stop codon and the final exon-exon junction must exceed 50 nucleotides. Ultimately, each transcript is assigned an NMD sensitivity score reflecting its anticipated susceptibility to degradation.

To predict additional consequences of isoform switches, built-in functions of IsoformSwitchAnalyzeR were used to translate the predicted ORF nucleotide sequences into

their corresponding amino acid sequences. These sequences were then subjected to analysis for coding potential, protein domains, intrinsically disordered regions (IDR) and intrinsically disordered binding regions (IDBR), and signal peptides using the external sequence analysis tools CPC2 [134], Pfam [135], IUPred2A [136], SignalP-6.0 [137].

Identification of DoGs and read-in transcripts

To identify genes with transcription readthrough beyond polyA signal-dependent cleavage sites (termed downstream-of-gene-containing transcripts, or DoGs) and those displaying read-in transcription, we employed the Automatic Readthrough Transcription Detection (ARTDeco) software in its "readthrough" and "get_dogs" modes [138]. This approach efficiently detects DoGs within a size range of 4-15 kb by employing a 500 bp rolling window analysis initiated from the furthest downstream transcription termination site. To ensure robust and reproducible DoG and read-in calls, only events consistently identified across replicates within each condition and reliably detected in the filtered output of DESeq2 were considered.

Data analysis

Data analysis was conducted in the R environment (v4.3.0). Box plots, mean-abundance (MA) plots, and pie charts were generated using the ggplot2 package v3.4.4 [139]. Bar plots were constructed using the ggpubr package v0.6.0 [140]. Violin plots, paired with p-values derived from two-sided Welch's t-tests, were generated through the ggstatsplot package v0.12.1 [141]. To correct for multiple testing, the Benjamin-Hochberg False Discovery Rate (P_{FDR}) adjustment was implemented [158]. Venn diagrams were plotted using the eulerr package v7.0.0 [142]. Heatmaps were constructed with the ComplexHeatmap package v2.18.0 [143]. For heatmaps, z-score transformation was applied to EBV gene expression levels across samples.

Hierarchical clustering was subsequently calculated using the Euclidean distance method, both on samples and on genes within subgroups defined by viral gene expression kinetics.

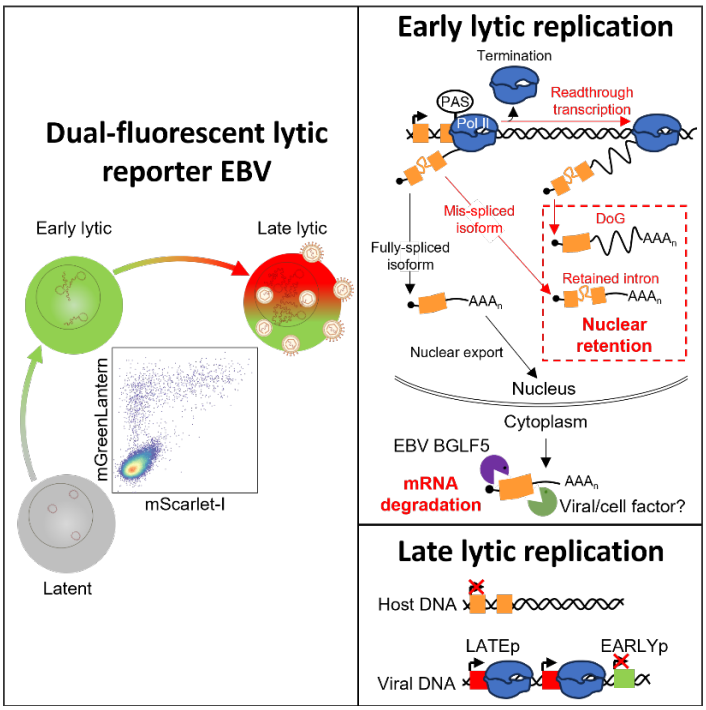
Lytic Raji Burkitt lymphoma RNA-seq data acquisition and analysis

Previously published raw RNA-seq data from Raji cells harboring a doxycycline-inducible co-expression plasmid for EBV BZLF1 and a truncated low-affinity nerve growth receptor (tLNGFR) receptor variant was acquired from the ArrayExpress database under the Accession number E-MTAB-7820 [41]. These cells were maintained in either an uninduced ("latent") state or induced with doxycycline and subsequently sorted for BZLF1-expressing cells using magnetic-activated cell sorting (MACS) targeting the tLNGFR receptor ("early lytic"). This study also incorporated external RNA controls (ERCC spike-in RNAs) at equal amounts per cell across samples. The data were analyzed using the GDC GRCh38.ERCC references with the previously published NC_007605 annotated genome [86] as described in the 'Read alignment and differential expression analysis' methods section.

Chapter 4

Investigating the Transcriptomic Landscape during EBV Lytic Replication and BGLF5's Role in Host Shutoff

This chapter is composed primarily of a research article entitled "EBV Induces Host Shutoff Extensively via BGLF5-Independent Mechanisms" by Alejandro Casco, Makoto Ohashi, and Eric Johannsen, manuscript in progress.



Introduction

Epstein-Barr virus (EBV) is a human cancer virus associated with multiple types of lymphoma and carcinomas [1,159]. Additionally, EBV is implicated in the pathogenesis of autoimmune diseases with especially strong data supporting a causal role in multiple sclerosis [3]. A member of the gammaherpesvirus family, EBV establishes life-long latent infection in memory B lymphocytes and reactivates from tonsillar B cells and spreads via saliva to other hosts. Approximately 95% of healthy adults are chronically infected with EBV, making it among the most successful human pathogens.

Although EBV infection is usually well tolerated, high seroprevalence and lifelong infection amplify exposure to this carcinogenic virus. Unlike most viral diseases, EBV-associated cancers are linked with latent viral infection. Nevertheless, EBV lytic replication plays a crucial indirect role by maintaining its high seroprevalence. Additionally, accumulating evidence suggests that EBV lytic replication also plays a direct role in oncogenesis [4,160]. A distinctive feature of EBV reactivation is that viral replication is asymptomatic, eliciting no apparent inflammatory response despite robust cellular and humoral immunity to EBV lytic antigens. This covert reactivation is associated with virus-mediated host shutoff that impairs the host cell's ability to synthesize new proteins during EBV replication, thereby pre-empting an antiviral response (reviewed in Casco, et al.) [161].

EBV lacks a homolog to the HSV virion host shutoff protein and instead is believed to cause host shutoff primarily via the BGLF5 exonuclease [48]. This exonuclease is conserved in all herpesviruses and processes viral DNA in nuclear replication compartments for packaging into capsids [49-51]. The BGLF5 exonuclease, and its homologs in other gammaherpesviruses, also localizes to the cytoplasm and unexpectedly proved to have RNA endonuclease activity that

affects host shutoff [56,57]. BGLF5 overexpression is sufficient to induce host mRNA degradation and has been widely employed to predict which mRNAs are targeted by (or resistant to) host shutoff during EBV reactivation. These studies have revealed that virtually all mRNAs, even viral mRNAs, are susceptible to BGLF5 cleavage. This finding highlights the importance of studying BGLF5 effects under physiologic conditions as viral mRNAs accumulate, illustrating that not all susceptible mRNAs are fully "shutoff." A further motivation for studying BGLF5 effects in the context of EBV reactivation from latency is that this approach will account for effects of other viral proteins that may modulate BGLF5 activity and/or exert direct effects on host mRNA stability.

A key barrier to studying host shutoff in EBV-infected cells is that reactivation from latency is both inefficient and asynchronous. As a consequence, a limited number of EBV-infected tumor cell lines with unusually high rates of replication have been extensively studied, while more physiologic models such as lymphoblastoid cell lines (LCLs) have been studied less. In addition, most analyses have relied on inefficient purification schemes to enrich cells supporting EBV replication.

To overcome these limitations, we constructed a recombinant EBV which expresses a green fluorescent protein with early kinetics and a red fluorescent protein with late kinetics. This dual-fluorescent lytic reporter (DFLR) EBV can be used to infect a wide variety of cells and, when coupled with FACS, allows isolation of highly purified subpopulations that are latently infected with EBV, at the early stage of lytic replication, or at the late stage of replication. We describe the transformation of primary B cells via infection with DFLR EBV into LCLs and their changes in the host and viral transcriptomes during EBV reactivation. Our results present the most detailed analysis to date of host transcriptomic changes during EBV replication, revealing a

remarkably dynamic transcriptional landscape. Notably, the observed patterns of host mRNA suppression deviate from established mechanisms, requiring a role for other viral proteins and/or the dysregulation of additional host regulatory pathways.

Results

Development and characterization of dual-fluorescent lytic reporter (DFLR) EBV transformed LCLs.

To study EBV lytic replication, we developed a dual-fluorescent lytic reporter virus termed DFLR EBV (Figure 4.1A) based on the M81 strain of EBV. Unlike the B95-8 laboratory strain, M81 contains no deletions and is thought to encode all the proteins and microRNAs of clinical EBV isolates [106]. In this recombinant virus, we have replaced the non-essential BXLF1 ORF with the GFP-derivative mGreenLantern (mGL) for early-phase visualization. For late-phase detection, the RFP-derivative mScarlet-I (mSI) is expressed downstream of the BILF2 ORF using "self-cleaving" peptide linkers (tandem P2A and T2A, tPT2A). Initially, we introduced DFLR EBV via bactofection into HEK293.MSCV-ZHT cells which we used to produce cell free virus. In order to study DFLR EBV in a physiological model of EBV reactivation, we transformed primary B cells with DFLR EBV into lymphoblastoid cell lines (LCLs). DFLR LCLs expressed EBV latent proteins at levels similar to and grew at rates comparable to WT LCLs. Another potential advantage of the M81 BACmid is that the artificial F-factor sequence is inserted into its terminal repeats and prone to deletion. While the exact mechanism is controversial, this deletion likely occurs through stochastic cleavage during encapsidation [162] or homologous recombination-mediated viral genome circularization during infection [163]. We could detect BAC sequences by PCR in early passage (~2 months) DFLR

LCLs (Figure 4.2B), but not after 6 months of culture (Figure 4.2A), suggesting that F-factor deletion did occur and likely provided a growth advantage to the DFLR LCLs.

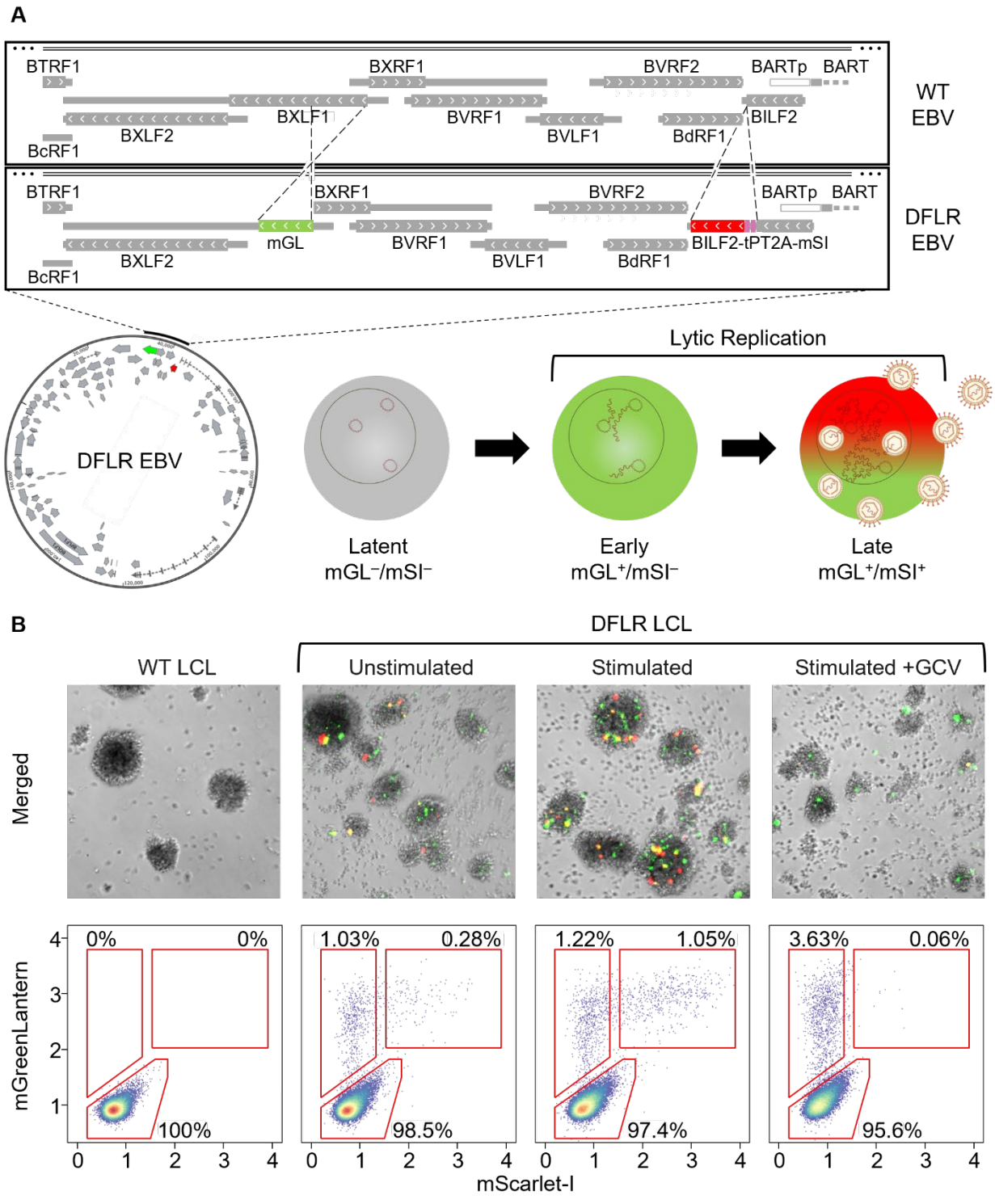


Figure 4.1 Establishment and validation of a dual-fluorescent lytic reporter Epstein-Barr virus (DFLR EBV). (A) Schematic of EBV M81 strain BACmid detailing the insertion of fluorescent reporter genes. The non-essential BXLF1 early ORF was replaced for the mGreenLantern (mGL) protein. Additionally, the mScarlet-I (mSI) was fused to the BILF2 late

ORF using the synthetic tPT2A "self-cleaving" peptides. The resulting virus expresses mGL during early lytic replication and mGL + mSI during late lytic replication. The BACmid map was generated using the SnapGene® software (from Dotmatics; available at snapgene.com). (B) Merged microscopic images of LCLs transformed with WT EBV (left) or with DFLR EBV without BCR stimulation (middle left), 72 h after pan-Ig treatment (middle right), and 72 h post-combined pan-Ig and ganciclovir (GCV) treatment (right). (C) Corresponding flow data for each of the four conditions from (B).

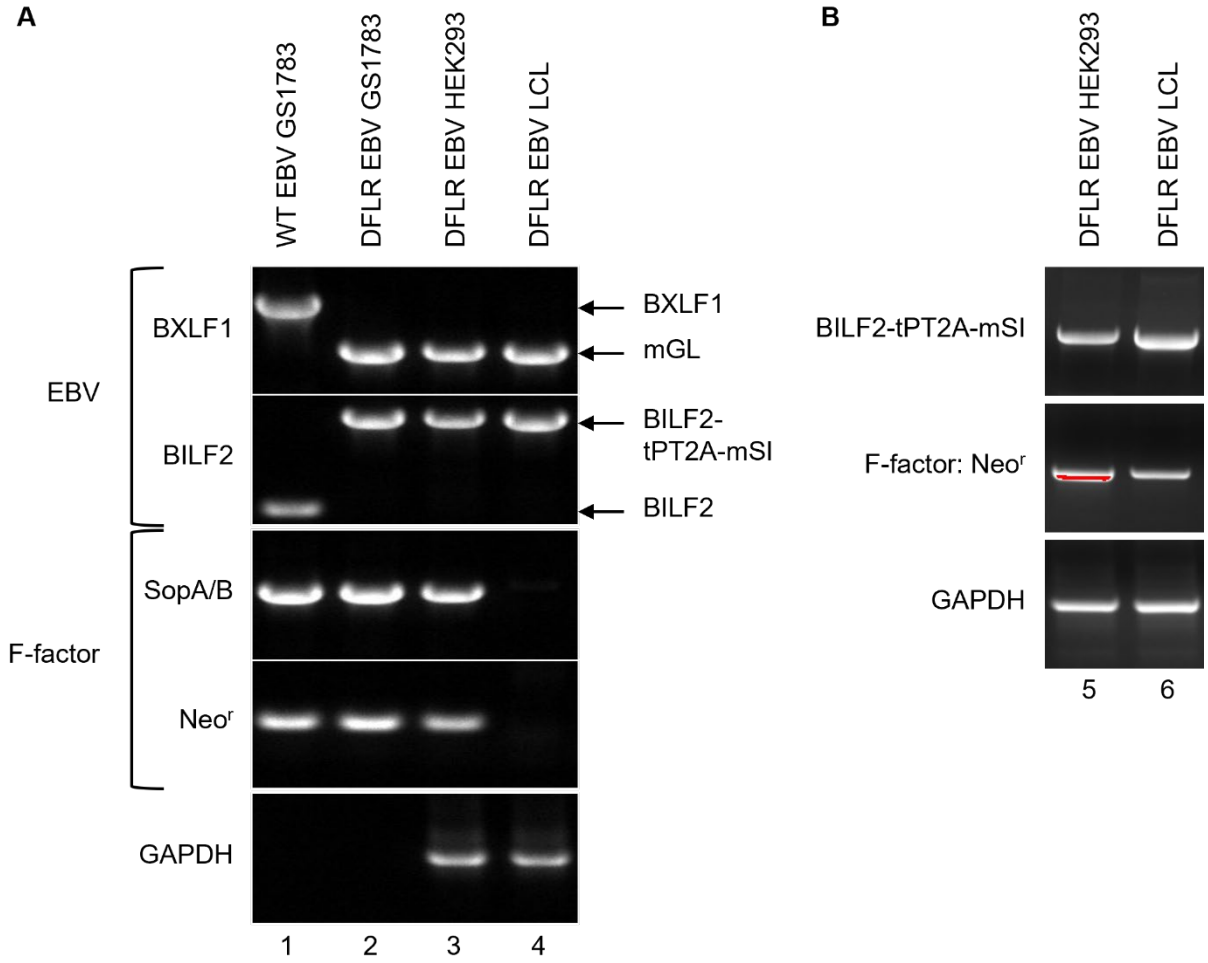


Figure 4.2 The F-factor present in short-cultured LCLs is lost in longer cultures. (A) PCR analysis of F-factor maintenance in LCL cultures harboring the DFLR EBV M81 strain BACmid. Amplified regions include the modified EBV genes (BXLF1 and BILF2), F-factor genes (Neor and SopA/B), and the human GAPDH gene. Template DNA was harvested from WT EBV BACmid GS1783 E. coli (lane 1) or from various sources harboring the DFLR EBV BACmid: GS1783 E. coli (lane 2), HEK293 cells (lane 3), and LCLs (lane 4) established from virus produced from the HEK293 cells after long-term (6 months) culture. (B)PCR of the same DFLR LCLs as in (A) but after short-term (2 months) culture, alongside the same DFLR HEK293 cells.

As an initial test of their reporter fidelity, DFLR LCLs were subjected to flow cytometry analysis. A minority of cells expressed the mGL and mSI lytic reporter proteins in the absence of stimulation (Figure 4.1B and Figure 4.1C), consistent with spontaneous lytic reactivation. No signal was observed in non-reporter WT LCLs. Upon BCR stimulation with pan anti-human immunoglobulin, the proportion of DFLR LCLs in the late phase increased from 0.28% to 1.05%. Intriguingly, the proportion of cells in the early phase did not significantly change, suggesting that BCR stimulation was mainly promoting progression to the lytic cascade. As a control, we treated BCR-stimulated DFLR LCLs with ganciclovir, a viral DNA polymerase inhibitor, and observed the expected loss of expression of our late gene mSI reporter. Based on this data we predicted that DFLR LCLs could be efficiently segregated into latent (mGL-/mSI-), early lytic (mGL+/mSI-), and late lytic (mGL+/mSI+) fractions.

DFLR LCLs can be efficiently sorted into latent, early lytic, and late lytic subpopulations.

To characterize the transcriptomes of DFLR LCLs undergoing lytic EBV replication, RNA was extracted from BCR-stimulated DFLR LCLs and FAC-sorted into latent, early lytic, and late lytic fractions (Figure 4.3A). Samples were subjected to polyA-enriched short-read RNA sequencing from two independent clones and a total of five biological replicates. Importantly, ERCC spike-in RNAs were added to allow normalization of mRNAs on a per cell basis to address artifacts that could be introduced by conventional RNA-seq normalization which might be inapplicable to cells undergoing host shutoff [41,108]. mRNA abundance was measured by creating a standard curve with known ERCC spike-in RNA quantities and expressed as mRNA molecules per cell.

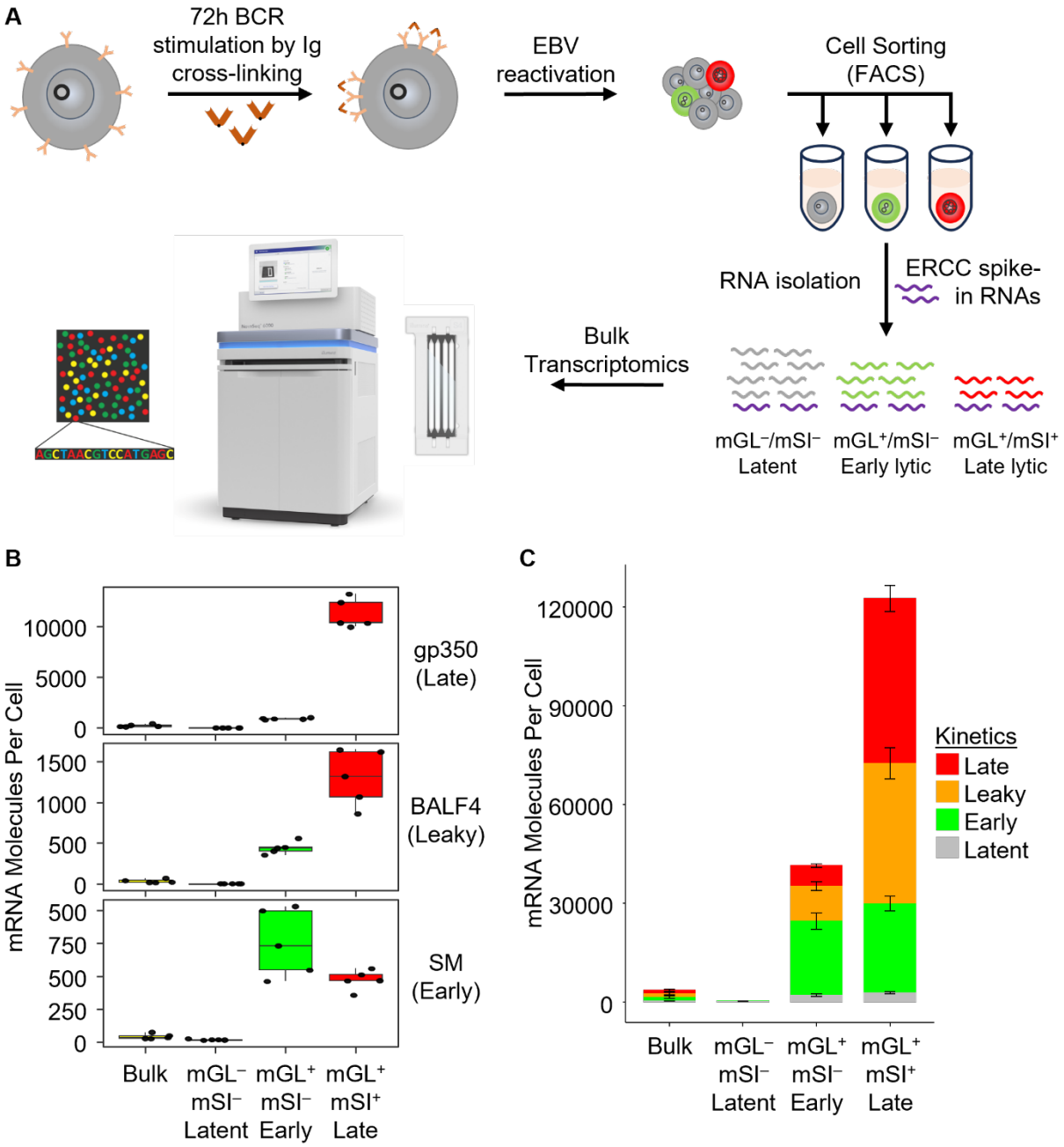


Figure 4.3 Early lytic and late lytic populations can be purified from DFLR LCLs induced for EBV replication. (A) Experimental workflow used to characterize transcriptomes of early and late lytic populations. 72-hours after induction of EBV replication via BCR stimulation with pan-Ig, DFLR LCLs were subjected to FACS to isolate latent (mGL-/mSI-), early lytic (mGL+/mSI-), and late lytic (mGL+/mSI-) fractions. An equal number of cells was sorted for each condition into TRIzol containing equimolar amounts of ERCC spike-in RNAs. Strand-specific polyA-enriched libraries were constructed for each subpopulation and from bulk DFLR LCLs and sequenced as detailed in the materials and methods. (B) Boxplots quantifying representative mRNAs from each EBV kinetic class (latent, early, leaky late, late) within each

fraction. (C) Stacked bar graphs quantifying total EBV mRNAs (latent and 3 lytic classes) within each fraction. Data are represented as mean ± SEM.

To assess the purity of each fraction, three viral genes with known kinetics: early (SM), leaky (BALF4), and late (gp350) were examined [87]. As expected, the latent (mGL-/mSI-) fraction lacked lytic gene expression, early (mGL+/mSI-) cells exhibited high levels of SM and moderate levels of BALF4 expression, and gp350 expression was only observed in late lytic (mGL+/mSI+) cells (Figure 4.3B). Using this data, we calculated the total levels of viral mRNAs (stratified by kinetic class) for cells in each fraction, excluding the EBER RNAs which lack polyadenylation and are therefore not captured by our analysis (Figure 4.3C). This analysis highlighted the efficiency of our DFLR LCL enrichment procedure. Whereas the stimulated bulk population contained, on average, 3,482 viral mRNAs/cell, the early (mGL+/mSI-) fraction had 41,424 viral mRNAs/cell and the late (mGL+/mSI+) fraction had 122,615 viral mRNAs/cell. For comparison, latent (mGL-/mSI-) DFLR LCLs expressed an average of 305 EBV RNAs/cell. These RNAs were predominantly latent mRNAs, further attesting to the efficiency of our fractionation. Notably, we did observe a small quantity of late mRNAs in the early fraction. This likely reflects our reliance on protein expression. Consistent with this explanation, we observed low levels of BILF2-tPT2A-mSI transcript in the early (mGL+/mSI-) fraction, presumably from cells that had begun the transition to late gene expression but had yet to translate those late mRNAs into protein. Interestingly, most viral genes with early kinetics did not show increased expression during the late phase and many were downregulated (Figure 4.4). Collectively, these results demonstrate that sorting of DFLR LCLs produces highly pure latent, early lytic, and late lytic fractions suitable for detailed transcriptomic analysis of changes induced by EBV reactivation from latency in B cells.

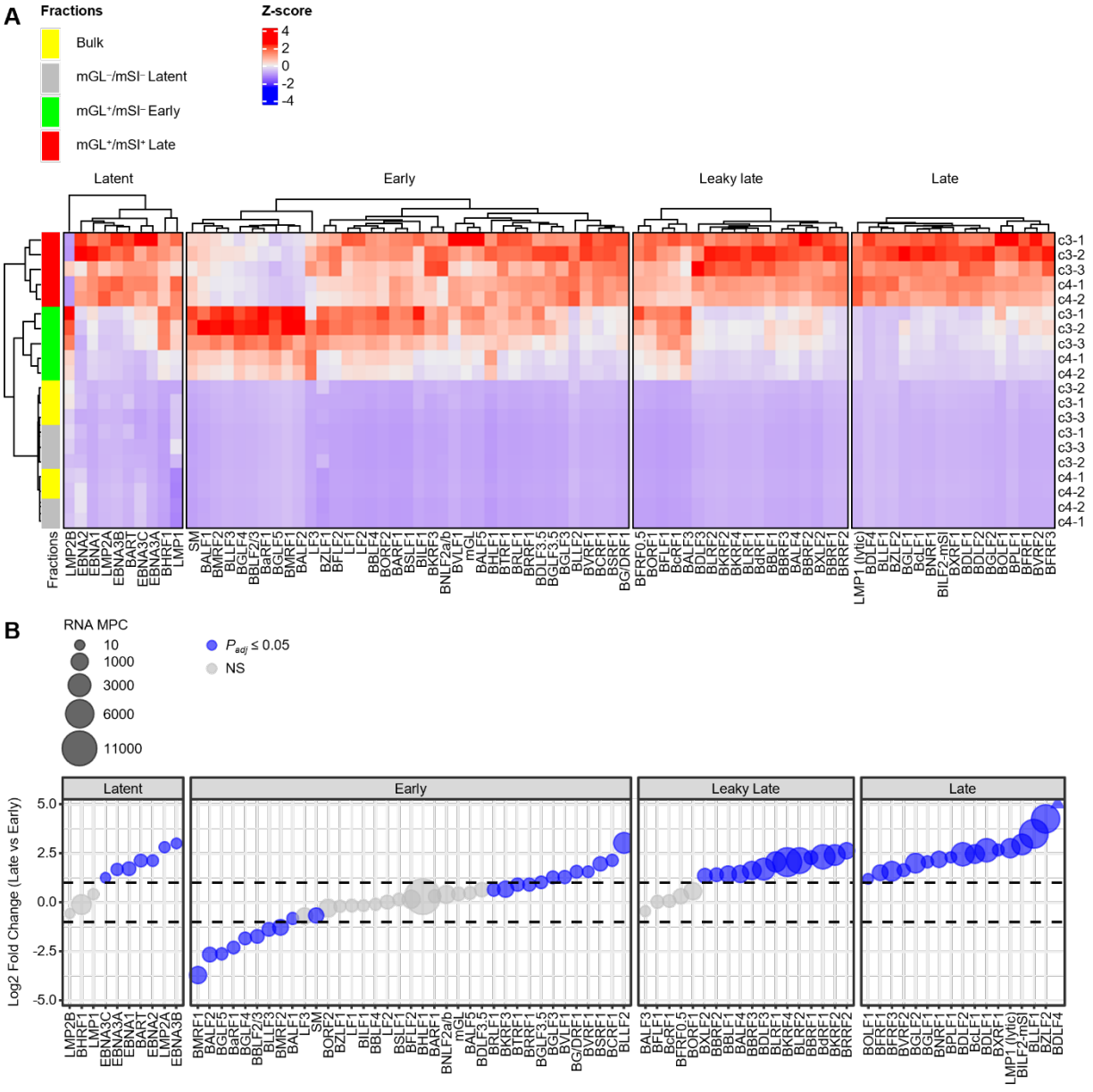


Figure 4.4 Profiling of EBV gene expression during early and late lytic replication. (A) Heatmap of standardized expression levels of all EBV genes across each sorted fraction. Genes are grouped by EBV kinetics (latent, early, leaky late, and late). (B) Dot plot showing the changes in viral gene expression during the late phase of lytic replication relative to the early phase, with dot size representing the mRNA molecules per cell (MPC) between conditions and color the significance (blue = $P_{adjusted} \leq 0.05$; gray = nonsignificant). Genes are grouped by EBV kinetics (latent, early, leaky late, and late).

Antibody staining for gp350 expression is less sensitive and specific than DFLR LCLs for purifying cells in the late lytic phase of EBV replication.

We also compared DFLR LCL lytic detection to that achieved with anti-gp350 staining, a commonly used method for isolating cells undergoing EBV replication. For this approach we used the 72A1 monoclonal antibody [164] which binds to the glycan-free region of gp350 [165], a characteristic believed to underpin its well-documented specificity. To our surprise, we found many mScarlet-I negative DFLR LCLs stained positive for gp350 (Figure 4.5A). By flow cytometry analysis, gp350 staining reliably identified late lytic (mGL+/mSI+) cells; however, a large number of cells in the latent (mGL-/mSI-) fraction were also gp350 positive. This detection was not attributable to endogenous expression of gp350, as the transcript was entirely undetectable in the latent fraction (Figure 4.5B). Overall, gp350 staining was 79% sensitive and 86% specific for late lytic cells relative to the mScarlet-I gold standard (Figure 4.5B), resulting in a positive predictive value of only 2.4% in this experiment. By contrast, the absence of a mGL-/mSI+ population (i.e., late gene positivity in the absence of early gene expression) suggests DFLR LCL sorting produces few false positives. We speculate that the false positives seen with gp350 staining may arise from viral particles and secreted exosomes adhering to latently infected cells (Figure 4.5C) as has been proposed by others examining this issue [105-107].

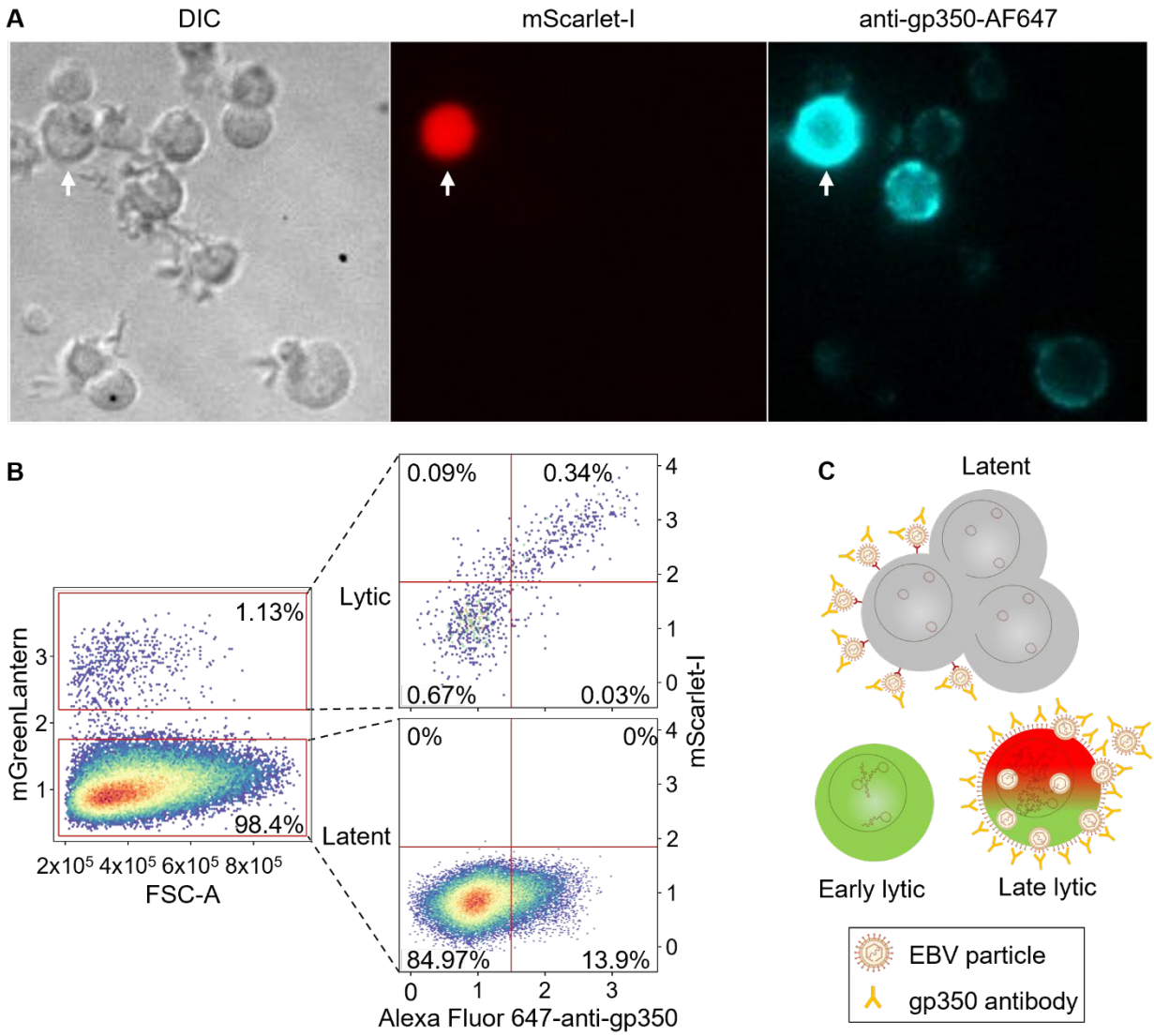


Figure 4.5 GP350 staining is sensitive, but not specific for cells in late phase. (A) Microscopy images of DFLR LCLs BCR stimulated for 72 hours with pan-Ig and stained with gp350 72a1. Visualized by differential interference contrast (DIC, left panel), mScarlet-I fluorescence for late gene expression (middle panel), and far-red fluorescence for anti-gp350 staining (right panel). (B) Flow cytometry analysis comparing mScarlet-I expression (late lytic reporter) to anti-gp350 staining in latent (mGL-) and lytic (mGL+) fractions. (C) Schematic depicting mechanism consistent with flow data: binding of gp350 containing virions to latently infected cells would produce gp350 staining in cells lacking lytic gene expression.

EBV lytic replication is associated with marked decrease in almost all host mRNAs.

In parallel with increases of viral gene expression, host mRNA abundance progressively declined during EBV lytic replication (Figure 4.6A). During the early phase, the host mRNA pool was depleted by 67.7%. This host shutoff phenomenon intensified during the late phase (mean log2 fold change: -3.73 late vs. -2.09 early), corresponding to depletion of 87.8% of host mRNA (Figure 4.6A and Figure 4.6B). At the individual gene level, 85.2% of host genes were significantly reduced early and 95.1% were significantly reduced by the late phase (Figure 4.6C). We also examined the extent to which RNA abundance and half-life influenced transcript susceptibility to host shutoff. Abundant mRNAs were slightly more affected with negative correlations during both the early and late phases (R values of -0.19 and -0.33, respectively; Figure 4.7A). Unexpectedly, we found that mRNAs with longer half-lives in LCLs [114] tended to exhibit greater decreases in abundance during EBV reactivation than those with shorter half-lives (R values of -0.17 and -0.22, Figure 4.7B). Thus, mRNAs with the highest abundance and longest half-lives were more susceptible to host shutoff; however, these factors only accounted for 6.5% and 15.7% of the variance in the observed downregulation within the early and late lytic subpopulations, respectively.

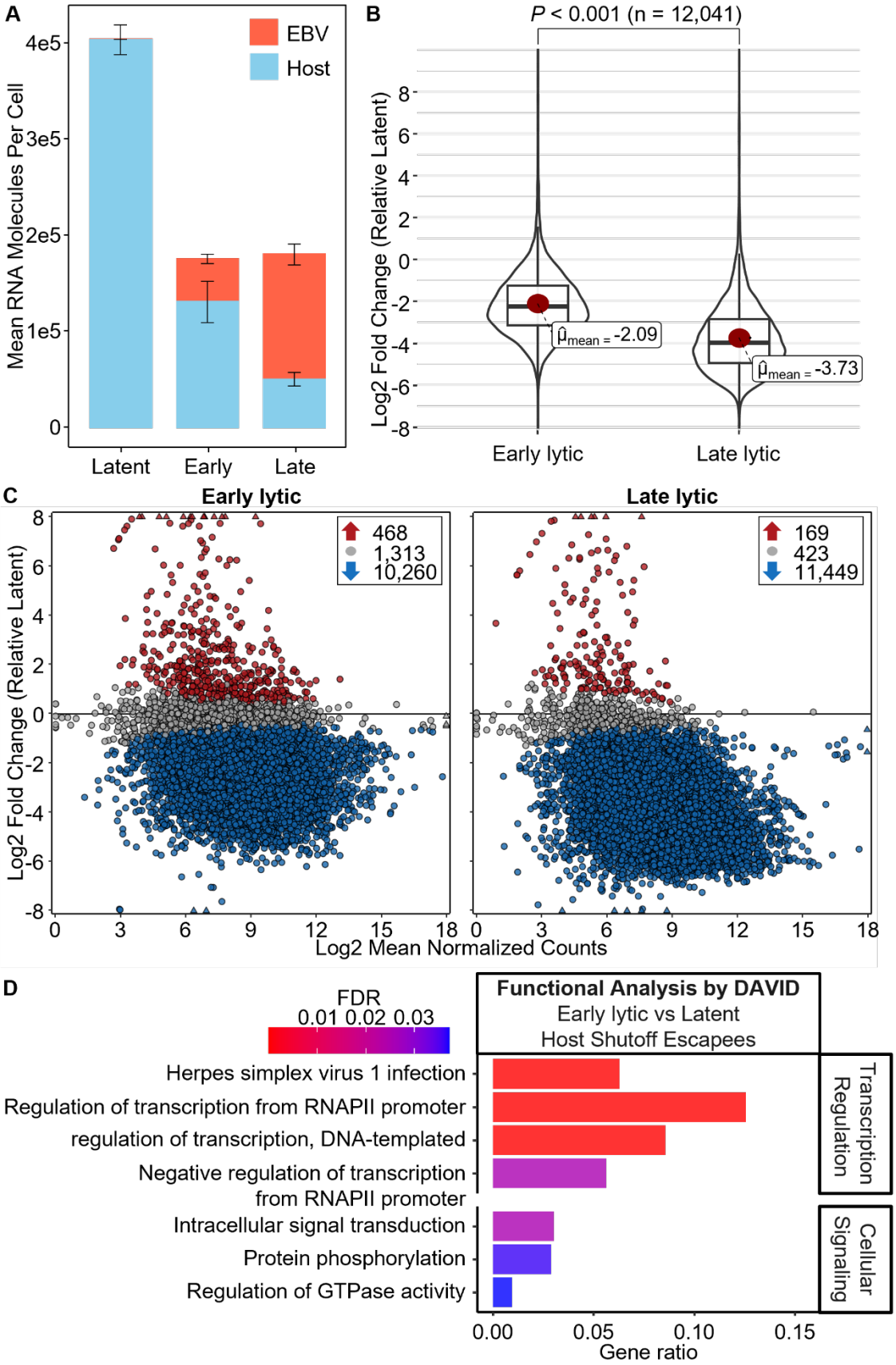

A
Mean RNA Molecules Per Cell
4e5
3e5
2e5
1e5
0
EBV
Host
Latent
Early
Late
B
P < 0.001 (n = 12,041)
Log2 Fold Change (Relative Latent)
8
6
4
2
0
-2
-4
-6
-8
μ̂mean = -2.09
μ̂mean = -3.73
Early lytic
Late lytic
C
Early lytic
Late lytic
Log2 Fold Change (Relative Latent)
468
1,313
10,260
169
423
11,449
0
3
6
9
12
15
18
Log2 Mean Normalized Counts
D
FDR
0.01 0.02 0.03
Functional Analysis by DAVID
Early lytic vs Latent
Host Shutoff Escapees
Herpes simplex virus 1 infection
Regulation of transcription from RNAPII promoter
regulation of transcription, DNA-templated
Negative regulation of transcription from RNAPII promoter
Intracellular signal transduction
Protein phosphorylation
Regulation of GTPase activity
Transcription Regulation
Cellular Signaling
0.00
0.05
0.10
0.15
Gene ratio

Figure 4.6 Extensive host shutoff during EBV replication. (A) Stacked bar graphs contrasting quantities of Host versus EBV mRNAs in latent, early lytic, and late lytic fractions. Data are represented as mean ± SEM. (B) Distribution of fold change in host gene expression levels in early and late lytic fractions relative to the latent fraction visualized through superimposed box and violin plots. The mean log2 fold change in gene expression for each plot is denoted by red dots. The p-value is derived from a two-sided Welch's test. (C) MA plots demonstrating host gene expression fold changes in early (left panel) and late (right panel) lytic fractions relative to the latent fraction. The x-axis depicts average levels of expression for each gene between conditions being compared. Differential expressed genes ($P_{adjusted} \leq 0.05$) are indicated in red (upregulated) or blue (downregulated); genes without significant expression change are colored gray. Triangles indicate data points exceeding the displayed range. (D) DAVID gene ontology (GO) and KEGG pathway analyses performed on genes not significantly downregulated during early lytic replication relative to latency. Functional themes were identified by DAVID annotation clustering and manually designated as "Transcription Regulation" and "Cellular Signaling".

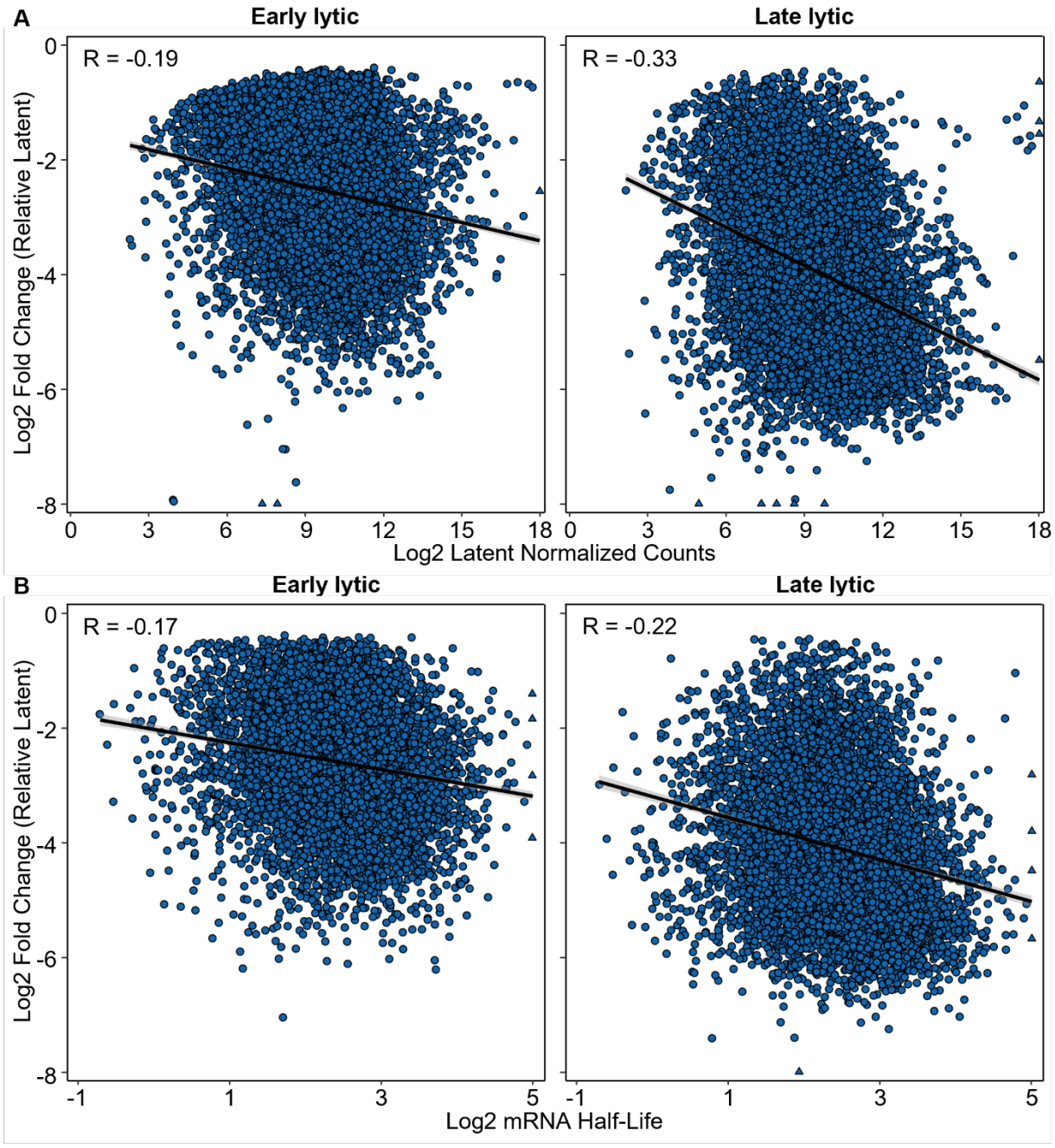


Figure 4.7 Relationship of mRNA levels during latency or mRNA stability in LCLs with mRNA downregulation during early or late lytic EBV replication in DFLR LCLs. (A-B) Scatterplots showing fold changes of significantly downregulated genes ($P_{adjusted} \leq 0.05$) in early (left panel) and late (right panel) lytic DFLR LCLs relative to latently infected DFLR LCLs. The x-axis depicts (A) expression levels for each gene during the latent state or (B) mRNA half-lives in LCLs from Duan et al. The Pearson correlation coefficient (R) is indicated. Triangles indicate data points exceeding the displayed range.

A recent report by Buschle et al. [41] characterized early-phase host shutoff in the highly lytic Raji Burkitt lymphoma (BL) cell line. This cell line's defective EBV genome blocks viral DNA amplification and late-phase entry, allowing early phase analysis without interference from late-phase effects. In order to assess the extent to which host shutoff targets are conserved between LCLs and the Raji BL cells, we reanalyzed their data using our pipeline. Of the genes that were expressed in both cell lines, 84.9% were significantly downregulated ($P_{adjusted} \leq 0.05$) in both systems during the early phase, but the correlation of fold change was weak (R = 0.59, Figure 4.8A and Figure 4.8B). This weakness was mainly due to a much greater degree of downregulation in the DFLR LCLs than in the Raji cells (mean log2 fold change: -2.15 vs. -1.5, Figure 4.8C). Although this could represent a true difference in the extant of host shutoff between the two cell lines, a more probable explanation is that this difference is attributable to differences in the purity of lytic population. Raji cells will enter lytic replication with relatively high efficiency, reported to be as high as 56% [166]. DFLR LCLs, when sorted, can yield an early lytic cell purity of ~99%.

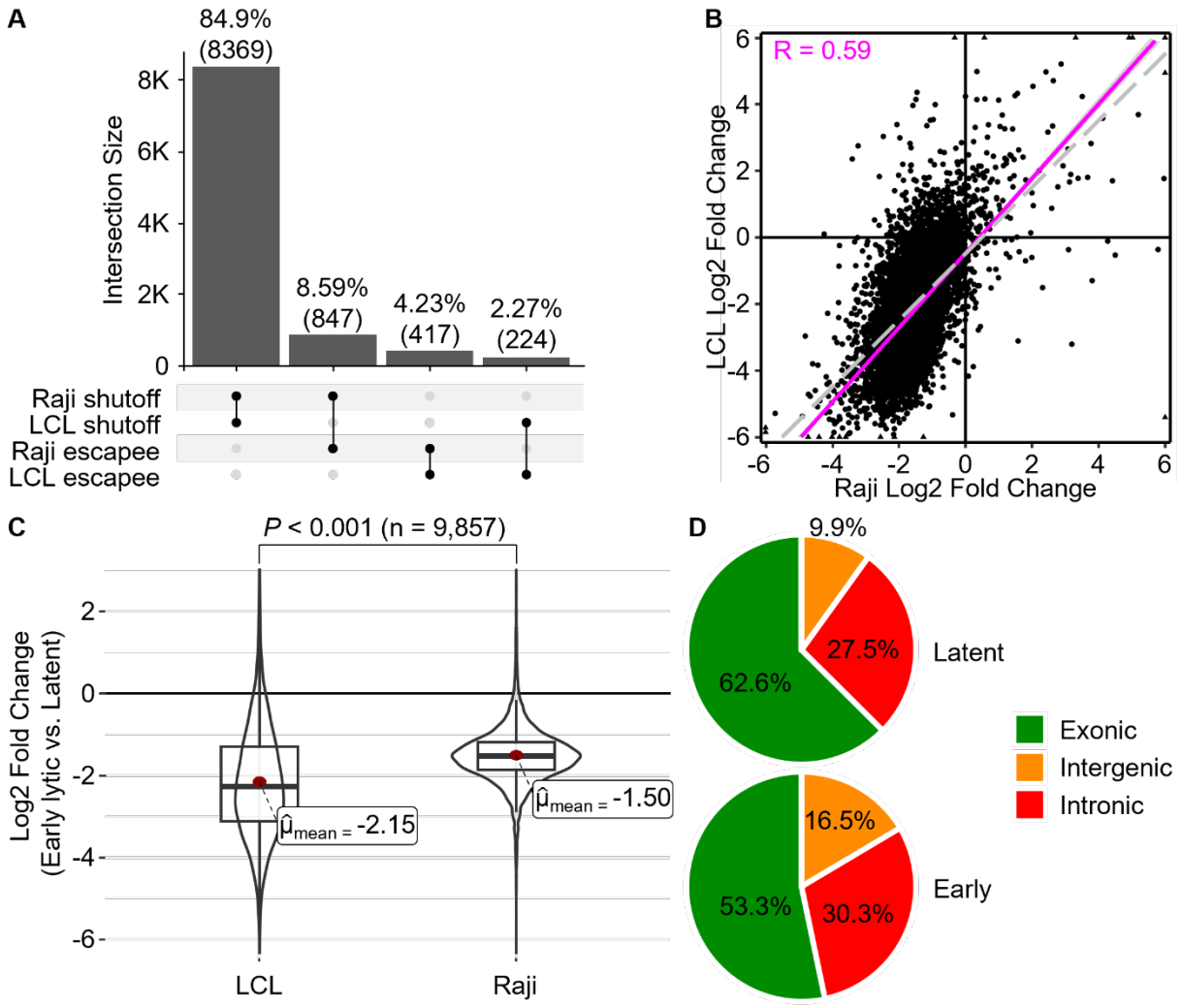


Figure 4.8 Early lytic replication gene expression changes in DFLR LCLs versus Buschle et al. Raji Burkitt lymphoma cells. (A) UpSet plot demonstrating the gene overlaps of host shutoff targets (downregulated and $P_{adjusted} \leq 0.05$) or escapees in both cell lines. (B) Log-log scatterplot showing the relationship between cell types undergoing early lytic replication relative to latency. The Pearson correlation coefficient (R) is indicated. (C) Fold change distributions in host gene expression levels during early lytic replication relative to latency in both cell types, visualized through superimposed box and violin plots. The mean gene expression log2 fold changes for each plot is denoted by red dots. The p-value is derived from a two-sided Welch's test. (D) Pie charts depicting distribution of reads mapping to exonic, intronic, and intergenic regions of host chromosomes in RNA-seq data from latent and early lytic Raji cells.

EBV replication is associated with marked qualitative disruption of the host transcriptome.

The host transcriptome was not merely reduced by EBV reactivation, we also observed extensive dysregulation of the remaining host mRNAs. This dysregulation was evident, for example, in the location to which host reads were mapping. In latently infected cells, a substantial majority of reads (77.7%) mapped to gene exons, 18.1% mapped to introns, and only 4.2% to intergenic regions (Figure 4.9A). This distribution is typical for polyA-enriched RNA-seq data, where some intronic coverage is expected due to incompletely spliced pre-mRNAs and intergenic coverage has been ascribed primarily to "fuzzy" transcription beyond gene annotation boundaries as well as unannotated long non-coding RNAs [167]. By contrast, in DFLR LCLs undergoing early lytic replication, only 48.4% of reads were exonic, 37.5% were intronic, and 14.1% mapped to intergenic regions. By late lytic replication, this distribution was skewed even further with reads being only 34.2% exonic, 43.5% intronic, and 22.3% intergenic. By contrast, in Raji BL cells EBV reactivation produced only a minor increase in intergenic reads and intronic read were not increased (Figure 4.8D).

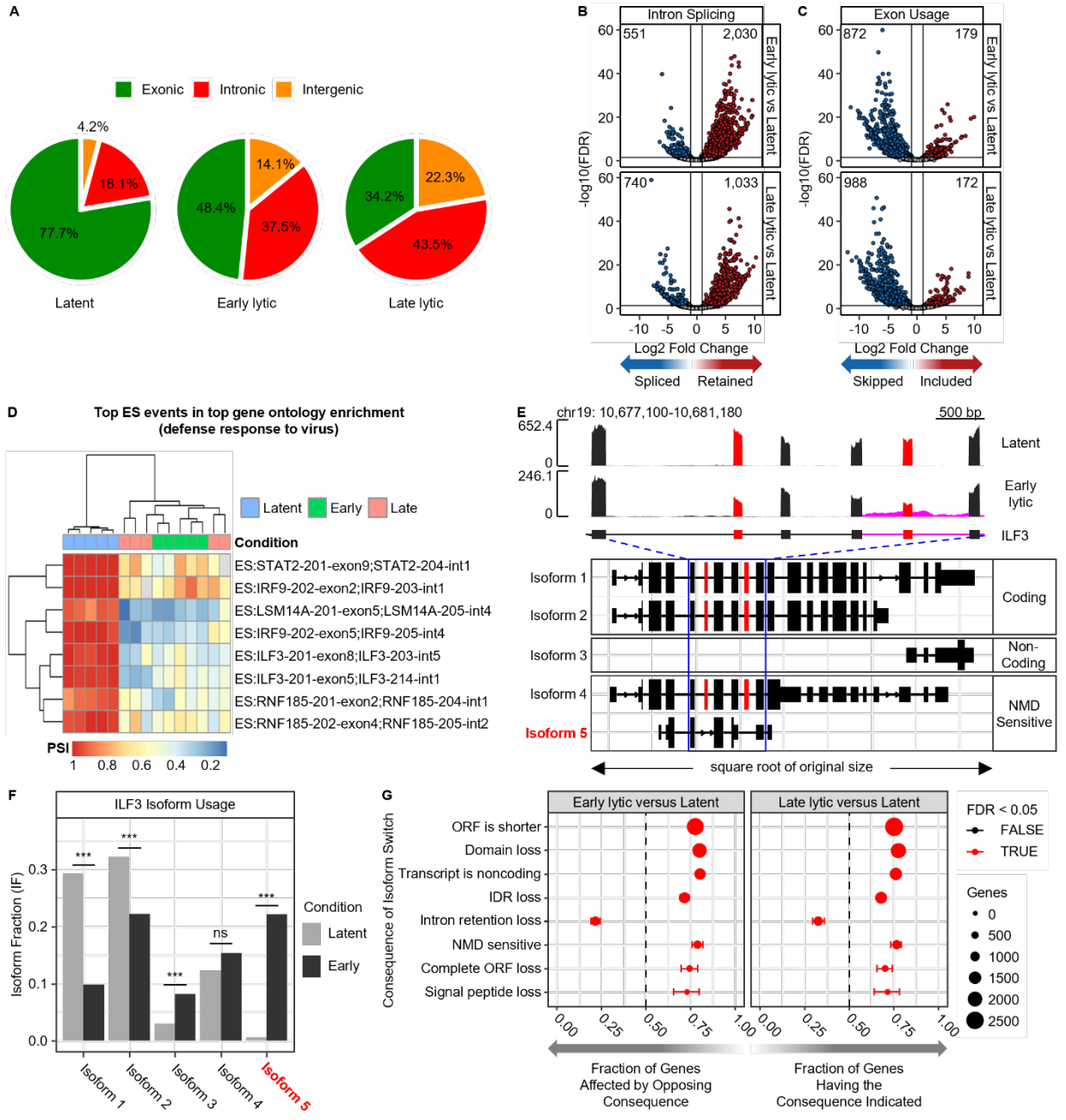


Figure 4.9 Comprehensive analysis of EBV lytic replication-induced splicing alterations reveals a systemic shift towards host shutoff programs. (A) Pie charts depicting distribution of reads mapping to exonic, intronic, or intergenic regions of host chromosomes in RNA-seq data from latent, early lytic, or late lytic fractions. (B-C) Volcano plots of differential (B) intron splicing or (C) exon usage events (fold change ≥ 2 and $P_{adjusted} \leq 0.05$) between early or late lytic fractions relative to latent fraction. (D) Heatmap presenting the top differential exon usage events of genes associated with the 'defense response to virus' category identified through GO enrichment analysis of genes with at least one differential exon usage event. ES exon skip; PSI percent spliced in. (E-F) (E) Coverage plot illustrating ILF3 gene exon skipping events (red)

identified in the preceding heatmap, accompanied by examples of intron retention (pink). Schematics of expressed isoforms are shown with their coding potential and NMD sensitivity. (F) Isoform switching in the ILF3 gene. The bar represents the standard error of the mean. (G) Genome-wide enrichment of isoform switch consequences in early or late lytic fractions relative to latent fraction. The opposing consequence of intron retention gain, intron retention loss, is significantly more likely during lytic replication relative to latency. The bar represents the standard error of the mean. *** p-value < 0.05, *ns* – no significant difference. *Isoform 1* ENST00000588657.6; *Isoform 2* ENST00000589998.6; *Isoform* 3 ENST00000586544.1; *Isoform 4* ENST00000587928.5; *Isoform 5* ENST00000589416.5; *ORF* open reading frame; *IDR* intrinsically disordered regions; *NMD* nonsense-mediated decay.

The observation of increased read alignment at intronic regions during EBV reactivation suggested inefficient splicing of host transcripts. To investigate whether the spliceosome was failing to remove introns properly, we used SpliceWiz [129] to identify differential alternative splicing events (fold change ≥ 2 and $P_{adjusted} \leq 0.05$) between lytic replication and latency (Figure 4.9B and Figure 4.9C). A significant increase in intron retention (IR) events was observed during the early lytic phase, with 2,030 IR events identified across 829 host genes. Interestingly, this trend subsided in late lytic cells, where only 549 host genes manifested 1,033 IR events. We also observed alterations in exon usage throughout the early and late lytic phases. In total, 642 host genes exhibited 872 exon skipping (ES) events in the early phase, while 697 host genes displayed 988 ES events in the late phase.

In an effort to define whether any specific pathways were affected by alternative splicing events, we conducted functional enrichment analysis of genes affected by IR and ES events. This analysis revealed that genes in the Gene Ontology (GO) category “defense response to virus” were significantly enriched ($P \leq 0.05$) among genes with ES events. The top differential ES events driving this enrichment are displayed as a heatmap in Figure 4.9D. This analysis identified several genes with known antiviral roles exhibiting altered splicing patterns. For example, coverage plots illustrated exon skipping and intron retention events within the ILF3

gene, a key antiviral effector required for efficient translation of type I IFNs and a subset of ISGs during the PKR response (Figure 4.9E) [168]. In an effort to predict the functional consequences of these alternative splicing events, we performed transcript-based analysis with IsoformSwitchAnalyzeR [131]. ILF3 alternative splicing events were predicted to be ILF3 inactivating due to frame-shifting mutants and susceptibility to nonsense-mediated decay (Figure 4.9E and Figure 4.9F). Global consequence analysis of host transcript isoform switches during EBV reactivation revealed a systemic bias towards non-functional transcripts (Figure 4.9G). Together, these findings suggest that in addition to quantitative displacement of host transcripts, EBV reactivation is associated with dysregulation of mRNA processing such that many of the host transcripts that persist no longer encode functional proteins.

We also investigated the basis of the increased intergenic transcription observed in early and late lytic cells. A likely mechanism for this phenomenon would be transcriptional readthrough by RNA polymerase II past usual polyadenylation sites into downstream regions, creating so called downstream-of-gene (DoG)-containing transcripts. DoGs observed during HSV-1 replication have been attributed to interference with 3' RNA processing by the HSV ICP27 protein [169,170]. Given that the SM protein (EBV homolog of ICP27) is expressed during early lytic replication, we sought to determine the extent of readthrough transcription in the early and late lytic cells using Automatic Readthrough Detection (ARTDeco) [138]. This analysis revealed extensive readthrough transcription, as evidenced by the TBC1D10A gene (Figure 4.10A) where transcription originating from TBC1D10A extends into the downstream CASTOR1 gene, which itself exhibits transcription termination failure and further readthrough into the downstream OIP5 gene. These transcriptional readthroughs into downstream genes, termed read-ins, have the additional consequence of artificially inflating apparent expression

levels of CASTOR1 and OIP5 (3.10B). On a global scale, early lytic replication induced a four-fold increase (554 vs. 136) in host DoG formation relative to latency, with a subsequent 14-fold decrease in DoGs during the late phase (Figure 4.10C). Nearly half of these DoGs during the early lytic phase gave rise to read-in transcripts (Figure 4.10D). While the decrease in DoGs in the late phase may suggest a reduced readthrough, it also suggests that these transcripts are highly unstable and unlikely to produce functional protein products.

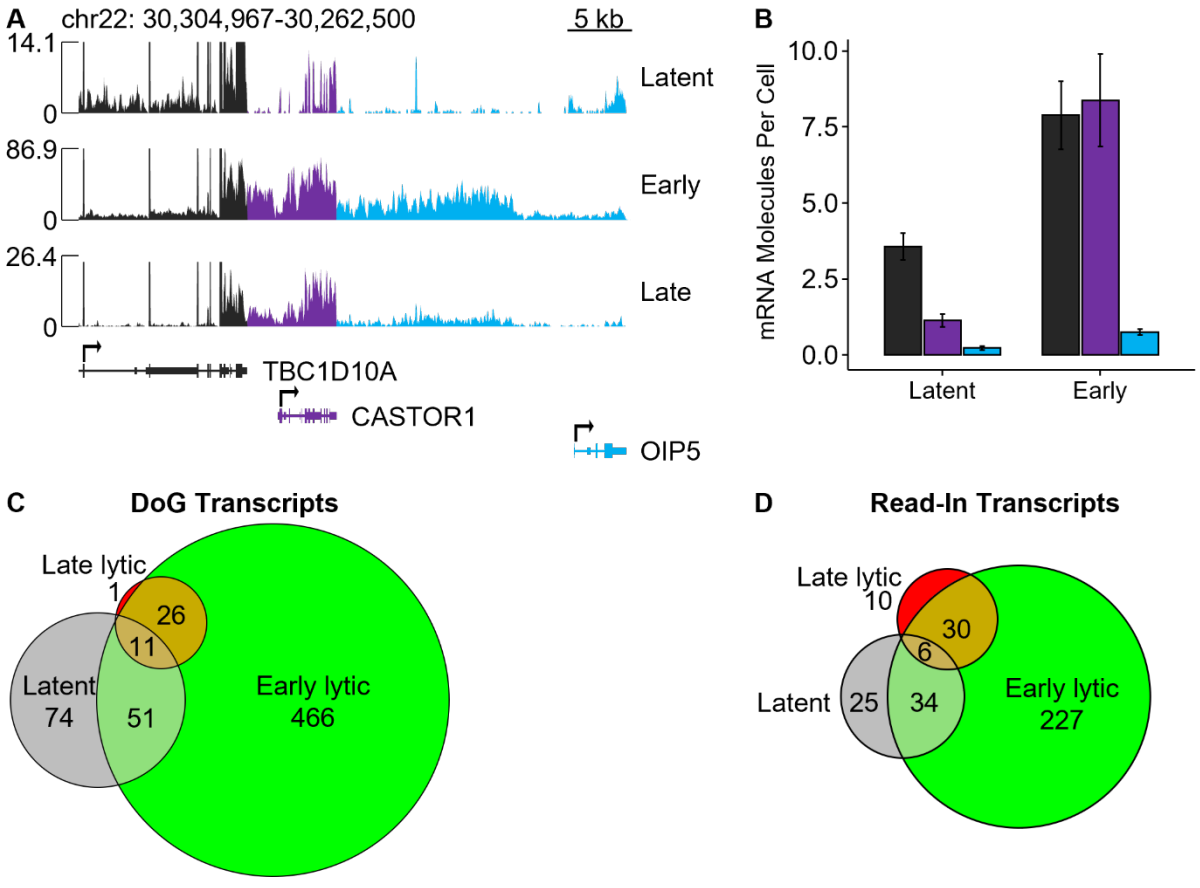


Figure 4.10 Global disruption of cellular transcription termination during EBV lytic replication. (A) RNA-seq read coverage showing example of TBC1D10A DoG transcript (purple) resulting in transcription read-in of the downstream CASTOR1 gene, and continued transcription readthrough (light blue) into the farther downstream OIP5 gene. Note that the coordinates of this locus were flipped to depict rightward transcription of these negative-stranded genes. (B) Expression of the TBC1D10A DoG transcript and the downstream read-in genes, CASTOR1 and OIP5. Data are represented as mean ± SEM. (C-D) Venn diagrams comparing

overlaps of host genes exhibiting (C) DoG-formation and (D) transcriptional read-in for each fraction.

This data reveals that EBV reactivation not only reduces host transcript levels but is associated with extensive dysregulation of transcript processing. Thus, many of the transcripts that appear to "survive" host shutoff may not encode functional protein products. Additionally, our data demonstrate that the expression levels of some genes may be overestimated as a consequence of readthrough transcription into downstream genes. As a result, even the markedly reduced levels of host mRNAs observed in early and late lytic cells (Figure 4.6) understate the extent to which the cell's ability to express new protein products in response to EBV reactivation is impaired.

Host shutoff induced by EBV reactivation is predominantly independent of BGLF5.

Although BGLF5 is believed to be the primary effector of EBV-induced host shutoff, multiple other mechanisms have been proposed to contribute to it (reviewed in [161]). To define the specific contribution of BGLF5 to the observed reductions in host transcripts during EBV reactivation, we constructed a BGLF5 knockout in the context of the DFLR EBV genome. Because of the extensive overlap of transcripts in this region of the EBV genome, we used a previously characterized BGLF5 deletion which removes amino acids 7 to 306, specifically abrogating all BGLF5 activity [88]. We then introduced DFLR EBVΔBGLF5 into HEK293 cells by bactofection which were used as a producer line for virions. Using these virions, we transformed primary B lymphocytes into DFLRΔBGLF5 LCLs which were used to study changes in viral and host transcription during EBV reactivation with our established methodology (Figure 4.3A).

Analysis of the viral transcriptome confirmed the BGLF5 deletion within the DFLR EBV genome and revealed expected alterations in viral gene expression patterns (Figure 4.11). Consistent with prior findings that EBV lytic transcripts are susceptible to BGLF5-dependent degradation [62,63,88], we observed a slight increase in early gene expression compared to WT virus. While late gene expression was preserved in the DFLRΔBGLF5 LCLs, there was a significant decrease in late gene expression levels compared to WT. Although BGLF5 is not known to play a direct role in late gene transcription, it is essential for processing of viral DNA for encapsidation and dysregulation of viral DNA processing may impact the efficiency of late gene transcription from newly replicated viral DNA [88,171-174].

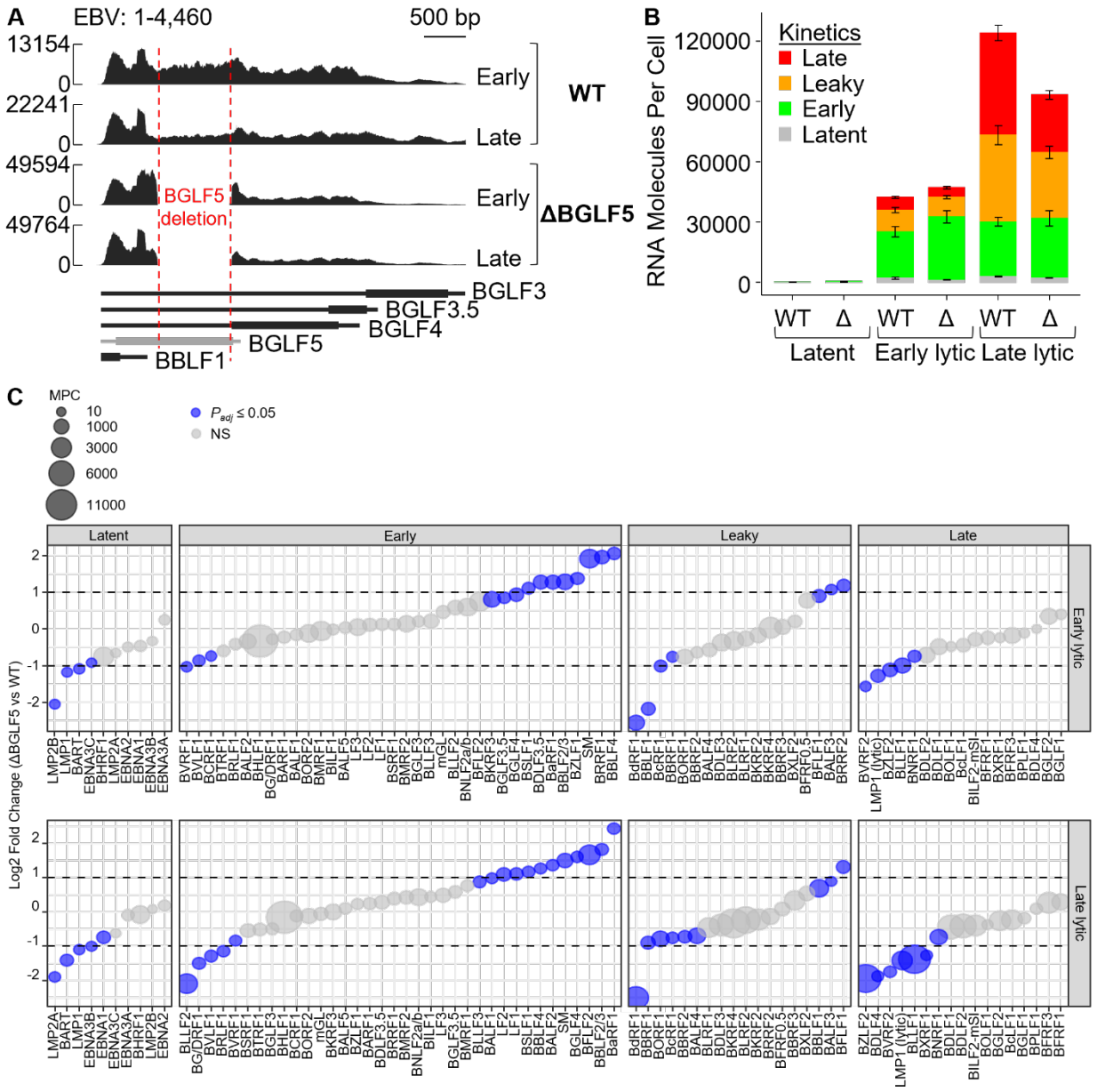


Figure 4.11 BGLF5-KO schematic and validation. (A) Schematic of leftward transcripts in the BamHI G region of the inverted EBV M81 strain genome (bottom) and averaged RNA-seq reads mapping to this region from lytic fractions of DFLR LCLs and DFLRΔBGLF5 LCLs (top) confirming the absence of BGLF5 in the latter. (B) Stacked bar graphs quantifying total EBV mRNAs (latent and 3 lytic classes) within latent, early lytic, and late lytic fractions from DFLR LCLs (WT) and DFLRΔBGLF5 LCLs (Δ). Data are represented as mean ± SEM. (C) Dot plot depicting differences in EBV encoded mRNAs in early or late lytic fractions of DFLRΔBGLF5 LCLs (ΔBGLF5) relative to DFLR LCLs (WT). Dot size represents the mRNA molecules per cell (MPC) between conditions and color the significance (blue = $P_{adjusted} \leq 0.05$; gray = nonsignificant). Genes are grouped by EBV kinetics (latent, early, leaky late, and late).

We next compared host transcript levels in early lytic and late lytic DFLRΔBGLF5 LCLs to those seen in WT DFLR LCLs. Remarkably, extensive host shutoff was observed even in the absence of BGLF5 (Figure 4.12A). This shutoff occurred at both early and late stages of lytic replication, with 83.4% and 90.3% of host genes downregulated, respectively. Interestingly, while the overall number of downregulated genes was similar to that observed with WT (85.2% early and 95.1% late), the magnitude of the reduction was significantly stronger with WT (Figure 4.12B). This difference was evident during the early stage (mean log2 fold change: -2.25 WT vs. -1.87 ΔBGLF5) and even larger during the late stage (mean log2 fold change: -3.97 WT vs. -2.88 ΔBGLF5).

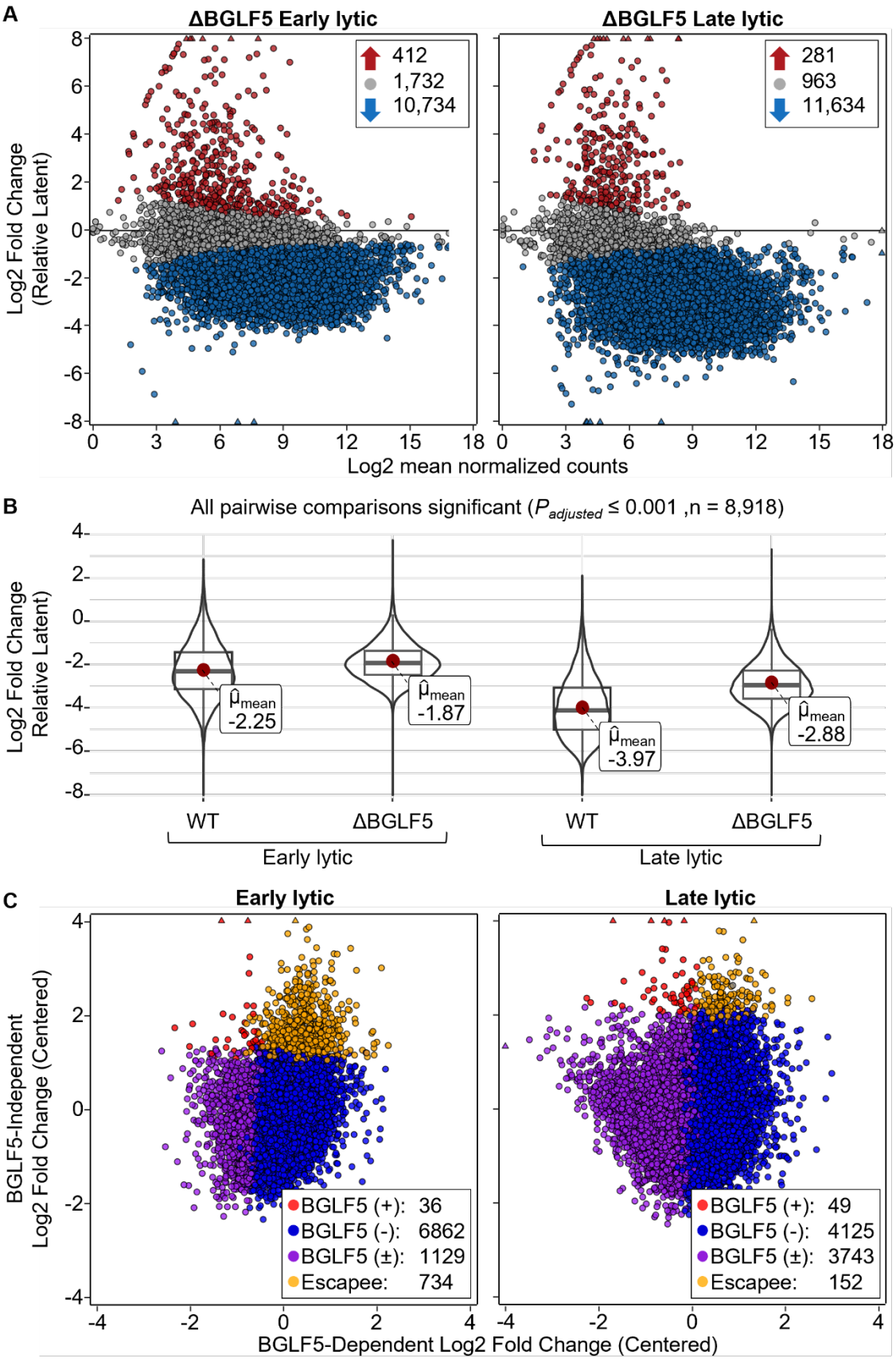

A
ΔBGLF5 Early lytic
ΔBGLF5 Late lytic
412
1,732
10,734
281
963
11,634
Log2 Fold Change
(Relative Latent)
Log2 mean normalized counts
B
All pairwise comparisons significant ($P_{adjusted}$ ≤ 0.001 ,n = 8,918)
Log2 Fold Change
Relative Latent)
$\hat{\mu}_{mean}$ -2.25
$\hat{\mu}_{mean}$ -1.87
$\hat{\mu}_{mean}$ -3.97
$\hat{\mu}_{mean}$ -2.88
WT
ΔBGLF5
WT
ΔBGLF5
Early lytic
Late lytic
C
Early lytic
Late lytic
BGLF5-Independent
Log2 Fold Change (Centered)
BGLF5 (+): 36
BGLF5 (-): 6862
BGLF5 (±): 1129
Escapee: 734
BGLF5 (+): 49
BGLF5 (-): 4125
BGLF5 (±): 3743
Escapee: 152
BGLF5-Dependent Log2 Fold Change (Centered)

Figure 4.12 Extensive BGLF5-independent host shutoff during EBV lytic replication. (A) MA plots demonstrating host gene expression fold changes in early (left panel) and late (right panel) lytic fractions of DFLRΔBGLF5 LCLs relative to the latent fraction of DFLRΔBGLF5 LCLs. The x-axis depicts average levels of expression for each gene between conditions being compared. Differential expressed genes ($P_{adjusted} \leq 0.05$) are indicated in red (upregulated) or blue (downregulated); genes without significant expression change are colored gray. (B) Distribution of fold change in host gene expression levels in early and late lytic fractions relative to the latent fraction for DFLR LCLs (WT) and DFLRΔBGLF5 LCLs (ΔBGLF5) visualized through superimposed box and violin plots. Data was filtered to genes expressed across all conditions. The mean log2 fold change in gene expression for each plot is denoted by red dots. The FDR-adjusted p-value is derived from a two-sided Welch's test. (C) Scatter plots comparing BGLF5-dependent (x-axis) and BGLF5-independent (y-axis) log2 fold changes (centered) in host gene expression in the early and late lytic fractions. BGLF5-dependent: DFLR LCLs versus DFLRΔBGLF5 LCLs; BGLF5-independent: early or late lytic fraction relative to latent fraction in DFLRΔBGLF5 LCLs. Genes are colored based on their BGLF5-dependence: red (BGLF5-dependent), blue (BGLF5-independent), purple (both), yellow (not significantly downregulated). Triangles indicate data points exceeding the displayed range.

To further quantify the contribution of BGLF5 to host shutoff, we imputed the relative magnitudes of BGLF5-dependent changes (by comparing WT lytic vs. ΔBGLF5 lytic) and BGLF5-independent changes (comparing ΔBGLF5 lytic vs. ΔBGLF5 latent) and plotted these results as a log-log plot (Figure 4.12C). As before, BGLF5-independent mechanisms emerged as the primary drivers of host shutoff, contributing to the downregulation 99.6% and 99.4% genes during the early and late phases, respectively. By contrast, the BGLF5 contribution was relatively modest, accounting for only 14.5% downregulation in early replication. These data also suggest that BGLF5 makes a larger contribution to host gene downregulation during the late phase (47.9%); however, we cannot exclude the possibility that some of this effect is not specifically attributable to impaired host shutoff as it might also be a consequence of impaired late gene expression observed with the BGLF5 mutant (Figure 4.11B and Figure 4.11C). Notably, we still observed extensive defects in mRNA transcript processing (Figure 4.13), suggesting that intron retention, exon skipping, and readthrough transcription that occur during EBV replication occur independent of BGLF5. Interestingly, when we examined the role of

RNA half-life and abundance on BGLF5-independent shutoff, there was no longer a correlation with RNA half-life (Figure 4.14). Thus, BGLF5 itself appears to be primarily responsible for the elimination of long-lived mRNA transcripts.

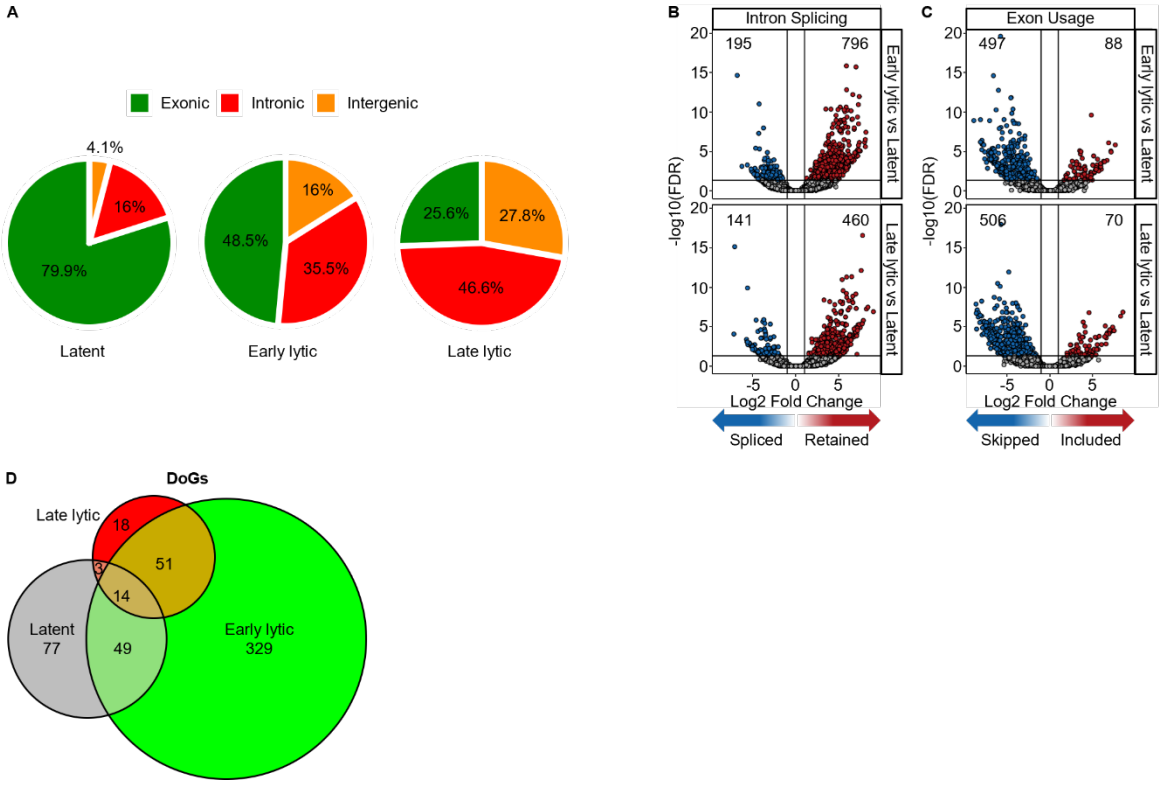


Figure 4.13 Global dysregulation of mRNA processing in DFLRΔBGLF5 LCLs. (A) Pie charts depicting distribution of reads mapping to exonic, intronic, or intergenic regions of host chromosomes in RNA-seq data from latent, early lytic, or late lytic fractions. (B-C) Volcano plots of differential (B) intron splicing or (C) exon usage events (fold change ≥ 2 and $P_{adjusted}$ ≤ 0.05) between early or late lytic fractions relative to latent fraction. (D) Venn diagrams comparing overlaps of host genes exhibiting DoG-formation for each fraction. Triangles indicate data points exceeding the displayed range.

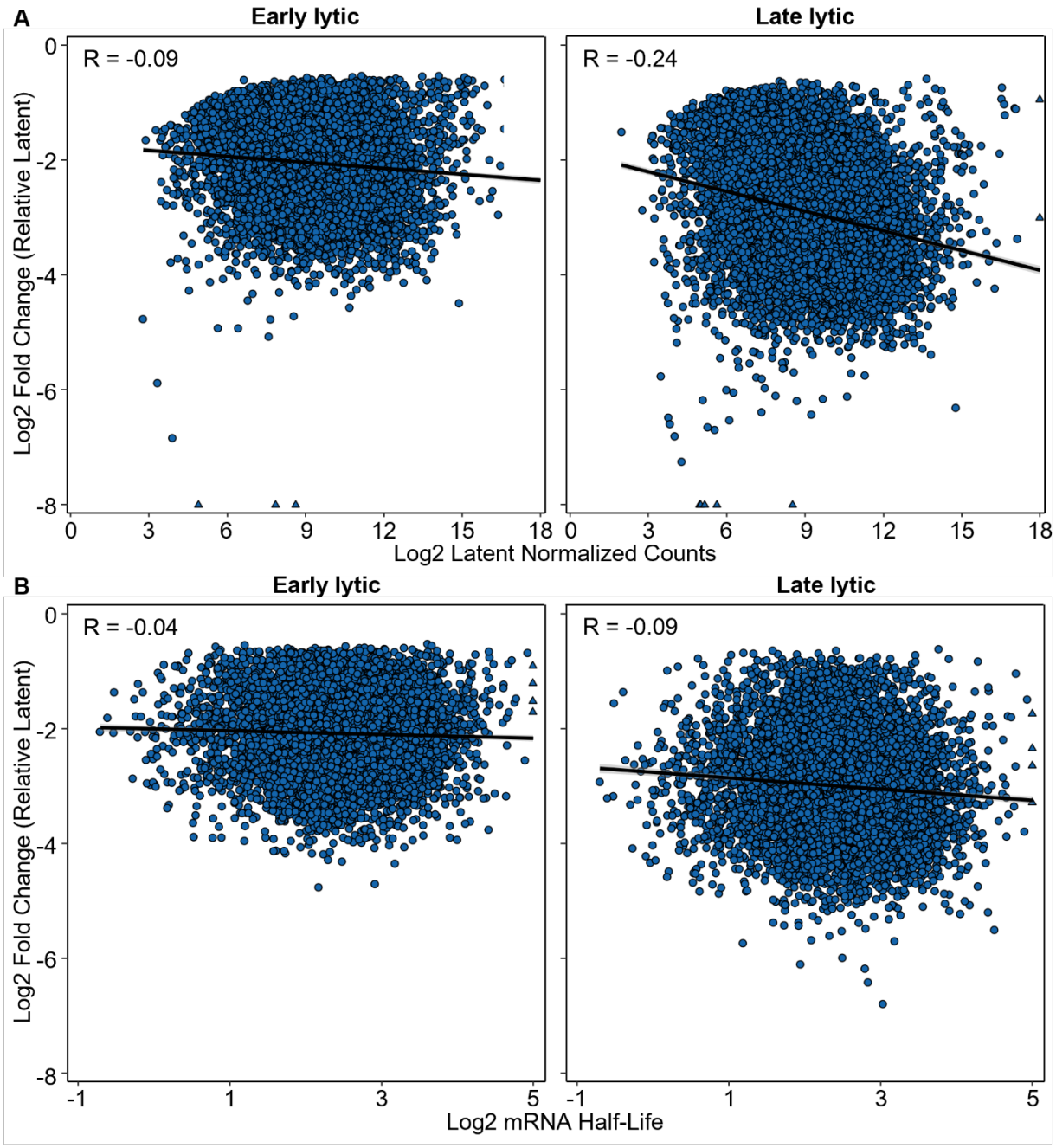


Figure 4.14 Relationship of mRNA levels during latency or mRNA stability in LCLs with mRNA downregulation during early or late lytic EBV replication in DFLRΔBGLF5 LCLs. (A-B) Scatterplots showing fold changes of significantly downregulated genes ($P_{adjusted} \leq 0.05$) in early (left panel) and late (right panel) lytic DFLRΔBGLF5 LCLs relative to latently infected DFLRΔBGLF5 LCLs. The x-axis depicts (A) expression levels for each gene during the latent state or (B) mRNA half-lives in LCLs from Duan et al. The Pearson correlation coefficient (R) is indicated.

In summary, although we did observe some host genes to be specifically shutoff by BGLF5, our most striking observation is that host shutoff is largely preserved when BGLF5 is deleted from the EBV genome. Thus, BGLF5-independent pathways emerge as the predominant mediators of host shutoff during EBV reactivation from their B cell latency reservoir.

Chapter 5

Discussion and Future Directions

Discussion

A major barrier to the study of EBV lytic replication is its inefficient and asynchronous reactivation from latency. Although single cell approaches can address some of these limitations, this generally incurs the cost of reduced sensitivity/depth. In this manuscript we report a powerful tool for conducting such investigations under physiologic conditions: a recombinant dual-fluorescent lytic reporter (DFLR) EBV constructed in the M81 strain. To the best of our knowledge, DFLR EBV retains all of the activities of wild-type EBV and this allowed us to transform primary B lymphocytes into DFLR LCLs. We demonstrate that FACS can be used to isolate subpopulations of latent, early lytic, and late lytic DFLR LCLs that are substantially purer than obtained with conventional gp350 based sorting. We then used RNA-seq to produce the most detailed characterization reported to date of the transcriptome remodeling that occurs in LCLs during EBV reactivation. We have previously argued that host shutoff must be studied under physiologic conditions because BGLF5 overexpression studies may overestimate the degree of host shutoff and miss key viral and cellular regulatory factors [161]. Indeed, we identified specific host transcripts that persist (and a few that are even upregulated) in LCLs during EBV reactivation. Although this number was not large, our results also demonstrated that the vast majority of host transcripts are decreased by BGLF5-independent mechanisms. Additionally, we observed that many host transcripts are rendered non-coding as a consequence of extensive dysregulation of mRNA processing. These additional effects underscore the importance of studying the host shutoff phenomenon in the context of the whole viral genome and the robustness of EBV's ability to counter host responses that require transcription.

EBV induces widespread host shutoff during lytic replication

Host shutoff has evolved independently in multiple virus families and the mechanisms and effects vary widely, even among the human herpesviruses [175]. For example, in HSV-1 the host shutoff protein is packaged in virions, in gammaherpesviruses the SOX/BGLF5 shutoff factors are expressed in cells supporting lytic replication, and betaherpesviruses are not known to exert host shutoff. Thus, characterizing host shutoff is essential to understanding its unique role in a given virus. Our study of EBV reactivation in LCLs revealed widespread downregulation of most host genes, resulting in a significant decrease in host mRNA abundance (Figure 4.6A-C). Even accounting for the presence of viral transcripts, there was a substantial decrease in the total mRNA in the cell, suggesting that EBV host shutoff is not functioning primarily to divert resources from the cell. Instead, EBV host shutoff appears to act primarily to prevent changes in the cellular proteome in response to reactivation from latency. A widespread, but less pronounced host shutoff was observed in Raji BL cells [41] (see Figure 4.8A-C for comparison). This discrepancy is likely technical as the estimated fraction of Raji cells that can be induced into lytic replication (~56%) is far less than the purity achieved with DFLR LCLs (~99%) [166]. As we have shown, the presence of non-lytic cells substantially reduces sensitivity for detecting reductions in host transcripts that accompany lytic replication. Moreover, our DFLR EBV allows effects in early and late lytic cells to be studied separately and we indeed observed that host shutoff intensifies in the late phase commensurate with a marked increase in late viral gene expression.

BGLF5-independent mechanisms are the predominant drivers of host shutoff

The most surprising finding of our study is that the viral BGLF5 nuclease does not account for the majority of host shutoff observed during early lytic replication. We constructed

the BGLF5 knockout in DFLR EBV with the intention of demonstrating the specific actions of BGLF5, independent of other mechanisms which might also interfere with host transcription. The finding that BGLF5 deletion resulted in only a modest reduction in host gene downregulation compared to WT virus during reactivation implicates BGLF5-independent mechanisms as the primary drivers of host shutoff, a finding that necessitates a reevaluation of the current paradigm.

Our findings underscore the importance of studying effects of viral genes in the context of the entire viral genome. Expression of individual viral genes led to the crucial discovery of SOX/BGLF5 as mediators of host shutoff in gammaherpes viruses, but also to the impression that other factors were bit players in this effect. Many viral proteins function as members of a viral protein complex or require modification (e.g., phosphorylation) by other viral factors for their activity. Thus, screens examining effects of individual viral proteins, while valuable, are no substitute for studying a virus in context. An important avenue for future studies will be to define the specific mechanisms of BGLF5-independent host shutoff. Likely candidates include disruption of mRNA processing by EBV SM protein, alterations of chromatin accessibility by Zta, and the reorganization of chromatin that accompanies formation of viral replication compartments. In addition, cellular stress or antiviral responses to EBV replication likely make important contributions to host shutoff. For example, although EBV encodes factors such as BHRF1 to counter apoptosis, early apoptotic events, such as mitochondrial outer membrane depolarization (MOMP) that release the PNPT1 exonuclease into the cytoplasm may contribute to host mRNA degradation [176]. Alternatively, activation of antiviral responses such as the OAS/RNase L pathway induced by dsRNA can lead to host shutoff through mRNA degradation [177]. Indeed, dsRNA has been observed in cells undergoing EBV lytic replication [178].

Though beyond the scope of this study, we believe that DFLR EBV will be an important tool for investigating these and other BGLF5-independent mechanisms.

Many host mRNAs expressed during EBV reactivation are dysfunctional

We observed widespread dysregulation of host splicing and mRNA processing during EBV reactivation from LCLs. We confirmed that the increased read mapping to intronic regions was due to intron retention and exon skipping (Figure 4.9A-C), resulting in the production of thousands of non-functional transcripts (Figure 4.9D-G). Notably, this splicing inhibition contrasts with the effects of other host shutoff-inducing viruses like HSV-1 and influenza. While HSV-1 doesn't seem to affect splicing [169,179], influenza infection increases it [180]. Our observation that splicing dysregulation specifically affects genes involved in the cellular defense against viruses (Figure 4.9D) is striking. It seems unlikely that this set of genes is specifically targeted for disruption, but rather these genes are expressed in response to viral replication. Nascent transcripts are subject to splicing disruption, whereas many of the "normal" transcripts from other pathways may have accumulated in the cytoplasm prior to EBV reactivation. Thus, splicing defects (and readthrough transcription discussed below) via various mechanisms including defective nuclear export, impaired stability, or compromised coding potential also serve to "shutoff" host transcriptional responses to EBV reactivation [181,182].

We observed a marked increase in intergenic reads induced by EBV reactivation that was partially attributable to readthrough transcription (Figure 4.10). Readthrough transcription has been observed in cells infected with many different viruses [169,179,183-185]. In the case of HSV-1 infection, this phenomenon has been linked to ICP27 [169], a protein conserved across herpesviruses, which disrupts 3' mRNA processing by interfering with the host cleavage factor CPSF [170]. The extent to which this activity is shared with SM, the EBV homolog of ICP27,

remains an open question. Notably, in contrast to HSV-1 infection where readthrough transcription increases at late infection stages, we observed a decrease in readthrough during the late phase of EBV lytic replication.

Readthrough transcription also occurs in response to a variety of cellular stressors, such as heat shock [186-189], osmotic stress [182,183,187,189], RNase L activation [190], and transcription inhibition [191]. Interestingly, heat shock [188,192] and RNase L activation also trigger splicing inhibition, especially intron retention [190]. These intron-retaining mRNAs, along with readthrough transcripts (DoGs), are held untranslated within the nucleus [181,182]. These observations call into question the extent to which disrupted splicing during EBV reactivation is instigated by the virus to thwart cellular responses versus an innate immune response by the cell. In this latter case, proteins such as ICP27 and SM may have evolved primarily to overcome these cell defenses and ensure viral transcripts are exported rather than impede export of cell transcripts.

Transcriptional readthrough is unlikely to account for all the intergenic reads we observed during EBV reactivation. For one thing, many of these reads map to gene deserts and are not easily attributed to failure to terminate normal transcription. Additionally, many readthrough transcripts appear to have very short half-lives and are less abundant in late lytic cells than in early (see Figure 4.10A, for example), yet the proportion of intergenic reads increases in late lytic cells (Figure 4.9A). One potential mechanism to explain these reads was recently reported by Ungerleider et al., who demonstrated that the viral pre-initiation complex (vPIC) responsible for late gene transcription can also promote transcription from TATT boxes in intergenic regions of the host cell [193]. This has also been proposed as an additional

mechanism of shutoff via transcriptional interference. Notably, this mechanism may account for some of the BGLF5-independent effects on transcript shutoff we report here.

Evidence of host and early viral gene transcriptional suppression during the late phase

One unique feature of our study is DFLR EBV allowed us to characterize a pure population of cells undergoing EBV late lytic transcription without contamination from cells still in the early lytic phase. Our data suggest there is little to no early gene transcription during the late phase, but a marked increase in late and leaky late transcripts (Figure 4.3C). This finding is consistent with a model in which early and late transcription are mutually exclusive and we have previously observed that early transcripts accumulate to greater levels when late transcription is blocked [87]. Because late transcription occurs from unchromatinized, newly synthesized viral DNA within replication compartments (reviewed in [30]), key transcriptional machinery may be usurped from chromatinized compartments of the nucleus to support late gene expression as several studies of late gene transcription have suggested [29,194]. Alternatively, the loss of early gene expression may be a consequence of viral DNA replication. A recent study proposes a transcription-replication handover model in which early gene expression ceases upon onset of DNA replication due to epigenetic mechanisms [103]. In any event, the late phase of EBV replication is also characterized by extensive reorganization of chromatin such that it becomes progressively compacted against the nuclear periphery as viral replication compartments enlarge [28,35]. It is unlikely that this highly compacted chromatin could support early gene expression even if some EBV genome copies remained in this nuclear compartment.

Some of our observations suggest host gene expression also ceases with the onset of late gene transcription. First, the accelerated decline of host transcripts in our late cells suggests an impairment of transcript production is superimposed upon host transcript degradation. Second, as

previously mentioned, the marked decrease in transcripts with splicing defects and readthrough in late cells relative to early cells is more likely to represent reduced transcription initiation than normalization of transcript processing. Finally, progressive compaction of the chromatin during the late phase should also lead to cessation of host transcription. While our results do not definitely establish this proposition, DFLR EBV enabled sorting combined with SLAM-seq or other measures of active transcription should allow this issue to be definitively addressed. Further, mutations in DFLR EBV that impair DNA replication or late gene transcription should be able to distinguish between the early-late versus the transcription-replication handover models discussed above.

A proteomic and transcriptomic discrepancy

The extensive host shutoff observed during lytic replication raises an important question: to what extent can the transcriptome be regarded as an accurate proxy for protein expression during EBV lytic replication? While alterations in mRNA can be helpful in predicting changes at the protein level, unique aspects of lytic replication cast doubt on this frequent assumption. A downregulation of protein mirroring the drastic decline in mRNA would almost certainly be incompatible with virion morphogenesis. Indeed, viruses employing host shutoff tend to have short replication cycles which may reflect the need to outpace the cellular protein degradation machinery. For example, within herpesviruses, the beta subfamily including CMV lack host shutoff and exhibit markedly slower replication (~72-120 hours). In contrast, viruses that exhibit host shutoff all exhibit relatively short replication cycles (<24 h), including alpha [195] and gamma [104] herpesviruses. Furthermore, a proteomic analysis comparing gp350-enriched and gp350-depleted EBV-infected B cells revealed minimal proteome-level alterations [71]. We hope that future studies using DFLR EBV will be more sensitive for detecting proteomic changes

during EBV reactivation, but currently available data offer clear indications that they are less extensive than transcriptomic changes.

DFLR EBV is a versatile tool

Although the current study is focused on host shutoff, we believe that DFLR EBV based sorting will be broadly applicable to the study of EBV replication. Inefficient and asynchronous entry into lytic replication are ubiquitous in all models of EBV infection. We have demonstrated that DFLR EBV offers substantial improvements in enrichment sensitivity and specificity for lytic cells. Furthermore, its dual-fluorescent reporters allow sub-fractionation into cells supporting early or late lytic replication. A key advantage of incorporating these reporter genes into the EBV genome is that DFLR EBV can, with relative ease, be introduced into any cell type or animal model of infection. A particularly promising avenue is that DFLR EBV can be used in high throughput screens, offering the potential to identify drugs and/or drug targets for lytic induction therapy of EBV tumors. We anticipate that a broad range of transcriptomic, proteomic, and biochemical assays will benefit from this new tool.

Future directions

A key finding of my research was the extensive downregulation of the host transcriptome during EBV lytic replication even without the viral BGFL5 nuclease. This suggests that BGLF5 is only one host shutoff mechanism among many; the nature of these warrants further investigation. Two cellular nucleases, RNase L and PNPT1, are promising candidates due to their host shutoff-like effects during viral infections and apoptosis, respectively. Using CRISPR-Cas9 to generate DFLRΔBGLF5 LCLs deficient in either RNase L or PNPT1, followed by RNA-seq analysis, will directly assess these nucleases' contribution to host shutoff during EBV lytic replication much as the experiments described here defined the specific role of BGLF5. By

inhibiting host shutoff, future studies can explore key questions such as: Is host shutoff (BGLF5-mediated or otherwise) essential for successful viral replication? If so, what are the critical host defenses that each mechanism subverts to sustain viral replication?

Another significant finding is the global disruption of cellular mRNA processing during EBV lytic replication, manifesting as readthrough transcription and aberrant splicing events. To determine the fate and potential functionality of these dysregulated transcripts, future studies should combine subcellular fractionation and RNA-seq to pinpoint their localization, with ribosome profiling to identify actively translated transcripts. To uncover the mechanism behind splicing inhibition, target RNA capture followed by mass spectrometry offers a sensitive approach for detecting changes in splicing factor binding patterns. Functional validation can be achieved through knockdown/knockout or overexpression of candidate splicing factors to observe if they can replicate or restore splicing patterns, respectively.

To investigate readthrough transcription, a comparative study is necessary, examining readthrough levels during EBV lytic replication, SM overexpression, and cellular stress. This will clarify SM's role, whether the pattern is virus-specific, and if EBV exploits cellular stress response pathways. Additionally, the role of the viral pre-initiation complex (vPIC) in aberrant transcription warrants exploration. Knockout of vPIC will shed light on the extent of its contribution to the observed mis-splicing and readthrough events.

Lastly, a vast knowledge gap persists regarding the late phase of EBV lytic replication. While our study suggests ongoing cellular transcription during the early phase, a dramatic shutoff occurs by the late phase. Nascent transcriptomics comparing latent, early lytic, and late lytic stages can directly quantify the extent of transcriptional shutoff during these phases. Additionally, coupling this method with ATAC-seq analysis will reveal how chromatin

reorganization observed in the late phase impacts chromatin accessibility and potentially suppresses transcription. Finally, proteomic studies are crucial for a deeper understanding of EBV lytic replication overall, particularly the poorly understood formation of viral replication compartments. These combined approaches will significantly enhance our understanding of the intricate mechanisms underlying EBV-mediated disruption of cellular processes during lytic replication.

References

1. Wong, Y., Meehan, M.T., Burrows, S.R., Doolan, D.L., and Miles, J.J. (2022). Estimating the global burden of Epstein-Barr virus-related cancers. J Cancer Res Clin Oncol *148*, 31-46. 10.1007/s00432-021-03824-y.
2. de Martel, C., Georges, D., Bray, F., Ferlay, J., and Clifford, G.M. (2020). Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. Lancet Glob Health *8*, e180-e190. 10.1016/S2214-109X(19)30488-7.
3. Bjornevik, K., Cortese, M., Healy, B.C., Kuhle, J., Mina, M.J., Leng, Y., Elledge, S.J., Niebuhr, D.W., Scher, A.I., Munger, K.L., and Ascherio, A. (2022). Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. Science *375*, 296-301. 10.1126/science.abj8222.
4. Rosemarie, Q., and Sugden, B. (2020). Epstein-Barr Virus: How Its Lytic Phase Contributes to Oncogenesis. Microorganisms *8*. 10.3390/microorganisms8111824.
5. Ma, S.D., Hegde, S., Young, K.H., Sullivan, R., Rajesh, D., Zhou, Y., Jankowska-Gan, E., Burlingham, W.J., Sun, X., Gulley, M.L., et al. (2011). A new model of Epstein-Barr virus infection reveals an important role for early lytic viral protein expression in the development of lymphomas. J Virol *85*, 165-177. 10.1128/JVI.01512-10.
6. Hong, G.K., Gulley, M.L., Feng, W.H., Delecluse, H.J., Holley-Guthrie, E., and Kenney, S.C. (2005). Epstein-Barr virus lytic infection contributes to lymphoproliferative disease in a SCID mouse model. J Virol *79*, 13993-14003. 10.1128/JVI.79.22.13993-14003.2005.
7. Yiu, S.P.T., Dorothea, M., Hui, K.F., and Chiang, A.K.S. (2020). Lytic Induction Therapy against Epstein-Barr Virus-Associated Malignancies: Past, Present, and Future. Cancers (Basel) *12*. 10.3390/cancers12082142.
8. Miyashita, E.M., Yang, B., Lam, K.M., Crawford, D.H., and Thorley-Lawson, D.A. (1995). A novel form of Epstein-Barr virus latency in normal B cells in vivo. Cell *80*, 593-601. 10.1016/0092-8674(95)90513-8.
9. Young, L.S., and Rickinson, A.B. (2004). Epstein-Barr virus: 40 years on. Nat Rev Cancer *4*, 757-768. 10.1038/nrc1452.
10. Tugizov, S.M., Herrera, R., and Palefsky, J.M. (2013). Epstein-Barr virus transcytosis through polarized oral epithelial cells. J Virol *87*, 8179-8194. 10.1128/JVI.00443-13.
11. Tugizov, S.M., Berline, J.W., and Palefsky, J.M. (2003). Epstein-Barr virus infection of polarized tongue and nasopharyngeal epithelial cells. Nat Med *9*, 307-314. 10.1038/nm830.
12. Babcock, G.J., Hochberg, D., and Thorley-Lawson, A.D. (2000). The expression pattern of Epstein-Barr virus latent genes in vivo is dependent upon the differentiation stage of the infected B cell. Immunity *13*, 497-506. 10.1016/s1074-7613(00)00049-2.
13. SoRelle, E.D., Dai, J., Reinoso-Vizcaino, N.M., Barry, A.P., Chan, C., and Luftig, M.A. (2022). Time-resolved transcriptomes reveal diverse B cell fate trajectories in the early response to Epstein-Barr virus infection. Cell Rep *40*, 111286. 10.1016/j.celrep.2022.111286.
14. Thorley-Lawson, D.A. (2001). Epstein-Barr virus: exploiting the immune system. Nat Rev Immunol *1*, 75-82. 10.1038/35095584.
15. Laichalk, L.L., Hochberg, D., Babcock, G.J., Freeman, R.B., and Thorley-Lawson, D.A. (2002). The dispersal of mucosal memory B cells: evidence from persistent EBV infection. Immunity *16*, 745-754. 10.1016/s1074-7613(02)00318-7.

16. Laichalk, L.L., and Thorley-Lawson, D.A. (2005). Terminal differentiation into plasma cells initiates the replicative cycle of Epstein-Barr virus in vivo. J Virol *79*, 1296-1307. 10.1128/JVI.79.2.1296-1307.2005.
17. Borza, C.M., Morgan, A.J., Turk, S.M., and Hutt-Fletcher, L.M. (2004). Use of gHgL for attachment of Epstein-Barr virus to epithelial cells compromises infection. J Virol *78*, 5007-5014. 10.1128/jvi.78.10.5007-5014.2004.
18. Temple, R.M., Zhu, J., Budgeon, L., Christensen, N.D., Meyers, C., and Sample, C.E. (2014). Efficient replication of Epstein-Barr virus in stratified epithelium in vitro. Proc Natl Acad Sci U S A *111*, 16544-16549. 10.1073/pnas.1400818111.
19. Hayman, I.R., Temple, R.M., Burgess, C.K., Ferguson, M., Liao, J., Meyers, C., and Sample, C.E. (2023). New insight into Epstein-Barr virus infection using models of stratified epithelium. PLoS Pathog *19*, e1011040. 10.1371/journal.ppat.1011040.
20. Jiang, R., Scott, R.S., and Hutt-Fletcher, L.M. (2006). Epstein-Barr virus shed in saliva is high in B-cell-tropic glycoprotein gp42. J Virol *80*, 7281-7283. 10.1128/JVI.00497-06.
21. Valencia, S.M., and Hutt-Fletcher, L.M. (2012). Important but differential roles for actin in trafficking of Epstein-Barr virus in B cells and epithelial cells. J Virol *86*, 2-10. 10.1128/JVI.05883-11.
22. Chesnokova, L.S., Jiang, R., and Hutt-Fletcher, L.M. (2015). Viral Entry. Curr Top Microbiol Immunol *391*, 221-235. 10.1007/978-3-319-22834-1_7.
23. Chen, J., and Longnecker, R. (2019). Epithelial cell infection by Epstein-Barr virus. FEMS Microbiol Rev *43*, 674-683. 10.1093/femsre/fuz023.
24. SoRelle, E.D., Dai, J., Bonglack, E.N., Heckenberg, E.M., Zhou, J.Y., Giamberardino, S.N., Bailey, J.A., Gregory, S.G., Chan, C., and Luftig, M.A. (2021). Single-cell RNA-seq reveals transcriptomic heterogeneity mediated by host-pathogen dynamics in lymphoblastoid cell lines. Elife *10*. 10.7554/eLife.62586.
25. Young, L.S., Lau, R., Rowe, M., Niedobitek, G., Packham, G., Shanahan, F., Rowe, D.T., Greenspan, D., Greenspan, J.S., Rickinson, A.B., and Farrell, P.J. (1991). Differentiation-associated expression of the Epstein-Barr virus BZLF1 transactivator protein in oral hairy leukoplakia. J Virol *65*, 2868-2874. 10.1128/JVI.65.6.2868-2874.1991.
26. Kenney, S.C., and Mertz, J.E. (2014). Regulation of the latent-lytic switch in Epstein-Barr virus. Semin Cancer Biol *26*, 60-68. 10.1016/j.semcancer.2014.01.002.
27. Caragliano, E., Brune, W., and Bosse, J.B. (2022). Herpesvirus Replication Compartments: Dynamic Biomolecular Condensates? Viruses *14*. 10.3390/v14050960.
28. Rosemarie, Q., Kirschstein, E., and Sugden, B. (2023). How Epstein-Barr Virus Induces the Reorganization of Cellular Chromatin. mBio *14*, e0268622. 10.1128/mbio.02686-22.
29. Djavadian, R., Chiu, Y.F., and Johannsen, E. (2016). An Epstein-Barr Virus-Encoded Protein Complex Requires an Origin of Lytic Replication In Cis to Mediate Late Gene Transcription. PLoS Pathog *12*, e1005718. 10.1371/journal.ppat.1005718.
30. Chakravorty, A., Sugden, B., and Johannsen, E.C. (2019). An Epigenetic Journey: Epstein-Barr Virus Transcribes Chromatinized and Subsequently Unchromatinized Templates during Its Lytic Cycle. J Virol *93*. 10.1128/JVI.02247-18.
31. Gruffat, H., Kadjouf, F., Mariame, B., and Manet, E. (2012). The Epstein-Barr virus BcRF1 gene product is a TBP-like protein with an essential role in late gene expression. J Virol *86*, 6023-6032. 10.1128/JVI.00159-12.
32. Wyrwicz, L.S., and Rychlewski, L. (2007). Identification of Herpes TATT-binding protein. Antiviral Res *75*, 167-172. 10.1016/j.antiviral.2007.03.002.

33. Aubry, V., Mure, F., Mariame, B., Deschamps, T., Wyrwicz, L.S., Manet, E., and Gruffat, H. (2014). Epstein-Barr virus late gene transcription depends on the assembly of a virus-specific preinitiation complex. J Virol *88*, 12825-12838. 10.1128/JVI.02139-14.
34. Watanabe, T., Narita, Y., Yoshida, M., Sato, Y., Goshima, F., Kimura, H., and Murata, T. (2015). The Epstein-Barr Virus BDLF4 Gene Is Required for Efficient Expression of Viral Late Lytic Genes. J Virol *89*, 10120-10124. 10.1128/JVI.01604-15.
35. Chiu, Y.F., Sugden, A.U., and Sugden, B. (2013). Epstein-Barr viral productive amplification reprograms nuclear architecture, DNA replication, and histone deposition. Cell Host Microbe *14*, 607-618. 10.1016/j.chom.2013.11.009.
36. Frey, T.R., Brathwaite, J., Li, X., Burgula, S., Akinyemi, I.A., Agarwal, S., Burton, E.M., Ljungman, M., McIntosh, M.T., and Bhaduri-McIntosh, S. (2020). Nascent Transcriptomics Reveal Cellular Prolytic Factors Upregulated Upstream of the Latent-to-Lytic Switch Protein of Epstein-Barr Virus. J Virol *94*. 10.1128/JVI.01966-19.
37. Yuan, J., Cahir-McFarland, E., Zhao, B., and Kieff, E. (2006). Virus and cell RNAs expressed during Epstein-Barr virus replication. J Virol *80*, 2548-2565. 10.1128/JVI.80.5.2548-2565.2006.
38. Arvey, A., Tempera, I., Tsai, K., Chen, H.S., Tikhmyanova, N., Klichinsky, M., Leslie, C., and Lieberman, P.M. (2012). An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. Cell Host Microbe *12*, 233-245. 10.1016/j.chom.2012.06.008.
39. Ramasubramanyan, S., Osborn, K., Al-Mohammad, R., Naranjo Perez-Fernandez, I.B., Zuo, J., Balan, N., Godfrey, A., Patel, H., Peters, G., Rowe, M., et al. (2015). Epstein-Barr virus transcription factor Zta acts through distal regulatory elements to directly control cellular gene expression. Nucleic Acids Res *43*, 3563-3577. 10.1093/nar/gkv212.
40. Auburn, H., Zuckerman, M., and Smith, M. (2016). Analysis of Epstein-Barr virus and cellular gene expression during the early phases of Epstein-Barr virus lytic induction. J Med Microbiol *65*, 1243-1252. 10.1099/jmm.0.000352.
41. Buschle, A., Mrozek-Gorska, P., Cernilogar, F.M., Ettinger, A., Pich, D., Krebs, S., Mocanu, B., Blum, H., Schotta, G., Straub, T., and Hammerschmidt, W. (2021). Epstein-Barr virus inactivates the transcriptome and disrupts the chromatin architecture of its host cell in the first phase of lytic reactivation. Nucleic Acids Res *49*, 3217-3241. 10.1093/nar/gkab099.
42. Fenwick, M.L. (1984). The Effects of Herpesviruses on Cellular Macromolecular-Synthesis. Compr Virol *19*, 359-390.
43. Ball, C.B., Parida, M., Li, M., Spector, B.M., Suarez, G.A., Meier, J.L., and Price, D.H. (2022). Human Cytomegalovirus Infection Elicits Global Changes in Host Transcription by RNA Polymerases I, II, and III. Viruses *14*. 10.3390/v14040779.
44. Bechtel, J.T., Winant, R.C., and Ganem, D. (2005). Host and viral proteins in the virion of Kaposi's sarcoma-associated herpesvirus. J Virol *79*, 4952-4964. 10.1128/JVI.79.8.4952-4964.2005.
45. Bortz, E., Whitelegge, J.P., Jia, Q., Zhou, Z.H., Stewart, J.P., Wu, T.T., and Sun, R. (2003). Identification of proteins associated with murine gammaherpesvirus 68 virions. J Virol *77*, 13425-13432. 10.1128/jvi.77.24.13425-13432.2003.
46. Johannsen, E., Luftig, M., Chase, M.R., Weicksel, S., Cahir-McFarland, E., Illanes, D., Sarracino, D., and Kieff, E. (2004). Proteins of purified Epstein-Barr virus. Proc Natl Acad Sci U S A *101*, 16286-16291. 10.1073/pnas.0407320101.

47. Zhu, F.X., Chong, J.M., Wu, L., and Yuan, Y. (2005). Virion proteins of Kaposi's sarcoma-associated herpesvirus. J Virol *79*, 800-811. 10.1128/JVI.79.2.800-811.2005.
48. Rowe, M., Glaunsinger, B., van Leeuwen, D., Zuo, J., Sweetman, D., Ganem, D., Middeldorp, J., Wiertz, E.J., and Ressing, M.E. (2007). Host shutoff during productive Epstein-Barr virus infection is mediated by BGLF5 and may contribute to immune evasion. Proc Natl Acad Sci U S A *104*, 3366-3371. 10.1073/pnas.0611128104.
49. Glaunsinger, B., Chavez, L., and Ganem, D. (2005). The exonuclease and host shutoff functions of the SOX protein of Kaposi's sarcoma-associated herpesvirus are genetically separable. J Virol *79*, 7396-7401. 10.1128/JVI.79.12.7396-7401.2005.
50. Richner, J.M., Clyde, K., Pezda, A.C., Cheng, B.Y., Wang, T., Kumar, G.R., Covarrubias, S., Coscoy, L., and Glaunsinger, B. (2011). Global mRNA degradation during lytic gammaherpesvirus infection contributes to establishment of viral latency. PLoS Pathog *7*, e1002150. 10.1371/journal.ppat.1002150.
51. Zuo, J., Thomas, W., van Leeuwen, D., Middeldorp, J.M., Wiertz, E.J., Ressing, M.E., and Rowe, M. (2008). The DNase of gammaherpesviruses impairs recognition by virus-specific CD8+ T cells through an additional host shutoff function. J Virol *82*, 2385-2393. 10.1128/JVI.01946-07.
52. Covarrubias, S., Richner, J.M., Clyde, K., Lee, Y.J., and Glaunsinger, B.A. (2009). Host shutoff is a conserved phenotype of gammaherpesvirus infection and is orchestrated exclusively from the cytoplasm. J Virol *83*, 9554-9566. 10.1128/JVI.01051-09.
53. Bagneris, C., Briggs, L.C., Savva, R., Ebrahimi, B., and Barrett, T.E. (2011). Crystal structure of a KSHV-SOX-DNA complex: insights into the molecular mechanisms underlying DNase activity and host shutoff. Nucleic Acids Res *39*, 5744-5756. 10.1093/nar/gkr111.
54. Buisson, M., Geoui, T., Flot, D., Tarbouriech, N., Ressing, M.E., Wiertz, E.J., and Burmeister, W.P. (2009). A bridge crosses the active-site canyon of the Epstein-Barr virus nuclease with DNase and RNase activities. J Mol Biol *391*, 717-728. 10.1016/j.jmb.2009.06.034.
55. Abernathy, E., Gilbertson, S., Alla, R., and Glaunsinger, B. (2015). Viral Nucleases Induce an mRNA Degradation-Transcription Feedback Loop in Mammalian Cells. Cell Host Microbe *18*, 243-253. 10.1016/j.chom.2015.06.019.
56. Covarrubias, S., Gaglia, M.M., Kumar, G.R., Wong, W., Jackson, A.O., and Glaunsinger, B.A. (2011). Coordinated destruction of cellular messages in translation complexes by the gammaherpesvirus host shutoff factor and the mammalian exonuclease Xrn1. PLoS Pathog *7*, e1002339. 10.1371/journal.ppat.1002339.
57. Gaglia, M.M., Covarrubias, S., Wong, W., and Glaunsinger, B.A. (2012). A common strategy for host RNA degradation by divergent viruses. J Virol *86*, 9527-9530. 10.1128/JVI.01230-12.
58. Abernathy, E., Clyde, K., Yeasmin, R., Krug, L.T., Burlingame, A., Coscoy, L., and Glaunsinger, B. (2014). Gammaherpesviral gene expression and virion composition are broadly controlled by accelerated mRNA degradation. PLoS Pathog *10*, e1003882. 10.1371/journal.ppat.1003882.
59. Gaglia, M.M., Rycroft, C.H., and Glaunsinger, B.A. (2015). Transcriptome-Wide Cleavage Site Mapping on Cellular mRNAs Reveals Features Underlying Sequence-Specific Cleavage by the Viral Ribonuclease SOX. PLoS Pathog *11*, e1005305. 10.1371/journal.ppat.1005305.

60. Mendez, A.S., Vogt, C., Bohne, J., and Glaunsinger, B.A. (2018). Site specific target binding controls RNA cleavage efficiency by the Kaposi's sarcoma-associated herpesvirus endonuclease SOX. Nucleic Acids Res *46*, 11968-11979. 10.1093/nar/gky932.
61. Lee, H., Patschull, A.O.M., Bagneris, C., Ryan, H., Sanderson, C.M., Ebrahimi, B., Nobeli, I., and Barrett, T.E. (2017). KSHV SOX mediated host shutoff: the molecular mechanism underlying mRNA transcript processing. Nucleic Acids Res *45*, 4756-4767. 10.1093/nar/gkw1340.
62. Horst, D., Burmeister, W.P., Boer, I.G., van Leeuwen, D., Buisson, M., Gorbalenya, A.E., Wiertz, E.J., and Ressing, M.E. (2012). The "Bridge" in the Epstein-Barr virus alkaline exonuclease protein BGLF5 contributes to shutoff activity during productive infection. J Virol *86*, 9175-9187. 10.1128/JVI.00309-12.
63. Feederle, R., Mehl-Lautscham, A.M., Bannert, H., and Delecluse, H.J. (2009). The Epstein-Barr virus protein kinase BGLF4 and the exonuclease BGLF5 have opposite effects on the regulation of viral protein production. J Virol *83*, 10877-10891. 10.1128/JVI.00525-09.
64. Glaunsinger, B., and Ganem, D. (2004). Highly selective escape from KSHV-mediated host mRNA shutoff and its implications for viral pathogenesis. J Exp Med *200*, 391-398. 10.1084/jem.20031881.
65. Hutin, S., Lee, Y., and Glaunsinger, B.A. (2013). An RNA element in human interleukin 6 confers escape from degradation by the gammaherpesvirus SOX protein. J Virol *87*, 4672-4682. 10.1128/JVI.00159-13.
66. Muller, M., and Glaunsinger, B.A. (2017). Nuclease escape elements protect messenger RNA against cleavage by multiple viral endonucleases. PLoS Pathog *13*, e1006593. 10.1371/journal.ppat.1006593.
67. Rodriguez, W., Srivastav, K., and Muller, M. (2019). C19ORF66 Broadly Escapes Virus-Induced Endonuclease Cleavage and Restricts Kaposi's Sarcoma-Associated Herpesvirus. J Virol *93*. 10.1128/JVI.00373-19.
68. Clyde, K., and Glaunsinger, B.A. (2011). Deep sequencing reveals direct targets of gammaherpesvirus-induced mRNA decay and suggests that multiple mechanisms govern cellular transcript escape. PLoS One *6*, e19655. 10.1371/journal.pone.0019655.
69. Macveigh-Fierro, D., Cicerchia, A., Cadorette, A., Sharma, V., and Muller, M. (2022). The m(6)A reader YTHDC2 is essential for escape from KSHV SOX-induced RNA decay. Proc Natl Acad Sci U S A *119*. 10.1073/pnas.2116662119.
70. Muller, M., Hutin, S., Marigold, O., Li, K.H., Burlingame, A., and Glaunsinger, B.A. (2015). A ribonucleoprotein complex protects the interleukin-6 mRNA from degradation by distinct herpesviral endonucleases. PLoS Pathog *11*, e1004899. 10.1371/journal.ppat.1004899.
71. Ersing, I., Nobre, L., Wang, L.W., Soday, L., Ma, Y., Paulo, J.A., Narita, Y., Ashbaugh, C.W., Jiang, C., Grayson, N.E., et al. (2017). A Temporal Proteomic Map of Epstein-Barr Virus Lytic Replication in B Cells. Cell Rep *19*, 1479-1493. 10.1016/j.celrep.2017.04.062.
72. Kumar, G.R., and Glaunsinger, B.A. (2010). Nuclear import of cytoplasmic poly(A) binding protein restricts gene expression via hyperadenylation and nuclear retention of mRNA. Mol Cell Biol *30*, 4996-5008. 10.1128/MCB.00600-10.

73. Park, R., El-Guindy, A., Heston, L., Lin, S.F., Yu, K.P., Nagy, M., Borah, S., Delecluse, H.J., Steitz, J., and Miller, G. (2014). Nuclear translocation and regulation of intranuclear distribution of cytoplasmic poly(A)-binding protein are distinct processes mediated by two Epstein Barr virus proteins. PLoS One *9*, e92593. 10.1371/journal.pone.0092593.
74. Kumar, G.R., Shum, L., and Glaunsinger, B.A. (2011). Importin alpha-mediated nuclear import of cytoplasmic poly(A) binding protein occurs as a direct consequence of cytoplasmic mRNA depletion. Mol Cell Biol *31*, 3113-3125. 10.1128/MCB.05402-11.
75. van Gent, M., Gram, A.M., Boer, I.G.J., Geerdink, R.J., Lindenbergh, M.F.S., Lebbink, R.J., Wiertz, E.J., and Ressing, M.E. (2015). Silencing the shutoff protein of Epstein-Barr virus in productively infected B cells points to (innate) targets for immune evasion. J Gen Virol *96*, 858-865. 10.1099/jgv.0.000021.
76. van Gent, M., Griffin, B.D., Berkhoff, E.G., van Leeuwen, D., Boer, I.G., Buisson, M., Hartgers, F.C., Burmeister, W.P., Wiertz, E.J., and Ressing, M.E. (2011). EBV lytic-phase protein BGLF5 contributes to TLR9 downregulation during productive infection. J Immunol *186*, 1694-1702. 10.4049/jimmunol.0903120.
77. Ariza, M.E., Glaser, R., Kaumaya, P.T., Jones, C., and Williams, M.V. (2009). The EBV-encoded dUTPase activates NF-kappa B through the TLR2 and MyD88-dependent signaling pathway. J Immunol *182*, 851-859. 10.4049/jimmunol.182.2.851.
78. Fiola, S., Gosselin, D., Takada, K., and Gosselin, J. (2010). TLR9 contributes to the recognition of EBV by primary monocytes and plasmacytoid dendritic cells. J Immunol *185*, 3620-3631. 10.4049/jimmunol.0903736.
79. Gaudreault, E., Fiola, S., Olivier, M., and Gosselin, J. (2007). Epstein-Barr virus induces MCP-1 secretion by human monocytes via TLR2. J Virol *81*, 8016-8024. 10.1128/JVI.00403-07.
80. Ladell, K., Dorner, M., Zauner, L., Berger, C., Zucol, F., Bernasconi, M., Niggli, F.K., Speck, R.F., and Nadal, D. (2007). Immune activation suppresses initiation of lytic Epstein-Barr virus infection. Cell Microbiol *9*, 2055-2069. 10.1111/j.1462-5822.2007.00937.x.
81. Zauner, L., Melroe, G.T., Sigrist, J.A., Rechsteiner, M.P., Dorner, M., Arnold, M., Berger, C., Bernasconi, M., Schaefer, B.W., Speck, R.F., and Nadal, D. (2010). TLR9 triggering in Burkitt's lymphoma cell lines suppresses the EBV BZLF1 transcription via histone modification. Oncogene *29*, 4588-4598. 10.1038/onc.2010.203.
82. Gumperz, J.E. (2006). The ins and outs of CD1 molecules: bringing lipids under immunological surveillance. Traffic *7*, 2-13. 10.1111/j.1600-0854.2005.00364.x.
83. Chung, B.K., Tsai, K., Allan, L.L., Zheng, D.J., Nie, J.C., Biggs, C.M., Hasan, M.R., Kozak, F.K., van den Elzen, P., Priatel, J.J., and Tan, R. (2013). Innate immune control of EBV-infected B cells by invariant natural killer T cells. Blood *122*, 2600-2608. 10.1182/blood-2013-01-480665.
84. Long, H.M., Meckiff, B.J., and Taylor, G.S. (2019). The T-cell Response to Epstein-Barr Virus-New Tricks From an Old Dog. Front Immunol *10*, 2193. 10.3389/fimmu.2019.02193.
85. Quinn, L.L., Zuo, J., Abbott, R.J., Shannon-Lowe, C., Tierney, R.J., Hislop, A.D., and Rowe, M. (2014). Cooperation between Epstein-Barr virus immune evasion proteins spreads protection from CD8+ T cell recognition across all three phases of the lytic cycle. PLoS Pathog *10*, e1004322. 10.1371/journal.ppat.1004322.

86. Casco, A., Gupta, A., Hayes, M., Djavadian, R., Ohashi, M., and Johannsen, E. (2022). Accurate Quantification of Overlapping Herpesvirus Transcripts from RNA Sequencing Data. J Virol *96*, e0163521. 10.1128/JVI.01635-21.
87. Djavadian, R., Hayes, M., and Johannsen, E. (2018). CAGE-seq analysis of Epstein-Barr virus lytic gene transcription: 3 kinetic classes from 2 mechanisms. PLoS Pathog *14*, e1007114. 10.1371/journal.ppat.1007114.
88. Feederle, R., Bannert, H., Lips, H., Muller-Lantzsch, N., and Delecluse, H.J. (2009). The Epstein-Barr virus alkaline exonuclease BGLF5 serves pleiotropic functions in virus replication. J Virol *83*, 4952-4962. 10.1128/JVI.00170-09.
89. Traylen, C., Ramasubramanyan, S., Zuo, J., Rowe, M., Almohammad, R., Heesom, K., Sweet, S.M., Matthews, D.A., and Sinclair, A.J. (2015). Identification of Epstein-Barr Virus Replication Proteins in Burkitt's Lymphoma Cells. Pathogens *4*, 739-751. 10.3390/pathogens4040739.
90. Zhu, J., Liao, G., Shan, L., Zhang, J., Chen, M.R., Hayward, G.S., Hayward, S.D., Desai, P., and Zhu, H. (2009). Protein array identification of substrates of the Epstein-Barr virus protein kinase BGLF4. J Virol *83*, 5219-5231. 10.1128/JVI.02378-08.
91. El-Guindy, A., Lopez-Giraldez, F., Delecluse, H.J., McKenzie, J., and Miller, G. (2014). A locus encompassing the Epstein-Barr virus bglf4 kinase regulates expression of genes encoding viral structural proteins. PLoS Pathog *10*, e1004307. 10.1371/journal.ppat.1004307.
92. Mure, F., Panthu, B., Zanella-Cleon, I., Delolme, F., Manet, E., Ohlmann, T., and Gruffat, H. (2018). Epstein-Barr Virus Protein EB2 Stimulates Translation Initiation of mRNAs through Direct Interactions with both Poly(A)-Binding Protein and Eukaryotic Initiation Factor 4G. J Virol *92*. 10.1128/JVI.01917-17.
93. Batisse, J., Manet, E., Middeldorp, J., Sergeant, A., and Gruffat, H. (2005). Epstein-Barr virus mRNA export factor EB2 is essential for intranuclear capsid assembly and production of gp350. J Virol *79*, 14102-14111. 10.1128/JVI.79.22.14102-14111.2005.
94. Farjot, G., Buisson, M., Duc Dodon, M., Gazzolo, L., Sergeant, A., and Mikaelian, I. (2000). Epstein-Barr virus EB2 protein exports unspliced RNA via a Crm-1-independent pathway. J Virol *74*, 6068-6076. 10.1128/jvi.74.13.6068-6076.2000.
95. Gruffat, H., Batisse, J., Pich, D., Neuhierl, B., Manet, E., Hammerschmidt, W., and Sergeant, A. (2002). Epstein-Barr virus mRNA export factor EB2 is essential for production of infectious virus. J Virol *76*, 9635-9644. 10.1128/jvi.76.19.9635-9644.2002.
96. Han, Z., Marendy, E., Wang, Y.D., Yuan, J., Sample, J.T., and Swaminathan, S. (2007). Multiple roles of Epstein-Barr virus SM protein in lytic replication. J Virol *81*, 4058-4069. 10.1128/JVI.02665-06.
97. Ruvolo, V., Gupta, A.K., and Swaminathan, S. (2001). Epstein-Barr virus SM protein interacts with mRNA in vivo and mediates a gene-specific increase in cytoplasmic mRNA. J Virol *75*, 6033-6041. 10.1128/JVI.75.13.6033-6041.2001.
98. Ruvolo, V., Wang, E., Boyle, S., and Swaminathan, S. (1998). The Epstein-Barr virus nuclear protein SM is both a post-transcriptional inhibitor and activator of gene expression. Proc Natl Acad Sci U S A *95*, 8852-8857. 10.1073/pnas.95.15.8852.
99. Murata, T., Sugimoto, A., Inagaki, T., Yanagi, Y., Watanabe, T., Sato, Y., and Kimura, H. (2021). Molecular Basis of Epstein-Barr Virus Latency Establishment and Lytic Reactivation. Viruses *13*. 10.3390/v13122344.

100. Ressing, M.E., van Leeuwen, D., Verreck, F.A., Keating, S., Gomez, R., Franken, K.L., Ottenhoff, T.H., Spriggs, M., Schumacher, T.N., Hutt-Fletcher, L.M., et al. (2005). Epstein-Barr virus gp42 is posttranslationally modified to produce soluble gp42 that mediates HLA class II immune evasion. J Virol *79*, 841-852. 10.1128/JVI.79.2.841-852.2005.
101. Lee, M.Y., and Lufkin, T. (2012). Development of the "Three-step MACS": a novel strategy for isolating rare cell populations in the absence of known cell surface markers from complex animal tissue. J Biomol Tech *23*, 69-77. 10.7171/jbt.12-2302-003.
102. Wattrus, S.J., and Zon, L.I. (2020). A Transgenic System for Rapid Magnetic Enrichment of Rare Embryonic Cells. Zebrafish *17*, 354-357. 10.1089/zeb.2020.1904.
103. Xu, H., Akinyemi, I.A., Haley, J., McIntosh, M.T., and Bhaduri-McIntosh, S. (2023). ATM, KAP1 and the Epstein-Barr virus polymerase processivity factor direct traffic at the intersection of transcription and replication. Nucleic Acids Res *51*, 11104-11122. 10.1093/nar/gkad823.
104. Nagaraju, T., Sugden, A.U., and Sugden, B. (2019). Four-dimensional analyses show that replication compartments are clonal factories in which Epstein-Barr viral DNA amplification is coordinated. Proc Natl Acad Sci U S A *116*, 24630-24638. 10.1073/pnas.1913992116.
105. Lopez-Montanes, M., Alari-Pahissa, E., Sintes, J., Martinez-Rodriguez, J.E., Muntasell, A., and Lopez-Botet, M. (2017). Antibody-Dependent NK Cell Activation Differentially Targets EBV-Infected Cells in Lytic Cycle and Bystander B Lymphocytes Bound to Viral Antigen-Containing Particles. J Immunol *199*, 656-665. 10.4049/jimmunol.1601574.
106. Tsai, M.H., Raykova, A., Klinke, O., Bernhardt, K., Gartner, K., Leung, C.S., Geletneky, K., Sertel, S., Munz, C., Feederle, R., and Delecluse, H.J. (2013). Spontaneous lytic replication and epitheliotropism define an Epstein-Barr virus strain found in carcinomas. Cell Rep *5*, 458-470. 10.1016/j.celrep.2013.09.012.
107. Vallhov, H., Gutzeit, C., Johansson, S.M., Nagy, N., Paul, M., Li, Q., Friend, S., George, T.C., Klein, E., Scheynius, A., and Gabrielsson, S. (2011). Exosomes containing glycoprotein 350 released by EBV-transformed B cells selectively target B cells through CD21 and block EBV infection in vitro. J Immunol *186*, 73-82. 10.4049/jimmunol.1001145.
108. Evans, C., Hardin, J., and Stoebel, D.M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. Brief Bioinform *19*, 776-792. 10.1093/bib/bbx008.
109. Schork, K., Podwojski, K., Turewicz, M., Stephan, C., and Eisenacher, M. (2021). Important Issues in Planning a Proteomics Experiment: Statistical Considerations of Quantitative Proteomic Data. Methods Mol Biol *2228*, 1-20. 10.1007/978-1-0716-1024-4_1.
110. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. 10.1186/s13059-014-0550-8.
111. Tischer, B.K., Smith, G.A., and Osterrieder, N. (2010). En passant mutagenesis: a two step markerless red recombination system. Methods Mol Biol *634*, 421-430. 10.1007/978-1-60761-652-8_30.

112. Grillot-Courvalin, C., Goussard, S., Huetz, F., Ojcius, D.M., and Courvalin, P. (1998). Functional gene transfer from intracellular bacteria to mammalian cells. Nat Biotechnol *16*, 862-866. 10.1038/nbt0998-862.
113. O'Grady, T., Baddoo, M., and Flemington, E.K. (2017). Analysis of EBV Transcription Using High-Throughput RNA Sequencing. Methods Mol Biol *1532*, 105-121. 10.1007/978-1-4939-6655-4_7.
114. Duan, J., Shi, J., Ge, X., Dolken, L., Moy, W., He, D., Shi, S., Sanders, A.R., Ross, J., and Gejman, P.V. (2013). Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. Sci Rep *3*, 1318. 10.1038/srep01318.
115. Gregory, C.D., Murray, R.J., Edwards, C.F., and Rickinson, A.B. (1988). Downregulation of cell adhesion molecules LFA-3 and ICAM-1 in Epstein-Barr virus-positive Burkitt's lymphoma underlies tumor cell escape from virus-specific T cell surveillance. J Exp Med *167*, 1811-1824. 10.1084/jem.167.6.1811.
116. Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S., Hahn, W.C., Sharp, P.A., et al. (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. RNA *9*, 493-501. 10.1261/rna.2192803.
117. Neuhierl, B., Feederle, R., Hammerschmidt, W., and Delecluse, H.J. (2002). Glycoprotein gp110 of Epstein-Barr virus determines viral tropism and efficiency of infection. Proc Natl Acad Sci U S A *99*, 15036-15041. 10.1073/pnas.232381299.
118. Team, R.C. (2023). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
119. Hahne, F., LeMeur, N., Brinkman, R.R., Ellis, B., Haaland, P., Sarkar, D., Spidlen, J., Strain, E., and Gentleman, R. (2009). flowCore: a Bioconductor package for high throughput flow cytometry. BMC Bioinformatics *10*, 106. 10.1186/1471-2105-10-106.
120. Monaco, G., Chen, H., Poidinger, M., Chen, J., de Magalhaes, J.P., and Larbi, A. (2016). flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. Bioinformatics *32*, 2473-2480. 10.1093/bioinformatics/btw191.
121. Van, P., Jiang, W., Gottardo, R., and Finak, G. (2018). ggCyto: next generation open-source visualization software for cytometry. Bioinformatics *34*, 3951-3953. 10.1093/bioinformatics/bty441.
122. Shumate, A., and Salzberg, S.L. (2021). Liftoff: accurate mapping of gene annotations. Bioinformatics *37*, 1639-1643. 10.1093/bioinformatics/btaa1016.
123. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
124. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21. 10.1093/bioinformatics/bts635.
125. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323. 10.1186/1471-2105-12-323.
126. Zhu, A., Ibrahim, J.G., and Love, M.I. (2019). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics *35*, 2084-2092. 10.1093/bioinformatics/bty895.
127. Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165. 10.1093/nar/gkw257.

128. Kurtenbach, S., and Harbour, J.W. (2019). SparK: A Publication-quality NGS Visualization Tool. bioRxiv, 845529. 10.1101/845529.
129. Wong, A.C.H., Wong, J.J., Rasko, J.E.J., and Schmitz, U. (2023). SpliceWiz: interactive analysis and visualization of alternative splicing in R. Brief Bioinform *25*. 10.1093/bib/bbad468.
130. Vitting-Seerup, K., and Sandelin, A. (2017). The Landscape of Isoform Switches in Human Cancers. Mol Cancer Res *15*, 1206-1220. 10.1158/1541-7786.MCR-16-0459.
131. Vitting-Seerup, K., and Sandelin, A. (2019). IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. Bioinformatics *35*, 4469-4471. 10.1093/bioinformatics/btz247.
132. Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome Res *22*, 2008-2017. 10.1101/gr.133744.111.
133. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47. 10.1093/nar/gkv007.
134. Kang, Y.J., Yang, D.C., Kong, L., Hou, M., Meng, Y.Q., Wei, L., and Gao, G. (2017). CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res *45*, W12-W16. 10.1093/nar/gkx428.
135. Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. Nucleic Acids Res *40*, D290-301. 10.1093/nar/gkr1065.
136. Meszaros, B., Erdos, G., and Dosztanyi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res *46*, W329-W337. 10.1093/nar/gky384.
137. Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gislason, M.H., Pihl, S.I., Tsirigos, K.D., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol *40*, 1023-1025. 10.1038/s41587-021-01156-3.
138. Roth, S.J., Heinz, S., and Benner, C. (2020). ARTDeco: automatic readthrough transcription detection. BMC Bioinformatics *21*, 214. 10.1186/s12859-020-03551-0.
139. Wickham, H. (2016). Ggplot2 : elegant graphics for data analysis (Springer Science+Business Media, LLC).
140. Kassambara, A. (2023). ggpubr: 'ggplot2' Based Publication Ready Plots.
141. Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach. Journal of Open Source Software *6*, 3167. 10.21105/joss.03167.
142. Larsson, J. (2022). eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses.
143. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 10.1093/bioinformatics/btw313.
144. Campbell, B.C., Nabel, E.M., Murdock, M.H., Lao-Peregrin, C., Tsoulfas, P., Blackmore, M.G., Lee, F.S., Liston, C., Morishita, H., and Petsko, G.A. (2020). mGreenLantern: a bright monomeric fluorescent protein with rapid expression and cell filling properties for neuronal imaging. Proc Natl Acad Sci U S A *117*, 30710-30721. 10.1073/pnas.2000942117.
145. Meng, Q., Hagemeier, S.R., Fingeroth, J.D., Gershburg, E., Pagano, J.S., and Kenney, S.C. (2010). The Epstein-Barr virus (EBV)-encoded protein kinase, EBV-PK, but not the

thymidine kinase (EBV-TK), is required for ganciclovir and acyclovir inhibition of lytic viral production. J Virol *84*, 4534-4542. 10.1128/JVI.02487-09.

146. Shimizu, N., Yoshiyama, H., and Takada, K. (1996). Clonal propagation of Epstein-Barr virus (EBV) recombinants in EBV-negative Akata cells. J Virol *70*, 7260-7263. 10.1128/JVI.70.10.7260-7263.1996.
147. Bindels, D.S., Haarbosch, L., van Weeren, L., Postma, M., Wiese, K.E., Mastop, M., Aumonier, S., Gotthard, G., Royant, A., Hink, M.A., and Gadella, T.W., Jr. (2017). mScarlet: a bright monomeric red fluorescent protein for cellular imaging. Nat Methods *14*, 53-56. 10.1038/nmeth.4074.
148. Liu, Z., Chen, O., Wall, J.B.J., Zheng, M., Zhou, Y., Wang, L., Vaseghi, H.R., Qian, L., and Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci Rep *7*, 2193. 10.1038/s41598-017-02460-2.
149. Hutt-Fletcher, L.M. (2015). EBV glycoproteins: where are we now? Future Virol *10*, 1155-1162. 10.2217/fvl.15.80.
150. Chen, H., Huang, J., Wu, F.Y., Liao, G., Hutt-Fletcher, L., and Hayward, S.D. (2005). Regulation of expression of the Epstein-Barr virus BamHI-A rightward transcripts. J Virol *79*, 1724-1733. 10.1128/JVI.79.3.1724-1733.2005.
151. Kanda, T., Otter, M., and Wahl, G.M. (2001). Coupling of mitotic chromosome tethering and replication competence in epstein-barr virus-based plasmids. Mol Cell Biol *21*, 3576-3588. 10.1128/MCB.21.10.3576-3588.2001.
152. Ali, A.K., Saito, S., Shibata, S., Takada, K., and Kanda, T. (2009). Distinctive effects of the Epstein-Barr virus family of repeats on viral latent gene promoter activity and B-lymphocyte transformation. J Virol *83*, 9163-9174. 10.1128/JVI.01979-08.
153. Andersen, M.N., Al-Karradi, S.N., Kragstrup, T.W., and Hokland, M. (2016). Elimination of erroneous results in flow cytometry caused by antibody binding to Fc receptors on human monocytes and macrophages. Cytometry A *89*, 1001-1009. 10.1002/cyto.a.22995.
154. Parks, D.R., Roederer, M., and Moore, W.A. (2006). A new "Logicle" display method avoids deceptive effects of logarithmic scaling for low signals and compensated data. Cytometry A *69*, 541-551. 10.1002/cyto.a.20258.
155. Overbey, E.G., Saravia-Butler, A.M., Zhang, Z., Rathi, K.S., Fogle, H., da Silveira, W.A., Barker, R.J., Bass, J.J., Beheshti, A., Berrios, D.C., et al. (2021). NASA GeneLab RNA-seq consensus pipeline: standardized processing of short-read RNA-seq data. iScience *24*, 102361. 10.1016/j.isci.2021.102361.
156. Vitting-Seerup, K., Porse, B.T., Sandelin, A., and Waage, J. (2014). spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. BMC Bioinformatics *15*, 81. 10.1186/1471-2105-15-81.
157. Weischenfeldt, J., Waage, J., Tian, G., Zhao, J., Damgaard, I., Jakobsen, J.S., Kristiansen, K., Krogh, A., Wang, J., and Porse, B.T. (2012). Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. Genome Biol *13*, R35. 10.1186/gb-2012-13-5-r35.
158. Hochberg, Y., and Benjamini, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. JR Stat Soc *57*, 289-300.
159. Johannsen, E., and Kaye, K. (2020). Epstein-Barr Virus (Infectious Mononucleosis, Epstein-Barr Virus-Associated Malignant Diseases, and Other Diseases). In Mandell,

Douglas, and Bennett's Principles and Practice of Infectious Diseases, J. Bennett, R. Dolin, and M. Blaser, eds. (Elsevier), pp. 1872-1890.
160. Damania, B., Kenney, S.C., and Raab-Traub, N. (2022). Epstein-Barr virus: Biology and clinical disease. Cell *185*, 3652-3670. 10.1016/j.cell.2022.08.026.
161. Casco, A., and Johannsen, E. (2023). EBV Reactivation from Latency Is a Degrading Experience for the Host. Viruses *15*. 10.3390/v15030726.
162. Sato, H., Takimoto, T., Tanaka, S., Tanaka, J., and Raab-Traub, N. (1990). Concatameric replication of Epstein-Barr virus: structure of the termini in virus-producer and newly transformed cell lines. J Virol *64*, 5295-5300. 10.1128/JVI.64.11.5295-5300.1990.
163. Li, S., Liu, B., Tan, M., Juillard, F., Szymula, A., Alvarez, A.L., Van Sciver, N., George, A., Ramachandran, A., Raina, K., et al. (2024). Kaposi's sarcoma herpesvirus exploits the DNA damage response to circularize its genome. Nucleic Acids Res *52*, 1814-1829. 10.1093/nar/gkad1224.
164. Hoffman, G.J., Lazarowitz, S.G., and Hayward, S.D. (1980). Monoclonal antibody against a 250,000-dalton glycoprotein of Epstein-Barr virus identifies a membrane antigen and a neutralizing antigen. Proc Natl Acad Sci U S A *77*, 2979-2983. 10.1073/pnas.77.5.2979.
165. Szakonyi, G., Klein, M.G., Hannan, J.P., Young, K.A., Ma, R.Z., Asokan, R., Holers, V.M., and Chen, X.S. (2006). Structure of the Epstein-Barr virus major envelope glycoprotein. Nat Struct Mol Biol *13*, 996-1001. 10.1038/nsmb1161.
166. Woellmer, A., Arteaga-Salas, J.M., and Hammerschmidt, W. (2012). BZLF1 governs CpG-methylated chromatin of Epstein-Barr Virus reversing epigenetic repression. PLoS Pathog *8*, e1002902. 10.1371/journal.ppat.1002902.
167. Agostini, F., Zagalak, J., Attig, J., Ule, J., and Luscombe, N.M. (2021). Intergenic RNA mainly derives from nascent transcripts of known genes. Genome Biol *22*, 136. 10.1186/s13059-021-02350-x.
168. Watson, S.F., Bellora, N., and Macias, S. (2020). ILF3 contributes to the establishment of the antiviral type I interferon program. Nucleic Acids Res *48*, 116-129. 10.1093/nar/gkz1060.
169. Rutkowski, A.J., Erhard, F., L'Hernault, A., Bonfert, T., Schilhabel, M., Crump, C., Rosenstiel, P., Efstathiou, S., Zimmer, R., Friedel, C.C., and Dolken, L. (2015). Widespread disruption of host transcription termination in HSV-1 infection. Nat Commun *6*, 7126. 10.1038/ncomms8126.
170. Wang, X., Hennig, T., Whisnant, A.W., Erhard, F., Prusty, B.K., Friedel, C.C., Forouzmand, E., Hu, W., Erber, L., Chen, Y., et al. (2020). Herpes simplex virus blocks host transcription termination via the bimodal activities of ICP27. Nat Commun *11*, 293. 10.1038/s41467-019-14109-x.
171. Grady, L.M., Szczepaniak, R., Murelli, R.P., Masaoka, T., Le Grice, S.F.J., Wright, D.L., and Weller, S.K. (2017). The Exonuclease Activity of Herpes Simplex Virus 1 UL12 Is Required for Production of Viral DNA That Can Be Packaged To Produce Infectious Virus. J Virol *91*. 10.1128/JVI.01380-17.
172. Hartenian, E., Mendez, A.S., Didychuk, A.L., Khosla, S., and Glaunsinger, B.A. (2023). DNA processing by the Kaposi's sarcoma-associated herpesvirus alkaline exonuclease SOX contributes to viral gene expression and infectious virion production. Nucleic Acids Res *51*, 182-197. 10.1093/nar/gkac1190.

173. Martinez, R., Sarisky, R.T., Weber, P.C., and Weller, S.K. (1996). Herpes simplex virus type 1 alkaline nuclease is required for efficient processing of viral DNA replication intermediates. J Virol *70*, 2075-2085. 10.1128/JVI.70.4.2075-2085.1996.
174. Uppal, T., Meyer, D., Agarwal, A., and Verma, S.C. (2019). The DNase Activity of Kaposi's Sarcoma-Associated Herpesvirus SOX Protein Serves an Important Role in Viral Genome Processing during Lytic Replication. J Virol *93*. 10.1128/JVI.01983-18.
175. Stern-Ginossar, N., Thompson, S.R., Mathews, M.B., and Mohr, I. (2019). Translational Control in Virus-Infected Cells. Cold Spring Harb Perspect Biol *11*. 10.1101/cshperspect.a033001.
176. Duncan-Lewis, C., Hartenian, E., King, V., and Glaunsinger, B.A. (2021). Cytoplasmic mRNA decay represses RNA polymerase II transcription during early apoptosis. Elife *10*. 10.7554/eLife.58342.
177. Burke, J.M., Moon, S.L., Matheny, T., and Parker, R. (2019). RNase L Reprograms Translation by Widespread mRNA Turnover Escaped by Antiviral mRNAs. Mol Cell *75*, 1203-1217 e1205. 10.1016/j.molcel.2019.07.029.
178. Yiu, S.P.T., Guo, R., Zerbe, C., Weekes, M.P., and Gewurz, B.E. (2022). Epstein-Barr virus BNRF1 destabilizes SMC5/6 cohesin complexes to evade its restriction of replication compartments. Cell Rep *38*, 110411. 10.1016/j.celrep.2022.110411.
179. Hennig, T., Michalski, M., Rutkowski, A.J., Djakovic, L., Whisnant, A.W., Friedl, M.S., Jha, B.A., Baptista, M.A.P., L'Hernault, A., Erhard, F., et al. (2018). HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. PLoS Pathog *14*, e1006954. 10.1371/journal.ppat.1006954.
180. Ashraf, U., Benoit-Pilven, C., Navratil, V., Ligneau, C., Fournier, G., Munier, S., Sismeiro, O., Coppee, J.Y., Lacroix, V., and Naffakh, N. (2020). Influenza virus infection induces widespread alterations of host cell splicing. NAR Genom Bioinform *2*, lqaa095. 10.1093/nargab/lqaa095.
181. Monteuuis, G., Wong, J.J.L., Bailey, C.G., Schmitz, U., and Rasko, J.E.J. (2019). The changing paradigm of intron retention: regulation, ramifications and recipes. Nucleic Acids Res *47*, 11497-11513. 10.1093/nar/gkz1068.
182. Vilborg, A., Passarelli, M.C., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2015). Widespread Inducible Transcription Downstream of Human Genes. Mol Cell *59*, 449-461. 10.1016/j.molcel.2015.06.016.
183. Bauer, D.L.V., Tellier, M., Martinez-Alonso, M., Nojima, T., Proudfoot, N.J., Murphy, S., and Fodor, E. (2018). Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. Cell Rep *23*, 2119-2129 e2113. 10.1016/j.celrep.2018.04.047.
184. Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L., et al. (2018). Transcription Elongation Can Affect Genome 3D Structure. Cell *174*, 1522-1536 e1522. 10.1016/j.cell.2018.07.047.
185. Zhao, N., Sebastiano, V., Moshkina, N., Mena, N., Hultquist, J., Jimenez-Morales, D., Ma, Y., Rialdi, A., Albrecht, R., Fenouil, R., et al. (2018). Influenza virus infection causes global RNAPII termination defects. Nat Struct Mol Biol *25*, 885-893. 10.1038/s41594-018-0124-7.

186. Cardiello, J.F., Goodrich, J.A., and Kugel, J.F. (2018). Heat Shock Causes a Reversible Increase in RNA Polymerase II Occupancy Downstream of mRNA Genes, Consistent with a Global Loss in Transcriptional Termination. Mol Cell Biol *38*. 10.1128/MCB.00181-18.
187. Hadar, S., Meller, A., Saida, N., and Shalgi, R. (2022). Stress-induced transcriptional readthrough into neighboring genes is linked to intron retention. iScience *25*, 105543. 10.1016/j.isci.2022.105543.
188. Shalgi, R., Hurt, J.A., Lindquist, S., and Burge, C.B. (2014). Widespread inhibition of posttranscriptional splicing shapes the cellular transcriptome following heat shock. Cell Rep *7*, 1362-1370. 10.1016/j.celrep.2014.04.044.
189. Vilborg, A., Sabath, N., Wiesel, Y., Nathans, J., Levy-Adam, F., Yario, T.A., Steitz, J.A., and Shalgi, R. (2017). Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. Proc Natl Acad Sci U S A *114*, E8362-E8371. 10.1073/pnas.1711120114.
190. Burke, J.M., Ripin, N., Ferretti, M.B., St Clair, L.A., Worden-Sapper, E.R., Salgado, F., Sawyer, S.L., Perera, R., Lynch, K.W., and Parker, R. (2022). RNase L activation in the cytoplasm induces aberrant processing of mRNAs in the nucleus. PLoS Pathog *18*, e1010930. 10.1371/journal.ppat.1010930.
191. Rios, F., Uriostegui-Arcos, M., and Zurita, M. (2024). Transcriptional Stress Induces the Generation of DoGs in Cancer Cells. Noncoding RNA *10*. 10.3390/ncrna10010005.
192. Jacob, A.G., and Smith, C.W.J. (2017). Intron retention as a component of regulated gene expression programs. Hum Genet *136*, 1043-1057. 10.1007/s00439-017-1791-x.
193. Ungerleider, N.A., Roberts, C., O'Grady, T.M., Nguyen, T.T., Baddoo, M., Wang, J., Ishaq, E., Concha, M., Lam, M., Bass, J., et al. (2024). Viral reprogramming of host transcription initiation. Nucleic Acids Res. 10.1093/nar/gkae175.
194. Sugimoto, A., Sato, Y., Kanda, T., Murata, T., Narita, Y., Kawashima, D., Kimura, H., and Tsurumi, T. (2013). Different distributions of Epstein-Barr virus early and late gene transcripts within viral replication compartments. J Virol *87*, 6693-6699. 10.1128/JVI.00219-13.
195. Dembowski, J.A., Dremel, S.E., and DeLuca, N.A. (2017). Replication-Coupled Recruitment of Viral and Cellular Factors to Herpes Simplex Virus Type 1 Replication Forks for the Maintenance and Expression of Viral Genomes. PLoS Pathog *13*, e1006166. 10.1371/journal.ppat.1006166.

Appendix I

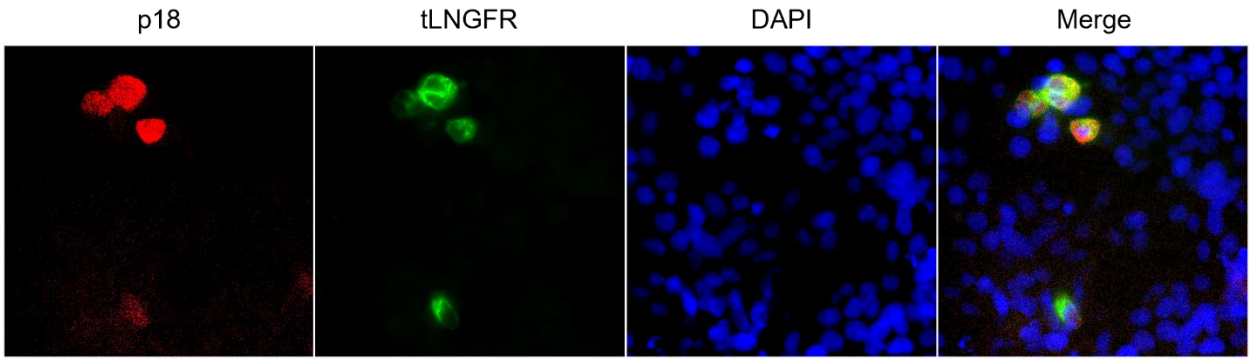
A MACS-compatible EBV mutant for isolation of late lytic cells

Introduction

To address the limitations of inefficient and asynchronous EBV lytic replication, we initially pursued magnetic-activated cell sorting (MACS) to enrich the lytic subpopulations. MACS utilizes antibody-conjugated paramagnetic beads to target specific surface markers on the cell of interest. These labeled cells are then isolated within a magnetic field, while unlabeled cells pass through. We selected MACS over fluorescence-activated cell sorting (FACS) due to its advantages [1]. Firstly, MACS allows enrichment of target populations *en masse*, reducing processing time compared to the single-cell sorting of FACS. Secondly, MACS utilizes less sophisticated and costly equipment, often available as benchtop sorters that do not require specialized operators. Finally, MACS may exert reduced cell stress in comparison to FACS [1].

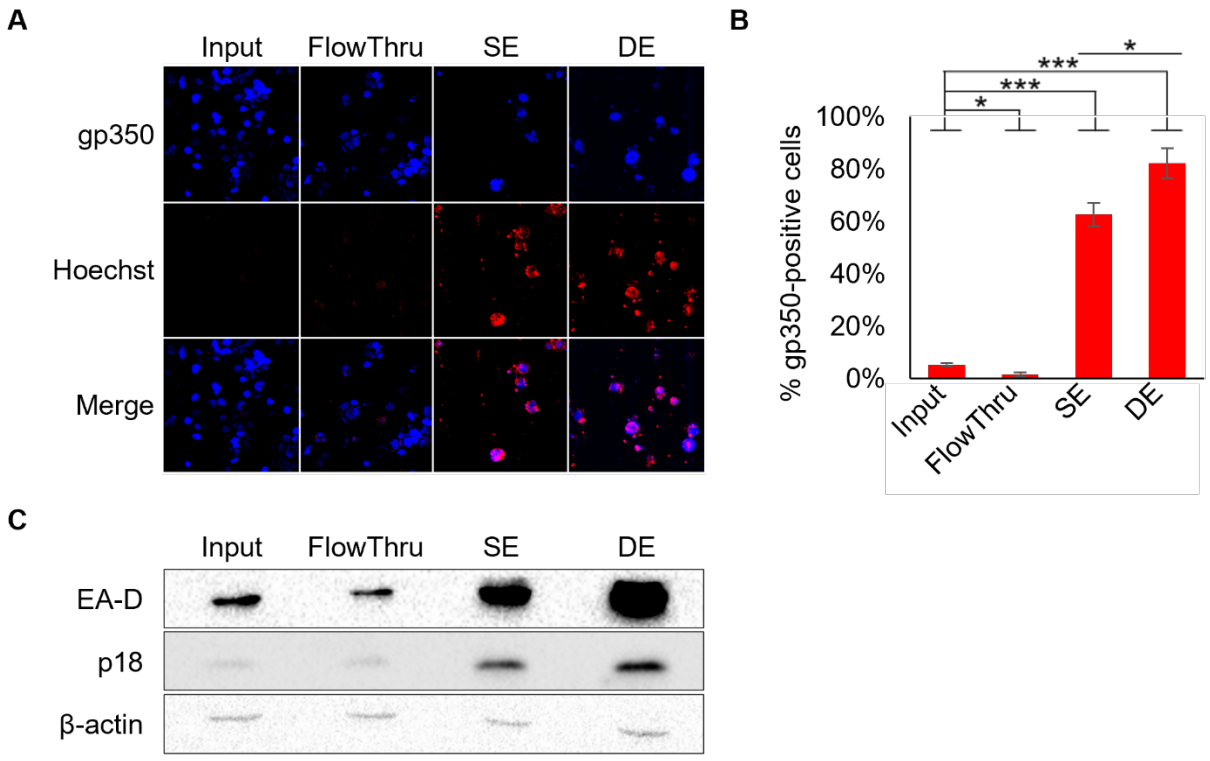
Results

We developed a membrane reporter EBV to address enable enrichment of late lytic cells. To achieve this, the BILF2 ORF in the B95-8 strain BACmid was replaced with a truncated low-affinity nerve growth receptor (tLNGFR). This reporter (EBV BILF2p-tLNGFR) was used to infect HEK293 cells, and lytic replication was induced via Rta and Zta transfection. Confirmation of late-kinetic tLNGFR expression was obtained through immunostaining, demonstrating co-expression with the late viral capsid antigen p18 (Appendix Figure I-1).



Appendix Figure I-1 tLNGFR expression in lytic induced EBV BILF2p-tLNGFR HEK293 cells reflects late kinetics. Images of HEK293 cells infected with EBV BILF2p-tLNGFR HEK293 cells induced into lytic replication by Rta and Zta transfection. Cells were fixed and probed for the late gene product p18, the cell surface reporter marker tLNGFR, and nuclei staining with DAPI. Cells were scored 48 hours post transfection.

Next, I assessed the efficiency of enriching late lytic cells using tLNGFR as a marker. Following antibody-conjugated MicroBead labeling of lytic induced cells, MACS separation yielded several fractions: input, flow-through (FlowThru), single enrichment (SE), and double enrichment (DE). Live-cell imaging and immunoblotting revealed a significant increase in both late membrane protein (gp350) positive cells and lytic protein expression (early EA-D and late p18) within the enriched fractions (SE: up to 66% gp350+; DE: up to 86% gp350+) (Appendix Figure I-2). These results demonstrate successful enrichment of cells in the late lytic phase of EBV infection.



Appendix Figure I-2 Enrichment efficiency of EBV BILF2p-tLNGFR HEK293 cells by MACS. HEK293 cells infected with EBV BILF2p-tLNGFR were induced into lytic replication by transfection with Rta and Zta followed by MACS separation using MACSelect™ LNGFR System (Miltenyi Biotech) at 48 hours post transfection. Four MACS-enriched cell fractions were collected: Input, flow through (FlowThru), single enrichment (SE), and double enrichment (DE). (A) Images of live cells from the four fractions stained for the late expressed cell surface glycoprotein gp350 (top row) or nuclei with Hoechst 33342 (middle row). Merged images are shown in the bottom row. (B) Quantification of gp350-positive cells for each fraction relative to Hoechst 33342 staining. (C) Immunoblotting of MACS-enriched fractions for the early lytic protein EA-D (top row), late lytic protein p18 (middle row), and loading control β-actin (bottom row). Data are represented as mean ± standard deviation ($n = 2$). * $p<0.05$, *** $p<0.001$.

Discussion

We initially used a membrane reporter EBV expressing tLNGFR from the late BILF2 promoter, for cost-effective and rapid enrichment of cells in the late lytic phase. This reporter construct was used in conjunction with the MACS system to achieve enrichment *en masse* of these rare cells. While successful in increasing late lytic cell purity to 60-80% (from 5%), significant contamination with non-late-lytic cells persisted (20-40%). Achieving maximum

purity (~80%) necessitated a double enrichment protocol, which not only extended processing time but also significantly diminished overall yield.

Our proposed strategy to enrich early lytic cells involved a dual-membrane reporter EBV, designed to express both tLNGFR (late lytic) and a truncated CD4 variant (early lytic). However, this approach presented several limitations. The requirement for multiple enrichment rounds and resulting low yields were compounded by the presence of tLNGFR+ cells in the flow-through, complicating subsequent early lytic population isolation. Importantly, the MACS technique is less effective with extremely rare target populations [2,3]. In physiologic models like lymphoblastoid cell lines where lytic replication frequency drops below 1%, the limitations of this approach become even more pronounced.

These considerations prompted us to reconsider our enrichment strategy. The dual-fluorescent lytic reporter EBV discussed in Chapter 3 and Chapter 4, combined with FACS-based separation, offers significantly improved purity (approaching 100%) even in scenarios with exceptionally low target cell frequencies.

Methods

Development of late lytic membrane reporter EBV.

The parental EBV BILF2p-tLNGFR BACmid was constructed using the p2089 BACmid. The BILF2 ORF was replaced with a truncated version of the low-affinity nerve growth receptor (tLNGFR) lacking the intracellular signaling domain using the GS1783 *E. coli*-based *En Passant* mutagenesis method [4] with the BILF2-to-tLNGFR gene block found in Appendix Table II-2. This EBV BILF2p-tLNGFR BAC was transferred to the diaminopimelate (DAP) auxotroph and invasive *E. coli* strain BM2710 [5]. Bactofection and colony selection of HEK293 cells was

performed as described in the Chapter 3 section entitled ‘Generation of WT and ΔBGLF5 DFLR EBV LCLs’.

Indirect immunofluorescence.

EBV BILF2p-tLNGFR HEK293 cells were grown on MatTek dishes (MatTek Corporation P35G-1.5-10-C; dish diameter 35 mm, glass thickness no. 1.5, microwell diameter 10 mm) and lytic reactivation was induced by transfection with R and Z expressing plasmids. 48 hours post-induction, cells were washed with 1X DPBS (Dulbecco’s Phosphate-Buffered Saline) and incubated 4% paraformaldehyde (PFA) in 1X DPBS for 10 minutes at room temperature. Cells were washed then permeabilized with 0.1% Triton X-100 in 1X DPBS for 10 minutes at 37°C. Cells were washed then incubated with 2% bovine serum albumin (BSA) in 1X DPBS for 1 hour at room temperature. Cells were labeled with a cocktail of primary antibodies in 2% BSA in 1X DPBS and incubated overnight at 4°C. The following antibodies were used: anti-goat EBV VCA-p18 (Thermo Scientific, PA1-73003; 1:2,000) and anti-mouse p75 tLNGFR Receptor (Abcam, ab245134; 2 ug/mL). The following day, cells were washed and incubated with a cocktail of secondary antibodies in 2% BSA in 1X DPBS for 1 hour at room temperature in the dark. The following secondary antibodies were used: Alex Fluor 646-donkey anti-goat (Abcam) and 546-goat anti-mouse (Abcam). Cells were washed and embedded with Vectashield antifade reagent (Vector Laboratories) containing 2 ug/mL DAPI. Digital images of fluorescent light were acquired using an epifluorescence microscope.

Magnetic-activated cell sorting (MACS).

Two biological replicates were performed. 48 hours post-transfection with Rta and Zta expressing plasmids, EBV BILF2p-tLNGFR HEK293 were labeled with anti-LNGFR conjugated MicroBeads (Miltenyi Biotec) in PBE buffer (PBS supplemented with 0.5% BSA and

5 mM EDTA) for 15 minutes at 4°C in dark. Cells were split into an input fraction and the remaining cells were enriched with the MACSelect LNGFR^{TM} System (Miltenyi Biotec) by aliquoting 500 uL of the magnetically labelled cell suspension through the LS separation column using the MidiMACS separation unit (Miltenyi Biotec) and collecting the effluent as the flow through fraction. The column was washed with PBE buffer, removed from the separator and the magnetically labeled cells were flushed out using a plunger as the SE fraction. A portion of the SE fraction was aliquoted through an MS separation column using the MiniMACS separation unit (Miltenyi Biotec). The column was washed with PBE buffer, removed from the separator and the magnetically labeled cells were flushed out using a plunger as the DE fraction.

Immunoblotting.

Each MACS fraction was harvested 48 hours post-transfection in RIPA Buffer and separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a nitrocellulose membrane. Membranes were blocked in Tris-buffered saline supplemented with 0.1% Tween 20 (TBST) containing 5% dry non-fat milk and incubated with appropriate primary antibodies overnight at 4°C. The following primary antibodies were used: anti-mouse EBV EA-D p52/50 (EMD Milipore, MAB8186; 1,3000), anti-goat EBV p18 (Thermo Scientific, PA1-73003; 1,1000), and anti-mouse β-actin (Santa Cruz Biotechnologies, C4; 1,200). Following treatment with primary antibodies, membranes were washed with TBST and incubated with appropriate antibodies for 1 hour at room temperature. The following secondary antibodies were used at a concentration of 1:2,000: goat anti-mouse poly-HRP (Fisher Scientific) and donkey anti-goat (Fisher Scientific). Membranes were washed and visualized with ECL chemiluminescent kit (Thermo Scientific) according to manufacturer.

Live-cell direct immunofluorescence.

10^6 cells from each MACS fraction was pelleted and brought up in 90 uL PBE buffer with 10 uL of primary anti-mouse EBV gp350 monoclonal antibody (72A1) and incubated for 20 minutes at room temperature. Cells were washed with 1 mL PBE buffer and incubated for 20 minutes with 1.5 uL of Alexa Fluor 546-goat anti-mouse (Abcam) in 96.5 uL PBE. 10 minutes into incubation, 2 uL of 200 ug/mL Hoechst 33342 (Thermo Scientific, H357) was added and incubated for the remaining 10 minutes. Cells were washed in 1 mL PBE and brough up in appropriate volume of PBE. Cells were mounted on a cover slide with cover slip and examined by fluorescence microscopy for gp350-positive cells. In the first replicate, the following gp350-positive cells were counted relative to Hoechst-positive cells per MACS fraction: input (32/563), flow through (13/624), SE (100/168), DE (100/116). Second replicate: input (30/620), flow through (9/860), SE (100/152), DE (100/128).

Statistics.

Microsoft Excel was used to perform two-sample t-tests assuming equal variances on MACS fractions relative to input between SE and DE fractions [6].

References

1. Sutermaster, B.A., and Darling, E.M. (2019). Considerations for high-yield, high-throughput cell enrichment: fluorescence versus magnetic sorting. Sci Rep 9, 227. 10.1038/s41598-018-36698-1.
2. Lee, M.Y., and Lufkin, T. (2012). Development of the "Three-step MACS": a novel strategy for isolating rare cell populations in the absence of known cell surface markers from complex animal tissue. J Biomol Tech 23, 69-77. 10.7171/jbt.12-2302-003.
3. Wattrus, S.J., and Zon, L.I. (2020). A Transgenic System for Rapid Magnetic Enrichment of Rare Embryonic Cells. Zebrafish 17, 354-357. 10.1089/zeb.2020.1904.
4. Tischer, B.K., Smith, G.A., and Osterrieder, N. (2010). En passant mutagenesis: a two step markerless red recombination system. Methods Mol Biol 634, 421-430. 10.1007/978-1-60761-652-8_30.
5. Grillot-Courvalin, C., Goussard, S., Huetz, F., Ojcius, D.M., and Courvalin, P. (1998). Functional gene transfer from intracellular bacteria to mammalian cells. Nat Biotechnol 16, 862-866. 10.1038/nbt0998-862.
6. Arthur, J. (2009). Test two sample assuming equal variances in Excel. Retrieved from: http://www.qimacros.com/qiwizard/t-test-two-sample.html

Appendix II

Primers and oligonucleotides

Appendix Table II-1 List of primer sequences used in PCR assays. Lowercase letters indicate the target binding site, while uppercase letters denote overhangs for Gibson assembly.

Primer	Sequence
FP-F	TAGAGCTAGCTGTACAAACATAAGCAAGCGatggtgagcaagggcgag
FP-R	CTTCGCGGCCGCGGATCCGATTTAAATTCGTTActtgtacagctcgtccatg
FR-F	gttatgcagtgccaccaattc
FR-R	cgatctggaggacaagttacac
Neo^r-F	tgctctgatgccgccgtgttc
Neo^r-R	gtgagatgacaggagatcctgc
SopA/B-F	ctcctgcctttctatcttg
SopA/B-R	gaggtcttctcaacactatc
BXLF1-F	tcggaaaggctgtatgacttc
BXLF1-R	gagagagcagcttggagtaac
BILF2-F	aaggccaagtctgtgtcag
BILF2-R	tctaaacgtcccacgcataag
GAPDH-F	tccatgacaactttggtatc
GAPDH-R	gtgacgggtctctctcttc

Appendix Table II-2 Gene block sequences. List of gene block (gBlock) sequences used for Gibson assembly or BAC recombineering via *En Passant* methodology.

gBlock	Sequence
MCS	tgtgtcccctggatacccgtattctagagctagctgtacaaacataagcaagcgaattcgaatttaaatcggatccgc ggccgcgaaggatctgcgatcgctccggtgcccgtcagtgggcagagcgcacatcgcccacagtccccgagaa gttgggggggaggggtcggcaattgaacgggtgcctagagaaggtggcgcggggtaaactgggaaagtgatgtc gtgtactggctccgccttttttcccgagggtgggggagaaccgtatataagtgcagtagtcgccgtgaacgttctttttc gcaacgggtttgccgccagaacacagctgaagcttcgaggggctcgcatctctccttcacgcgcccgccgccccta cctgaggccgccatccacgccggttgagtcgcgttctgccgcctcccgcctgtggtgcctcctgaactgcgtccgc cgtctaggtaagtttaaagctcaggtcgagaccgggcctttgtccggcgctcccttggagcctacctagactcagcc ggctctccacgctttgcctgaccctgcttgctcaactctacgtctttgtttcgttttctgttctgcgccgttacagatccaa gctgtgaccggcgcctacgctagacgccaccatggagagcgacgagagcggcctgcccgcccgtacgatgacc gagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccagggccgtgcgcaccctcgccgccgcg ttcgccgactaccc
hyg^{r}-to-neo^{r}	**tgctggctgaggcgttctcgaaatcagctcttgttcggtcggcatctact**tcagaagaactcgtcaagaaggcg atagaaggcgatgcgctgcgaatcgggagcggcgataccgtaaagcacgaggaagcggtcagcccattcgccg ccaagctcttcagcaatatcacgggtagccaacgctatgtcctgatagcggtccgccacacccagccggccacag tcgatgaatccagaaaagcggccattttccaccatgatattcggcaagcaggcatcgccatgggtcacgacgagat cctcgccgtcgggcatgctcgccttgagcctggcgaacagttcggctggcgcgagcccctgatgctcttcgtccag atcatcctgatcgacaagaccggcttccatccgagtacgtgctcgctcgatgcgatgtttcgcttggtggtcgaatgg gcaggtagccggatcaagcgtatgcagccgccgcattgcatcagccatgatggatactttctcggcaggagcaag gtgagatgacaggagatcctgccccggcacttcgcccaatagcagccagtcccttcccgcttcagtgacaacgtc gagcacagctgcgcaaggaacgcccgtcgtggccagccacgatagccgcgctgcctcgtcttgcagttcattcag ggcaccggacaggtcggtcttgacaaaaagaaccgggcgcccctgcgctgacagccggaacacggcggcatc agagcagccgattgtctgttgtgcccagtcatagccgaatagcctc***tccacccaagcggccggagaacctgcgt*** ***gcaatccatcttgttcaatcat***tagggataacagggtaatttaccaatgcttaatcagtgaggcacctatctcagcgat ctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggc cccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccgga agggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagt aagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggta tggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagct ccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattct cttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcg gcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatc attggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgt gcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgc aaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagg gttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccg***t*** ***ccacccaagcggccggagaacctgcgtgcaatccatcttgttcaatcat*****ggtggagggcccccccgcagatct** **gagctttttgcaaaagccta**
GFP-KO	**tttaaagcaagtaaaacctctacaaatgtgg*tatggctgattatgatcagggtggcgaccggtagcgcta***tag ggataacagggtaatttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttg cctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcg agacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggt cctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagttt gcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcc

gBlock	Sequence
	caacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtc agaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaa gatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgccc ggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggc gaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcat cttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcga cacggaaatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggata catatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccg***gtatggctgattatgatcagg*** ***gtggcgaccggtagcgcta***gcggatctgacggttcactaaaccagctct
BXLF1-to-mGL	**agaccggcacataccttcacaatgtcctcgggggcaaaatactgtgttag*ctacttgtacagctcgtccatgtc*** ***atgtgtaatcccggcggcggtcaccc***cggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaa atatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaaca tttccgtgtcgcccttattccctttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaa gatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttt tcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgcc gggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagc atcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaactta cttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttga tcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaa caacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcg gataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtga gcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacg gggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaaa ttaccctgttatccct***a******ctacttgtacagctcgtccatgtcatgtgtaatcccggcggcggtcaccc***tctccttcagga ccatgtgatcgcgcttctcgttggggtctttgctcagtttggactgatggctcaggtagtggttgtcgggcagcagca cggggccgtcgccgatgggggtgttctgctggtagtggtcggcgagctgcacgccgccgtcctcaacgttgtggc gggtcttgaagttagccttgatgccgttcttctgcttgtcggccgtgatatagaccttgtggctgttgaagttgtactcca gcttgtgccccaggatgttgccgtcctccttgaagtcgatgcccttcagcacgatgcggttcaccagggtgtcgccc tcgaacttcacctcggcgcgggtcttgtaggtaccgtcgtccttgaaagagatggtgcgctcctggacgtagccttc gggcatggcggacttgaagaagtcgtgctgcttcatgtggtcggggtagcgggcgaagcaggccacgccgtagc ctaaggtggtcacgagggtgggccagggcacgggcagcttgccggtggtgcagatgaacttcagggtcagcttg ccgttggtggcatcgccctcgccctcgccgcggacgctgaacttgtggccgtttacgtcgccgtccagctcgacca ggatgggcaccaccccggtgaacagctcctcgcccttgctcaccat**ggatcccacccgaggtctgtgtgccctct** **ccacgcacgacctggcaaaat**
BILF2-tPT2A-mSI	**acgcgtggcttgaaagtaaactttattgcgtgttagtaactgtccattca**cttgtacagctcgtccatgccgccg gtggagtggcggccctcggagcgttcgtactgttccaccacggtgtagtcctcgttgtgggaggtgatgtccaactt gcggtcgacgttgtaggcgccgggcatctgcacgggcttcttggccttgtaggtggtcttgaagtccgccaggtag cggccgccgtccttcaggcgcagggccatcttaatgtcgcccttcagcacgccgtcctcggggtacaaccgctcg gtggacgcttcccagcccattgtcttcttctgcattacggggccgtcaggagggaagttggtgccgcggagcttcac cttgtagatcagggtgccgtcctccagggaggtgtcctgggtcacggtcacggcgccgccgtcctcgaagttcatc acgcgctcccacttgaagccctcggggaaggactgcttatagtagtcggggatgtcggcggggtgcttgatgaag gccctggagccgtacatgaactgaggggacaggatgtcccaggagaagggcaggggggccacccttggtcacct tcagcttggcggtctgggtgccctcgtaggggcggccctcgccctcgccctcgatctcgaactcgtggccgttcat ggagccctccatgtgcaccttgaaccgcatgaactccttgatcactgcctcgcccttgctcaccatctcgagtgggc cgggattttcctccacgtccccgcatgttagaagacttcccctgccctcgccggagccaagctttccagcactagta

gBlock	Sequence
	ggcccggggttttcttcaacatcacctgcttgctttagcagagagaagtttgtggcgccgctgcccaggggtatcca gcccttgcgccgcctcccccc***agcccgccagccacccccagacaggagatgataatgatgaggagcaccgga***t agggataacagggtaatttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagt tgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccg cgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtg gtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagt ttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttc ccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgt cagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgta agatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcc cggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcgggg cgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagca tcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcga cacggaaatgttgaatactcatactcttccttttttcaatattattgaagcatttatcagggttattgtctcatgagcggata catatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccg***agcccgccagccaccccag*** ***acaggagatgataatgatgaggagcaccgga*****gccaccacagcacaagtgattaggagcagggcccagtg** **cacccaggtggt**
BGLF5-KO	**agcatagactccacccccctgttggcagctaggctgtctgccaccaggtcat**cactattaactatcattaatcatt actactcatccacgtcggccat***ctggaccaaagaagtcgtctgccaagagttcagctaccag***tagggataacag ggtaatttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccc cgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacg ctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaacttt atccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgtt gttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaa ggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagt tggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctg tgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaata cgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctc aaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttca ccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatg ttgaatactcatactcttccttttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatg tatttagaaaaataaacaaatagggggttccgcgcacatttccccg***ctggaccaaagaagtcgtctgccaagagttc*** ***agctaccag***acctggaaga
BILF2-to-tLNGFR	**cacatatttaggcctgagaagcagggagttttaaaactgacagtttagga**atgggggcaggtgccaccggcc gcgccatggacgggccgcgcctgctgctgttgctgcttctgggggtgtcccttggaggtgccaaggaggcatgcc ccacaggcctgtacacacacagcggtgagtgctgcaaagcctgcaacctgggcgagggtgtggcccagccttg tggagccaaccagaccgtgtgtgagccctgcctggacagcgtgacgttctccgacgtggtgagcgcgaccgagc cgtgcaagccgtgcaccgagtgcgtggggctccagagcatgtcggcgccgtgcgtggaggccgacgacgccgt gtgccgctgcgcctacggctactaccaggatgagacgactgggcgctgcgaggcgtgccgcgtgtgcgaggcgg gctcgggcctcgtgttctcctgccaggacaagcagaacaccgtgtgcgaggagtgccccgacggcacgtattcc gacgaggccaaccacgtggacccgtgcctgccctgcaccgtgtgcgaggacaccgagcgccagctccgcgag tgcacacgctgggccgacgccgagtgcgaggagatccctggccgttggattacacggtccacacccccagagg gctcggacagcacagcccccagcacccaggagcctgaggcacctccagaacaagacctcatagccagcac ggtggcaggtgtggtgaccacagtgatgggcagctcccagcccgtggtgacccgaggcaccaccgacaacctc atccctgtctattgctccatcctggctgctgt***ggttgtgggtcttgtggcctacatagccttcaagaggtga***tagg gataacagggtaatttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagtt

gBlock	Sequence
	gcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccg cgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagt ggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaat agtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccg gttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgat cgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgcc atccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagtt gctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaa cgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcaccca actgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaa agggaataagggcgacacggaaatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggtt attgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccg***g gttgtgggtcttgtggcctacatagccttcaagaggtga***atggacaggtactaacacgcaataaagtttactt tcaagccacgcgttta
BXLF1-to-eGFP	**attttgccaggtcgtgtgtggagagggcacacagaccccgggtgggatcc**atggtgagcaagggcgaggag ctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccgg cgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtg ccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaag cagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacg gcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggca tcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatat catggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgt gcagctcgccgaccactaccagcagaacaccccccatcggcgacggccccgtgctgctgcccgacaaccact acctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagt***tcgt gaccgccgccgggatcactctcggcatggacgagctgtacaagtaa***tagggataacagggtaatttaccaat gcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtaga taactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggc tccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcct ccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgcca ttgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcga gttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggc cgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtg actggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatac gggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactct caaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttact ttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacgg aaatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacat atttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccg***tcgtgaccgccgccgggatc actctcggcatggacgagctgtacaagtaa***ctaacacagtattttgccccgaggacattgtgaaggtatgt gccggtct
BILF2-to-mCherry	**cacatatttaggcctgagaagcagggagttttaaaactgacagtttagga**atggtgagcaagggcgaggagg ataacatggccatcatcaaggagttcatgcgcttcaaggtgcacatggagggctccgtgaacggccacgagttc gagatcgagggcgagggcgagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaaggg tggcccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaaggcctacgtgaagcacc ccgccgacatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaacttcgag gacggcggcgtggtgaccgtgacccaggactcctccctgcaggacggcgagttcatctacaaggtgaagctgcg cggcaccaacttcccctccgacggccccgtaatgcagaagaagaccatgggctgggaggcctcctccgagcg

gBlock	Sequence
	gatgtaccccgaggacggcgccctgaagggcgagatcaagcagaggctgaagctgaaggacggcggccacta cgacgctgaggtcaagaccacctacaaggccaagaagcccgtgcagctgcccggcgcctacaacgtcaaca tcaagttggacatcacctcccacaacgaggactacaccatcgtggaacagtacgaacgcgccgaggg***ccgcc*** ***actccaccggcggcatggacgagctgtacaagtag***tagggataacagggtaatttaccaatgcttaatcagtg aggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgata cgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatc agcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtct attaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggc atcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatc ccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttat cactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagta ctcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataatac cgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatctt accgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcg tttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaat actcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtat ttagaaaaataaacaaatagggttccgcgcacatttccccg***ccgccactccaccggcggcatggacgagc*** ***tgtacaagtag*atggacaggtactaacacgcaataaagtttactttcaagccacgcgttta**

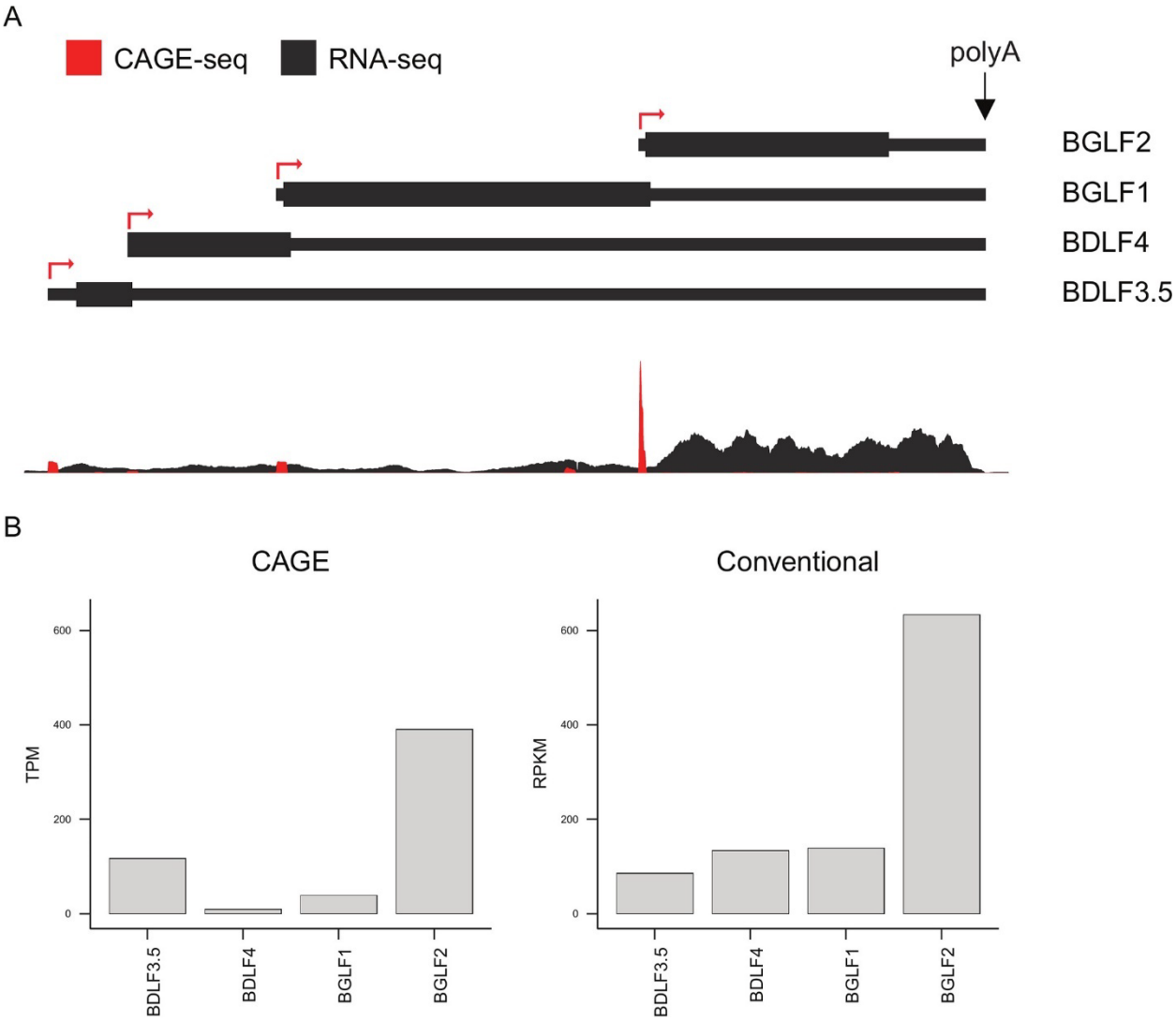
Appendix III

A Computational Method for Quantifying Herpesvirus Gene Expression

This chapter is composed primarily of a research article entitled “Accurate Quantification of Overlapping Herpesvirus Transcripts from RNA Sequencing Data” by Alejandro Casco, Akansha Gupta, Mitchell Hayes, Reza Djavadian, Makoto Ohashi, and Eric Johannsen published in the *Journal of Virology* January 2022.

Introduction

Epstein-Barr virus (EBV) encodes more than 100 transcripts within its 172-kb genome. While this transcript density is typical for a herpesvirus, coding regions in host genomes are orders of magnitude less compact. As a consequence, bioinformatic approaches that accurately quantify host gene transcripts may present unexpected difficulties when applied to viral gene analysis. For example, the noncoding BART transcripts expressed at high levels in most EBV epithelial tumors are frequently misattributed to the overlapping BALF3, BALF4, BALF5, and LF2 genes [1–4]. Although this error can be readily avoided using strand-specific RNA sequencing (RNA-seq) analysis, EBV lytic transcription analysis presents challenges that are more difficult to address. The principal challenge arises from the extensive use of shared poly(A) sites in herpesvirus lytic gene groupings, referred to below as "gene clusters" (Appendix Figure III-1A). While each gene is expressed from a unique promoter, all but the shortest transcript contain additional open reading frames (ORFs) in their 3′ untranslated regions (3′ UTRs). In general, these transcripts lack an internal ribosome entry site (IRES), and therefore, only the first ORF is translated. We have demonstrated previously the inaccurate results of quantitative PCR (qPCR)-based transcript quantification that arise from this overlapping, nonpolycistronic gene structure [5]. Likewise, conventional short-read RNA-seq analysis, where fragments are first mapped to the genome and then assigned to specific genes based on ORF annotations, can result in marked overestimation of the 3′ gene's abundance due to incorrect attribution to its transcript of reads from upstream transcripts (Appendix Figure III-1B, right). Although techniques that quantify only the 5′ region of each transcript, such as cap analysis of gene expression sequencing (CAGE-seq), are not susceptible to these artifacts (Appendix Figure III-1B, left), their expense and availability limit general application.



Appendix Figure III-1 Herpesvirus gene clusters, which share a polyadenylation signal, result in overlapping transcripts that confound conventional RNA-seq read assignment to genes. (A) (Top) Schematic of EBV genomic region containing the leftward (reverse-stranded) overlapping gene cluster consisting of BDLF3.5, BDLF4, BGLF1, and BGLF2, which share a single polyadenylation signal (vertical arrow). mRNA transcripts encoding each protein product are shown, including coding sequences (wide solid bars) and untranslated regions (narrow solid bars). (Bottom) RNA-seq read depth is displayed in black, and CAGE-seq signals in red, from experiments performed in Akata cells induced for lytic replication (11). (B) Bar plots displaying transcript abundance estimates derived from CAGE-seq data (left), expressed as tags per million (TPM), or from conventional analysis of RNA-seq data (right), expressed as reads per kilobase per million mapped reads (RPKM).

Previously, Bruce et al. [6] designed a method utilizing unique coding sequences (UCDS) to estimate the expression of overlapping genes in Kaposi's sarcoma-associated herpesvirus

(KSHV). This method involves subdividing the genome into regions that contain a unique set of transcripts (i.e., no transcripts begin or end within a UCDS region). Subsequently, the read depth difference between two UCDS regions can be used to infer the expression level of the transcript that is present in only one of the UCDS regions. Despite its recent use in the herpesvirus field [6–9], the accuracy of this method, to our knowledge, has not been empirically validated. The availability of duplicate cell pellets from our recently published CAGE-seq analysis of EBV lytic transcription [5] presented such an opportunity. RNA libraries were prepared from three different experimental conditions, and Illumina short-read RNA-seq was performed. Using these data and our previously published CAGE-seq results, we assessed the accuracy of the UCDS method. In addition, using transcription start site (TSS) data from that study combined with recent data characterizing lytic poly(A) sites [10], we assembled an accurate map of EBV lytic transcripts. This map should allow even more accurate quantification of EBV lytic gene expression via a method we designate the Unique TranScript (UTS) method. In this article, we analyze the abilities of the UCDS and UTS methods to accurately quantify EBV lytic gene expression by using CAGE-seq as a gold standard.

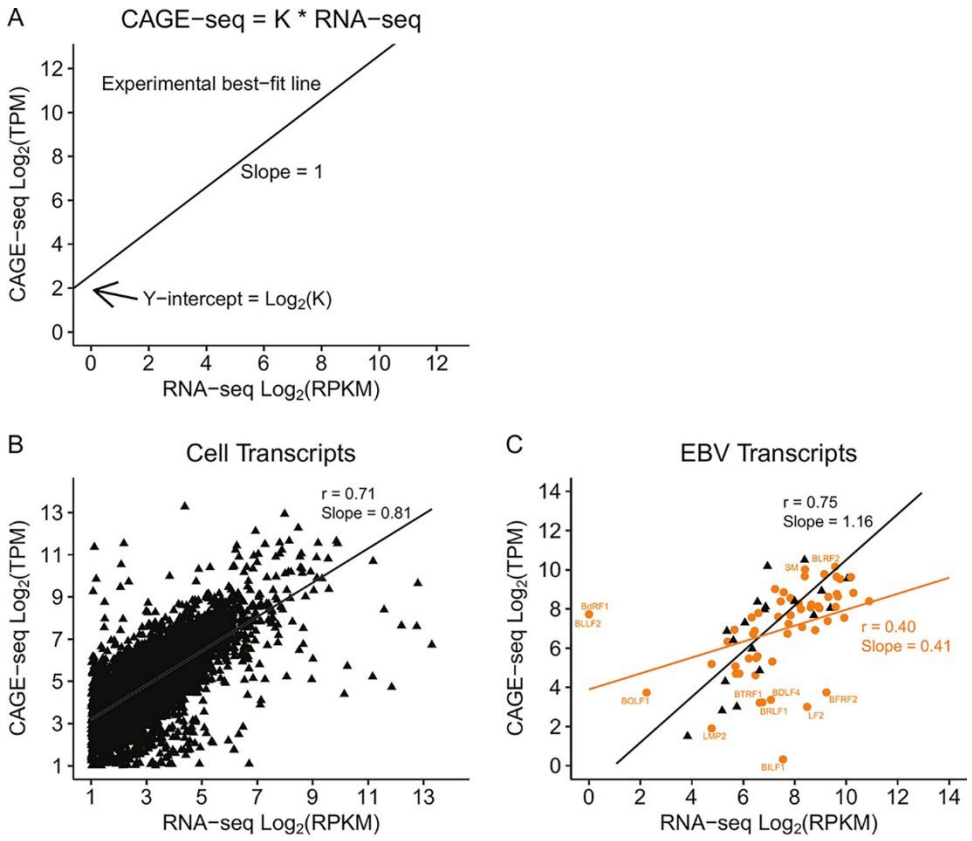
Results

Conventional (ORF based) RNA-seq read assignment is inaccurate for viral gene quantification but performs well for host genes.

The example presented in Appendix Figure III-1 illustrates a viral gene cluster whose transcripts are not directly quantifiable by conventional RNA-seq. However, which genes are subject to the greatest measurement error when this overlap is not accounted for is unknown. As an initial assessment, we reanalyzed a published data set from the Akata cell line for which both CAGE-seq and RNA-seq data were available [11]. In this analysis, we estimated the normalized

expression level of each host and EBV gene from the RNA-seq data in reads per kilobase per million reads mapped (RPKM), as detailed in Materials and Methods. As a gold standard, we also used CAGE-seq data to quantify the expression level of each gene in transcripts per million (TPM). Although RPKM and TPM cannot be readily interconverted, a valid RNA-seq analysis should at least produce RPKM values that are proportional to the TPM values obtained by CAGE-seq [i.e., CAGE-seq (TPM) = K × RNA-seq (RPKM)]. The constant K can reflect both universal and experiment-specific factors that result in differences in absolute quantification between methods. Such factors could include differences in library preparation chemistries and the efficiency of rRNA depletion. As shown in Appendix Figure III-2A, plotting this comparison as a log-log plot of CAGE-seq versus RNA-seq data should result in a slope of 1, with a nonzero y-intercept related to K. This transformation also minimizes the effect of variance increasing with gene expression levels (i.e., heteroscedasticity) commonly seen in RNA-seq data. We quantify the strength of the correlation between the two estimates across all genes by calculating the Pearson correlation coefficient, r. For host genes, in agreement with previous studies [12, 13], we observed a strong correlation between the CAGE-seq and RNA-seq estimates, with a slope of 0.81 and a correlation coefficient of 0.71 (Appendix Figure III-2B). Likewise, nonoverlapping EBV gene expression estimates derived from conventional RNA-seq analysis correlated strongly with those obtained from CAGE-seq, with a slope of 1.16 and a correlation coefficient of 0.75 (Appendix Figure III-2C). In contrast, expression level estimates for clustered EBV lytic genes whose transcripts overlap partially or completely with other lytic transcripts correlated poorly with CAGE-seq estimates (slope, 0.41; correlation coefficient, 0.40) (Appendix Figure III-2C). These results demonstrate that conventional RNA-seq analysis with read assignment based solely on ORF annotations is unable to quantify the expression levels of

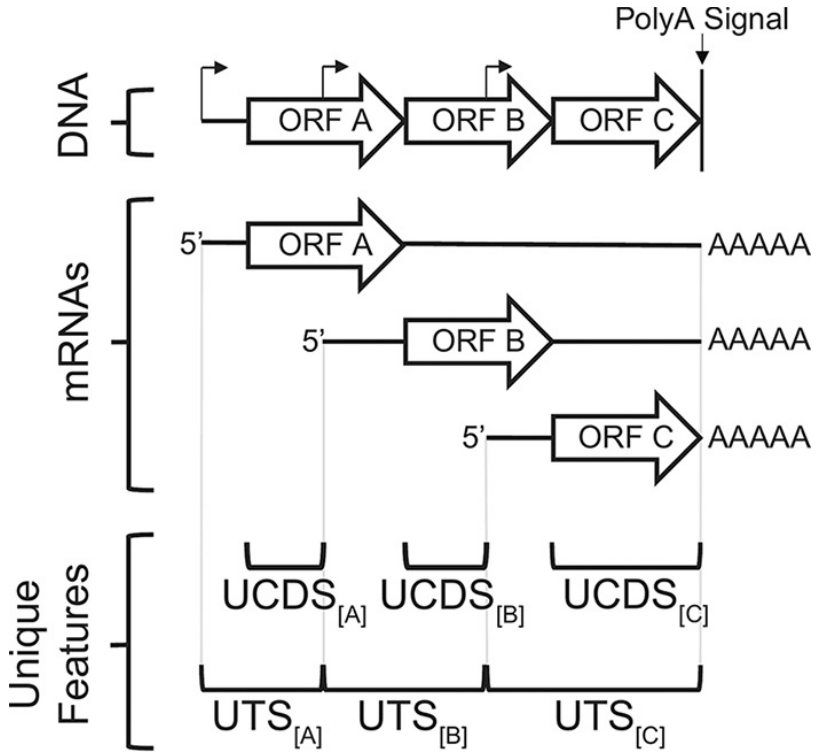
overlapping EBV genes accurately, despite accurately estimating host gene and nonoverlapping EBV gene expression.



Appendix Figure III-2 Conventional RNA-seq analysis produces inaccurate estimates of viral gene expression relative to the gold standard, CAGE-seq. (A) Expected log-log plot of CAGE-seq versus RNA-seq data when both methods produce accurate relative transcript quantification [i.e., CAGE-seq (TPM) = K × RNA-seq (RPKM)]. K is a constant value that reflects both universal and experiment-specific factors that give different absolute quantifications by the CAGE-seq and RNA-seq methods (see the text for discussion of specific examples). (B) Log-log dot plot comparing expression level estimates of host genes determined by CAGE-seq (y axis) versus conventional RNA-seq (x axis) analysis from EBV-positive Akata Burkitt lymphoma cells (11). Note that the slope is close to the idealized value of 1, with a high Pearson correlation coefficient (r) indicating good agreement between CAGE-seq and RNA-seq relative abundance estimates for host genes. (C) Dot plot for EBV lytic genes from the same analysis as described for panel B. Potential outliers are labeled and were scored using Cook's distance with a threshold of $1/N$ ($N = 75$). Note that the slope for overlapping EBV genes (orange circles and line) deviates significantly from 1 and has a lower Pearson correlation coefficient (r), whereas nonoverlapping EBV genes (black triangles and line) are comparable to host genes. This indicates that conventional RNA-seq analysis produces inaccurate estimates of EBV lytic genes with overlapping transcripts.

Accounting for viral transcript overlap substantially improves estimates of lytic gene expression from RNA-seq data.

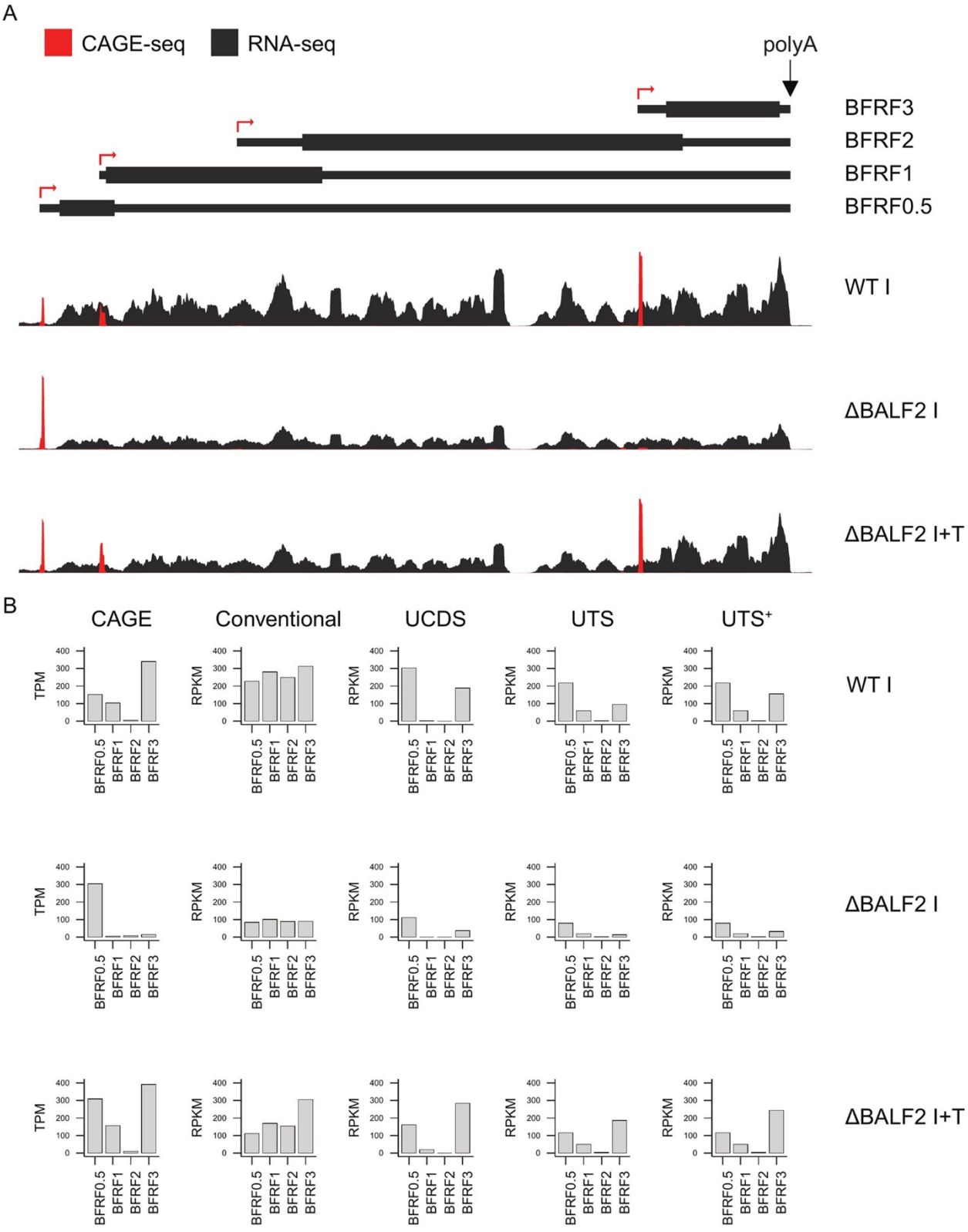
Bruce et al. [6] have previously proposed an approach to account for viral transcript overlap, which they call the unique coding sequence (UCDS) method. This method requires knowledge of which transcripts share a polyadenylation signal but does not require detailed information regarding transcription start site (TSS) locations. Instead, regions of the genome likely to be overlapped by a specific number of transcripts are defined and are used sequentially to estimate the abundance of each transcript, starting from the most 5′ (i.e., longest) transcript (illustrated in Appendix Figure III-3). These regions, based on ORF annotations, are trimmed to maximize the likelihood that 5′ UTRs from downstream genes are excluded. Due to this trimming, some RNA-seq reads are not assignable to any UCDS region, decreasing effective sequencing depth in an effort to improve accuracy. We assessed the extent to which this method improves accuracy in EBV lytic gene abundance estimates. In addition, because we have detailed information regarding the TSSs of EBV lytic transcripts, we also assessed a novel method that builds on the UCDS approach, without discarding many of the reads that the UCDS method requires to be discarded. We termed this approach the Unique TranScript (UTS) method (Appendix Figure III-3).



Appendix Figure III-3 Schematic of the Unique TranScript (UTS) and unique coding sequence (UCDS) methods for abundance estimation of overlapping genes. Shown is an example of a transcription unit with multiple coterminal transcripts, including some with no unique sequence. In this case, the RPKM for UTS/UCDS regions with no unique sequence are calculated by subtracting the RPKM attributable to the longer overlapping transcripts, as detailed in the text.

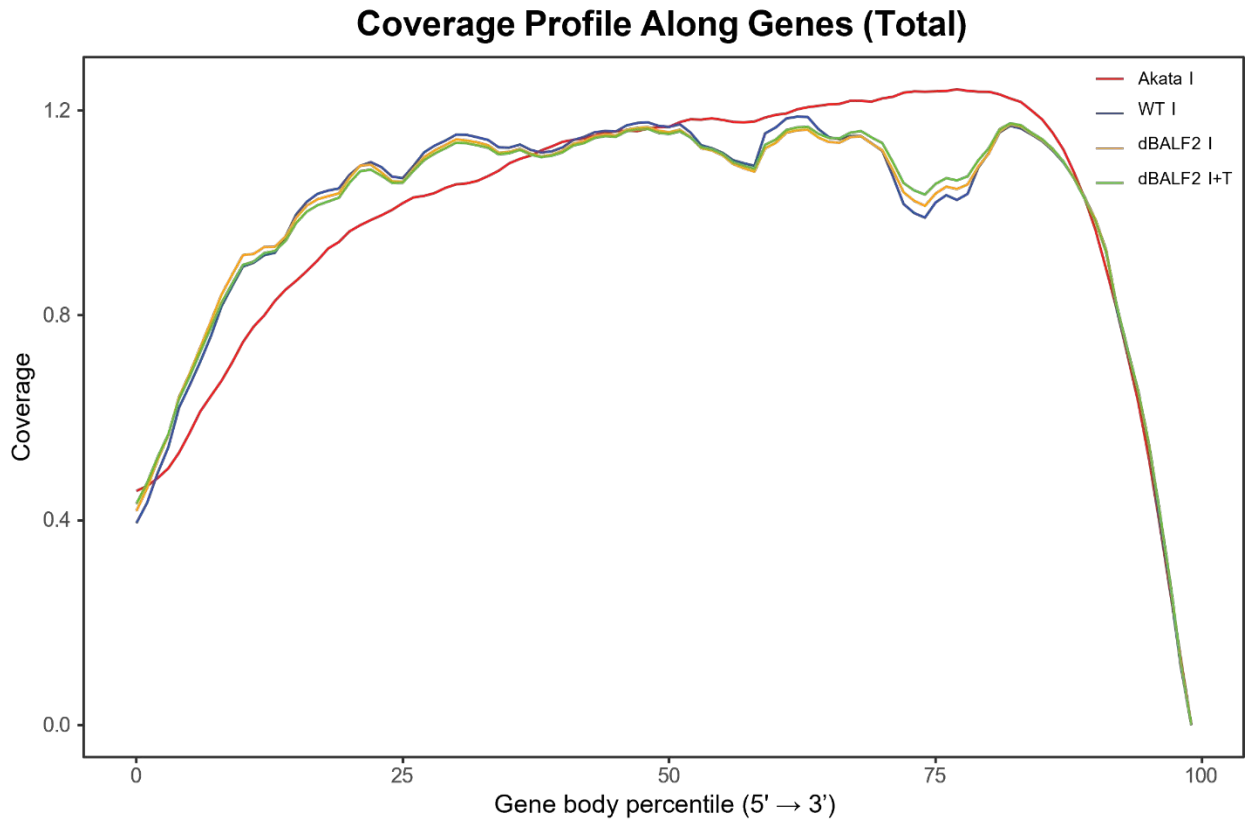
To evaluate the UCDS and UTS methods, we performed RNA-seq on duplicate cell pellets corresponding to conditions we had previously analyzed by CAGE-seq [5]. Three conditions were chosen, two of which express the full complement of EBV lytic genes (wild-type EBV induced [WT I] and EBVΔBALF2 induced and trans-complemented [ΔBALF2 I+T]) and one of which is restricted to early gene expression (EBVΔBALF2 induced [ΔBALF2 I]). RNA-seq data were aligned to the GDC GRCh38 index using STAR; reads were assigned to EBV genes using the UCDS or UTS approach; and abundance estimates for each transcript were calculated (in RPKM) as detailed in Materials and Methods. The genomic region containing the overlapping BFRF0.5–BFRF1–BFRF2–BFRF3 genes sharing a poly(A) site is instructive of the

differences among these approaches (Appendix Figure III-4A). The conventional (ORF based) method resulted in the least accurate estimates of gene expression across all three data sets (Appendix Figure III-4B, compare CAGE to Conventional). In all cases, the expression level of BFRF2 is greatly overestimated due to the misattribution of reads from the upstream BFRF0.5 and BFRF1 transcripts to this gene. In the worst case, conventional read assignment falsely measures the expression of the BFRF1 and BFRF3 late gene transcripts from an EBV mutant defective in late gene expression at levels comparable to those of early genes (Appendix Figure III-4B, compare CAGE to Conventional for the ΔBALF2 I condition). In contrast, both the UTS and UCDS methods improved the gene expression estimates of overlapping genes across all three data sets, as demonstrated by patterns of gene expression similar to those with CAGE-seq (Appendix Figure III-4B, compare CAGE to UCDS and UTS). However, the UTS method underestimated the abundance of BFRF3, the most 3′ transcript. We speculated that this may be due to coverage bias observed at transcript ends (Appendix Figure III-5). In order to account for this bias, we modified our UTS method to use an effective transcript length for calculating read depth (see Materials and Methods for details). This modification of the UTS method was designated UTS^{+} and substantially improved estimates for BFRF3 (Appendix Figure III-4B, compare CAGE to UTS^{+}) as well as for other overlapping EBV lytic transcripts whose transcription start sites are closest to the shared poly(A) site (i.e., the shortest transcript in each cluster).



Appendix Figure III-4 Example of improved EBV transcript quantification by the UTS and UCDS methods. (A) Schematic of the EBV genomic region containing the rightward

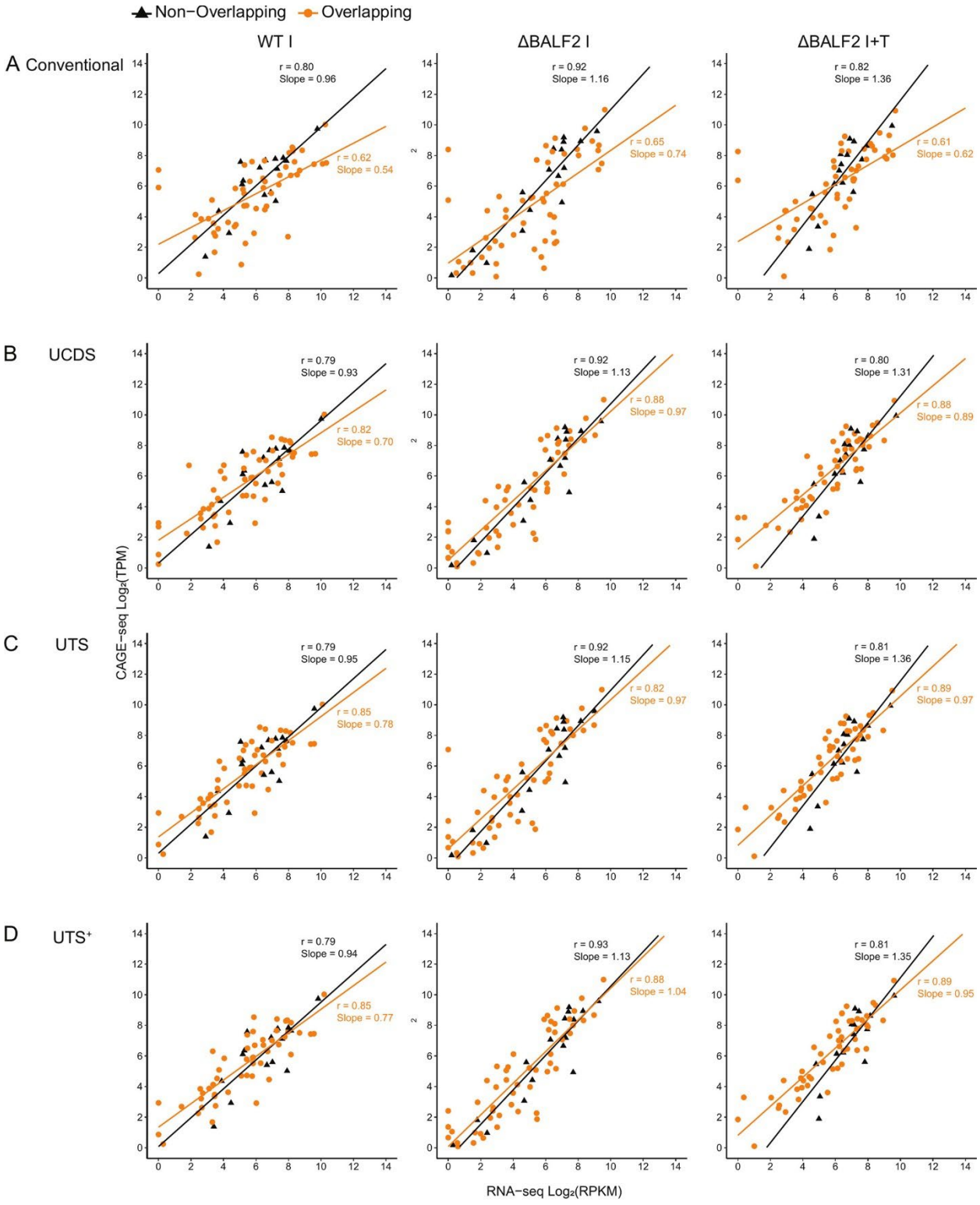
(forward-stranded) overlapping genes BFRF0.5, BFRF1, BFRF2, and BFRF3, which share a single polyadenylation signal (vertical arrow). RNA-seq read depth is displayed in black, and CAGE-seq signals in red, from three different experiments (WT-I, ΔBALF2-I, and ΔBALF2-I+T), described in more detail below. mRNA transcripts encoding each protein product are shown, including coding sequences (wide solid bars) and UTRs (narrow solid bars). (B) Bar plots displaying EBV transcript abundance estimates derived from CAGE-seq (in TPM) or various RNA-seq analysis methodologies (conventional, UCDS, UTS, and UTS^{+}, all expressed in RPKM) for the following three experiments: HEK293 cells infected with EBV induced for lytic replication (WT-I), HEK293 cells infected with EBV with the BALF2 gene deleted ($EBV_{\Delta BALF2}$), induced for lytic replication (ΔBALF2 I), and HEK293 cells infected with $EBV_{\Delta BALF2}$ induced for lytic replication and trans-complemented with BALF2 (ΔBALF2 I+T).



Appendix Figure III-5 Read depth plots as a function of transcript position for RNA-seq data. Total count depth for all host transcripts was calculated using Qualimap v2.2.1 (29) and was plotted for each of three RNA-seq experiments from this manuscript (WT-I, ΔBALF2 I, and ΔBALF2 I+T) and published Akata RNA-seq data as a function of position within each transcript (30). Note the relative decrease in depth at the 5' and 3' transcript ends.

To analyze more comprehensively how each method affected estimates of lytic gene expression, we plotted the normalized expression of most EBV genes obtained by CAGE-seq

versus RNA-seq (Appendix Figure III-6). In this comparison, we censored genes subject to artifacts arising from the transfection of BZLF1, BRLF1, and BALF2. For nonoverlapping genes across all three conditions, each RNA-seq read assignment method resulted in strong correlation with CAGE-seq. In contrast, analysis of overlapping genes by the conventional method resulted in poor correlation with CAGE-seq for each condition (Appendix Figure III-6, top, orange circles and lines). CAGE-seq estimates were well approximated by both the UCDS (slopes, 0.70 to 0.97; r, 0.82 to 0.88) and UTS (slopes, 0.78 to 0.97; r, 0.82 to 0.89) methods. This result suggests that the UTS method markedly improves the expression level estimates of overlapping EBV genes. Further, the UTS^{+} method, which also corrects for read distribution bias, produced abundance estimates that correlated most closely with the CAGE-seq gold standard (slopes, 0.77 to 1.04; r, 0.85 to 0.89).

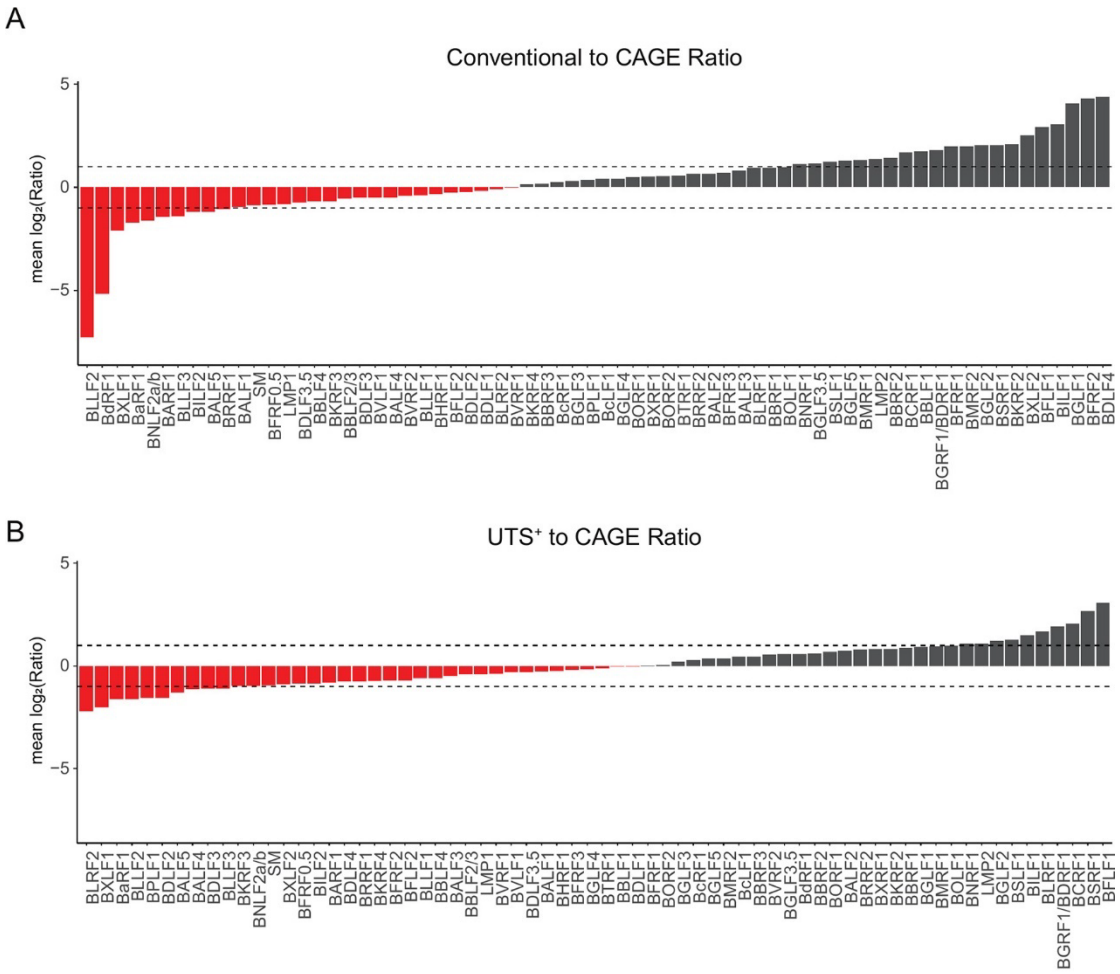


Appendix Figure III-6 Global comparison of EBV gene transcript quantification of RNA-seq data by various methodologies relative to the gold-standard, CAGE-seq. (A) Log-log dot plots of estimated EBV gene expression levels by CAGE-seq (y axis, in TPM) versus

conventional RNA-seq (x axis, in RPKM) analysis from three different experiments (WT I, ΔBALF2 I, and ΔBALF2 I+T). Linear regression of data is drawn for nonoverlapping genes (black lines) and for genes with nested overlapping structure (orange lines); Pearson's correlation coefficient (*r*) and the slope are indicated. (B through D) Log-log dot plots and linear regression analyses as described for panel A, comparing CAGE-seq to RNA-seq analyzed by UCDS (B), UTS (C), or UTS$^+$ (D) methodology.

Identification of EBV genes and gene clusters most affected by the RNA-seq read assignment method.

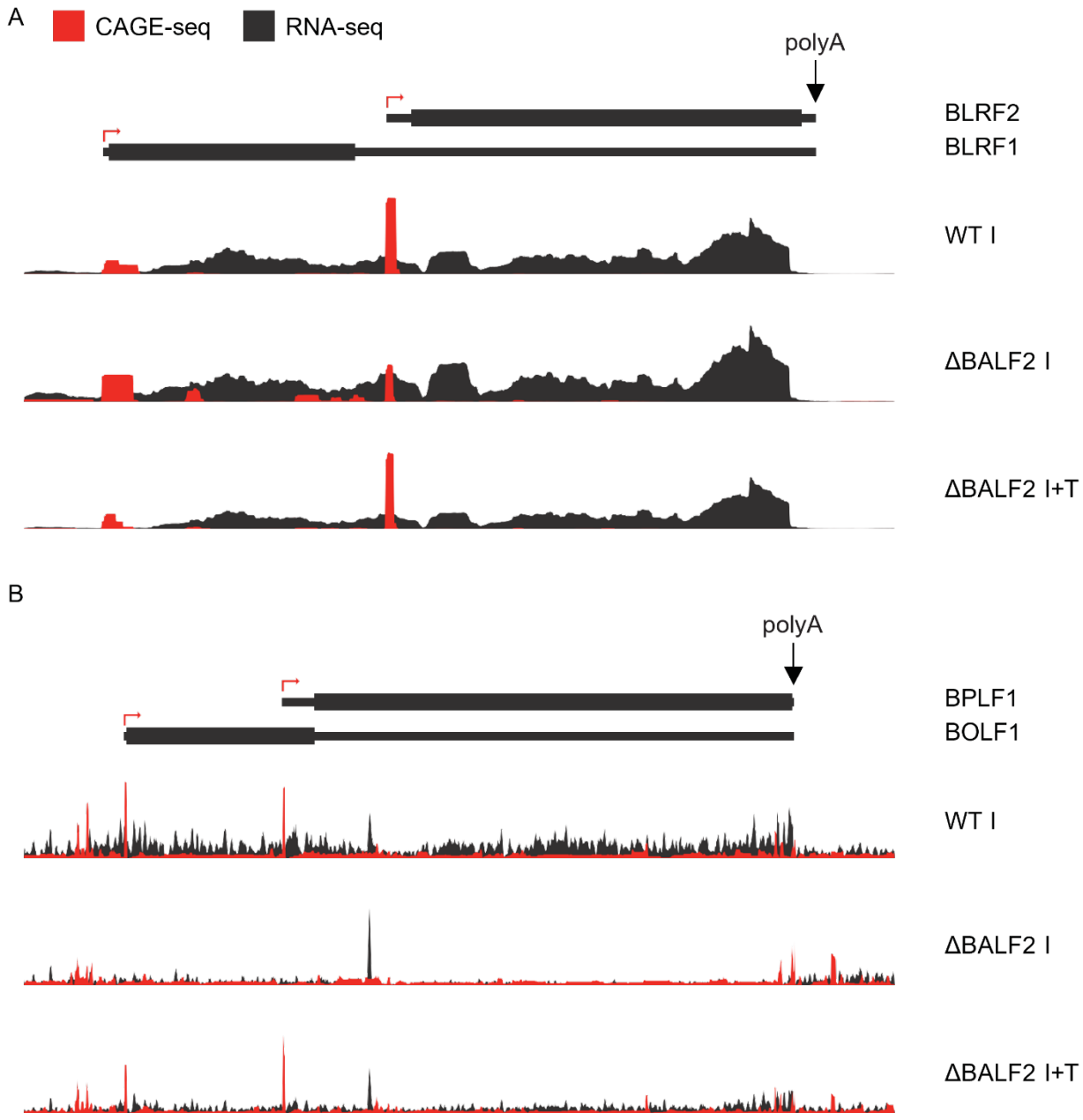
To determine which EBV genes were most subject to errors in measurement of their abundances, we first calculated the expected expression level for the gene in RPKM (using the linear relationship determined from the nonoverlapping EBV gene linear regression [Appendix Figure III-6, black lines]). This expected value was then compared to the abundance estimate derived from the conventional and UTS$^+$ methods (Appendix Figure III-7A and B). With the conventional approach, about 55% of the genes were within 1 log unit (2-fold) of the CAGE-seq estimate. This improved to about 70% of the genes with the UTS$^+$ methodology. The most striking improvements were seen with BLLF2 and BdRF1, genes that are completely overlapped by BLLF1 and BVRF2, respectively. Since these ORFs have no unique sequence, the conventional method does not assign any RNA-seq reads to either gene, creating the false impression that they are not expressed. We also examined the gene clusters that were most sensitive to the read assignment method (Appendix Figure III-7C). Both UCDS and UTS$^+$ produced improvements for many clusters; the most dramatic results were seen for the BLLF2- and BdRF1-containing clusters, as well as the BXLF1–BXLF2 cluster. At least four other clusters showed more-modest improvements in correlations with CAGE-seq data. In the case of the BGLF3–BGLF3.5–BGLF4–BGLF5–BBLF1 cluster, the improved performance of the UTS$^+$ method resulted primarily from the utilization of many reads that did not fall within a UCDS region and therefore had to be discarded by the UCDS method.


A
Conventional to CAGE Ratio
mean log2(Ratio)
5
0
−5
B
UTS+ to CAGE Ratio
mean log2(Ratio)
5
0
−5
C
Conventional
UCDS
UTS+
Cluster
NON-OVERLAPPING
LMP2--BNRF1
BFRF0.5--BFRF1--BFRF2--BFRF3
BORF1--BORF2
BaRF1--BMRF1--BMRF2
BLRF1--BLRF2
BKRF2--BKRF3--BKRF4
BBRF1--BBRF2
BcRF1--BTRF1
BVRF1--BXRF1
BdRF1--BVRF2
LMP1--BNLF2a/b
BALF3--BALF4
BALF5--BILF1
BXLF1--BXLF2
BDLF1--BDLF2--BDLF3
BDLF3.5--BDLF4--BGLF1--BGLF2
BGLF3--BGLF3.5--BGLF4--BGLF5--BBLF1
BLLF1--BLLF2
BSLF1--SM
BOLF1--BPLF1
BFLF1--BFLF2
Pearson's Correlation
-1
0
1

Appendix Figure III-7 Performance of UTS versus conventional methods relative to CAGE-seq across all EBV genes and by gene cluster. (A) Bar plots showing the log of the ratio of transcript abundance estimates for conventional RNA-seq (top) or the UTS$^+$ method (bottom) to those for CAGE-seq. Transcripts whose abundance is underestimated relative to CAGE-seq (red bars) appear on the left, whereas transcripts whose abundance is overestimated relative to CAGE-seq (black bars) appear on the right. (B) Heat map showing the Pearson correlation coefficient between CAGE-seq and the indicated read assignment methods for RNA-seq. Each row corresponds to either all nonoverlapping EBV lytic genes (the top row) or the indicated cluster of EBV lytic genes that share a poly(A) signal. Correlation coefficients were calculated by Pearson's correlation on CAGE-seq and RNA-seq genes within each cluster.

Two clusters, BOLF1–BPLF1 and BLRF1–BLRF2, were notable in that the conventional method appeared to outperform both the UCDS and the UTS$^+$ method. To better understand these discrepancies, we compared the CAGE-seq signal strengths with the read depth from RNA-seq mapping for each experimental condition (Appendix Figure III-8). In the case of BOLF1–BPLF1, the most straightforward explanation would be that these two genes do not actually share a polyadenylation signal (and therefore do not overlap as is assumed by the UCDS and UTS$^+$ methods). We searched several RNA-seq data sets for evidence for this putative BOLF1 polyadenylation site but could not confirm this lack of overlap of BOLF1 and BPLF1. However, it should be noted that the low abundance of the BOLF1 transcripts does not permit us to exclude the possibility that BOLF1 has an unannotated polyadenylation site. In the case of the BLR1–BLRF2 cluster, the UCDS and UTS$^+$ methods appeared to be identifying bona fide differences between the RNA-seq and CAGE-seq measurements of transcript abundance (i.e., the apparent correlation between the conventional method and CAGE-seq belies actual discrepancies in our data). However, we did not observe these discrepancies for this cluster between RNA-seq read depth and CAGE-seq in the published Akata data [11]. Thus, it is unclear whether the BLRF1-BLRF2 cluster is generally problematic for RNA-seq quantification or is merely an outlier in our data set. Finally, a third cluster, BaRF1–BMRF1–BMRF2, was remarkable in that none of the

RNA-seq read assignment methods produced a satisfactory correlation with CAGE-seq data. We examine the cause of this phenomenon in the next section.

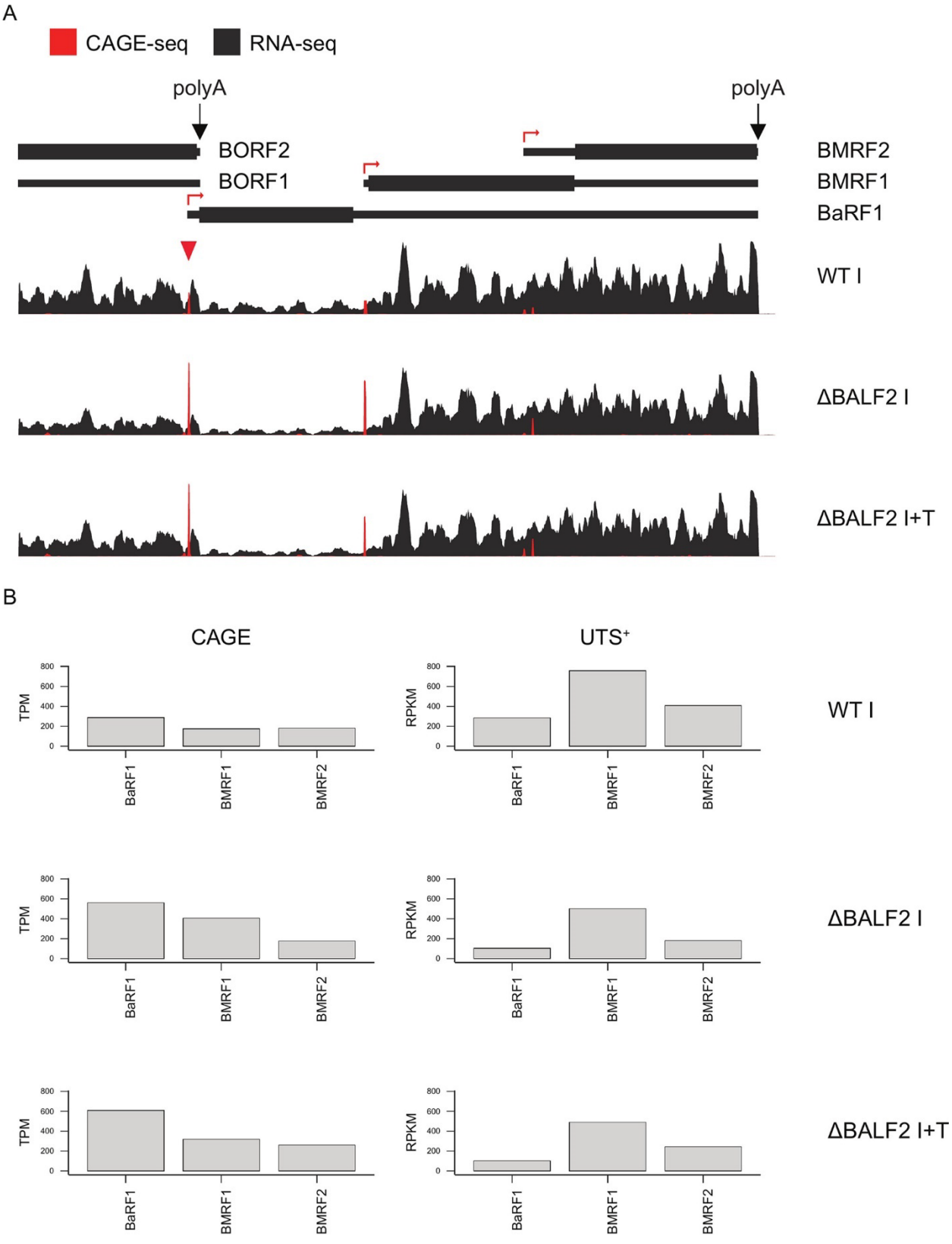


Appendix Figure III-8 RNA-seq read depth and CAGE-seq data for BLRF1-BLRF2 and BOLF1-BPLF1 clusters. (A) Genomic region corresponding to the rightward overlapping genes BLRF1 and BLRF2 that share a polyA site. RNA-seq read depth is displayed in black and CAGE-seq signals in red from the 3 experiments (WT-I, dBALF2-I, and dBALF2-I+T) – described in Appendix Figure III-4. mRNA transcripts encoding each protein product are shown including coding sequences (wide black bars) and UTRs (narrow black bars). Note that the sequencing depth corresponding to the BLRF2 transcript is much less than would be expected

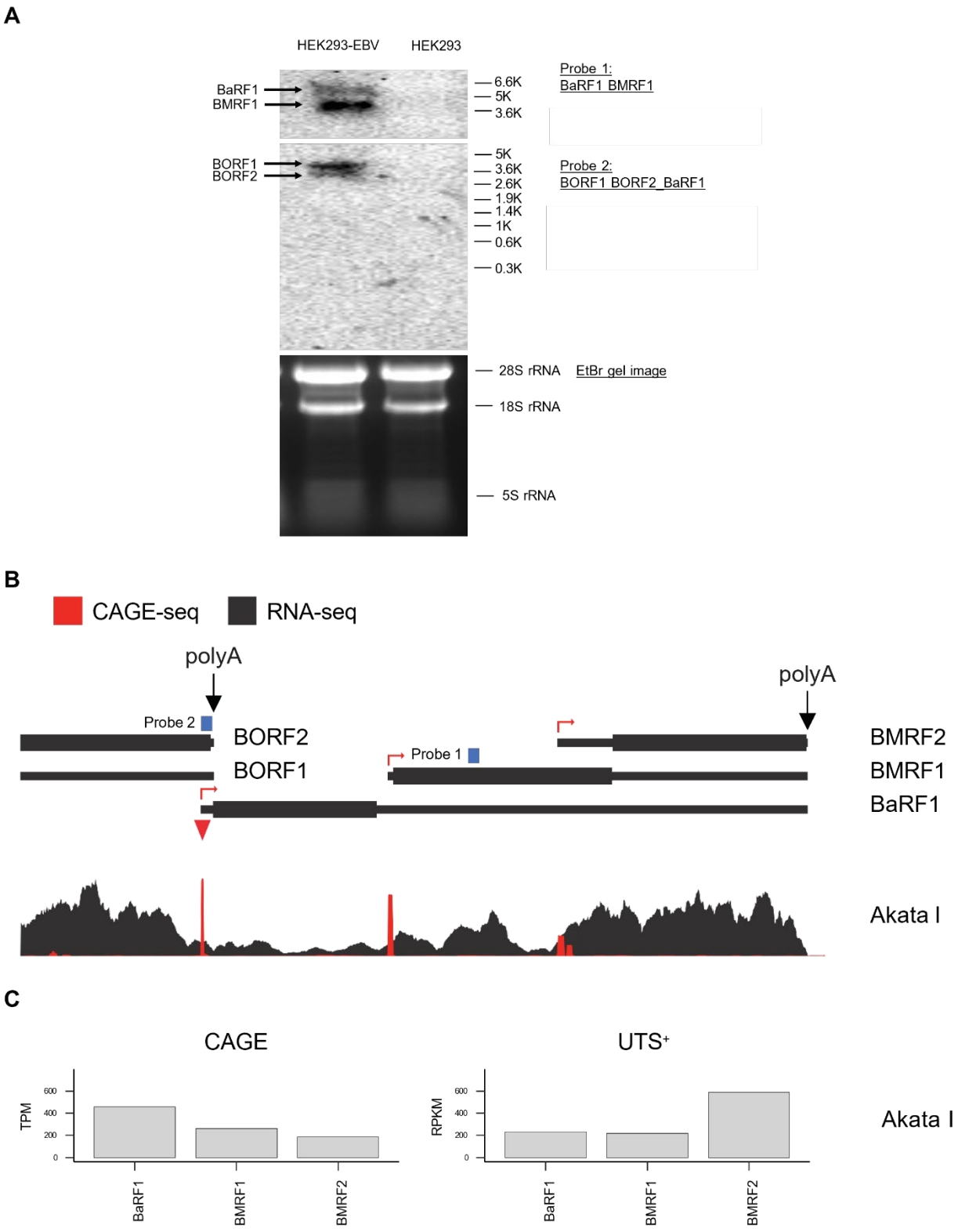
given the strength of the CAGE-seq signal at the BLRF2 TSS in our dataset. In contrast, previously published data from Akata cells does show an increase in RNA-seq read depth that more closely mirrors the CAGE-seq results (30). (B) Genomic region, RNA-seq, and CAGE-seq data corresponding to the leftward overlapping genes BOLF1 and BPLF1 that share a polyA site. As with the BLRF1-BLRF2 cluster, no increase in RNA-seq depth is observed as would be expected from the 3' transcript (BPLF1). In this instance, the extremely low read depth may be responsible for this apparent discrepancy. Alternatively, an unannotated BOLF1 polyA site could explain these data.

Premature polyadenylation may explain the low abundance of BaRF1 seen by RNA-seq.

An unusual feature of the BaRF1–BMRF1–BMFR2 gene cluster is that it partially overlaps with a second gene cluster consisting of BORF1 and BORF2 (Appendix Figure III-9A). As a result, the BORF1–BORF2 shared polyadenylation signal is just downstream of the BaRF1 transcription start site. This raises the possibility that a fraction of the BaRF1 primary transcripts is cleaved and polyadenylated prematurely and thus lacks the BaRF1 open reading frame. In this case, the CAGE-seq TSS signal would exceed the RNA-seq read depth, as observed in the ΔBALF2-I and ΔBALF2-I+T experiments and, to a lesser degree, in the WT-I experiment (Appendix Figure III-9A and B). We attempted to confirm the existence of this putative 75-bp BaRF1 truncation product by Northern blotting but were unsuccessful (Appendix Figure III-10A). Because premature polyadenylation frequently results in transcript instability [14], we cannot exclude the possibility that the steady-state level of this 75-bp product was below the threshold of detection for our Northern blot. Notably, we also observed this discrepancy between CAGE-seq and RNA-seq measurements for BaRF1 in our reanalysis of published Akata data [11] (Appendix Figure III-10B and C). Thus, BaRF1 transcript levels do appear to be consistently higher when measured by CAGE-seq than when measured by other methods, but we cannot exclude the possibility that factors other than premature polyadenylation are responsible.


A
CAGE-seq
RNA-seq
polyA
polyA
BORF2
BORF1
BMRF2
BMRF1
BaRF1
WT I
ΔBALF2 I
ΔBALF2 I+T
B
CAGE
UTS+
TPM
RPKM
800
600
400
200
0
BaRF1
BMRF1
BMRF2
WT I
ΔBALF2 I
ΔBALF2 I+T

Appendix Figure III-9 Example of CAGE-seq discrepancy contributing to observed variance between CAGE-seq and RNA-seq. Shown is the genomic region corresponding to the rightward overlapping genes BaRF1, BMRF1, and BMRF2, which share a poly(A) site. RNA-seq read depth is displayed in black, and CAGE-seq signals in red, from the three experiments (WT-I, dBALF2-I, and dBALF2-I+T) described in the legend to Appendix Figure III-4. mRNA transcripts encoding each protein product are shown, including coding sequences (wide solid bars) and UTRs (narrow solid bars). Note that for the BaRF1 gene, the CAGE-seq TSS signal (red arrow) is much higher than the resultant RNA-seq read depth downstream of this TSS. As discussed in the text, this discrepancy may be due to premature polyadenylation and cleavage of a significant proportion of these transcripts at the BORF1/BORF2 poly(A) signal. (B) Bar plots displaying EBV transcript abundance estimates derived from CAGE-seq (in TPM) or RNA-seq analyzed by UTS^{+} (expressed in RPKM). Note the discrepancy between CAGE-seq and UTS+ RNA-seq estimates of BaRF1 transcript abundance (especially under the ΔBALF2 I and ΔBALF2 I+T conditions).



Appendix Figure III-10 Examination of the discrepancy between RNA-seq read depth and CAGE-seq signal strength for the BaRF1 transcript. (A) Northern blot for EBV transcripts in the BaRF1-BMRF1-BMRF2 and BORF1-BORF2 gene clusters in EBV infected HEK-293 cells

induced for replication (left lane) versus uninfected control (right lane). Blots were probed with the indicated DIG-labeled probes (see materials and methods for details) and detected transcripts are as indicated to the left. (B) Genomic region corresponding to the rightward overlapping genes BaRF1, BMRF1, and BMRF2 that share a polyA site. Location of overlapping BORF1/BORF2 polyA site and Norther blot probes (blue boxes) are also indicated. Below this schematic, RNA-seq read depth is displayed in black and CAGE-seq signals in red from published Akata strain data (30). Note that the sequencing depth corresponding to the BaRF1 transcript is much less than would be expected given the strength of the CAGE-seq signal at the BaRF1 TSS as was seen in our own data (Appendix Figure III-9). (C) Bar plots displaying EBV transcript abundance estimates derived from CAGE-seq (in TPM) or RNA-seq analyzed by UTS^{+} (expressed in RPKM).

Discussion

The study of how viruses subvert cell processes in the service of their replication has provided seminal insights into cell biology, including the discovery of introns, oncogenes, and enhancer elements. The application of next-generation sequencing (NGS) technology to the study of viral infection has greatly increased the scope and depth of our understanding of how viral pathogens accomplish these feats. However, because viruses are under different constraints and often violate the norms of their hosts, it is essential to reexamine assumptions made by NGS bioinformatic pipelines developed for the study of cellular processes when one is applying them to the study of viruses. Here, using Epstein-Barr virus lytic replication as a model, we examine the performance of a conventional RNA-seq pipeline in quantifying EBV and host gene expression relative to that of CAGE-seq. The impetus for this analysis was that viruses, due to their limited coding capacity, exhibit extensively overlapping and bidirectional transcription that confounds conventional read assignments to genes. Although others have voiced this concern and proposed strategies to address it [6], we are not aware of prior systematic attempts to quantify either the magnitude of the resulting errors in measurement or the effectiveness of proposed remedies. In this study, by analyzing paired RNA-seq and CAGE-seq samples from cells infected with different EBV mutants, we attempt to address this gap in our knowledge. Our

results confirm that conventional read assignment (ORF-based) methods perform well for quantification of host and nonoverlapping viral transcripts but are less accurate for the quantification of overlapping transcripts. These errors are not merely quantitative: we identify instances where conventional pipelines incorrectly report the expression of late genes from a virus defective in late transcription. The extents to which these errors affect the validity of a study will differ, but our results clearly document these errors and validate approaches to mitigate them.

A principal goal of this study was to assess whether read assignment methods that consider transcript overlap improve the concordance between RNA-seq and CAGE-seq measurements. The first approach we examined was the UCDS method, introduced by Bruce et al. [6], which has been widely used for KSHV transcription and has also been applied to EBV. Our results demonstrate that UCDS significantly improves the accuracy of RNA-seq abundance measurements of overlapping EBV lytic transcripts. Our UTS method, an extension of UCDS that requires knowledge of every TSS, resulted in similar improvements in accuracy. Because the UTS method discards fewer RNA-seq reads, it may produce more-accurate results from data sets with less sequencing depth. This appears to be responsible for the improved quantitation of the BGLF3 cluster by UTS relative to UCDS (Appendix Figure III-7C). We achieved further improvements in UTS accuracy by incorporating an effective transcript length (UTS^+ method). Because this was incorporated into our method post hoc, it remains to be prospectively determined whether the UTS^+ method produces better correlation with CAGE-seq than the UTS method. However, it is unlikely that the improvements seen with UTS^+ are due to mere overfitting. The incorporation of the effective length did not arise from trial and error but rather was chosen specifically to address read-mapping biases at transcript ends (Appendix Figure

III-5) and our observation that the largest errors were present in the most 3′ transcript of a given cluster.

Differential expression algorithms are only as good as the genome annotations they use as inputs. The availability of high-quality TSS and polyadenylation data for type 1 EBV was essential for our analysis. It is unlikely that these annotations are perfect, and our identification of clusters with discrepant CAGE-seq/RNA-seq measurements highlights gene clusters whose annotations may require further refinements. These include BaRF1–BMRF1–BMRF2, BLRF1–BLRF2, and BOLF1–BPLF1. Conversely, given the increasingly well-documented variation in EBV genome sequences, it will be important to determine the extent to which prototype annotations reflect transcription from real-world EBV strains.

Although we have documented that the UTS^+ and UCDS methods result in substantial improvements in the accuracy of EBV lytic transcript quantification, there are several important limitations to be considered. First, because both methods primarily address the phenomenon of gene clusters, they are unlikely to result in substantial improvements in the quantification of latency genes, because latency transcripts rarely share polyadenylation sites. An additional caveat is that the transcript annotation used here was conservative and excluded many recently reported novel transcripts [11, 15]. Our rationale for doing this was that we do not know whether these transcripts are expressed by all EBV strains or are unique to the virus in which they were discovered. We also did not know whether the abundances of these transcripts were sufficient to justify their inclusion. Nevertheless, we acknowledge that the inclusion of these transcripts, at least the most abundant species, may lead to further improvements in read assignment accuracy. It should also be noted that we do not expect that the UTS^+ or UCDS method will perform well with RNA-seq data that lacks uniform read distribution, a key assumption of both

methodologies. In general, this assumption holds well for ribodepleted RNA-seq data but must be carefully verified for poly(A) signal-enriched libraries, including most single-cell RNA-seq protocols. In these instances, tools such as the isoform-aware aligner RSEM may be required. We found that RSEM performed very well on our paired-end data (Appendix Table III-1) but required substantial computing resources. Interestingly, for our UTS^{+} and UCDS analyses, we used the paired-end data only for library insert size estimation, relying on the 5′ single-end read data for quantification. Thus, it is likely that UTS^{+} and UCDS will perform well with single-end read data sets, provided that an accurate estimate of the average library size can be furnished. Additionally, these methods are much less computationally intensive than RSEM and can be performed on a laptop computer if reads have already been aligned to reference genomes.

Appendix Table III-1Summary statistics of Conventional, UCDS, UTS, UTS+ and RSEM RNA-seq quantification of RNA-seq data compared to CAGE-seq for overlapping genes. For each of three RNA-seq experiments (WT-I , ΔBALF2 I , and ΔBALF2 I+T) the Person and Spearman's correlation coefficients were calculated between RNA-seq data analyzed by one of the indicated methods (Conventional, UCDS, UTS, UTS+, or RSEM) versus the CAGE-seq gold standard.

		Overlapping Genes						*Non-Overlapping Genes*					
		Pearson's Test		Spearman's Test		Linear Regression		Pearson's Test		Spearman's Test		Linear Regression	
Condition	Method	corr	p-value	corr	p-value	R-squared[1]	Slope[2]	corr	p-value	corr	p-value	R-squared[1]	Slope[2]
WT I	Conventional	0.62	1.01e-06	0.69	1.81e-08	0.37	0.55 ± 0.20	0.80	1.01e-06	0.69	2.85e-03	0.61	0.96 ± 0.36
	UCDS	0.82	1.10e-13	0.81	3.81e-13	0.67	0.70 ± 0.14	0.79	1.10e-13	0.68	3.79e-03	0.60	0.93 ± 0.39
	UTS	0.85	1.27e-15	0.84	9.89e-15	0.72	0.79 ± 0.14	0.79	1.27e-15	0.68	3.79e-03	0.60	0.95 ± 0.39
	UTS+	0.85	7.42e-16	0.84	1.12e-14	0.73	0.77 ± 0.13	0.79	7.42e-16	0.74	1.17e-03	0.60	0.94 ± 0.39
	RSEM	0.85	3.34e-15	0.83	1.32e-13	0.72	0.68 ± 0.12	0.81	3.34e-15	0.74	7.24e-04	0.63	0.95 ± 0.36
dBALF2 I	Conventional	0.65	1.82e-07	0.66	1.27e-07	0.41	0.74 ± 0.24	0.92	3.09e-07	0.89	4.33e-06	0.84	1.16 ± 0.26
	UCDS	0.88	6.41e-18	0.87	2.50e-17	0.77	0.97 ± 0.15	0.92	4.33e-07	0.82	8.94e-05	0.84	1.13 ± 0.26
	UTS	0.83	4.60e-14	0.82	5.97e-14	0.68	0.97 ± 0.19	0.92	3.49e-07	0.81	1.64e-04	0.84	1.15 ± 0.26
	UTS+	0.88	1.26e-17	0.86	1.40e-16	0.77	1.04 ± 0.16	0.93	2.80e-07	0.86	1.54e-05	0.86	1.13 ± 0.25
	RSEM	0.86	8.62e-16	0.86	1.18e-15	0.74	0.96 ± 0.16	0.93	6.68e-08	0.89	2.05e-06	0.86	1.13 ± 0.23
dBALF2 I+T	Conventional	0.61	1.68e-06	0.71	5.41e-09	0.36	0.62 ± 0.23	0.82	1.71e-04	0.78	6.27e-04	0.65	1.36 ± 0.52
	UCDS	0.88	3.56e-17	0.87	4.67e-17	0.76	0.89 ± 0.14	0.80	3.44e-04	0.76	9.07e-04	0.61	1.31 ± 0.54
	UTS	0.89	4.01e-18	0.87	2.09e-16	0.78	0.98 ± 0.14	0.82	2.13e-04	0.77	7.57e-04	0.64	1.36 ± 0.54
	UTS+	0.89	1.86e-18	0.88	3.41e-17	0.79	0.95 ± 0.14	0.81	2.82e-04	0.76	9.07e-04	0.62	1.35 ± 0.55
	RSEM	0.87	4.95e-16	0.85	2.03e-14	0.75	0.81 ± 0.13	0.82	1.92e-04	0.80	2.18e-04	0.65	1.37 ± 0.51

[1] Adjusted R-squared
[2] Slope of the regression with 95% confidence interval

While our study focused on EBV, the UCDS and UTS^{+} methods are generally applicable to all herpesviruses. We have included open-source scripts (available at https://github.com/mhayesbio/uts) to make this method more accessible. Given suitable

transcript annotation and reads aligned to a given viral genome, the program will output read assignments adjusted for transcript overlap that can be used for downstream analyses such as differential expression. In theory, the UTS script could be adapted to the study of nonherpesviruses as well, although it will be important to prospectively evaluate its performance against appropriate benchmarks if the transcripts' structures are substantially different from those studied here.

Materials and Methods

CAGE-seq data acquisition and analysis.

No-amplification nontagging Illumina cap analysis of gene expression sequencing (CAGE-seq) data sets from HEK293 cells infected with wild-type EBV or EBVΔBALF2 induced into lytic replication (WT I or ΔBALF2 I, respectively) and EBVΔBALF2 induced and trans-complemented with BALF2 (ΔBALF2 I+T) have been described previously (5) and are available under BioProject accession number PRJNA471349. CAGE-seq data from Akata cells induced for lytic replication [11] was obtained from NCBI SRA (accession numbers SRR3233929 and SRR3233931). CAGE-seq data were aligned to the NCI Genomic Data Commons (GDC) reference genome. Read count summarization was performed with mmquant [16] using an EBV1 annotation file (based on GenBank accession number NC_007605). For host genes, most CAGE tags are detected within 500 bp from the annotated TSS [12]. Therefore, the gencode.v22.annotation.gtf file was edited to capture reads within 500 bp of the annotated TSS. For EBV genes, a window 200 bp upstream and 50 bp downstream of each gene's TSS was used. In rare instances where the 200-bp upstream region extended into another window, it was truncated to ensure a unique window for each gene. Transcript abundance was then normalized

to millions of reads per tags mapped (to host and virus genes combined) and expressed as tags per million (TPM).

RNA sequencing and alignment.

Duplicate cell pellets from HEK293 cells infected with wild-type EBV or EBVΔBALF2 previously analyzed by CAGE-seq were stored at –150°C for approximately 1 year. Frozen cells were directly lysed into TRIzol and total RNA isolated with the Direct-zol MiniPrep kit (Zymo). Stranded library prep and NovaSeq SP PE50 sequencing were performed by the Oklahoma Medical Research Foundation's Clinical Genomics Center. RNA-seq data from Akata cells induced for lytic replication [11] were downloaded from NCBI SRA, accession number SRR3233907. RNA-seq reads were aligned to the GDC GRCh38-based genome index using STAR, version 2.6.1d [17].

Northern blotting.

Approximately 5 million cells were lysed in TRIzol and total RNA extracted by using Direct-zol RNA MiniPrep kits (Zymo) according to the manufacturer's instructions. Twenty micrograms of total RNA was denatured in 50% formamide and 2.2 M formaldehyde at 65°C for 15 min. Denatured RNAs were loaded onto a 1% agarose gel containing 2.2 M formaldehyde and electrophoresed with a 1× morpholinepropanesulfonic acid (MOPS) buffer (20 mM MOPS, 2 mM sodium acetate trihydrate, 1 mM EDTA [pH 7.0]). The RNA was transferred to a BrightStar-Plus positively charged nylon membrane (Fisher Scientific) by capillary transfer with 20× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and was then UV cross-linked.

The BMRF1 oligonucleotide (GGTGGCTCTTGGACACCTTATCATCATATT) or the BORF2_BaRF1 oligonucleotide (AAGATTCACAGGCTCGTTCAGTAAAACCGTCAGTCTCCAACAGAT) was digoxigenin

(DIG) labeled using a DIG Oligonucleotide Tailing kit (MilliporeSigma), and labeling efficiency was determined according to the manufacturer's instructions. Membranes were prehybridized in DIG Easy Hyb buffer (MilliporeSigma) containing poly(A) and poly(dA) (MilliporeSigma) at 46°C for 2 h. The prehybridization solution was replaced with fresh DIG Easy Hyb buffer containing the DIG-labeled oligonucleotide, and hybridization was continued at 46°C overnight. The membrane was washed with 2× SSC wash buffer (2× SSC, 0.1% SDS) at room temperature and with 0.5× SSC wash buffer (0.5× SSC, 0.1% SDS) at 46°C.

The hybridized membrane was detected using a DIG Luminescent Detection kit (MilliporeSigma) according to the manufacturer's instructions. Briefly, the membrane was blocked for 2 h at room temperature and was incubated with an alkaline phosphatase-conjugated anti-DIG antibody (1:10,000 dilution) at 4°C overnight. The membrane was washed with washing buffer (0.1 M maleic acid, 0.15 M NaCl, 0.3% Tween 20), equilibrated with detection buffer (0.1 M Tris-HCl, 0.1 M NaCl [pH 9.5]), and treated with CSPD substrate. The signal was detected with a ChemiDoc imaging system (Bio-Rad).

Conventional RNA-seq analysis.

Read count summarization of host and viral gene annotated coding sequence (CDS) features was performed using mmquant with the single-end strandedness -s and the default -l 1 option to emulate the "intersection-strict" mode of htseq-count [16]. The gencode.v22.annotation.gtf was used for host gene annotation, and a curated EBV annotation based on GenBank accession number NC_007605 (File S1) was used for viral genes. Any reads mapping to overlapping CDS features were excluded as ambiguously mapped reads. Gene expression levels were normalized to CDS feature length and millions of reads mapped and were expressed as reads per kilobase per million mapped reads (RPKM).

Implementation of the UCDS method.

Using the same read alignment as for conventional RNA-seq, EBV read count summarization was performed using mmquant with the single-end strandedness -s and the -l 1 option [16]. The -l 1 option emulates the "union" mode option of htseq-count, which is equivalent to the "intersection-nonempty" mode of htseq-count, since the UCDS features (Appendix Table III-2) were generated to avoid overlap using the approach described by Bruce et al. [6]. Briefly, each UCDS feature begins at the ATG start codon and extends until any of the following is encountered: (i) the stop codon for the respective gene, (ii) 1 bp prior to the TSS of the next gene on the same strand, (iii) 1 bp plus one read length prior to the ATG of the next gene on the same strand (52 bp for this analysis [read length of 51 bp + 1]). In rare cases where two gene clusters were in close proximity to each other, a UCDS feature was truncated at the 5′ end to ensure that all UCDS features were separated by at least one read length. For each UCDS region, read depth was calculated as follows: depthN = raw counts(UCDSN)/length(UCDSN). For nested transcripts, the read depth was adjusted to account for the overlapping portion of the larger transcript B using the following formula: depthA = depthA − depthB (where "=" is intended as an assignment operator). Gene expression levels were then normalized to UCDS feature length and millions of reads mapped and expressed as RPKM.

Appendix Table III-2 Summary of UCDS and UTS regions used in this study. For EBV each gene (column 1), the corresponding UCDS (column 2), UTS (column 3), and UTS^{+} (column 4) sizes and the strandedness (column 5) are indicated. Genes that share a polyA site are defined as a gene cluster and number sequentially (column 7). Genes that do not share a polyA site are designated "non-overlapping" (also column 7). For genes within clusters, the UCDS or UTS zone whose depth is subtracted to calculate the true sequencing depth is indicated (column 6). The sole exception to this is cluster #1 as most of the overlapping BNRF1 sequence

falls within an LMP2 intron. In this case, the LMP2 UTS zone was merely truncated (i.e., similar to the approach used for UCDS).

Gene	UCDS Size (bp)	UTS Size (bp)	UTS' Size (bp)	Strand	Primary Transcript (Subtract)[1]	Cluster
LMP2	1623	1924	1823	+		1
BNRF1	3672	3699	3598	+		1
BFRF0.5	166	279	279	+		2
BFRF1	611	642	642	+	BFRF0.5	2
BFRF2	1564	1870	1870	+	BFRF1	2
BFRF3	531	716	615	+	BFRF2	2
BORF1	960	1151	1050	+		3
BORF2	2424	2633	2607	+	BORF1	3
BaRF1	903	963	937	+		4
BMRF1	912	942	942	+	BaRF1	4
BMRF2	1074	1388	1287	+	BMRF1	4
BLRF1	309	354	253	+		5
BLRF2	489	538	437	+	BLRF1	5
BKRF2	217	217	217	+		6
BKRF3	571	749	749	+	BKRF2	6
BKRF4	654	883	782	+	BKRF3	6
BBRF1	1590	1888	1787	+		7
BBRF2	837	1015	914	+	BBRF1	7
BcRF1	2076	2294	2193	+		8
BTRF1	1215	1450	1349	+	BcRF1	8
BXRF1	474	723	622	+		9
BVRF1	1713	1856	1755	+	BXRF1	9
BVRF2	724	898	797	+		10
BdRF1	1038	1098	997	+	BVRF2	10
LMP1	1161	1908	1807	-		11
BNLF2a/b	480	576	475	-	LMP1	11
BALF3	1618	1651	1651	-		12
BALF4	2446	2464	2464	-	BALF3	12
BALF5	2983	3582	3582	-		13
BILF1	947	947	947	-	BALF5	13
BXLF1	1578	1837	1736	-		15
BXLF2	2121	2403	2302	-	BXLF1	15
BDLF1	841	841	774	-		16
BDLF2	1263	1370	1370	-	BDLF1	16
BDLF3	705	741	640	-	BDLF2	16
BDLF3.5	161	329	228	-		17
BDLF4	596	617	617	-	BDLF3.5	17
BGLF1	1450	1502	1502	-	BDLF4	17
BGLF2	1011	1439	1338	-	BGLF1	17
BGLF3	852	1053	952	-		18
BGLF3.5	83	216	216	-	BGLF3	18
BGLF4	1185	1442	1442	-	BGLF3.5	18
BGLF5	1038	1126	1126	-	BGLF4	18
BBLF1	228	556	455	-	BGLF5	18
BRLF1	1818	2052	1951	-		19
BZLF1	738	857	759	-	BRLF1	19
BLLF1	1501	1545	1444	-		20
BLLF2	447	623	522	-	BLLF1	20
BSLF1	2542	2806	2705	-		21
SM	1440	1495	1394	-	BSLF1	21
BOLF1	3076	3131	3030	-		22
BPLF1	9450	10116	10015	-	BOLF1	22
BFLF1	1395	1409	1308	-		23
BFLF2	957	1169	1068	-	BFLF1	23
EBER1	167	167	167	+		Non-overlapping
EBER2	173	173	173	+		Non-overlapping
BCRF1	513	622	421	+		Non-overlapping
BHRF1	576	1340	1139	+		Non-overlapping
BSRF1	657	703	502	+		Non-overlapping
BRRF1	933	1091	890	+		Non-overlapping
BRRF2	1597	1665	1464	+		Non-overlapping
BBRF3	1218	1249	1048	+		Non-overlapping

Gene	UCDS Size (bp)	UTS Size (bp)	UTS' Size (bp)	Strand	Primary Transcript (Subtract)[1]	Cluster
BGRF1/BDRF1	2073	2173	1972	+		Non-overlapping
BARF1	666	690	489	+		Non-overlapping
BHLF1	1983	2525	2324	-		Non-overlapping
BLLF3	837	900	699	-		Non-overlapping
BZLF2	672	622	424	-		Non-overlapping
BBLF4	2430	2635	2434	-		Non-overlapping
BBLF2/3	2130	2242	2041	-		Non-overlapping
BcLF1	4126	4360	4193	-		Non-overlapping
BVLF1	819	1258	1057	-		Non-overlapping
BILF2	747	809	608	-		Non-overlapping
BALF2	3133	3149	3049	-		Non-overlapping
BALF1	549	589	388	-		Non-overlapping

[1] For quantification of a transcript sharing a polyadenylation site, the normalized transcript count from the indicated upstream UTS/UCDS region is subtracted

Implementation of the UTS method.

For the UTS approach, the same read alignment as for conventional RNA-seq was used. EBV read count summarization was performed with mmquant with the single-end strandedness -s and the default -l -1 option to emulate the "intersection-strict" mode of htseq-count and to prevent the quantification of ambiguous reads [16] using an annotation of UTS features derived from the same curated EBV annotation as for the conventional approach (Appendix Table III-2). See Results for a graphical illustration of UTS regions (Appendix Figure III-3). Briefly, each zone is defined by the TSS site of each gene and extends until another TSS site on the same strand or a poly(A) cleavage site is encountered. A small number of UTS features were truncated at their 5′ or 3′ ends, when necessary, to ensure that each UTS feature was separated by at least 1 bp. For each UTS region, read depth was calculated as follows: depthN = raw counts(UTSN)/length(UTSN). As with the UCDS method, read depth was then adjusted for nested transcripts to account for the overlapping portion of the larger transcript B using the following formula: depthA = depthA − depthB. Gene expression levels were then normalized to UTS feature length and millions of reads mapped and expressed as RPKM.

Implementation of effective transcript lengths for the UTS$^+$ method.

To account for 5′ and 3′ mapping bias, an effective length was used in RPKM calculations [18, 19]. This refinement uses the mean RNA-seq library fragment length (LImean), which can be input by the user (in our analysis, this was estimated from paired-end reads using deepTools bamPEFragmentSize [20] with default parameters). In cases where this is not known, this adjustment can be omitted (i.e., the UTS method can be used), or an LImean of 200 bp can be used, since this is a common optimal length for many RNA-seq protocols. For nonoverlapping transcripts, the effective length was defined as the transcript length minus LImean. For gene clusters, the lengths of the most 5′ UTS zone and the most 3′ UTS zone were decreased by ½ LImean. For UTS zones in the middle of a cluster, no effective length correction was implemented, since the 3′ end is not present in the UTS zone and reads derived from the upstream region are ambiguously mapped to the 5′ end bias region. Finally, for overlapping transcripts that are not part of a cluster, the UTS zone was decreased by ½ LImean if it included the 5′ or 3′ end of the transcript. No effective length was introduced for transcripts or UTS zones of <300 bp, since these small regions are difficult to quantify accurately with or without an effective length [21]. These effective length values (summarized in Appendix Table III-2) are then used instead of the UTS lengths to calculate the RPKM.

Availability and implementation of UTS/UTS$^+$ in Python.

To facilitate the implementation of the UTS method, a Python script was created that automatically calculates UTS regions from a GTF annotation file (e.g., File S1). The process was further automated with a Bash script that takes the GTF, an alignment file for both viral and host genes (BAM format), and, optionally, a library insert length (LIavg). The Bash script additionally requires installation of mmquant and SAMtools. Estimates of viral gene expression

levels in RPKM and expected counts are output as a tab-delimited file for evaluation and downstream analyses. These scripts and GTF files for EBV types 1 and 2 are available through the GitHub repository (https://github.com/mhayesbio/uts).

Visualization tracks.

Coverage from BAM files was calculated, and bedGraph files were generated, using deepTools bamCoverage [20] with the following parameters: -bs 1 -of bedgraph –region chrEBV –normalizeUsing RPGC –effectiveGenomeSize (122,535 for RNA-seq and 28,108 for CAGE-seq) –filterRNAstrand (forward or reverse). The Akata I CAGE-seq replicates were averaged using the deepTools bamCoverage parameter -ps averages. The effective EBV genome size is described as the total length of the EBV exome for the RNA-seq data or the total length of the EBV TSS windows for CAGE-seq data. NGS coverage graphs were then plotted using SparK (version 2.6.2) [22].

Statistical analysis.

Linear regression and correlation analyses were calculated using the stats (version 4.0.3) R package [23]. Linear regression and bar plots were generated using R packages ggplot2 (version 3.3.3) [24] and ggpubr (version 0.4.0) [25]. The heat map was generated using the gplots (version 3.1.1) R package [26]. The summary statistics (Appendix Table III-1) was generated using R packages kableExtra (version 1.3.2) [27] and tinytex (version 0.29) [28].

Data availability.

Sequence data have been deposited to NCBI SRA (BioProject accession number PRJNA771685).

References

1. Zhang R, Strong MJ, Baddoo M, Lin Z, Wang YP, Flemington EK, Liu YZ. 2017. Interaction of Epstein-Barr virus genes with human gastric carcinoma transcriptome. Oncotarget 8:38399–38412.
2. Chakravorty S, Yan B, Wang C, Wang L, Quaid JT, Lin CF, Briggs SD, Majumder J, Canaria DA, Chauss D, Chopra G, Olson MR, Zhao B, Afzali B, Kazemian M. 2019. Integrated pan-cancer map of EBV-associated neoplasms reveals functional host-virus interactions. Cancer Res 79:6010–6023.
3. Borozan I, Zapatka M, Frappier L, Ferretti V. 2018. Analysis of Epstein-Barr virus genomes and expression profiles in gastric adenocarcinoma. J Virol 92:e00330-18.
4. Hu L, Lin Z, Wu Y, Dong J, Zhao B, Cheng Y, Huang P, Xu L, Xia T, Xiong D, Wang H, Li M, Guo L, Kieff E, Zeng Y, Zhong Q, Zeng M. 2016. Comprehensive profiling of EBV gene expression in nasopharyngeal carcinoma through paired-end transcriptome sequencing. Front Med 10:61–75.
5. Djavadian R, Hayes M, Johannsen E. 2018. CAGE-seq analysis of Epstein-Barr virus lytic gene transcription: 3 kinetic classes from 2 mechanisms. PLoS Pathog 14:e1007114.
6. Bruce AG, Barcy S, DiMaio T, Gan E, Garrigues HJ, Lagunoff M, Rose TM. 2017. Quantitative analysis of the KSHV transcriptome following primary infection of blood and lymphatic endothelial cells. Pathogens 6:11.
7. Rose TM, Bruce AG, Barcy S, Fitzgibbon M, Matsumoto LR, Ikoma M, Casper C, Orem J, Phipps W. 2018. Quantitative RNAseq analysis of Ugandan KS tumors reveals KSHV gene expression dominated by transcription from the LTd downstream latency promoter. PLoS Pathog 14:e1007441.
8. Faure A, Hayes M, Sugden B. 2019. How Kaposi's sarcoma-associated herpesvirus stably transforms peripheral B cells towards lymphomagenesis. Proc Natl Acad Sci USA 116:16519–16528.
9. Watanabe T, Nishimura M, Izumi T, Kuriyama K, Iwaisako Y, Hosokawa K, Takaori-Kondo A, Fujimuro M. 2020. Kaposi's sarcoma-associated herpesvirus ORF66 is essential for late gene expression and virus production via interaction with ORF34. J Virol 94:e01300-19.
10. Majerciak V, Yang W, Zheng J, Zhu J, Zheng ZM. 2019. A genome-wide Epstein-Barr virus polyadenylation map and its antisense RNA to EBNA. J Virol 93:e01593-18.
11. O'Grady T, Wang X, Honer Z, Bentrup K, Baddoo M, Concha M, Flemington EK. 2016. Global transcript structure resolution of high gene density genomes through multi-platform data integration. Nucleic Acids Res 44:e145.
12. Nepal C, Hadzhiev Y, Previti C, Haberle V, Li N, Takahashi H, Suzuki AM, Sheng Y, Abdelhamid RF, Anand S, Gehrig J, Akalin A, Kockx CE, van der Sloot AA, van Ijcken WF, Armant O, Rastegar S, Watson C, Strahle U, Stupka E, Carninci P, Lenhard B, Muller F. 2013. Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. Genome Res 23:1938–1950.
13. Kawaji H, Lizio M, Itoh M, Kanamori-Katayama M, Kaiho A, Nishiyori-Sueki H, Shin JW, Kojima-Ishiyama M, Kawano M, Murata M, Ninomiya-Fukuda N, Ishikawa-Kato S, Nagao-Sato S, Noma S, Hayashizaki Y, Forrest AR, Carninci P, FANTOM Consortium.

2014. Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. Genome Res 24:708–717.
14. Kamieniarz-Gdula K, Proudfoot NJ. 2019. Transcriptional control by premature termination: a forgotten mechanism. Trends Genet 35:553–564.
15. Moldován N, Szenthe K, Bánáti F, Fülöp Á, Csabai Z, Tombácz D, Minárovits J, Boldogkői Z. 2020. Transcriptome profiling of Epstein-Barr virus using nanopore sequencing. Research Square https://doi.org/10.21203/rs.3.rs-109629/v1.
16. Zytnicki M. 2017. mmquant: how to count multi-mapping reads? BMC Bioinformatics 18:411.
17. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.
18. Roberts A, Pachter L. 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods 10:71–73.
19. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14:417–419.
20. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44:W160–W165.
21. Zhang C, Zhang B, Lin LL, Zhao S. 2017. Evaluation and comparison of computational tools for RNA-seq isoform quantification. BMC Genomics 18:583.
22. Kurtenbach S, Harbour JW. 2019. SparK: a publication-quality NGS visualization tool. bioRxiv https://doi.org/10.1101/845529.
23. R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
24. Wickham H. 2016. ggplot2: elegant graphics for data analysis, v3.3.3. Springer-Verlag, New York, NY. https://ggplot2.tidyverse.org.
25. Kassambara A. 2020. ggpubr: 'ggplot2' based publication ready plots, v0.4.0. https://CRAN.R-project.org/package=ggpubr.
26. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Mueller S, Schwartz M, Venables B. 2020. gplots: various R programming tools for plotting data, v3.1.1. https://CRAN.R-project.org/package=gplots.
27. Zhu H. 2021. kableExtra: construct complex table with 'kable' and pipe syntax, v1.3.2. https://CRAN.R-project.org/package=kableExtra.
28. Yihui X. 2021. tinytex: helper functions to install and maintain TeX Live, and compile LaTeX documents, v0.29. https://github.com/yihui/tinytex.
29. Garcia-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Gotz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. 2012. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics 28:2678–2679.
30. O'Grady T, Wang X, Honer Zu Bentrup K, Baddoo M, Concha M, Flemington EK. 2016. Global transcript structure resolution of high gene density genomes through multi-platform data integration. Nucleic Acids Res 44:e145.

Appendix IV

EBV Zta Recruits P-TEFb to Promote Rta-synergy Gene Expression

Introduction

Rta and Zta are the two key EBV transcription factors that control the switch between latency and lytic replication. Previous work in our lab employed an EBV mutant lacking the Rta and Zta genes to dissect their individual contributions in regulating viral gene transcription [1]. The study demonstrated that Rta acts as the principal activator of EBV early genes. Notably, for a specific subset of early genes, Rta only activated low-level expression and Zta co-expression was required for efficient expression. These genes were therefore designated $Rta^{synergy}$ genes. Interestingly, Zta did not increase the stability of $Rta^{synergy}$ transcripts, nor did it influence RNA polymerase II occupancy or the presence of activating histone marks at $Rta^{synergy}$ promoters. These findings suggest that Zta's role at the promoters of $Rta^{synergy}$ genes lies in facilitating the resolution of a transcriptional pause.

Pausing of RNA polymerase II (RNAPII) is a regulatory step mediated by negative elongation factors, resulting in elongation arrest approximately 20-100 nucleotides downstream of the transcription start site [2-4]. Release from this paused state requires the recruitment of positive transcription elongation factor b (P-TEFb), a complex containing cyclin-dependent kinase 9 (CDK9) and either cyclin T1 or T2 [5]. P-TEFb counteracts the negative elongation factors through its kinase activity and phosphorylates the C-terminal domain of RNAPII, promoting its release from the pause site and facilitating continued transcription elongation [6].

P-TEFb activity is highly regulated, with a significant portion (50-90%) sequestered in an inactive form within the 7SK small nuclear ribonucleoprotein (7SK snRNP) complex [7-9]. The stability of this complex relies on the 7SK snRNA scaffold and its interactions with MePCE and LARP7 [10]. HEXIM1 and HEXIM2 (primarily HEXIM1) bind to the 7SK snRNA, and upon P-TEFb binding to the 7SK snRNP complex, directly inhibit CDK9's kinase activity [11]. Various

stimuli can disrupt the 7SK snRNP complex and release P-TEFb. These stimuli include stress, environmental signals, cytokines, chromatin stress, and viral infections [7,12-17]. Once released, P-TEFb can be recruited to target genes by transcription factors like the super elongation complex (SEC) or BRD4.

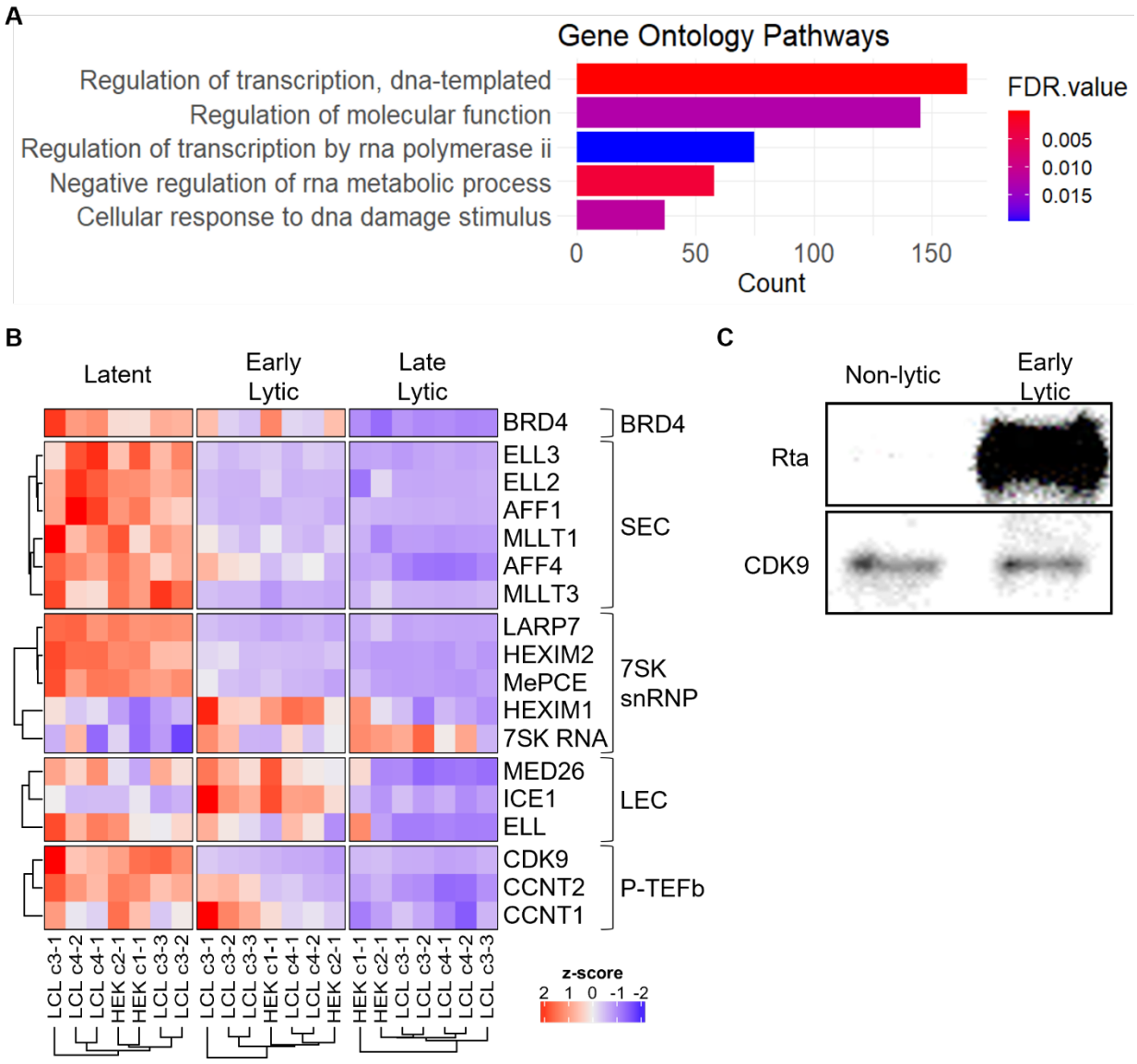
The observed RNAPII promoter occupancy at $Rta^{synergy}$ genes in the absence of Zta parallels the paused state prevalent across inactive and inducible genes [18,19]. This suggests transcriptional elongation regulation by P-TEFb as a plausible mechanism for Zta-mediated co-activation of $Rta^{synergy}$ genes.

Results

During my initial investigations, I developed a dual-fluorescent lytic reporter (DFLR) within the B95-8 EBV strain. This construct, analogous in function to the DFLR M81 EBV strain described in Chapter 4, utilizes eGFP and mCherry to replace the early BXLF1 and late BILF2 ORFs. To assess the functionality of this reporter system, a pilot experiment was conducted utilizing DFLR B95-8 HEK293.ZHT cells. These cells constitutively express the EBV immediate early protein Zta fused to a 4-hydroxytamoxifen (4-HT)-dependent mutant estrogen receptor binding domain (ZHT) [20]. Treatment with 4-HT induces translocation of ZHT to the nucleus, initiating lytic replication. DFLR B95-8 HEK293.ZHT cells were treated with 4-HT to induce lytic replication for 48 hours. Subsequently, fluorescence-activated cell sorting (FACS) and RNA sequencing were employed to isolate and characterize non-lytic ($eGFP^-/mCherry^-$), early lytic (GFP^+/RFP^-), and late lytic (GFP^+/RFP^+) cellular populations (as per the protocol outlined in Figure 4.3)

Pathway enrichment analysis was conducted on the subset of host genes escaping host shutoff during early lytic replication. This analysis revealed pathways primarily involved in

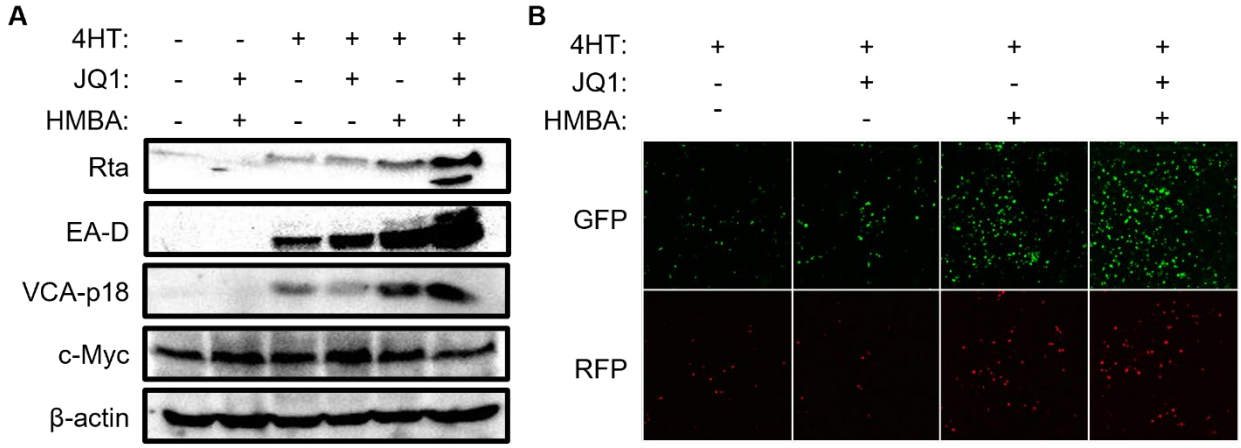
transcriptional regulation (Figure 4.6 and Appendix Figure IV-1A). Of particular interest, several transcripts within this host shutoff escapee subset encode regulators of P-TEFb (Appendix Figure IV-1B). Key components of the 7SK snRNP complex (HEXIM1 and 7SK snRNA), BRD4, and the little elongation complex (LEC) were either resistant to host shutoff or exhibited upregulation. While the transcript encoding the P-TEFb component cyclin T1 (CCNT1) remained stable during the early phase, the CDK9 transcript exhibited downregulation (Appendix Figure IV-1B). However, this downregulation of CDK9 at the transcript level was not mirrored at the protein level (Appendix Figure IV-1C). This discrepancy between transcript and protein expression levels during lytic replication is explored further in Chapter 4 (Discussion section entitled ‘A proteomic and transcriptomic discrepancy’). P-TEFb release from the 7SK snRNP complex initiates a negative feedback loop, increasing HEXIM1 expression. This, in turn, facilitates reassembly of the inhibitory 7SK snRNP complex restoring P-TEFb equilibrium [21]. Therefore, the observed increase in HEXIM1 levels during early lytic replication could potentially suggest increased P-TEFb activity.



Appendix Figure IV-1 P-TEFb regulation during early lytic replication. (A) Analysis of gene pathways enriched in biological processes using RNA-seq data from 4-HT lytically induced DFLR B95-8 HEK293.ZHT cells, with a focus on genes that escape host shutoff during the early lytic phase. (B) Heatmap showing Z-score transformed (by cell type) expression levels from RNA-seq data pf DFLR M81 LCLs and DFLR B95-8 HEK293.ZHT cells. (C) Western blot analysis of viral lytic protein, Rta, and cellular protein, CDK9, in sorted fractions isolated from DFLR B95-8 HEK293.ZHT cells lytically induced by 4-HT treatment.

To investigate whether EBV redirects P-TEFb activity towards viral transcription, I targeted its interactions with the 7SK snRNP complex and BRD4 using inhibitors (Appendix Figure IV-2A and B). BRD4 inhibition with JQ1 modestly increased the early lytic cell population, upregulating the immediate-early Rta and early EA-D proteins. In contrast, late-

phase progression was reduced, as evidenced by decreased mCherry-positive cells and lower p18 protein levels. This aligns with previous findings implicating BRD4 in lytic DNA replication, such that its inhibition disrupts DNA replication-dependent late gene expression [22]. Inhibition of the 7SK snRNP complex with hexamethylene bisacetamide (HMBA) had a more pronounced effect, increasing lytic reactivation and protein levels.

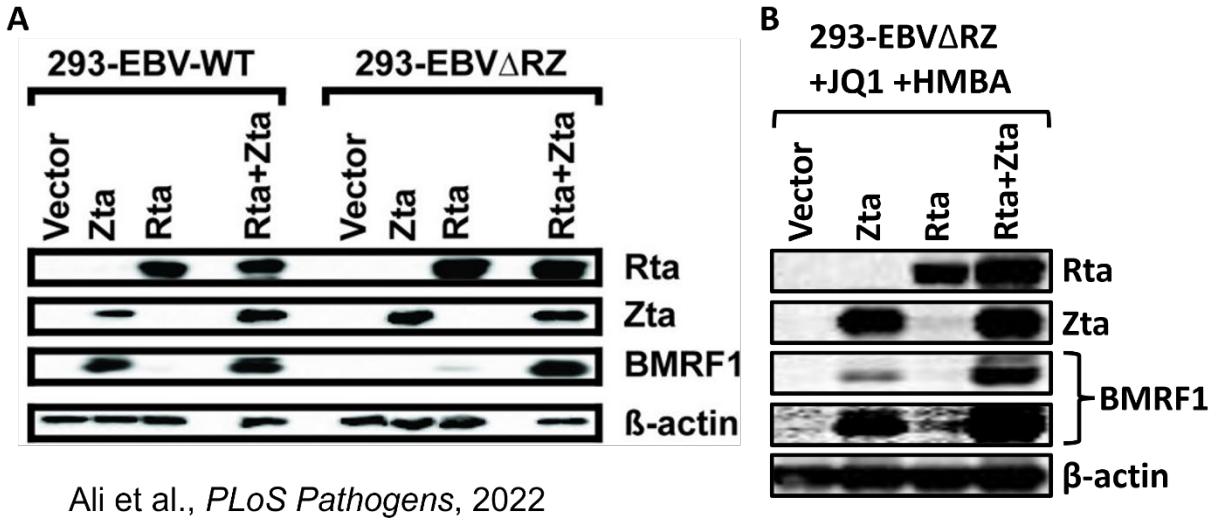


Appendix Figure IV-2 Impact of BRD4 and 7SK snRNP inhibition of EBV lytic reactivation. (A) Western blot or (B) live cell imaging of DFLR B95-8 HEK293.ZHT cells lytically induced with 4-HT and treated with DMSO (control [-]), JQ1 (BRD4 inhibitor), or HMBA (7SK snRNP inhibitor). Live cell imaging was performed with cells at equal confluency across conditions (~80%).

Intriguingly, the combined inhibition of both BRD4 and the 7SK snRNP complex with JQ1 and HMB1, respectively, led to a synergistic increase in lytic reactivation and protein levels. Importantly, this occurred without reducing c-Myc levels, a potent anti-lytic factor known to be downregulated by JQ1 treatment in specific cell types [23,24]. These data suggest active P-TEFb recruitment to EBV promoters, facilitating lytic reactivation.

Recent work in our lab demonstrated that Zta synergistically enhances expression of a subset of viral genes when co-acting with Rta, despite lacking a direct transcriptional effect itself [1]. This makes Zta a potential candidate for recruiting P-TEFb to viral promoters to facilitate

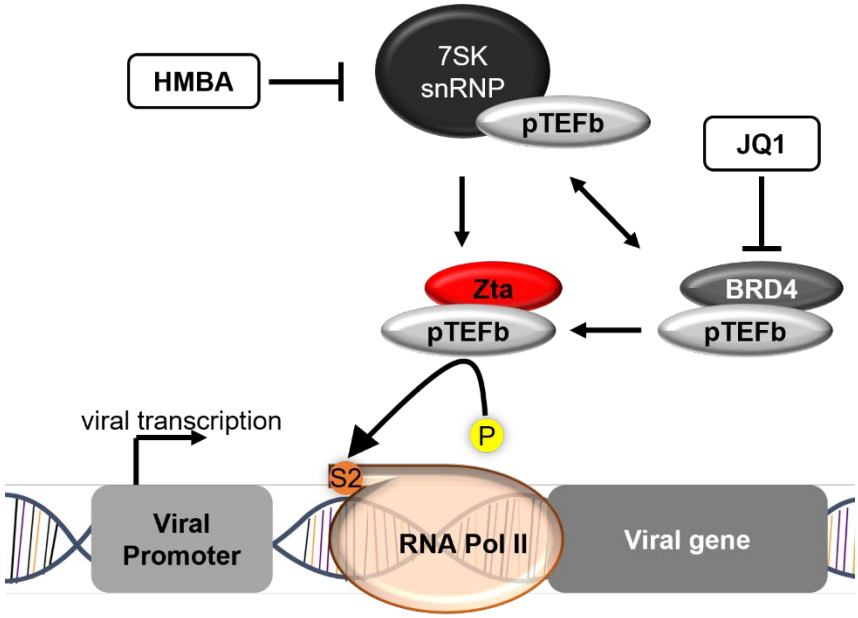
RNAPII release from the promoter-proximal paused state. Collaborating with Ahmed Ali, we tested this using HEK293 cells infected with a mutant EBV strain lacking Rta and Zta and treated with JQ1 and HMBA. Subsequently, these cells were transfected with control, Zta, Rta, or Zta + Rta expression vectors. Western blot analysis was performed to quantify the $Rta^{synergy}$ protein BMRF1 (Appendix Figure IV-3). Notably, JQ1 and HMBA treatment resulted in detectable BMRF1 levels when Zta was transfected alone. This observation is intriguing as our labs previous study has shown that Zta lacks a direct transcription effect by itself [1]. Furthermore, this treatment did not appear to affect BMRF1 levels with Rta alone.



Appendix Figure IV-3 Zta promotes activation of $Rta^{synergy}$ genes through P-TEFb. (A) Figure from Ali et al. [1] showing western blot for WT Akata strain EBV (293-EBV-WT) and mutant Akata strain EBV lacking Rta and Zta genes (293-EBVΔRZ) transfected with empty vector, or expression plasmid for Zta, Rta, or both (Rta+Zta). (B) 293-EBVΔRZ treated with JQ1 and HMBA and transfected with the same plasmids as in (A). The top and bottom panels of BMRF1 show short and long exposures, respectively.

This data suggest a model where disruption of the majority of cellular P-TEFb interactions via 7SK snRNP inhibition redirects a substantial pool of P-TEFb towards the viral transcription factor Zta, which subsequently promotes RNAPII promoter clearance at viral genes

(Appendix Figure IV-4). This redirection is further amplified by BRD4 inhibition, likely by preventing P-TEFb from associating with newly available BRD4 and by potentially releasing it from pre-existing BRD4 interactions.



Appendix Figure IV-4 Proposed model of enhanced EBV lytic reactivation through redirection of P-TEFb towards Zta-mediated viral transcription. The model proposes that enhanced EBV lytic reactivation can be achieved by redirecting P-TEFb toward Zta-mediated viral transcription. HMBA treatment inhibits P-TEFb interactions with 7SK snRNP, thereby redirecting P-TEFb toward cellular BRD4 and Zta. Conversely, JQ1 treatment inhibits P-TEFb interactions with BRD4, favoring binding with the 7SK snRNP inhibitor and, to a lesser degree, Zta. Simultaneous HMBA and JQ1 treatment is hypothesized to disrupt P-TEFb interactions with both 7SK snRNP and BRD4. This would liberate significant amounts of P-TEFb, facilitating robust Zta binding and recruitment to viral promoters to drive extensive viral gene transcription.

Methods

Cell lines

HEK293 cells and derivatives were cultured at 37°C in a humidified atmosphere with 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM Gebco) supplemented with 10% fetal bovine serum (FBS, HyClone) and 1% penicillin-streptomycin (Gibco). BACmid and lentiviral-

infected HEK293 cells were maintained with 200 μg/mL Hygromycin B and 1 μg/mL puromycin, respectively.

Construction of recombinant EBV BACmids

The dual-fluorescent lytic reporter (DFLR) EBV used in this study was derived from the EBV B95-8 BACmid (also known as p2089) [25]. *En Passant* mutagenesis in the GS1783 *E. coli* strain [26] using the gene blocks (gBlocks) found in Appendix Table II-2. To be compatible with a fluorescent reporter system, the CMV-driven eGFP present in the F-factor was deleted using the GFP-KO gBlock. The early expressed and non-essential BXLF1 ORF was then substituted with enhanced GFP (eGFP) [27-29] using the BXLF1-to-eGFP gBlock. The late-expressed orphan BILF2 ORF [27,30] was replaced with mCherry using the BILF2-to-mCherry gBlock.

The EBVΔRZ BACmid in HEK293 cells has been previously described [1,31].

Generation of B95-8 DFLR HEK293.ZHT cells

The diaminopimelate (DAP) auxotroph and invasive *E. coli* strain BM2710 was used to deliver the B95-8 DFLR BACMid into HEK293.MSCV-ZHT cells as described in the Chapter 3 section entitled 'Generation of WT and ΔBGLF5 DFLR EBV LCLs'. Cells were selected and maintained with 200 μg/mL Hygromycin B.

Cell chemical treatments and transfections

For lytic replication induction, B95-8 DFLR HEK293.MSCV-ZHT cells were treated with 200 nM 4-hydroxytamoxifen (4-HT) and EBVΔRZ HEK293 cells were transfected with 0.5 μg pSG5, pSG5-Zta, pSG5-Rta or 0.25 μg pSG5-Zta plus 0.25 μg pSG5-Rta vectors for 48 hours. Cells were treated with 10 mM hexamethylene bis-acetamide (HMBA, Sigma-Aldirch, #224235) and/or 1 μM (+/-)-JQ1 (Santa Cruz, sc-472932) at the same time as 4-HT treatment or transfection.

FACS, sample collection, and RNA-seq

48-hour 4-HT treated B95-8 DFLR HEK293.ZHT cells were washed with PBS, trypsinized, and resuspended in ice-cold FACS buffer containing (1x DPBS -Ca^{2+}/-Mg^{2+}; 2% FBS; 5 mM EDTA) containing 0.5 μg/mL of the vital dye DAPI to a final concentration of $5x10^6$ cells/mL. For compensation, unstained HEK293 cells served as the negative control and a DAPI-treated dead cell population (heat treated at 72°C for 5 minutes) controlled DAPI fluorescence. Additionally, HEK293 cells transfected with pEGFP1 or pCEP-mCherry provided eGFP and mCherry calibrations, respectively.

For RNA-seq, sorting was performed using the BD FACS Aria cell sorter (UW Flow Cytometry, Madison, WI) at 4°C using a 130 μm nozzle. Samples were collected by sorting 100,00 cells from each sub-population directly into RNase free tubes containing 3 mL TRIzol-LS (Thermo-Fisher) and the equivalent of a 1:125 dilution of 1 μL ERCC Spike-In RNA (Invitrogen, #4456740). For prolonged sorting, samples were vortexed every 10 minutes to ensure cell lysis. The volume of the lysate was then adjusted to 4 mL to maintain a 3:1 TRIzol to sample ratio, and lysates were thoroughly mixed and centrifuged before being stored at -80°C. Each replicate was lytic induced and sorted on different days. RNA sequencing was performed as described in the Chapter 3 section entitled ‘RNA-sequencing’. All conditions consisted of two biological replicates.

For collection of whole cell lysates, samples were collected by sorting 300,000 cells from each sub-population directly into 1.5 mL microcentrifuge tubes containing 500 μl FACS buffer. Samples were centrifuged at 300 x g for 5 minutes. Cells were then lysed and analyzed as described below.

Immunoblotting

Bulk or sorted cells were lysed with RIPA buffer (50 mM Tris HCL (pH 8.0), 150 mM NaCl, 1% NP-40 or TritonX-100, 0.5% Sodium Deoxycholate and 0.1% SDS) by incubation on ice for 30 minutes. Lysates were cleared by centrifugation at 13,000 RPM for 30 min at 4°C. The protein concentration of bulk lysates was quantified using the Coomassie (Bradford) assay kit (Thermo Scientific catalog #23200). Equal amounts of protein (bulk) or cells (sorted) lysates were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a nitrocellulose membrane. Membranes were blocked in Tris-buffered saline supplemented with 0.1% Tween 20 (TBST) containing 5% dry non-fat milk and incubated with appropriate antibodies overnight at 4°C. The following primary antibodies were used: anti-mouse Zta (Santa Cruz, sc-53904; 1:200), anti-rabbit Rta polyclonal antibody directed against the Rta peptide sequence DPDEETSSQAVKALREMAD (1:250), anti-mouse EA-D (EMD Millipore, BAB8186; 1:1000), anti-goat VCAp18 (Thermo Scientific, DA-73003; 1:1000), anti-mouse c-Myc (Santa Cruz, sc-41; 1:1000), anti-mouse CDK9 (Santa Cruz, sc-13130; 1:200), and anti-mouse β-actin (Santa Cruz, sc-4778; 1:200). Following primary antibody incubation, membranes were washed with TBST and incubated with appropriate HRP-conjugated antibodies (Thermo Fisher Scientific antibodies: 1:10000 mouse #A24524, 1:3000 rabbit #A10547, and 1:3000 goat #A24524). Membranes were washed with TBST and visualized with ECL chemiluminescent kit (Pierce ECL Western Thermo Fisher Scientific catalog #32106).

Data analysis

Read alignment and differential expression analysis was performed as in the Chapter 3 **s**ection entitled ‘Read alignment and differential expression analysis’ except using the type 1 EBV reference genome (GenBank accession number NC_007605). DAVID gene ontology (GO)

analysis was performed on genes not significantly downregulated ($P_{adjusted} \leq 0.05$). Heatmaps were constructed with the ComplexHeatmap package v2.18.0 [32] in the R environment (v4.3.0) [33]. For heatmaps, z-score transformation was applied to ERCC spike-in gene normalized expression levels across samples by cell type. Hierarchical clustering was subsequently calculated using the Euclidian distance method, both on samples and genes within subgroups defined by condition and protein complex.

References

1. Ali, A., Ohashi, M., Casco, A., Djavadian, R., Eichelberg, M., Kenney, S.C., and Johannsen, E. (2022). Rta is the principal activator of Epstein-Barr virus epithelial lytic transcription. PLoS Pathog 18, e1010886. 10.1371/journal.ppat.1010886.
2. Yamaguchi, Y., Inukai, N., Narita, T., Wada, T., and Handa, H. (2002). Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. Mol Cell Biol 22, 2918-2927. 10.1128/MCB.22.9.2918-2927.2002.
3. Adelman, K., and Henriques, T. (2018). Transcriptional speed bumps revealed in high resolution. Nature 560, 560-561. 10.1038/d41586-018-05971-8.
4. Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018). Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. Nature 560, 607-612. 10.1038/s41586-018-0440-4.
5. Egloff, S. (2021). CDK9 keeps RNA polymerase II on track. Cell Mol Life Sci 78, 5543-5567. 10.1007/s00018-021-03878-8.
6. Fujinaga, K., Huang, F., and Peterlin, B.M. (2023). P-TEFb: The master regulator of transcription elongation. Mol Cell 83, 393-403. 10.1016/j.molcel.2022.12.006.
7. Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol Cell 23, 297-305. 10.1016/j.molcel.2006.06.014.
8. Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature 414, 322-325. 10.1038/35104581.
9. Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature 414, 317-322. 10.1038/35104575.
10. Markert, A., Grimm, M., Martinez, J., Wiesner, J., Meyerhans, A., Meyuhas, O., Sickmann, A., and Fischer, U. (2008). The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. EMBO Rep 9, 569-575. 10.1038/embor.2008.72.
11. Michels, A.A., Fraldi, A., Li, Q., Adamson, T.E., Bonnet, F., Nguyen, V.T., Sedore, S.C., Price, J.P., Price, D.H., Lania, L., and Bensaude, O. (2004). Binding of the 7SK snRNA turns the HEXIM1 protein into a P-TEFb (CDK9/cyclin T) inhibitor. EMBO J 23, 2608-2619. 10.1038/sj.emboj.7600275.
12. AJ, C.Q., Bugai, A., and Barboric, M. (2016). Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. Nucleic Acids Res 44, 7527-7539. 10.1093/nar/gkw585.
13. Kim, Y.K., Mbonye, U., Hokello, J., and Karn, J. (2011). T-cell receptor signaling enhances transcriptional elongation from latent HIV proviruses by activating P-TEFb through an ERK-dependent pathway. J Mol Biol 410, 896-916. 10.1016/j.jmb.2011.03.054.
14. Contreras, X., Barboric, M., Lenasi, T., and Peterlin, B.M. (2007). HMBA releases P-TEFb from HEXIM1 and 7SK snRNA via PI3K/Akt and activates HIV transcription. PLoS Pathog 3, 1459-1469. 10.1371/journal.ppat.0030146.
15. Chen, R., Liu, M., Li, H., Xue, Y., Ramey, W.N., He, N., Ai, N., Luo, H., Zhu, Y., Zhou, N., and Zhou, Q. (2008). PP2B and PP1alpha cooperatively disrupt 7SK snRNP to

release P-TEFb for transcription in response to Ca2+ signaling. Genes Dev 22, 1356-1368. 10.1101/gad.1636008.
16. Bugai, A., Quaresma, A.J.C., Friedel, C.C., Lenasi, T., Duster, R., Sibley, C.R., Fujinaga, K., Kukanja, P., Hennig, T., Blasius, M., et al. (2019). P-TEFb Activation by RBM7 Shapes a Pro-survival Transcriptional Response to Genotoxic Stress. Mol Cell 74, 254-267 e210. 10.1016/j.molcel.2019.01.033.
17. Ji, X., Zhou, Y., Pandit, S., Huang, J., Li, H., Lin, C.Y., Xiao, R., Burge, C.B., and Fu, X.D. (2013). SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. Cell 153, 855-868. 10.1016/j.cell.2013.04.028.
18. Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130, 77-88. 10.1016/j.cell.2007.05.042.
19. Muse, G.W., Gilchrist, D.A., Nechaev, S., Shah, R., Parker, J.S., Grissom, S.F., Zeitlinger, J., and Adelman, K. (2007). RNA polymerase is poised for activation across the genome. Nat Genet 39, 1507-1511. 10.1038/ng.2007.21.
20. Johannsen, E., Luftig, M., Chase, M.R., Weicksel, S., Cahir-McFarland, E., Illanes, D., Sarracino, D., and Kieff, E. (2004). Proteins of purified Epstein-Barr virus. Proc Natl Acad Sci U S A 101, 16286-16291. 10.1073/pnas.0407320101.
21. Liu, P., Xiang, Y., Fujinaga, K., Bartholomeeusen, K., Nilson, K.A., Price, D.H., and Peterlin, B.M. (2014). Release of positive transcription elongation factor b (P-TEFb) from 7SK small nuclear ribonucleoprotein (snRNP) activates hexamethylene bisacetamide-inducible protein (HEXIM1) transcription. J Biol Chem 289, 9918-9925. 10.1074/jbc.M113.539015.
22. Keck, K.M., Moquin, S.A., He, A., Fernandez, S.G., Somberg, J.J., Liu, S.M., Martinez, D.M., and Miranda, J.L. (2017). Bromodomain and extraterminal inhibitors block the Epstein-Barr virus lytic cycle at two distinct steps. J Biol Chem 292, 13284-13295. 10.1074/jbc.M116.751644.
23. Guo, R., Jiang, C., Zhang, Y., Govande, A., Trudeau, S.J., Chen, F., Fry, C.J., Puri, R., Wolinsky, E., Schineller, M., et al. (2020). MYC Controls the Epstein-Barr Virus Lytic Switch. Mol Cell 78, 653-669 e658. 10.1016/j.molcel.2020.03.025.
24. Coleman, D.J., Gao, L., Schwartzman, J., Korkola, J.E., Sampson, D., Derrick, D.S., Urrutia, J., Balter, A., Burchard, J., King, C.J., et al. (2019). Maintenance of MYC expression promotes de novo resistance to BET bromodomain inhibition in castration-resistant prostate cancer. Sci Rep 9, 3823. 10.1038/s41598-019-40518-5.
25. Delecluse, H.J., Hilsendegen, T., Pich, D., Zeidler, R., and Hammerschmidt, W. (1998). Propagation and recovery of intact, infectious Epstein-Barr virus from prokaryotic to human cells. Proc Natl Acad Sci U S A 95, 8245-8250. 10.1073/pnas.95.14.8245.
26. Tischer, B.K., Smith, G.A., and Osterrieder, N. (2010). En passant mutagenesis: a two step markerless red recombination system. Methods Mol Biol 634, 421-430. 10.1007/978-1-60761-652-8_30.
27. Djavadian, R., Hayes, M., and Johannsen, E. (2018). CAGE-seq analysis of Epstein-Barr virus lytic gene transcription: 3 kinetic classes from 2 mechanisms. PLoS Pathog 14, e1007114. 10.1371/journal.ppat.1007114.
28. Meng, Q., Hagemeier, S.R., Fingeroth, J.D., Gershburg, E., Pagano, J.S., and Kenney, S.C. (2010). The Epstein-Barr virus (EBV)-encoded protein kinase, EBV-PK, but not the

thymidine kinase (EBV-TK), is required for ganciclovir and acyclovir inhibition of lytic viral production. J Virol 84, 4534-4542. 10.1128/JVI.02487-09.
29. Shimizu, N., Yoshiyama, H., and Takada, K. (1996). Clonal propagation of Epstein-Barr virus (EBV) recombinants in EBV-negative Akata cells. J Virol 70, 7260-7263. 10.1128/JVI.70.10.7260-7263.1996.
30. Hutt-Fletcher, L.M. (2015). EBV glycoproteins: where are we now? Future Virol 10, 1155-1162. 10.2217/fvl.15.80.
31. Djavadian, R., Chiu, Y.F., and Johannsen, E. (2016). An Epstein-Barr Virus-Encoded Protein Complex Requires an Origin of Lytic Replication In Cis to Mediate Late Gene Transcription. PLoS Pathog 12, e1005718. 10.1371/journal.ppat.1005718.
32. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32, 2847-2849. 10.1093/bioinformatics/btw313.
33. Team, R.C. (2023). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).