
Ribosome profiling of EBV-infected cells

2016

Maja Bencun

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics of the Ruperto-
Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by

Maja Bencun, M.Sc.

born in: Zenica; Bosnia and Herzegowina

Oral-examination: _____

Title:

Ribosome profiling of EBV-infected cells

Referees: Prof. Dr. Dr. Henri-Jacques Delecluse

Prof. Dr. Martin Müller

“The truth is rarely pure and never simple.”

— Oscar Wilde, *The Importance of Being Earnest*

Corrigendum

Maja Bencun:

Ribosome Profiling of EBV-infected cells

PhD Thesis, Ruperto-Carola University of Heidelberg, 2016

Last updated: April, 9th 2017.

Chapter 3: Results

The references in this section of the thesis do not match the references listed under section 7: References.

The publications listed below are the correct references for this section of the thesis and are listed according to the numbering in the text:

1. Tsai, M. H. *et al.* Spontaneous lytic replication and epitheliotropism define an Epstein-Barr virus strain found in carcinomas. *Cell reports* **5**, 458-470, doi:10.1016/j.celrep.2013.09.012 (2013).
2. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789-802, doi:10.1016/j.cell.2011.10.002 (2011).
3. Schneider-Poetsch, T. *et al.* Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nature chemical biology* **6**, 209-217, doi:10.1038/nchembio.304 (2010).
4. Fresno, M., Jimenez, A. & Vazquez, D. Inhibition of translation in eukaryotic systems by harringtonine. *European journal of biochemistry / FEBS* **72**, 323-330 (1977).
5. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics* **10**, 57-63, doi:10.1038/nrg2484 (2009).
6. Arvey, A. *et al.* An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. *Cell host & microbe* **12**, 233-245, doi:10.1016/j.chom.2012.06.008 (2012).
7. Fok, V., Mitton-Fry, R. M., Grech, A. & Steitz, J. A. Multiple domains of EBER 1, an Epstein-Barr virus noncoding RNA, recruit human ribosomal protein L22. *Rna* **12**, 872-882, doi:10.1261/rna.2339606 (2006).
8. Chew, G. L. *et al.* Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. *Development* **140**, 2828-2834, doi:10.1242/dev.098343 (2013).
9. Strong, M. J. *et al.* Latent Expression of the Epstein-Barr Virus (EBV)-Encoded Major Histocompatibility Complex Class I TAP Inhibitor, BNLF2a, in EBV-Positive Gastric Carcinomas. *Journal of virology* **89**, 10110-10114, doi:10.1128/JVI.01110-15 (2015).
10. Glickman, J. N., Howe, J. G. & Steitz, J. A. Structural analyses of EBER1 and EBER2 ribonucleoprotein particles present in Epstein-Barr virus-infected cells. *Journal of virology* **62**, 902-911 (1988).
11. Iwakiri, D. & Takada, K. Role of EBERs in the pathogenesis of EBV infection. *Advances in cancer research* **107**, 119-136, doi:10.1016/S0065-230X(10)07004-1 (2010).
12. Rennekamp, A. J. & Lieberman, P. M. Initiation of Epstein-Barr virus lytic replication requires transcription and the formation of a stable RNA-DNA hybrid molecule at OriLyt. *Journal of virology* **85**, 2837-2850, doi:10.1128/JVI.02175-10 (2011).
13. Sample, J., Hummel, M., Braun, D., Birkenbach, M. & Kieff, E. Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear proteins: a probable transcriptional initiation site. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 5096-5100 (1986).
14. Sample, J. & Kieff, E. Transcription of the Epstein-Barr virus genome during latency in growth-transformed lymphocytes. *Journal of virology* **64**, 1667-1674 (1990).
15. Jeang, K. T. & Hayward, S. D. Organization of the Epstein-Barr virus DNA molecule. III. Location of the P3HR-1 deletion junction and characterization of the NotI repeat units that form

part of the template for an abundant 12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript. *Journal of virology* **48**, 135-148 (1983).

16. Hummel, M. & Kieff, E. Epstein-Barr virus RNA. VIII. Viral RNA in permissively infected B95-8 cells. *Journal of virology* **43**, 262-272 (1982).
17. Al-Mozaini, M. *et al.* Epstein-Barr virus BART gene expression. *The Journal of general virology* **90**, 307-316, doi:10.1099/vir.0.006551-0 (2009).
18. Smith, P. R. *et al.* Structure and coding content of CST (BART) family RNAs of Epstein-Barr virus. *Journal of virology* **74**, 3082-3092 (2000).
19. Calderwood, M. A., Holthaus, A. M. & Johannsen, E. The Epstein-Barr virus LF2 protein inhibits viral replication. *Journal of virology* **82**, 8509-8519, doi:10.1128/JVI.00315-08 (2008).
20. Heilmann, A. M., Calderwood, M. A. & Johannsen, E. Epstein-Barr virus LF2 protein regulates viral replication by altering Rta subcellular localization. *Journal of virology* **84**, 9920-9931, doi:10.1128/JVI.00573-10 (2010).
21. Wu, L. *et al.* Epstein-Barr virus LF2: an antagonist to type I interferon. *Journal of virology* **83**, 1140-1146, doi:10.1128/JVI.00602-08 (2009).
22. Concha, M. *et al.* Identification of new viral genes and transcript isoforms during Epstein-Barr virus reactivation using RNA-Seq. *Journal of virology* **86**, 1458-1467, doi:10.1128/JVI.06537-11 (2012).
23. O'Grady, T. *et al.* Global bidirectional transcription of the Epstein-Barr virus genome during reactivation. *Journal of virology* **88**, 1604-1616, doi:10.1128/JVI.02989-13 (2014).
24. O'Grady, T. *et al.* Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucleic acids research*, doi:10.1093/nar/gkw629 (2016).
25. Dresang, L. R. *et al.* Coupled transcriptome and proteome analysis of human lymphotropic tumor viruses: insights on the detection and discovery of viral genes. *BMC genomics* **12**, 625, doi:10.1186/1471-2164-12-625 (2011).
26. Ossevoort, M. *et al.* The nested open reading frame in the Epstein-Barr virus nuclear antigen-1 mRNA encodes a protein capable of inhibiting antigen presentation in cis. *Molecular immunology* **44**, 3588-3596, doi:10.1016/j.molimm.2007.03.005 (2007).
27. Kwun, H. J. *et al.* Human DNA tumor viruses generate alternative reading frame proteins through repeat sequence recoding. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E4342-4349, doi:10.1073/pnas.1416122111 (2014).
28. Lu, P. D., Harding, H. P. & Ron, D. Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *The Journal of cell biology* **167**, 27-33, doi:10.1083/jcb.200408003 (2004).
29. Vatter, K. M. & Wek, R. C. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 11269-11274, doi:10.1073/pnas.0400541101 (2004).
30. Bodescot, M., Perricaudet, M. & Farrell, P. J. A promoter for the highly spliced EBNA family of RNAs of Epstein-Barr virus. *Journal of virology* **61**, 3424-3430 (1987).
31. Henderson, S. *et al.* Epstein-Barr virus-coded BHRF1 protein, a viral homologue of Bcl-2, protects human B cells from programmed cell death. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 8479-8483 (1993).
32. Pearson, G. R. *et al.* Identification of an Epstein-Barr virus early gene encoding a second component of the restricted early antigen complex. *Virology* **160**, 151-161 (1987).
33. Austin, P. J., Flemington, E., Yandava, C. N., Strominger, J. L. & Speck, S. H. Complex transcription of the Epstein-Barr virus BamHI fragment H rightward open reading frame 1 (BHRF1) in latently and lytically infected B lymphocytes. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 3678-3682 (1988).
34. Isaksson, A., Berggren, M. & Ricksten, A. Epstein-Barr virus U leader exon contains an internal ribosome entry site. *Oncogene* **22**, 572-581, doi:10.1038/sj.onc.1206149 (2003).
35. Spriggs, K. A., Stoneley, M., Bushell, M. & Willis, A. E. Re-programming of translation following cell stress allows IRES-mediated translation to predominate. *Biology of the cell / under the auspices of the European Cell Biology Organization* **100**, 27-38, doi:10.1042/BC20070098 (2008).
36. Ivanov, I. P., Firth, A. E., Michel, A. M., Atkins, J. F. & Baranov, P. V. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic acids research* **39**, 4220-4234, doi:10.1093/nar/gkr007 (2011)

37. Touriol, C. *et al.* Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biology of the cell / under the auspices of the European Cell Biology Organization* **95**, 169-178 (2003).
38. Chua, J. J. *et al.* Synthesis of two SAPAP3 isoforms from a single mRNA is mediated via alternative translational initiation. *Scientific reports* **2**, 484, doi:10.1038/srep00484 (2012).
39. Calkhoven, C. F., Muller, C. & Leutz, A. Translational control of C/EBPalpha and C/EBPbeta isoform expression. *Genes & development* **14**, 1920-1932 (2000).
40. Leissring, M. A. *et al.* Alternative translation initiation generates a novel isoform of insulin-degrading enzyme targeted to mitochondria. *The Biochemical journal* **383**, 439-446, doi:10.1042/BJ20041081 (2004).

Chapter 4: Discussion

The references in this section of the thesis do not match the references listed under section 7: References.

The publications listed below are the correct references for this section of the thesis and are listed according to the numbering in the text:

1. Brar, G. A. *et al.* High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335**, 552-557, doi:10.1126/science.1215110 (2012).
2. Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. & Lander, E. S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154**, 240-251, doi:10.1016/j.cell.2013.06.009 (2013).
3. Ingolia, N. T. *et al.* Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell reports* **8**, 1365-1379, doi:10.1016/j.celrep.2014.07.045 (2014).
4. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789-802, doi:10.1016/j.cell.2011.10.002 (2011).
5. Stern-Ginossar, N. *et al.* Decoding human cytomegalovirus. *Science* **338**, 1088-1093, doi:10.1126/science.1227919 (2012).
6. Arias, C. *et al.* KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS pathogens* **10**, e1003847, doi:10.1371/journal.ppat.1003847 (2014).
7. Popa, A., Lebrigand, K., Barbry, P. & Waldmann, R. Pateamine A-sensitive ribosome profiling reveals the scope of translation in mouse embryonic stem cells. *BMC genomics* **17**, 52, doi:10.1186/s12864-016-2384-0 (2016).
8. Rowe, M. *et al.* Host shutoff during productive Epstein-Barr virus infection is mediated by BGLF5 and may contribute to immune evasion. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 3366-3371, doi:10.1073/pnas.0611128104 (2007).
9. Ricci, E. P. *et al.* Translation of intronless RNAs is strongly stimulated by the Epstein-Barr virus mRNA export factor EB2. *Nucleic acids research* **37**, 4932-4943, doi:10.1093/nar/gkp497 (2009).
10. O'Grady, T. *et al.* Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucleic acids research*, doi:10.1093/nar/gkw629 (2016).
11. O'Grady, T. *et al.* Global bidirectional transcription of the Epstein-Barr virus genome during reactivation. *Journal of virology* **88**, 1604-1616, doi:10.1128/JVI.02989-13 (2014).
12. Tsurumi, T. Purification and characterization of the DNA-binding activity of the Epstein-Barr virus DNA polymerase accessory protein BMRF1 gene products, as expressed in insect cells by using the baculovirus system. *Journal of virology* **67**, 1681-1687 (1993).
13. Tsurumi, T., Daikoku, T., Kurachi, R. & Nishiyama, Y. Functional interaction between Epstein-Barr virus DNA polymerase catalytic subunit and its accessory subunit in vitro. *Journal of virology* **67**, 7648-7653 (1993).
14. Tsurumi, T., Daikoku, T. & Nishiyama, Y. Further characterization of the interaction between the Epstein-Barr virus DNA polymerase catalytic subunit and its accessory subunit with regard

to the 3'-to-5' exonucleolytic activity and stability of initiation complex at primer terminus. *Journal of virology* **68**, 3354-3363 (1994).

15. Quinlivan, E. B. *et al.* Direct BRLF1 binding is required for cooperative BZLF1/BRLF1 activation of the Epstein-Barr virus early promoter, BMRF1. *Nucleic acids research* **21**, 1999-2007 (1993).
16. Nakayama, S. *et al.* Epstein-Barr virus polymerase processivity factor enhances BALF2 promoter transcription as a coactivator for the BZLF1 immediate-early protein. *The Journal of biological chemistry* **284**, 21557-21568, doi:10.1074/jbc.M109.015685 (2009).
17. Zhang, Q. *et al.* Functional and physical interactions between the Epstein-Barr virus (EBV) proteins BZLF1 and BMRF1: Effects on EBV transcription and lytic replication. *Journal of virology* **70**, 5131-5142 (1996).
18. Laressergues, D. *et al.* Primary transcripts of microRNAs encode regulatory peptides. *Nature* **520**, 90-93, doi:10.1038/nature14346 (2015).
19. Dambaugh, T. R. & Kieff, E. Identification and nucleotide sequences of two similar tandem direct repeats in Epstein-Barr virus DNA. *Journal of virology* **44**, 823-833 (1982).
20. Gao, Y., Smith, P. R., Karran, L., Lu, Q. L. & Griffin, B. E. Induction of an exceptionally high-level, nontranslated, Epstein-Barr virus-encoded polyadenylated transcript in the Burkitt's lymphoma line Daudi. *Journal of virology* **71**, 84-94 (1997).
21. Hudewentz, J., Delius, H., Freese, U. K., Zimber, U. & Bornkamm, G. W. Two distant regions of the Epstein-Barr virus genome with sequence homologies have the same orientation and involve small tandem repeats. *The EMBO journal* **1**, 21-26 (1982).
22. Lieberman, P. M., Hardwick, J. M. & Hayward, S. D. Responsiveness of the Epstein-Barr virus NotI repeat promoter to the Z transactivator is mediated in a cell-type-specific manner by two independent signal regions. *Journal of virology* **63**, 3040-3050 (1989).
23. Nuebling, C. M. & Mueller-Lantzsch, N. Identification and characterization of an Epstein-Barr virus early antigen that is encoded by the NotI repeats. *Journal of virology* **63**, 4609-4615 (1989).
24. Xue, S. A., Jones, M. D., Lu, Q. L., Middeldorp, J. M. & Griffin, B. E. Genetic diversity: frameshift mechanisms alter coding of a gene (Epstein-Barr virus LF3 gene) that contains multiple 102-base-pair direct sequence repeats. *Molecular and cellular biology* **23**, 2192-2201 (2003).
25. Rennekamp, A. J. & Lieberman, P. M. Initiation of Epstein-Barr virus lytic replication requires transcription and the formation of a stable RNA-DNA hybrid molecule at OriLyt. *Journal of virology* **85**, 2837-2850, doi:10.1128/JVI.02175-10 (2011).
26. Goossen, B. & Hentze, M. W. Position is the critical determinant for function of iron-responsive elements as translational regulators. *Molecular and cellular biology* **12**, 1959-1966 (1992).
27. Kozak, M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Molecular and cellular biology* **9**, 5134-5142 (1989).
28. Poole, E., Kuan, W. L., Barker, R. & Sinclair, J. The human cytomegalovirus non-coding Beta2.7 RNA as a novel therapeutic for Parkinson's disease--Translational research with no translation. *Virus research* **212**, 64-69, doi:10.1016/j.virusres.2015.05.007 (2016).
29. Bergamini, G. *et al.* The major open reading frame of the beta2.7 transcript of human cytomegalovirus: in vitro expression of a protein posttranscriptionally regulated by the 5' region. *Journal of virology* **72**, 8425-8429 (1998).
30. McSharry, B. P., Tomasec, P., Neale, M. L. & Wilkinson, G. W. The most abundantly transcribed human cytomegalovirus gene (beta 2.7) is non-essential for growth in vitro. *The Journal of general virology* **84**, 2511-2516, doi:10.1099/vir.0.19298-0 (2003).
31. Lin, Z. *et al.* Whole-genome sequencing of the Akata and Mutu Epstein-Barr virus strains. *Journal of virology* **87**, 1172-1182, doi:10.1128/JVI.02517-12 (2013).
32. Geballe, A. P. & Mocarski, E. S. Translational control of cytomegalovirus gene expression is mediated by upstream AUG codons. *Journal of virology* **62**, 3334-3340 (1988).
33. de Turris, V., Nicholson, P., Orozco, R. Z., Singer, R. H. & Muhlemann, O. Cotranscriptional effect of a premature termination codon revealed by live-cell imaging. *Rna* **17**, 2094-2107, doi:10.1261/rna.02918111 (2011).
34. Smith, C. M. & Steitz, J. A. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Molecular and cellular biology* **18**, 6897-6909 (1998).

35. Tani, H., Torimura, M. & Akimitsu, N. The RNA degradation pathway regulates the function of GAS5 a non-coding RNA in mammalian cells. *PLoS one* **8**, e55684, doi:10.1371/journal.pone.0055684 (2013).
36. Yepiskoposyan, H., Aeschmann, F., Nilsson, D., Okoniewski, M. & Muhlemann, O. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *Rna* **17**, 2108-2118, doi:10.1261/rna.030247.111 (2011).
37. Nicholson, P. *et al.* Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cellular and molecular life sciences : CMLS* **67**, 677-700, doi:10.1007/s00018-009-0177-1 (2010).
38. Mendell, J. T., Sharifi, N. A., Meyers, J. L., Martinez-Murillo, F. & Dietz, H. C. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nature genetics* **36**, 1073-1078, doi:10.1038/ng1429 (2004).
39. Smith, J. E. *et al.* Translation of small open reading frames within unannotated RNA transcripts in *Saccharomyces cerevisiae*. *Cell reports* **7**, 1858-1866, doi:10.1016/j.celrep.2014.05.023 (2014).
40. Laux, G., Freese, U. K. & Bornkamm, G. W. Structure and evolution of two related transcription units of Epstein-Barr virus carrying small tandem repeats. *Journal of virology* **56**, 987-995 (1985).
41. Xue, S. & Barna, M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nature reviews. Molecular cell biology* **13**, 355-369, doi:10.1038/nrm3359 (2012).
42. Parker, B., Bankier, A., Satchwell, S., Barrell, B. & Farrell, P. Sequence and Transcription of Raji Epstein-Barr Virus DNA Spanning the B95-8 Deletion Region. *Virology* **179**, 339-346 (1990).
43. Shannon-Lowe, C. D., Neuhierl, B., Baldwin, G., Rickinson, A. B. & Delecluse, H. J. Resting B cells as a transfer vehicle for Epstein-Barr virus infection of epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 7065-7070, doi:10.1073/pnas.0510512103 (2006).
44. Calderwood, M. A. *et al.* Epstein-Barr virus and virus human protein interaction maps. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 7606-7611, doi:10.1073/pnas.0702332104 (2007).
45. Calderwood, M. A., Holthaus, A. M. & Johannsen, E. The Epstein-Barr virus LF2 protein inhibits viral replication. *Journal of virology* **82**, 8509-8519, doi:10.1128/JVI.00315-08 (2008).
46. Heilmann, A. M., Calderwood, M. A. & Johannsen, E. Epstein-Barr virus LF2 protein regulates viral replication by altering Rta subcellular localization. *Journal of virology* **84**, 9920-9931, doi:10.1128/JVI.00573-10 (2010).
47. Wu, L. *et al.* Epstein-Barr virus LF2: an antagonist to type I interferon. *Journal of virology* **83**, 1140-1146, doi:10.1128/JVI.00602-08 (2009).
48. Davison, A. J. & Stow, N. D. New genes from old: redeployment of dUTPase by herpesviruses. *Journal of virology* **79**, 12880-12892, doi:10.1128/JVI.79.20.12880-12892.2005 (2005).
49. Zhu, J. *et al.* Protein array identification of substrates of the Epstein-Barr virus protein kinase BGLF4. *Journal of virology* **83**, 5219-5231, doi:10.1128/JVI.02378-08 (2009).
50. Sadler, R. H. & Raab-Traub, N. Structural analyses of the Epstein-Barr virus BamHI A transcripts. *Journal of virology* **69**, 1132-1141 (1995).
51. Yamamoto, T. & Iwatsuki, K. Diversity of Epstein-Barr virus BamHI-A rightward transcripts and their expression patterns in lytic and latent infections. *Journal of medical microbiology* **61**, 1445-1453, doi:10.1099/jmm.0.044727-0 (2012).
52. Al-Mozaini, M. *et al.* Epstein-Barr virus BART gene expression. *The Journal of general virology* **90**, 307-316, doi:10.1099/vir.0.006551-0 (2009).
53. Edwards, R. H., Marquitz, A. R. & Raab-Traub, N. Epstein-Barr virus BART microRNAs are produced from a large intron prior to splicing. *Journal of virology* **82**, 9094-9106, doi:10.1128/JVI.00785-08 (2008).
54. Cai, X. *et al.* Epstein-Barr virus microRNAs are evolutionarily conserved and differentially expressed. *PLoS pathogens* **2**, e23, doi:10.1371/journal.ppat.0020023 (2006).
55. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563-547 (1996).
56. Anderson, D. M. *et al.* A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**, 595-606, doi:10.1016/j.cell.2015.01.009 (2015).
57. Kondo, T. *et al.* Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science* **329**, 336-339, doi:10.1126/science.1188158 (2010).

58. Magny, E. G. *et al.* Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* **341**, 1116-1120, doi:10.1126/science.1238802 (2013).
59. Barbosa, C., Peixeiro, I. & Romao, L. Gene expression regulation by upstream open reading frames and human disease. *PLoS genetics* **9**, e1003529, doi:10.1371/journal.pgen.1003529 (2013).
60. Wethmar, K. The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley interdisciplinary reviews. RNA* **5**, 765-778, doi:10.1002/wrna.1245 (2014).
61. Wethmar, K. *et al.* Comprehensive translational control of tyrosine kinase expression by upstream open reading frames. *Oncogene* **35**, 1736-1742, doi:10.1038/onc.2015.233 (2016).
62. Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7507-7512, doi:10.1073/pnas.0810916106 (2009).
63. Lee, S. *et al.* Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E2424-2432, doi:10.1073/pnas.1207846109 (2012).
64. Ye, Y. *et al.* Analysis of human upstream open reading frames and impact on gene expression. *Human genetics* **134**, 605-612, doi:10.1007/s00439-015-1544-7 (2015).
65. Shabman, R. S. *et al.* An upstream open reading frame modulates ebola virus polymerase translation and virus replication. *PLoS pathogens* **9**, e1003147, doi:10.1371/journal.ppat.1003147 (2013).
66. Vattam, K. M. & Wek, R. C. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 11269-11274, doi:10.1073/pnas.0400541101 (2004).
67. Palam, L. R., Baird, T. D. & Wek, R. C. Phosphorylation of eIF2 facilitates ribosomal bypass of an inhibitory upstream ORF to enhance CHOP translation. *The Journal of biological chemistry* **286**, 10939-10949, doi:10.1074/jbc.M110.216093 (2011).
68. Le Clorennec, C. *et al.* Molecular basis of cytotoxicity of Epstein-Barr virus (EBV) latent membrane protein 1 (LMP1) in EBV latency III B cells: LMP1 induces type II ligand-independent autoactivation of CD95/Fas with caspase 8-mediated apoptosis. *Journal of virology* **82**, 6721-6733, doi:10.1128/JVI.02250-07 (2008).
69. Isaksson, A., Berggren, M. & Ricksten, A. Epstein-Barr virus U leader exon contains an internal ribosome entry site. *Oncogene* **22**, 572-581, doi:10.1038/sj.onc.1206149 (2003).
70. Bieleski, L., Hindley, C. & Talbot, S. J. A polypyrimidine tract facilitates the expression of Kaposi's sarcoma-associated herpesvirus vFLIP through an internal ribosome entry site. *The Journal of general virology* **85**, 615-620, doi:10.1099/vir.0.19733-0 (2004).
71. Bieleski, L. & Talbot, S. J. Kaposi's sarcoma-associated herpesvirus vCyclin open reading frame contains an internal ribosome entry site. *Journal of virology* **75**, 1864-1869, doi:10.1128/JVI.75.4.1864-1869.2001 (2001).
72. Mehta, A., Trotta, C. R. & Peltz, S. W. Derepression of the Her-2 uORF is mediated by a novel post-transcriptional control mechanism in cancer cells. *Genes & development* **20**, 939-953, doi:10.1101/gad.1388706 (2006).
73. Child, S. J., Miller, M. K. & Geballe, A. P. Translational control by an upstream open reading frame in the HER-2/neu transcript. *The Journal of biological chemistry* **274**, 24335-24341 (1999).
74. Fernandez, J. *et al.* Regulation of internal ribosome entry site-mediated translation by eukaryotic initiation factor-2alpha phosphorylation and translation of a small upstream open reading frame. *The Journal of biological chemistry* **277**, 2050-2058, doi:10.1074/jbc.M109199200 (2002).
75. Yaman, I. *et al.* The zipper model of translational control: a small upstream ORF is the switch that controls structural remodeling of an mRNA leader. *Cell* **113**, 519-531 (2003).
76. Bonneau, A. M. & Sonenberg, N. Involvement of the 24-kDa cap-binding protein in regulation of protein synthesis in mitosis. *The Journal of biological chemistry* **262**, 11134-11139 (1987).
77. Tarnowka, M. A. & Baglioni, C. Regulation of protein synthesis in mitotic HeLa cells. *Journal of cellular physiology* **99**, 359-367, doi:10.1002/jcp.1040990311 (1979).
78. Huang, J. T. & Schneider, R. J. Adenovirus inhibition of cellular protein synthesis involves inactivation of cap-binding protein. *Cell* **65**, 271-280 (1991).
79. Pyronnet, S., Pradayrol, L. & Sonenberg, N. A cell cycle-dependent internal ribosome entry site. *Molecular cell* **5**, 607-616 (2000).

80. Adamson, A. L. *et al.* Epstein-Barr virus immediate-early proteins BZLF1 and BRLF1 activate the ATF2 transcription factor by increasing the levels of phosphorylated p38 and c-Jun N-terminal kinases. *Journal of virology* **74**, 1224-1233 (2000).
81. Gargouri, B., Van Pelt, J., El Feki Ael, F., Attia, H. & Lassoued, S. Induction of Epstein-Barr virus (EBV) lytic cycle in vitro causes oxidative stress in lymphoblastoid B cell lines. *Molecular and cellular biochemistry* **324**, 55-63, doi:10.1007/s11010-008-9984-1 (2009).
82. Cao, J. & Geballe, A. P. Inhibition of nascent-peptide release at translation termination. *Molecular and cellular biology* **16**, 7109-7114 (1996).
83. Cao, J. & Geballe, A. P. Coding sequence-dependent ribosomal arrest at termination of translation. *Molecular and cellular biology* **16**, 603-608 (1996).
84. Chen, A., Kao, Y. F. & Brown, C. M. Translation of the first upstream ORF in the hepatitis B virus pregenomic RNA modulates translation at the core and polymerase initiation codons. *Nucleic acids research* **33**, 1169-1181, doi:10.1093/nar/gki251 (2005).
85. Lu, C. C. *et al.* Characterization of the uracil-DNA glycosylase activity of Epstein-Barr virus BKRF3 and its role in lytic viral DNA replication. *Journal of virology* **81**, 1195-1208, doi:10.1128/JVI.01518-06 (2007).
86. Su, M. T. *et al.* Uracil DNA glycosylase BKRF3 contributes to Epstein-Barr virus DNA replication through physical interactions with proteins in viral DNA replication complex. *Journal of virology* **88**, 8883-8899, doi:10.1128/JVI.00950-14 (2014).
87. Curran, J. & Kolakofsky, D. Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA. *The EMBO journal* **7**, 245-251 (1988).
88. Prats, A. C., De Billy, G., Wang, P. & Darlix, J. L. CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus. *Journal of molecular biology* **205**, 363-372 (1989).
89. Touriol, C. *et al.* Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biology of the cell / under the auspices of the European Cell Biology Organization* **95**, 169-178 (2003).
90. Hopkins, B. D. *et al.* A secreted PTEN phosphatase that enters cells to alter signaling and survival. *Science* **341**, 399-402, doi:10.1126/science.1234907 (2013).
91. Liang, H. *et al.* PTENalpha, a PTEN isoform translated through alternative initiation, regulates mitochondrial function and energy metabolism. *Cell metabolism* **19**, 836-848, doi:10.1016/j.cmet.2014.03.023 (2014).
92. Furnari, F. B., Zacny, V., Quinlivan, E. B., Kenney, S. & Pagano, J. S. RAZ, an Epstein-Barr virus transdominant repressor that modulates the viral reactivation mechanism. *Journal of virology* **68**, 1827-1836 (1994).
93. Lee, D. Y., Lee, J. & Sugden, B. The unfolded protein response and autophagy: herpesviruses rule! *Journal of virology* **83**, 1168-1172, doi:10.1128/JVI.01358-08 (2009).
94. Lam, N., Sandberg, M. L. & Sugden, B. High physiological levels of LMP1 result in phosphorylation of eIF2 alpha in Epstein-Barr virus-infected cells. *Journal of virology* **78**, 1657-1664 (2004).
95. Zhao, Y. *et al.* LMP1 stimulates the transcription of eIF4E to promote the proliferation, migration and invasion of human nasopharyngeal carcinoma. *The FEBS journal* **281**, 3004-3018, doi:10.1111/febs.12838 (2014).
96. Moody, C. A. *et al.* Modulation of the cell growth regulator mTOR by Epstein-Barr virus-encoded LMP2A. *Journal of virology* **79**, 5499-5506, doi:10.1128/JVI.79.9.5499-5506.2005 (2005).
97. Shen, C. L. *et al.* Ribosome Protein L4 is essential for Epstein-Barr Virus Nuclear Antigen 1 function. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 2229-2234, doi:10.1073/pnas.1525444113 (2016).
98. Houmani, J. L., Davis, C. I. & Ruf, I. K. Growth-promoting properties of Epstein-Barr virus EBER-1 RNA correlate with ribosomal protein L22 binding. *Journal of virology* **83**, 9844-9853, doi:10.1128/JVI.01014-09 (2009).

Chapter 5: Materials

The references in this section of the thesis do not match the references listed under section 7: References.

The publications listed below are the correct references for this section of the thesis and are listed according to the numbering in the text:

1. Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *The Journal of general virology* **36**, 59-74, doi:10.1099/0022-1317-36-1-59 (1977).
2. Delecluse, H. J., Hilsendegen, T., Pich, D., Zeidler, R. & Hammerschmidt, W. Propagation and recovery of intact, infectious Epstein-Barr virus from prokaryotic to human cells. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 8245-8250 (1998).
3. Tsai, M. H. *et al.* Spontaneous lytic replication and epitheliotropism define an Epstein-Barr virus strain found in carcinomas. *Cell reports* **5**, 458-470, doi:10.1016/j.celrep.2013.09.012 (2013).

Chapter 6: Methods

Section 6.8.2. Read alignment & normalization, page 104: The correct citation is: Dillies, M. A. *et al.* A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in bioinformatics* **14**, 671-683, doi:10.1093/bib/bbs046 (2013).

Section 6.8.3. Identification of translation initiation sites, page 105: The correct citation is: Stern-Ginossar, N. *et al.* Decoding human cytomegalovirus. *Science* **338**, 1088-1093, doi:10.1126/science.1227919 (2012).

Table of Contents

| | |
|------------------------------|----------|
| Summary | 1 |
| Zusammenfassung | 2 |
| Acknowledgments | 3 |
| List of Abbreviations | 4 |

| | |
|---|-----------|
| 1. Introduction | 6 |
| 1.1. EBV is an oncogenic herpesvirus | 6 |
| 1.1.1. EBV particle structure | 6 |
| 1.1.2. EBV genome structure | 7 |
| 1.1.3. The EBV transcriptome | 10 |
| 1.1.4. The viral glycoproteins determine cell tropism | 10 |
| 1.1.5. Lytic replication | 12 |
| 1.1.6. B cell transformation <i>in vitro</i> | 13 |
| 1.1.7. EBV infection of the host | 13 |
| 1.1.8. EBV strain heterogeneity | 15 |
| 1.1.9. EBV-associated malignancies | 16 |
| 1.2. Translation in mammalian cells | 19 |
| 1.2.1. Translation initiation | 20 |
| 1.2.2. Translational regulation through 5'leaders of mRNAs | 22 |
| 1.2.3. Viruses depend on the translation machinery of their host cell | 26 |
| 1.2.4. Defining translomes by ribosome profiling | 26 |

| | |
|----------------------------|-----------|
| 2. Aim of my thesis | 28 |
|----------------------------|-----------|

| | |
|--|-----------|
| 3. Results | 29 |
| 3.1. Ribosome profiling of EBV-infected cells | 29 |
| 3.2. Differences in host gene translation between B95-8 and M81-infected LCLs | 34 |
| 3.3. Viral mRNA translation in B95.8 and M81-infected LCLs | 36 |
| 3.3.1. Latent genes | 38 |
| 3.3.2. Lytic genes | 40 |
| 3.4. Ribosomes associate with non-coding RNAs | 45 |
| 3.5. Putative EBV proteins are associated with ribosomes | 48 |

| | |
|---|-----------|
| 3.6. Ribosome coverage of newly identified EBV-derived polyA⁺ RNA transcripts | 51 |
| 3.7. Novel small open reading frames are encoded in the EBV genome | 53 |
| 3.7.1. Genes encoded in the EBNA transcription unit are regulated by several uORFs | 56 |
| 3.7.2. Viral protein translation is regulated by upstream open reading frames | 60 |
| 3.8. Alternative translation initiation | 62 |
| <hr/> | |
| 4. Discussion | 64 |
| 4.1. Ribosome profiling reveals differences in translation of cellular genes in two different EBV strains | 64 |
| 4.2. Ribosome profiling allows for the analysis of annotated viral proteins and the identification of novel viral genes | 65 |
| 4.3. Ribosomes associate with non-coding RNAs | 66 |
| 4.4. Putative EBV proteins are associated with ribosomes | 68 |
| 4.5. The EBV genome encodes small and upstream open reading frames | 69 |
| 4.6. EBV infection remodels cellular gene expression processes | 73 |
| 4.7. Future studies | 74 |
| <hr/> | |
| 5. Materials | 75 |
| 5.1. Eukaryotic cell lines and culture medium | 75 |
| 5.2. Bacterial strains, culture medium and antibiotics | 75 |
| 5.3. Plasmids | 75 |
| 5.4. 5' transcript leader sequences (TLS) synthesized for reporter construct experiments | 76 |
| 5.5. Recombinant EBV clones | 78 |
| 5.6. Oligonucleotides | 78 |
| 5.7. Chemicals & Reagents | 81 |
| 5.8. Enzymes | 83 |
| 5.9. Antibodies | 83 |
| 5.10. Commercial kits | 84 |
| 5.11. Buffers | 84 |
| 5.12. Consumables | 85 |

| | |
|--|-----------|
| 5.13. Equipment | 86 |
| 5.14. Software | 87 |
| <hr/> | |
| 6. Methods | 88 |
| 6.1. Eukaryotic experimental system | 88 |
| 6.1.1. Culture conditions & maintenance | 88 |
| 6.1.2. Freezing and thawing cells | 88 |
| 6.1.3. Induction of producer cells lines into virus production | 88 |
| 6.1.4. Isolation of primary B lymphocytes | 89 |
| 6.1.5. Infection of primary B lymphocytes and generation of LCLs | 90 |
| 6.1.6. Dual luciferase reporter assays | 90 |
| 6.2. Prokaryotic experimental system | 90 |
| 6.2.1. Culture conditions | 90 |
| 6.2.2. Preparation of chemically competent bacteria | 91 |
| 6.2.3. E.coli transformation by heat-shock | 91 |
| 6.3. Working with DNA | 91 |
| 6.3.1. Cloning of overexpression plasmids | 91 |
| 6.3.2. Extraction of DNA with Phenol:Butanol | 92 |
| 6.3.3. Plasmid DNA isolation (Mini & Midi preparation) | 93 |
| 6.3.4. Quantitative real-time PCR (qRT-PCR) | 94 |
| 6.4. Working with RNA | 96 |
| 6.4.1. RNA extraction from cell pellets | 96 |
| 6.4.2. RNA extraction from sucrose solutions after polysome profiling | 96 |
| 6.4.3. Reverse transcription of RNA | 97 |
| 6.5. Immunofluorescence analysis | 97 |
| 6.6. Polysome profiling | 98 |
| 6.6.1. Preparation of linear sucrose gradients | 98 |
| 6.6.2. Preparation of cells for ultracentrifugation | 98 |
| 6.6.3. Gradient fractionation | 98 |
| 6.7. Ribosome profiling | 99 |
| 6.7.1. Generation and purification of ribosome protected fragments (RPFs) | 99 |
| 6.7.2. RNA size selection | 99 |
| 6.7.3. Dephosphorylation & linker ligation | 100 |

| | |
|---|------------|
| 6.7.4. Reverse transcription | 101 |
| 6.7.5. Circularization | 102 |
| 6.7.6. Depletion of rRNA | 102 |
| 6.7.7. PCR amplification and barcoding | 103 |
| 6.7.8. Sequencing | 104 |
| 6.8. Bioinformatics | 104 |
| 6.8.1. Read alignment & normalization | 104 |
| 6.8.2. Metagene analysis | 105 |
| 6.8.3. Identification of translation initiation sites | 105 |
| 6.8.4. Calculation of ratios | 105 |
| 6.8.5. UORF conservation analysis | 105 |
| 6.8.6. IPA analysis | 105 |
| 6.9. Statistical analysis | 106 |
| 6.10. Work distribution | 106 |
| <hr/> | |
| 7. References | 107 |
| <hr/> | |
| 8. List of Figures | 127 |
| 9. List of Tables | 129 |
| 10. Appendix | 130 |

Summary

The Epstein-Barr virus (EBV) is an oncogenic γ -herpesvirus that establishes a life-long infection in humans. It is ubiquitous in the population and is responsible for the development of multiple diseases including cancer. Analysis of the EBV DNA and RNA sequences has predicted that the viral genome encodes approximately 100 protein-coding genes. However, the existence of many putative proteins has not been confirmed by biochemical methods yet. Furthermore, the identification of EBV open reading frames (ORFs) is difficult as viral genes are encoded on both strands of the double-stranded DNA genome and often overlap. Moreover, EBV encodes different types of non-coding RNAs.

In order to identify the full scope of EBV's coding capacity, ribosome profiling of replicating and non-replicating EBV strains was performed. Ribosome profiling combines classical ribosome footprinting experiments with current deep sequencing technology to map translating ribosomes on mRNA at single nucleotide resolution. This approach confirmed the majority of previously identified ORFs and has enabled the identification of 28 novel small open reading frames and of 8 alternative translation initiation sites. 25 of the 28 small ORFs were localized in the 5'leaders of several mRNA transcripts and are classified as upstream open reading frames (uORFs). Several of these uORFs were found to repress the translation of the downstream encoded main ORF.

In summary, ribosome profiling of EBV-infected cells has allowed a comprehensive identification and annotation of the EBV ORFs and has revealed a novel mode of viral gene expression regulation at the translational level.

Zusammenfassung

Das Epstein-Barr-Virus (EBV) ist ein onkogenes Virus, das eine lebenslange Infektion im Menschen etabliert. Es ist allgegenwärtig in der menschlichen Population und trägt zur Entstehung von mehreren Krankheiten bei, unter anderem auch Krebs. Die Analyse der EBV DNA und RNA Sequenzen hat vorhergesagt, dass das virale Genome ungefähr 100 proteinkodierende Gene enthält. Die Existenz vieler dieser putativen Proteine ist jedoch noch nicht mit biochemischen Methoden bestätigt worden. Zudem ist die Identifizierung von EBVs offenen Leserahmen (OLRs) schwierig, da virale Gene auf dem doppelsträngigen DNA-Genom auf beiden Strängen und oft überlappend kodiert sind. Darüber hinaus enthält EBV verschiedene Arten von nicht-kodierenden RNAs.

Um die gesamte Kodierungskapazität des EBV Genoms zu identifizieren, wurden Ribosomenprofile von replizierenden und nicht replizierenden EBV Stämmen erstellt. Ribosomenprofile verbinden klassische "*ribosome footprinting*" Experimente mit aktuellen Hochdurchsatzsequenzierungstechnologien, um translatierende Ribosomen auf Nukleotidebene auf mRNA zu identifizieren. Diese Methode bestätigte die Mehrheit der zuvor beschriebenen viralen OLRs und hat die Identifizierung von 28 neuen, kleinen offenen Leserahmen sowie 8 alternativer Translationsstart ermöglicht. 25 der 28 OLRs waren auf mRNAs im 5'Bereich vom Hauptleserahmen unterschiedlicher Gene lokalisiert. Für einige dieser kleinen OLRs konnte eine repressive Funktion auf die Translation des nachgeschalteten Gens nachgewiesen werden.

Zusammenfassend haben Ribosomenprofile von EBV-infizierten Zellen eine umfangreiche Identifizierung und Annotation der EBV OLRs ermöglicht. Zudem wurde eine neue Regulationsweise der viralen Genexpression auf Translationsebene aufgezeigt.

Acknowledgments

I would first like to thank my thesis supervisor Prof. Dr. Dr. Delecluse for giving me the opportunity to work in his lab and for providing constant support and guidance.

I also wish to acknowledge my examination committee Prof. Dr. Martin Müller, Prof. Dr. Sven Diederichs and Dr. Grant Hansman, who have kindly agreed to critically assess my work.

Furthermore, I thank my colleagues who have actively contributed to this work: Dr. Olaf Klinke und Dr. Agnes Hotz-Wagenblatt.

I would also like to acknowledge the previous and present members of F100 who have been a big part of my daily life and have in some way or another influenced this work. I especially thank Angelika and Helmut for always finding a solution to every problem. A special thank you also to all the technicians who have always lent a helping hand.

A separate thank you to Ming-Han, who despite his enormous workload always found the time to answer my questions and offer help. And thank you to Janina, for the scientific discussions, the proofreading of my thesis and for just being a friend.

Finally, I would like to thank my husband, my parents, my sister and my friends for providing the support and encouragement needed.

List of Abbreviations

| | |
|-------|--|
| aa | amino acid |
| aTIS | alternative translation initiation |
| BAC | bacterial artificial chromosome |
| BART | BamHI fragment A rightward transcript |
| BCR | B cell receptor |
| cDNA | copy DNA |
| CDS | coding DNA sequence |
| CHX | cycloheximide |
| Cp | C promoter |
| dsDNA | double-stranded DNA |
| E | early lytic genes |
| EBER1 | Epstein-Barr virus-encoded small RNA 1 |
| EBER2 | Epstein-Barr virus-encoded small RNA 2 |
| eBL | endemic Burkitt's lymphoma |
| EBNA | Epstein-Barr nuclear antigen |
| EBV | Epstein-Barr Virus |
| eIF | eukaryotic initiation factor |
| eRF | eukaryotic release factor |
| FBS | fetal bovine serum |
| FLuc | <i>Firefly</i> luciferase |
| fwd | forward |
| GC | germinal center |
| Harr | harringtonine |
| HCMV | human cytomegalovirus |
| HCV | hepatitis C virus |
| HH-4 | human herpesvirus 4 |
| IE | immediate early lytic genes |
| IPA | Ingenuity Pathway Analysis |
| IR | internal direct repeat |
| IRES | internal ribosome entry site |
| KSHV | Kaposi's sarcoma herpesvirus |

| | |
|----------------|------------------------------------|
| L | late late lytic genes |
| LCL | lymphblastoid cell lines |
| LMP | latent membrane protein |
| miRNA | microRNA |
| mORF | main open reading frame |
| NGS | next generation sequencing |
| NMD | nonsense-mediated mRNA decay |
| NPC | nasopharyngeal carcinoma |
| nt | nucleotide |
| OLR | offener Leserahmen |
| ORF | open reading frame |
| OriLyt | origin of lytic replication |
| OriP | origin of plasmid replication |
| PBMC | peripheral blood mononuclear cells |
| RBP | RNA-binding protein |
| rev | reverse |
| RGD | arginine-glycine-aspartate |
| RLuc | <i>Renilla</i> luciferase |
| RPF | ribosome-protected footprint |
| sisRNA | stable intronic sequence RNA |
| snoRNA | small nucleolar RNA |
| TR | terminal repeat |
| U _L | unique long region |
| uORF | upstream open reading frame |
| U _S | unique short region |
| UTR | untranslated region |
| v-snoRNA | viral small nucleolar RNA |
| Wp | W promoter |

1. Introduction

1.1. EBV is an oncogenic herpesvirus

Herpesviruses are a diverse group of DNA viruses. More than 100 herpesviruses have been described to date. Epstein-Barr virus (EBV) is classified as a human herpesvirus and belongs to the *gammaherpesvirus* subfamily. Within the systematic nomenclature established for human herpesviruses EBV is known as human herpesvirus 4 (HHV-4).

EBV was the first virus described as an etiologic agent for human tumors. In 1958 the pathologist Denis Burkitt firstly described multifocal lymphomas in the jaws of young children born and living in Sub-Saharan Africa. The epidemiological features of this tumor led him to hypothesize an infectious agent as the driving force for the disease. This led to the discovery of EBV, which was isolated and described by Sir Michael Epstein, Yvonne Barr and Bert Achong¹.

EBV infects about 90-95% of the world's population^{2,3}. The infection is asymptomatic in most individuals⁴. It is known as a frequent causative agent of an infectious mononucleosis syndrome and is estimated to cause up to 2% of tumors worldwide⁵.

1.1.1. EBV particle structure

The structure of the EBV particle is shown schematically in figure 1.1. The outermost layer of the particle is a host cell-derived lipid envelope with viral glycoproteins embedded within it⁶.

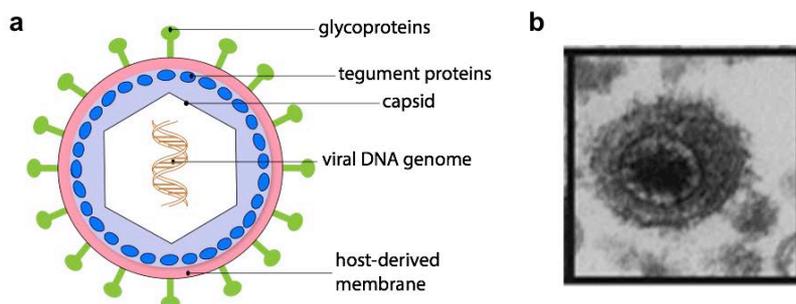


Figure 1.1. EBV particle structure. a. A schematic overview of the EBV particle is depicted. See text for details. b. Electron micrograph of an EBV particle. The electron-dense area contains the viral DNA genome. The EM image was taken from: Pavlova et al. 2013⁷.

Beneath the lipid layer of the virion is the tegument. The tegument is filled with RNA and protein complexes of viral (the tegument proteins) and cellular origin^{6,8}. These proteins modulate the infection process.

The tegument layer surrounds an icosahedral nucleocapsid composed of 162 capsomeres⁶. The capsid surrounds the linear, double-stranded DNA (dsDNA) genome. The genome is wrapped around a toroid-shaped protein core⁶.

1.1.2. EBV genome structure

The size of the dsDNA genome ranges from 172 to 186 kb⁶. As with other herpesviruses the genome of EBV can be divided into the unique short region (U_s) and the unique long region (U_L)⁹ (Figure 1.2.). Between these two domains lies a segment consisting of an internal direct repeat domain (IR1). Moreover, the U_L domain is subdivided into smaller segments (U2-U5) by further internal direct repeat domains (IRs)⁶. The genome is flanked by 0.5 kb long terminal repeats (TRs). The number of TRs varies between different viruses¹⁰. The variation comes from the random recombination of the TRs during genome circularization.

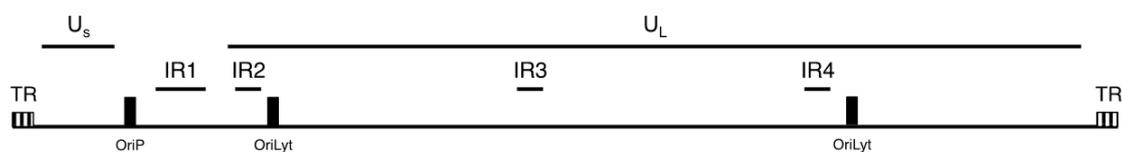


Figure 1.2. The linear U_s/U_L organization of the EBV genome. TR: terminal repeat. U_s: unique short region. U_L: unique long region. OriP: origin of plasmid replication. OriLyt: origin of lytic replication. IR1-4: internal direct repeats.

The origin of plasmid replication (oriP) maintains the episome in the infected cell. During virus particle production, also known as lytic replication, the viral DNA is copied from two origins of lytic replication (oriLyt) within the U_L domain.

The nomenclature for EBV open-reading frames (ORFs) is derived from the analysis and characterization of the B95-8 genome after BamHI restriction enzyme digestion (Fig. 1.3.). Each EBV transcript in the viral genome was given a name based on the location within one of the BamHI restriction fragments. For example: the gene BZLF1 is the acronym for a gene localized in the BamHI Z fragment with a leftward oriented open reading frame¹¹. Since it is the first open reading frame in the Z fragment it carries the number 1.

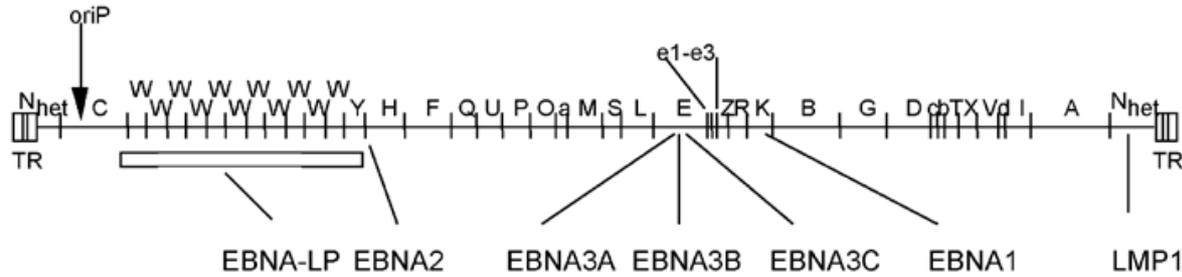


Figure 1.3. Organization of the viral genome according to BamHI restriction enzyme digestion. The BamHI restriction fragments are shown in the B95-8 genome. Restriction fragments are named according to their size with A being the largest fragment. The diagram indicates the location of viral latent proteins. In addition to the conventional BamHI fragment-derived nomenclature these proteins also have an alternative nomenclature as shown here. TR: terminal repeat. OriP: Origin of plasmid replication. Image was taken from: Young et al. 2003¹².

EBV-encoded genes are divided into two broad categories. The first category comprises the lytic genes which are responsible for virus particle production. The lytic genes are further subdivided into immediate early (IE) genes that initiate the lytic cascade, the early (E) genes that replicate the viral genome and the late (L) genes that constitute and assemble the virions.

The second category comprises the latent genes. Latent genes are expressed in infected cells that do not replicate. These cells are termed latently infected cells. The latent genes mediate the establishment of a persistent infection in the host. The number of latent genes expressed in latently infected cells varies and depends on the infected cell type. This leads to a variety of virus-host interactions that have different consequences for the infected cell. The latent genes are divided into two sub-types: (i) the Epstein-Barr nuclear antigens (EBNA) and (ii) the latent membrane proteins (LMPs). The EBNA family of proteins consists of: EBNA-LP, EBNA1, EBNA2, EBNA3A, EBNA3B and EBNA3C. These proteins maintain the viral genome in infected cells and activate viral and cellular transcription. The LMP proteins comprise: LMP1, LMP2A and LMP2B. LMP1 is a transmembrane protein mimicking constitutive CD40 signaling¹³. LMP2A confers constitutive B cell receptor (BCR) signaling¹⁴. LMP2B lacks the first coding exon of LMP2A and regulates LMP2A activity.

The expression of the complete set of the latent genes leads to active proliferation in infected B cells. This state is known as full latency.

Apart from protein-coding genes EBV also encodes several non-coding RNAs. The most abundant are the two Epstein-Barr virus encoded small RNAs (EBERs) EBER1 and EBER2⁶. They are localized in the nucleus¹⁵. Both EBER RNAs are reported to

interact with the La antigen¹⁵ and EBV1 has been found to interact with the ribosomal protein L22^{16,17}. Furthermore, EBV2 has recently been described to localize to the TRs of the viral genome and thereby modulate LMP2 expression¹⁸.

EBV further encodes 44 mature micro-RNAs (miRNAs)^{19,20} (reviewed in Klinke et al. 2014²¹). These are 19 to 25 nucleotides long non-coding RNAs involved in regulation of protein expression in infected cells. The miRNAs are grouped into two separate clusters within the EBV genome (Fig. 1.4.). The first miRNA cluster is located around the BamHI fragment H rightward open reading frame 1 (BHRF1) ORF and it encodes four miRNAs (Fig. 1.4.). The second cluster in the BamHI fragment A rightward transcript (BART) encodes all remaining miRNAs arranged in two sub-clusters (cluster I and cluster II) and one separately encoded miRNA (miR-BART2)²² (Fig. 1.4). BHRF1 miRNAs are important factors for EBV-mediated transformation of B-cells²³⁻²⁵. Some BART miRNAs are reported to limit lytic replication in infected cells^{26,27}.

EBV has also been reported to encode a small nucleolar RNA (v-snoRNA1)²⁸. Recent computational analysis of small RNA sequencing data has identified a stable intronic sequence RNA (sisRNA) encoded within the W repeats of the EBV genome²⁹.

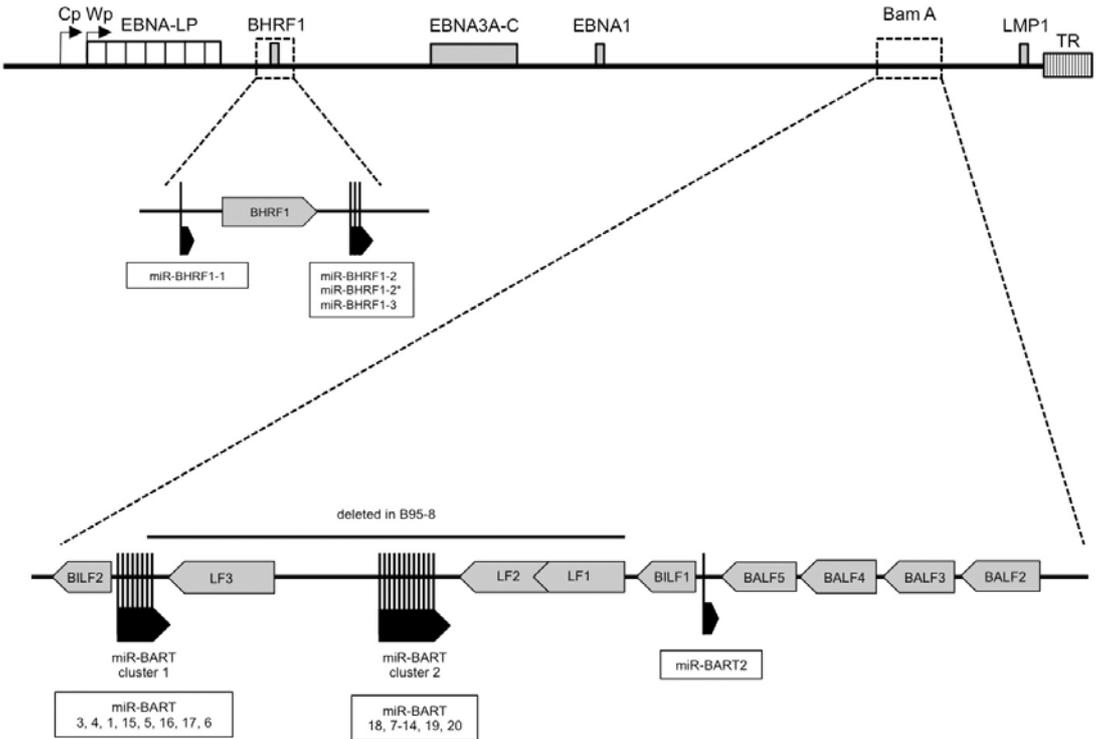


Figure 1.4. EBV-encoded miRNAs. Schematic illustration of the genomic regions encoding EBV miRNAs. Four miRNAs are encoded around the BHRF1 gene. One miRNA is processed from the 5' leader of the transcript while the other two are encoded in the 3'untranslated region of the gene. The BamHI A genomic region encodes the majority of the viral miRNAs that are derived from the introns of the highly spliced BamHI A rightward transcripts (BARTs). For the BART miRNAs only one of the miRNAs of the pre-miRNA duplex are listed but both strands of the processed pre-miRNA transcript can be loaded into the RNA-induced silencing complex.

1.1.3. The EBV transcriptome

Viruses in general are known for the efficient utilization of the coding capacity dictated by their genome sizes. Herpesviruses exploit this principle to the fullest and have evolved to use the limited coding space optimally. RNA sequencing experiments have highlighted the complex transcriptome of EBV^{30,31}. Genes are transcribed from both DNA strands and frequently overlap³². Non-canonical transcription initiation is often used among lytic genes and many genes possess alternatively spliced isoforms³³. Furthermore, a recent publication has identified close to 300 new EBV transcripts that are produced upon lytic replication initiation³³. These transcripts are in addition to the already annotated EBV transcripts.

1.1.4. The viral glycoproteins determine cell tropism

The viral lipid envelope harbours multiple glycoproteins which play a role in virus attachment, entry and egress. The gp350/220 glycoproteins mediate virus attachment to particular target cells. Gp110 (gB), gp85 (gH) and gp25 (gL) build up the core fusion complex. The gp42 can interact with gH/gL. It further interacts with human leukocyte antigen (HLA) class II. BMRF2 is a glycoprotein that contains an arginine-glycine-aspartate (RGD) motif which is a motif used by extracellular matrix proteins to bind to integrins^{34,35}. Integrins are transmembrane proteins expressed on cells that mediate the cell-cell or cell-extracellular matrix contact³⁵. Different combinations of the viral glycoproteins are necessary to infect different cell types.

The cell tropism of EBV is broad. EBV is reported to infect human B cells, epithelial cells, NK cells and T cells³⁶⁻³⁸.

EBV readily infects B cells *in vitro* and B cell infection by EBV is best studied compared to the other described target cells⁶. Infection of the host's B cells occurs through the interaction of viral gp350/220 receptors with the cellular CD21³⁹⁻⁴¹ or CD35 proteins⁴². Both proteins are complement receptors on B cells. The attachment of the virus to the target cell via CD21 leads to endocytosis of the particle.

Additionally the viral gp42/gH/gL trimer interacts with cellular MHC class II molecules and integrins to mediate fusion⁴³.

Epithelial cells are difficult to infect *in vitro*. Infection experiments with primary epithelial cells report less than 0.1% of cells being infected with EBV^{44,45}. The receptors used for viral B lymphocyte infection are only expressed at low levels on epithelial cells^{46,47}. *In vitro* studies have shed some light on the infection process by using EBV mutants that are devoid of specific glycoproteins. In epithelial cell infection gp350 and gp42 are not needed. The gH/gL complex on the other hand is essential⁴⁸. Treatment of viruses with gH/gL neutralizing antibodies reduces virus binding to epithelial cells. Furthermore, a recent publication by Wang et al. has identified neuropilin 1 (NRP1), a co-receptor for growth factors, as an interaction partner of gp110 and mediator of viral entry into epithelial cells⁴⁹. Additionally, BMRF2 is important for epithelial cell infection. The protein interacts with integrins and mediates the infection on the basolateral side of polarized epithelial cells³⁴.

A significantly higher infection efficiency can be achieved by using transfer infection. Transfer infection involves EBV bound to naïve B cells that can transfer the virus to epithelial cells⁵⁰. The crosslinking of the CD21 protein on the B cell surface leads to activation of adhesion molecules on the cell⁴⁵. This enables the formation of a virological synapse between the B cell and epithelial cell that facilitates infection⁴⁵. The percentage of EBV⁺ epithelial cells after transfer infection is reported to reach 20% of the culture.

Entry of the virion proceeds in epithelial cells at the cell surface. No prior endocytosis is involved.

Once infection is successful, 20-40% of the infected epithelial cells will enter lytic replication within days following infection⁵¹⁻⁵³. Long-term culture of EBV-infected epithelial cells has so far only been demonstrated in telomerase reverse transcriptase-immortalized epithelial cells that carry an additional deletion of the cyclin-dependent kinase inhibitor p16 or overexpress cyclin D1^{54,55}. Tsang et al. have shown that EBV-infected epithelial cells arrest growth after some time due to the activity of the cyclin-dependent kinase inhibitor⁵⁴. By deregulation of one of the cell cycle regulators the infected cells will continue proliferating.

1.1.5. Lytic replication

Lytic replication is necessary to generate progeny virions that spread the infection to other individuals and replenish the pool of infected cells.

Lytic replication begins with the expression of two immediate early genes; the lytic transactivators: BZLF1 (also known as Z or Zta) and BRLF1 (also known as R or Rta)^{56,57}. These two proteins bind to lytic promoters in the viral genome and activate the lytic replication cycle⁶.

The viral episome is replicated by the cellular DNA replication machinery during latency. Once lytic replication is initiated, the viral genome is replicated by the early lytic proteins. These include: BALF5, the viral DNA polymerase, BSLF1, the viral primase and BMRF1, the polymerase processivity factor. The genome is replicated by the rolling circle method as a large concatemer that is then cleaved within the TR regions to generate individual EBV genomes that can be packaged into nucleocapsids^{58,59}.

Genome replication is also needed for the production of the late lytic genes which are transcribed from newly replicated DNA⁶⁰. For example, the glycoprotein gp350 is not expressed in lytically replicating cells that have been treated with phosphonoacetic acid which inhibits the viral DNA polymerase and prevents viral genome replication⁶¹. Nucleocapsid assembly occurs in the nucleus. From there the nucleocapsid passes through several cellular membranes with multiple envelopment and de-envelopment processes taking place. Viral proteins such as BFRF1 direct this virion maturation process⁶². The mature virion is released from Golgi-derived vesicles at the plasma membrane⁶³.

Analysis of fractionated tonsils shows that EBV replicates in differentiated plasma cells *in vivo*⁶⁴. This is supported by histological analyses^{65,66}.

Apart from terminal differentiation, stress can also trigger lytic reactivation in infected cells. The link between stress-inducing agents and lytic replication has mainly been studied *in vitro*. Agents triggering lytic replication include chemotherapeutics and hypoxia⁶⁷. Stimuli that engage certain pathways such as TGF- β signaling, calcium-associated pathways or B cell receptor signaling also lead to reactivation (discussed in: Amon et al. 2005 and Tsurumi et al. 2005^{68,69}). Many lytic replication inducers can also be found in our daily environment. Some lytic cycle inducing agents are for example, phorbol esters, n-butyrate and nitrosamines present in preserved or grilled food.

1.1.6. B cell transformation *in vitro*

Naïve B cells can readily be infected with EBV *in vitro*. As the virus enters the cell, the crosslinking of CD21 triggers intracellular signaling that causes the cell to transition from G0 into the G1 phase of the cell cycle⁷⁰. Upon infection a “pre-latent” state is induced within the affected B cells. Before the viral DNA is transcribed, viral transcripts can already be detected which are probably delivered by the virions^{71,72}. This mRNA cargo allows immediate expression of viral proteins^{72,73}. Here, expression of a subset of lytic genes in parallel to latent genes occurs^{74,75}. The earliest latent transcripts detected are EBNA2 and EBNA-LP^{76,77}. The lytic viral genes transcribed are BHRF1, an anti-apoptotic bcl-2-like protein⁷⁸, the viral transactivators BZLF1 and BRLF1^{79,80}, the viral immune evasion proteins (BALF1, BCRF1, BNLF2a) and BMRF1, the viral processivity factor⁷². This early event of lytic replication is not well understood and the relevance of it still remains to be elucidated. Once the pre-latent stage is overcome and the viral DNA has entered the nucleus the cells progress to latency.

The first latency stage is latency IIb. This stage directs the initial round of cell division after infection⁸¹. Latency IIb is characterized by the additional expression of EBNA1, EBNA3A, EBNA3B, EBNA3C, the EBER transcripts and viral miRNAs. The infected cell then transits to the latency III stage also known as the growth program⁸². In latency III the complete set of latent genes is produced^{6,76,83,84}. These include the LMP1, LMP2A and LMP2B proteins. *In vitro* infected cells stay in latency III and continue proliferating indefinitely as immortalized lymphoblastoid cell lines (LCLs).

In vivo the infection process of EBV is more complex and is discussed in the following chapter.

1.1.7. EBV infection of the host

EBV spreads and enters the host via saliva⁸⁵. The virus crosses the epithelial barrier in the region of the Waldeyer tonsillar ring in the nasopharynx, probably by transcytosis⁸⁶. Another model proposes that the virus infects the epithelium of the Waldeyer tonsillar ring in the oropharynx and is then amplified through lytic replication. The second model is somewhat disputed as there is evidence that epithelial cells are resistant to EBV infection from the mucosal side of the epithelium³⁴. Once the viral particles have passed the epithelial barrier, they go on to infect naïve B cells situated in the underlying lymphoid tissues.

Following infection the B cells progress through the same latency stages as described in section 1.1.6. Primary EBV infection is countered by a strong immune response in the host. Therefore, infected cells do not stay in latency III but progressively downregulate the viral gene expression in order to avoid destruction by the immune system⁸⁷⁻⁸⁹.

Latency III *in vivo* is active for about three days and closely mimicks the initial burst of proliferation of antigen-activated B cells in germinal centers (GC)^{90,91}. Once infection is successfully established viral gene expression is sequentially silenced until the infected B cell is virtually invisible to the host's immune system. For that to occur the infected cells migrate to the follicle of the lymph node and initiate a GC reaction.

In the GC, the infected cells switch to the default transcription program also termed latency IIa. Latency IIa rescues the infected cells from the GC and allows them to differentiate into memory B cells. Infected B cells that have entered the GC phenotypically resemble true GC B cells. They express the characteristic surface markers (CD10⁺, CD77⁺, CD38⁺), chemokine receptors (CXCR4, CXCR5 and no CCR7) as well as the GC specific proteins: AID and bcl-6⁹². Within the GC only a subset of latent proteins are expressed: EBNA1, LMP1 and LMP2^{6,93}. EBNA1 is essential to ensure proper segregation of the episome to daughter cells as well as to recruit the cellular replication machinery to the viral genome during S phase⁹⁴. LMP1 and LMP2 provide the necessary survival signals to enable the transition from GC to memory B cell.

Following memory B cell differentiation the infected cells switch to latency 0. Latency 0 is characterized by complete absence of viral protein expression^{95,96}. This protects the virus from immune surveillance and allows long-term viral persistence.

Occasionally, the infected resting memory B cell will divide and the latency I program is initiated that expresses EBNA1 only.

Genes expressed during the different latency stages are summarized in figure 1.5.

| pre-latency | latency IIb | latency III | latency IIa | latency I | latency 0 |
|--|--|---|--|-------------------------------|--|
| viral gene expression: | | | | | |
| EBERs BHRF1 miRNAs BART miRNAs EBNA-LP EBNA2 BHRF1 BZLF1 BRLF1 BALF1 BCRF1 BNLF2a BMRF1 | EBERs BHRF1 miRNAs BART miRNAs EBNA-LP EBNA2 EBNA3s EBNA1 BHRF1 | EBERs BHRF1 miRNAs BART miRNAs EBNA-LP EBNA2 EBNA3s EBNA1 BHRF1 LMP1 LMP2s | EBERs BART miRNAs EBNA1 LMP1 LMP2s | EBERs BART miRNAs EBNA1 | EBERs BART miRNAs |
| | | | | | <div style="border: 1px solid black; padding: 5px; width: fit-content;"> non-coding RNAs latent genes lytic genes </div> |

Figure 1.5. EBV latency states during EBV infection. The figure summarizes the different latency states observed during EBV infection. The viral genes expressed during each latency stage are listed below and classified as non-coding RNAs (EBERs and miRNAs), latent genes or lytic genes.

1.1.8. EBV strain heterogeneity

The first EBV strain to be sequenced was B95-8¹¹. It was at that time the only strain in culture that produced enough virus to allow for restriction enzyme analysis and subsequent cloning of these fragments for sequencing. Originally, B95-8 was derived from a spontaneous human LCL arising from an IM patient⁹⁷. It was used to transform lymphocytes of the cotton top marmoset and the infected cells produced EBV spontaneously *in vitro*. This virus was then used to transform B cells and the resulting LCLs could be induced into lytic replication by treatment with phorbol esters and n-butyrate. Samples used for sequencing were produced by this chemical induction method⁹⁸. The B95-8 strain has for a long time been used as the prototypic laboratory strain despite a 13.6 kb deletion within its genome in comparison to other EBV strains⁹⁹. This deletion alters the expression of some genes¹⁰⁰. Moreover, it is missing a large part of the EBV viral miRNAs¹⁰¹ that have been described to influence cell growth, cell cycling and lytic replication.

The advent of cheaper whole-genome DNA sequencing has now made it possible to study EBV isolates and geographic variation on a larger scale. Palser et al. have sequenced and published 71 new EBV strains isolated from healthy EBV carriers and from multiple primary tumors taken from several distinct geographic regions¹⁰². They described several polymorphisms across viral open reading frames. The strongest polymorphisms were located within latent genes. Yet, their study makes no conclusions on possible links between the studied genotypes to observed phenotypes.

Most EBV genes have a sequence identity of about 90 - 95% between different isolates. EBNA2 on the other hand only has a sequence identity of 70% at nucleotide

level and 54% at protein level. This polymorphism was used early on to classify EBV isolates into type 1 or type 2 based on the EBNA2 gene sequence⁶. It was the first classification of EBV isolates that correlated a viral gene sequence to a phenotypic trait in infected cells. Type 1 EBV transforms B cells readily into LCLs, type 2 EBV on the other hand is less efficient¹⁰³. EBV type 1 is also the prevalent strain in the world. In sub-Saharan Africa, type 2 is equally abundant and mixed infections occur frequently. Type 2 is also often found in individuals suffering from some kind of immune-incompetence (due to immunosuppression in transplant recipients or immune ablation in HIV⁺ patients)¹⁰⁴⁻¹⁰⁸.

A recent study linking viral polymorphisms and the phenotypic behaviour of the virus stems from the characterization of the M81 EBV strain, a virus isolate retrieved from a nasopharyngeal carcinoma (NPC) patient in China¹⁰⁹.

The M81 isolate was derived from a NPC biopsy of a Chinese patient¹¹⁰. It was recently cloned into a bacterial artificial chromosome (BAC) and characterized in greater detail¹⁰⁹. The virus exhibits reduced B cell tropism compared to B95-8 but infects epithelial cells at higher efficiency¹⁰⁹. Infection of B cells with the M81 strain leads to the initiation of spontaneous lytic replication in cell culture in 3-6% of infected LCLs at around 3 weeks post infection^{109,111}. Infectious virus is produced even though most is found to be bound to neighboring LCLs. In comparison, the B95-8 strain does not mount lytic replication *in vitro*.

Genetically M81 and B95-8 are quite distant. Most viral genes harbor several polymorphisms. Tsai et al. used B95-8 and M81 to construct hybrid viruses by exchanging the BZLF1 open reading frame and promoter from M81 with the one from B95-8. BZLF1 is a viral transactivator that upon expression initiates lytic replication in EBV-infected cells. The study showed that the BZLF1 gene was one of the genes contributing to the described phenotype of M81.

1.1.9. EBV-associated malignancies

EBV infects humans all over the world. Nevertheless, some EBV-associated diseases are preferentially encountered in specific geographic regions. Furthermore, these diseases range from benign to malignant tumors. Especially in immunocompromised individuals the virus can be the cause of severe disease. Some of the EBV-associated cancers include endemic Burkitt's lymphoma (eBL), Hodgkin's

lymphoma, NK/T-cell lymphoma³⁶, nasopharyngeal carcinoma (NPC) and gastric carcinoma.

The genetic background of the host, co-infecting pathogens, local diet and life-style habits such as smoking are reported to influence the development of certain EBV-associated diseases¹¹². Moreover, several of EBV-associated cancers are accompanied by increased titers to lytic antigens. This indicates that ongoing and badly controlled virus replication could cause disease development. Support for lytic replication as a risk factor comes from a recent publication reporting that viral lytic genes are frequently co-expressed with cellular cancer-associated pathways¹¹³. Furthermore, there is a possibility that these diseases might be linked to particular EBV strains.

It is outside of the scope of this thesis to describe all EBV-associated malignancies. Therefore, three EBV-associated diseases will be discussed briefly: (i) infectious mononucleosis, (ii) endemic Burkitt's lymphoma and (iii) nasopharyngeal carcinoma.

Infectious mononucleosis

EBV can cause an infectious mononucleosis syndrome (IM). IM is characterized by a benign expansion of B lymphocytes in the presence of high fever as well as a sore throat and fatigue. The disease is usually self-limiting and IM patients recover completely. In rare cases, it can affect the liver causing hepatitis and in the worst case liver failure. EBV-associated IM is often caused by a delayed primary EBV infection¹¹⁴. The age of primary EBV infection has shifted into adolescence in the developed countries¹¹⁵⁻¹¹⁹. While by adulthood almost all individuals are seropositive for EBV, the prevalence in young children is much lower, ranging from 20 to 80% depending on age group, geographic region and ethnicity^{115-117,120,121}. This late encounter with EBV puts affected individuals at greater risk for IM during primary infection¹²². The reasons for this are not understood. The disease itself mainly stems from the inflammatory response triggered by the infection and once symptoms occur the infection is already residing¹²³.

Endemic Burkitt's lymphoma

Endemic Burkitt's lymphoma (eBL) was the first human cancer linked to a viral infection. EBV is detected in approximately 95% of eBL cells¹²⁴. The disease is geographically restricted to the malaria belt of the world implying that there is a

correlation between malaria prevalence and the development of this disease^{6,124}. Studies show that patients treated for malaria have high EBV blood titers¹²⁵. The hallmark of eBL is deregulated expression of the c-myc oncogene due to a translocation with immunoglobulin genes¹²⁶⁻¹²⁸. This translocation has been shown to be sufficient in causing Burkitt's lymphoma in mice¹²⁹. EBV is clonal in eBL hinting that the virus was already present before the translocation occurred and the lymphoma developed¹³⁰.

Nasopharyngeal carcinoma

The WHO has classified NPC into 3 subtypes based on their differentiation state. EBV can be found in all three subtypes¹³¹ but has a 100% association with the undifferentiated type¹³². The tumor has a high geographic restriction mainly occurring in Southeast Asia, Northern Africa and Alaska⁶.

A hallmark of NPC are the increased antibody titers to EBV¹³³⁻¹³⁵. They increase as the disease progresses and are mainly directed against lytic antigens^{136,137}.

The M81 strain described in the previous chapter exhibits increased epitheliotropism and propensity to initiate lytic replication in infected cells. Both features are prerequisites for disease development. This supports the hypothesis that distinct EBV strains cause specific EBV-associated diseases¹⁰⁹. Moreover, other EBV strains isolated from NPC patients cluster together with M81 in phylogenetic analyses.

The intrinsic characteristics of the NPC-derived viral strains together with environmental factors such as diet, might promote the development of the disease. Several studies have reported a link between the consumption of preserved food, containing high levels of phorbol esters and nitrosamines, and the development of NPC¹³⁸⁻¹⁴⁰. Campaigns against preserved foods have lead to a drop of NPC incidence rates in these populations¹⁴¹. Different Chinese herbs used in traditional medicine¹⁴², work-related exposure to toxic dust as well as fumes and smoking^{143,144} have also been identified as risk factors for NPC. Furthermore, a genetic component is involved as NPC incidence is seen clustered in families^{145,146}.

1.2. Translation in mammalian cells

Gene expression in mammalian cells is controlled on multiple levels starting in the nucleus with the modulation of transcription. Protein abundance is further regulated by altering transcript levels through degradation pathways and sequestering mRNAs in specialized storage granules (Fig. 1.6). Finally, protein levels are directly controlled by the translation rate of the mRNA and protein turnover. Translation is a dynamic process and studying the translome, the RNAs that are translated at a given timepoint in a cell, enables the distinction between transcripts that are physiologically important and the ones that constitute transcriptional noise.

In the next paragraph, I will describe the molecular and regulatory mechanisms that enable protein expression in mammalian cells.

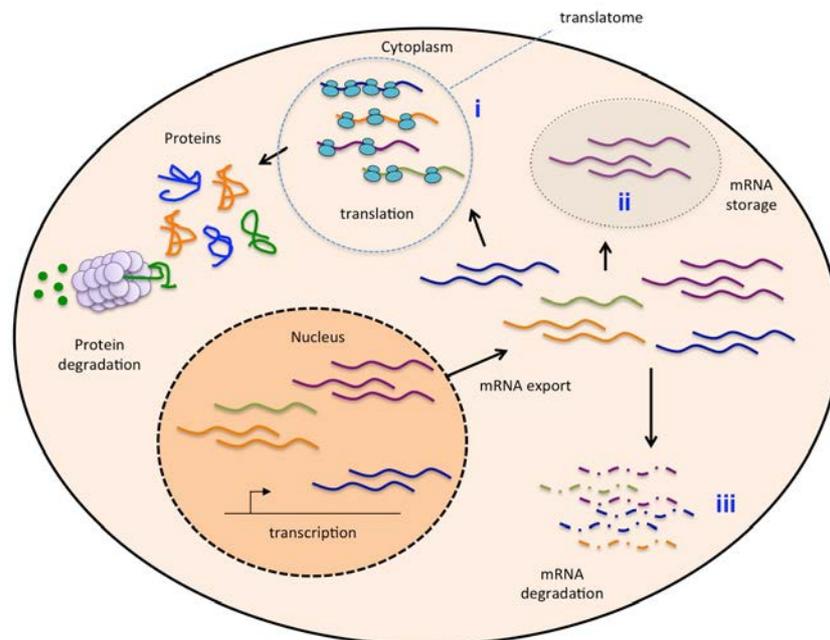


Figure 1.6. Post-transcriptional gene expression control. An overview over the main processes that lead to protein expression. Upon transcription in the nucleus the mRNA is exported into the cytoplasm. Here the mRNA can be destined to one of three pathways: (i) the mRNA associates with the translation machinery and leads to the production of protein, (ii) the mRNA is stored in specialized RNA granules or (iii) the mRNA is degraded. Each step contributes to the regulation of protein abundance in a cell. Protein abundance is further controlled by protein turnover rate.

1.2.1. Translation initiation

Translation can be divided into three distinct parts: (i) initiation, (ii) elongation and (iii) termination¹⁴⁷. All three stages are regulated although in general, translation is primarily regulated at the initiation stage as this is said to be the rate-limiting step¹⁴⁸.

(i) Initiation:

In eukaryotic cells the vast majority of translated mRNA is cap-dependent, meaning that the production of functional protein requires a complex that assembles on the 7-methylguanosine cap at the 5' end of the transcript (Fig. 1.7)^{147,148}. This complex is termed eukaryotic initiation factor 4F (eIF4F) and consists of the cap-binding protein eIF4E, the scaffold protein eIF4G and the RNA helicase eIF4A. It is responsible for regulating ribosome attachment and protein synthesis. At initiation, the 40S ribosomal subunit interacts with eIF3, which is bound to eIF4G. Here the 40S subunit forms a ternary complex with eIF2, the initiator transfer RNA^{met} and GTP, which in turn associates with eIF1, eIF1A and eIF3 to form the 43S preinitiation complex that scans the 5' leader region of the mRNA until it encounters a start codon.

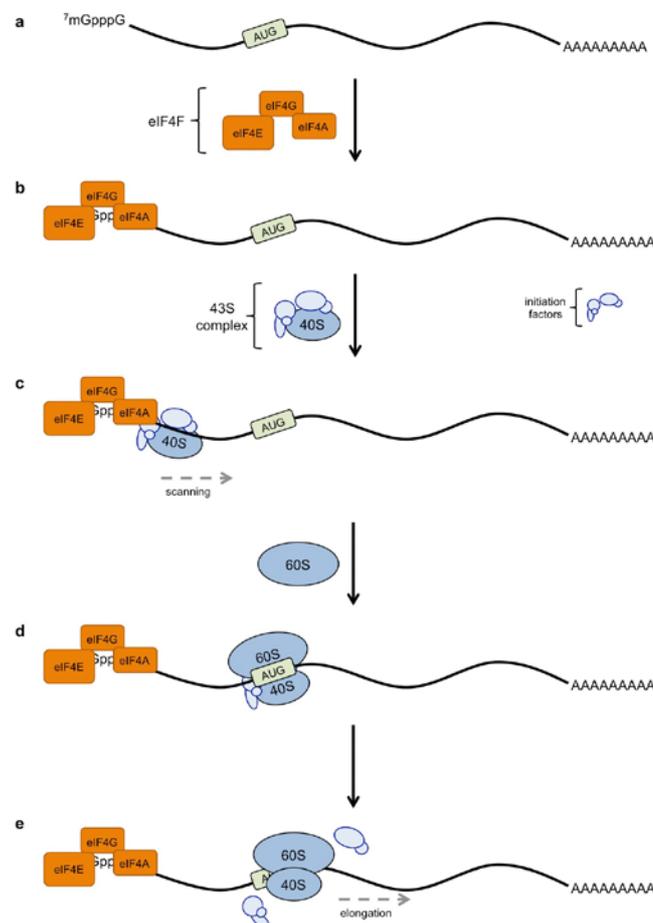


Figure 1.7. Cap-dependent translation initiation. A schematic of canonical translation initiation in mammalian cells is shown. a. A simplified view of a mRNA transcript is depicted. The 5'cap ($7mGpppG$) and poly(A)-tail are included. The AUG start codon of the open reading frame is boxed in green. b. The translation of a mRNA begins with the assembly of three of the eukaryotic translation initiation factors (eIFs) on the mRNA cap: the eIF4E (cap-binding protein), eIF4G (scaffold protein) and eIF4A (RNA helicase). These three proteins constitute the eIF4F complex. c. Assembly of the eIF4F complex on the cap leads to the binding of the small ribosomal subunit (40S). The 40S subunit is complexed with further initiation factors that make up the 43S pre-initiation complex. This complex scans the mRNA until it encounters a start codon in the right sequence context fit for initiation. d. Once a start codon is encountered, the 60S ribosomal subunit joins and with the 40S subunit the 80S translation-competent ribosome is formed. This enables the translation of the open reading frame. e. Translation initiation factors dissociate from the 80S ribosome and are replaced by elongation factors (not shown here).

Subsequently, eIF5 is recruited and triggers hydrolysis of the eIF2 α -bound GTP, thereby facilitating release of the initiation factors from 40S and allowing the association of the 60S ribosomal subunit to form the elongation-competent 80S ribosome. The initiation codon AUG in the context of a strong Kozak consensus sequence is the most efficient site for 80S ribosome assembly¹⁴⁹. The Kozak consensus sequence is a particular nucleotide sequence surrounding the start codon that determines the efficiency of start codon recognition by the ribosome¹⁴⁹⁻¹⁵¹ (Fig.1.8). It was identified by mutagenesis studies and subsequent sequence analysis of 699 vertebrate mRNAs^{149,152}. The adenosine base of the start codon AUG is numbered as +1. If one of the crucial bases does not match, the Kozak sequence is classified as intermediate. If there is no match on both of the consensus bases, then the Kozak sequence is classified as weak.

If the AUG is in a suboptimal Kozak sequence context, the scanning 43S preinitiation complex can skip it. This process is called leaky scanning^{151,153}. Once the ribosome begins translation of the ORF, the initiation factors dissociate and are replaced by eukaryotic elongation factors (eEFs).

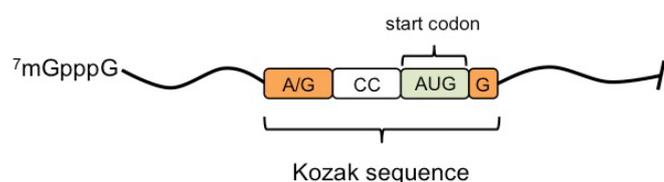


Figure 1.8. The Kozak consensus sequence. A schematic depiction of the Kozak consensus sequence is shown. The AUG start codon is highlighted in green. The crucial bases determining the Kozak sequence strength are shown in orange. The adenine base (A) of the AUG start codon is referred to as base position +1. The base in position -3 is either of the purine bases adenine (A) or guanine (G). The base in position +4 is ideally a G.

(ii) elongation:

Translation is mainly regulated at the initiation step¹⁴⁷ but there have also been reports showing that translational control and fine tuning can occur at the step of elongation¹⁵⁴⁻¹⁵⁶.

During elongation eEF1A delivers aminoacyl tRNAs to the ribosome's A site for peptide bond formation. eEF2 mediates the translocation process following the peptide bond formation. Elongation continues until a stop codon is encountered.

(iii) termination:

Stop codons are recognized by the eukaryotic release factor eRF1 which then induces hydrolysis of the tRNA bound to the new polypeptide¹⁵⁷. eRF3 removes eRF1 and the ribosome dissociates into the 40S and 60S subunits.

Apart from cap-dependent translation, a cap-independent internal ribosome entry site (IRES) can be used for initiation. Here, the 40S subunit can directly bind to the mRNA close to the start codon by recognizing a specific RNA secondary structure. This was first described in viral protein synthesis of picornaviruses^{158,159} but has also been identified in other viruses such as Hepatitis C Virus (HCV)¹⁶⁰ and EBV¹⁶¹. IRES-mediated translation initiation is also used by cellular genes such as BiP¹⁶², VEGF¹⁶³ and c-myc^{164,165}. Currently, 10% of mammalian mRNAs are predicted to contain an IRES^{166,167}. These elements are particularly important during mitosis, cell differentiation or cell stress when cap-dependent translation is compromised.

1.2.2. Translation regulation through 5'leaders of mRNAs

Translation is a highly energy-consuming process and thus needs to be carefully regulated in cells. Furthermore, control of protein production on the level of translation gives the cell a lot of flexibility allowing it to respond to signaling cues almost immediately. Translation is regulated by controlling the abundance of initiation factors and the availability of ribosomes^{147,148}.

Protein expression can further be regulated through several mechanisms guided by *cis*-acting elements encoded on the mRNA itself. These regulatory elements have been described in the 5'leaders and 3'untranslated regions (UTRs) of mRNA transcripts. I will focus here only on the regulation mediated by the 5'leader.

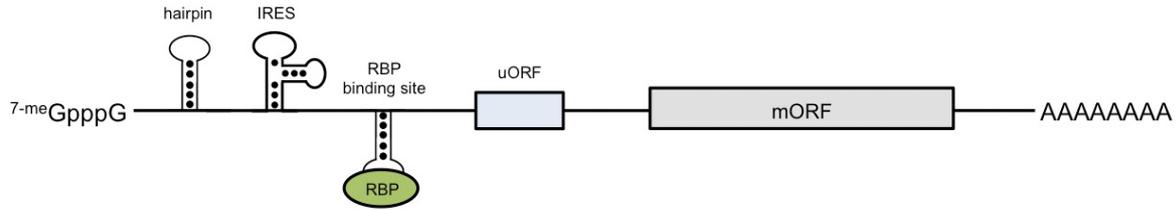


Figure 1.9. *Cis*-acting regulatory elements in the 5' leaders of mammalian mRNAs. A schematic depiction of the regulatory elements identified in the 5' leaders of mammalian mRNAs. The 5' leader can encode several elements that affect the downstream translation of the main open reading frame (mORF). From left to right: stable hairpin structures can stall the 43S preinitiation complex from scanning the mRNA. Internal ribosome entry sites (IRES) direct cap-independent translation and preclude scanning of the mRNA. RNA-binding proteins (RBP) bind to sequence-dependent sites on the mRNA transcripts and regulate mRNA stability and translation rates. Upstream open reading frames (uORFs) are small open reading frames in the 5' leaders that modulate the translation of the mORF. 7-meGpppG: 7-methylguanosine cap.

Secondary structures

Within the 5' leader, particular nucleotide compositions can favour the formation of stable secondary structures such as hairpins (Fig. 1.9)¹⁶⁸. Especially GC-rich sequences form complex structures that can inhibit or stall 43S preinitiation movement within the transcript leader^{150,169-171}.

RNA-binding proteins

RNA-binding proteins (RBPs) interact with specific RNA sequences present in the 5' leaders and 3'UTRs of mRNA transcripts. Their binding can directly affect the stability and translation rates of mRNAs (Fig. 1.9)¹⁷²⁻¹⁷⁴.

Upstream open reading frames

Upstream open reading frames (uORFs) are short open reading frames localized within the 5' leaders of mRNA transcripts (Fig. 1.9)¹⁷⁵. They are common in mRNAs coding for proteins that need to be tightly controlled in their expression level. These include oncogenes, cell differentiation factors, growth factors and cell cycle proteins^{176,177}. The short ORFs interfere with the translation of the downstream encoded main ORF (mORF)^{178,179}. uORFs regulate translation by either of the following mechanisms (Fig. 1.10):

1. Upon translation of the uORF the ribosome dissociates and does not reach the mORF (Fig. 1.10b)¹⁸⁰⁻¹⁸².
2. The uORF encodes a functional peptide and upon translation the peptide inhibits progression of the ribosome (Fig. 1.10c)¹⁸³⁻¹⁸⁵.

3. In a subset of uORF-encoding mRNAs the stop codon of the uORF is recognized as a premature stop codon that triggers nonsense-mediated mRNA decay and leads to degradation of the transcript (Fig. 1.10d)^{186,187}.

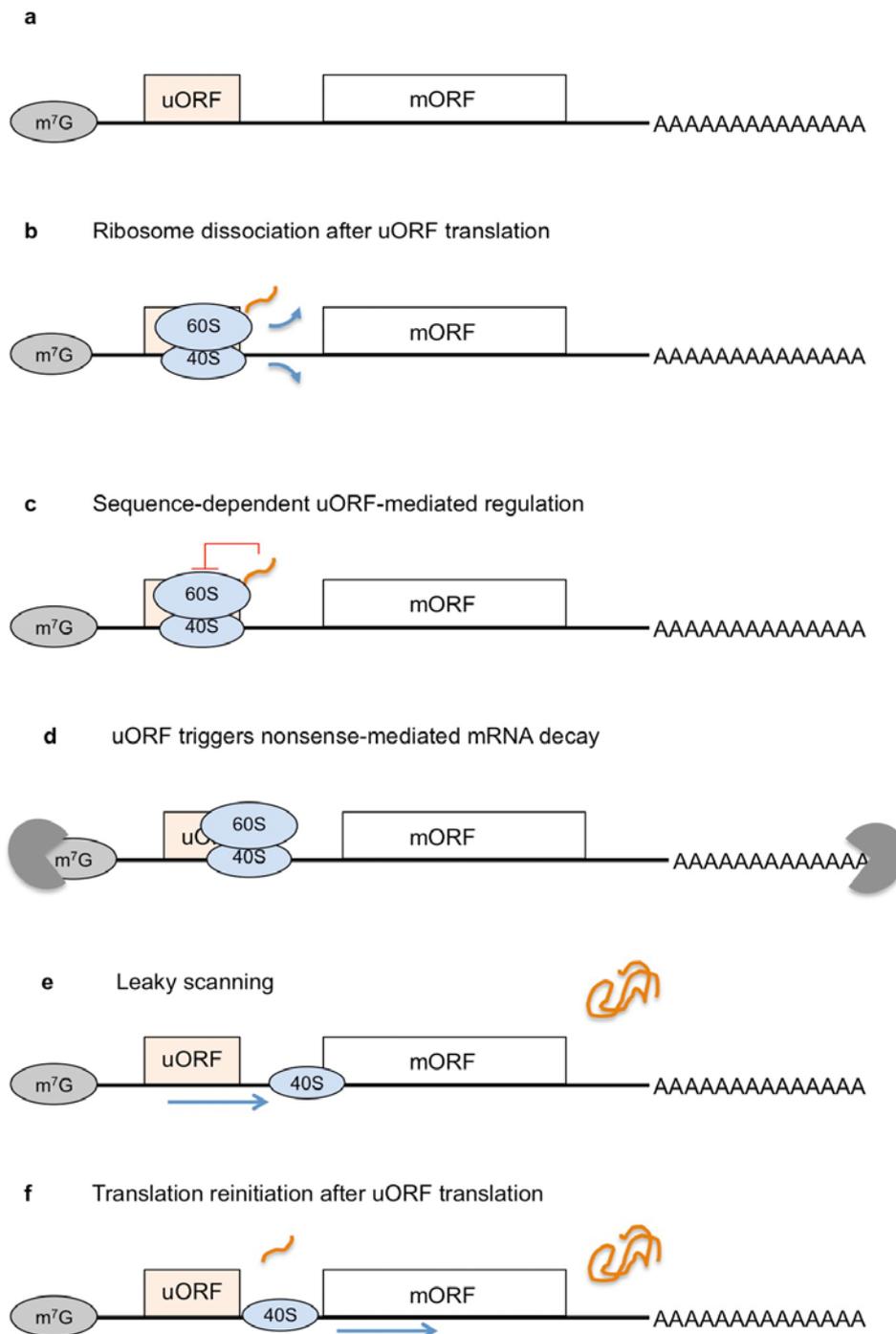


Figure 1.10. Upstream open reading frames regulate translation. a. Upstream open reading frames (uORFs) are short open reading frames encoded upstream of the main open reading frame (mORF) on mRNA transcripts. Depicted here is a schematic overview of the organization of a mRNA containing an uORF. m⁷G: 7-methylguanosine cap. The regulation of translation by uORFs occurs via different mechanisms: b. Ribosomes

recognize the translation initiation codon of the uORF, translate it and dissociate after termination, never reaching the mORF. c. The uORF encodes a functional peptide that upon translation stalls ribosome progression on the mRNA. d. The uORF translation triggers nonsense-mediated mRNA decay and the mRNA transcript is degraded before the mORF is translated. e. The 40S subunit skips the uORF translation initiation codon and initiates translation at the mORF. f. The uORF is recognized by the ribosome and translated, but upon translation the ribosome does not dissociate completely from the mRNA. The 40S subunit remains attached, scans down to the mORF and initiates translation there.

Nevertheless, uORFs do not always completely abolish translation of the mORF. The mORF is translated when the ribosomes scan across the uORF initiation codon (also called leaky scanning) (Fig. 1.10e)¹⁵³. Moreover, the post-termination 40S subunit may remain associated with the mRNA after uORF translation and can resume scanning downwards to the start codon of the mORF (Fig. 1.10f)^{188,189}. The efficiency of reinitiation is dependent on the length of the translated uORF, the distance of the uORF termination codon to the start codon of the mORF and also the physiological state of the cell¹⁸⁸⁻¹⁹¹ (Fig. 1.11.). The shorter the translated uORF the more likely it is that downstream reinitiation will occur^{192,193}. This is mainly due to the initiation factors not having dissociated completely by the time the uORF stop codon is reached. But also an increase in the distance between the uORF stop codon and the mORF start codon has been described to positively influence mORF translation^{190,194}.

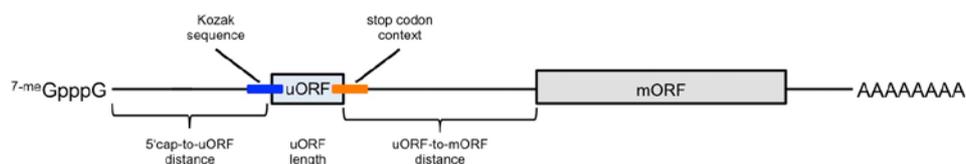


Figure 1.11. Structural characteristics influencing the functionality of uORFs. A schematic depiction of the different parameters that influence the efficiency of uORF-mediated translation regulation is shown. UORFs are located in the 5' leaders of mRNA transcripts between the 5' cap (7-meGpppG) and the main open reading frame (mORF). The regulatory potential of the uORF depends on several parameters: (i) the position within the 5' leader relative to the 5' cap and the start codon of the mORF, (ii) the length of the encoded ORF, (iii) the strength of the Kozak sequence around the uORF start codon (blue line) and (iv) the sequence surrounding the termination codon (orange line).

In plants and yeast it was shown that additionally the intercistronic sequence affects the efficiency of reinitiation^{195,196}. Single nucleotide exchanges in the intercistronic space between the uORF and the C/EBP α ORF strongly influence the reinitiation rate of the scanning ribosomes¹⁹⁷. The distance between the 5' cap and the uORF start codon is also important. If the uORF start codon is too close to the 5' cap the

preinitiation complex will not have formed during the scanning process to initiate translation on the start codon of the uORF¹⁹⁸.

1.2.3. Viruses depend on the translation machinery of their host cell

Viruses fully rely on the translation apparatus of the host cell. The protein synthesis machinery is used to produce the viral tools needed to hijack the cell, replicate the genome, generate progeny virions and in the case of herpesviruses, establish a persistent infection.

EBV-derived mRNAs contain the conserved structural features of host cell transcripts: the 5' 7-methylguanosine cap and the polyadenylated 3' tail. Therefore, ribosome loading and translation occur in the same manner.

The densely packed genome of EBV makes precise annotation of translated ORFs from genomic information alone particularly difficult. Generally, the large viral genomes of herpesviruses have been annotated by the usage of sequence homology to already described genes in other subfamilies and *in silico* predictions with defined ORF annotation parameters. This approach is useful at predicting canonical ORFs of defined sizes but does not include small ORFs or ORFs with alternative translation initiation sites. In recent years the development of a new approach to study translation in greater detail has led to the discovery of small ORFs and alternative translation initiation in different virus families¹⁹⁹⁻²⁰¹. The method used to identify these new genes is discussed in the following section.

1.2.4. Defining translatoemes by ribosome profiling

Recently, translation has been utilized to annotate complex viral genomes more precisely and to gain mechanistic insight into gene expression regulation¹⁹⁹⁻²⁰¹. Dr. Nick Ingolia together with Prof. Jonathan Weissman have developed an experimental system to analyse translated transcripts at a given time point in a cell at single nucleotide resolution²⁰² (Fig. 1.12.). The method exploits the property of the ribosome to physically enclose the mRNA at the site it is currently translating. This enclosure protects this particular mRNA fragment from digestion by nucleases. These ribosome protected footprints (RPFs) can subsequently be converted into cDNA libraries for sequencing. Combined with total RNA sequencing it can deliver detailed insights into which mRNAs produced at any given time in a cell are actually also translated. Furthermore, by using drugs that block ribosome progression directly after initiation

the method can be used to map previously unidentified translation initiation sites²⁰². By applying this method to different cell types under different physiological conditions Ingolia *et al.* have identified a plentitude of alternative translation initiation sites (aTIS) and previously uncharacterized open reading frames^{203,204}. The method has also contributed to the identification of novel ORFs in several different viruses and has helped to elucidate the complex and dynamic translational events employed by these viruses during the infection process^{199-201,205}.

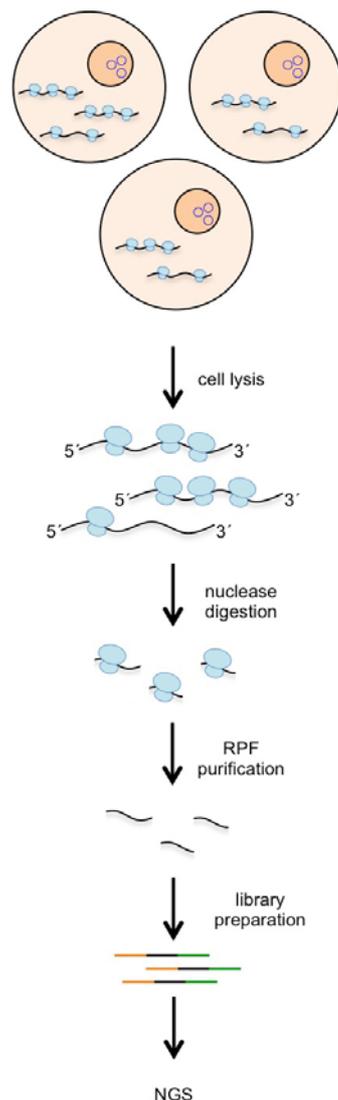


Figure 1.12: An overview over the ribosome profiling method. Cells of interest are lysed and treated with RNase. The RNA that had been loaded with ribosomes at time of lysis is incompletely digested due to the bound ribosomes. These ribosome-protected mRNA fragments (RPFs) can be purified. They are then converted into a cDNA library for next generation sequencing (NGS).

2. Aim of my thesis

EBV is a widespread human pathogen that can cause cancers of lymphoid and epithelial origin. Fifty years have passed since the initial discovery of EBV and 30 years since the sequencing of the B95-8 EBV strain. The development of next generation sequencing approaches has led to a deeper understanding of genetic polymorphisms found in different virus isolates and has increased our understanding of the complex transcriptional processes that take place in infected cells during lytic replication. However, it is unclear whether the protein-coding potential of this complex DNA virus is completely understood. This is, in part, due to the complex transcriptional landscape observed in these cells but also due to the complex translational processes employed by mammalian cells.

The aim of my thesis was to define the translation products of EBV-infected cells in the two EBV strains M81 and B95-8. The specific questions I addressed were:

- (i) What are the differences between the two strains?
- (ii) Do annotated open reading frames initiate at the sites annotated in data bases?
- (iii) Are there translated open reading frames that have not been recognized yet?
- (iv) Are the transcripts identified as non-coding truly non-coding?

3. Results

3.1. Ribosome profiling of EBV-infected cells

In order to map the full scope of translated ORFs in established LCLs, ribosome profiling libraries were generated of B cells transformed with either the B95-8 or the M81 EBV strain. The two different EBV strains were chosen for analysis as they differ substantially in their *in vitro* and *in vivo* characteristics¹. In *in vitro* cultures of M81-infected LCLs a small percentage of cells initiates spontaneous lytic replication. The B95-8 strain on the other hand is largely latent in cell culture. Therefore, an additional harringtonine-treated library was generated from HEK293 producer cell lines of B95-8 that had been induced into lytic replication by transfecting a BZLF1-encoding plasmid.

Ribosome profiling is based on the sequencing of 28-30 nucleotide (nt)-sized RNA fragments from nuclease-treated cell lysates. These fragments are protected from nuclease digestion by bound ribosomes and are called ribosome protected fragments (RPF). These RPFs are converted into a cDNA library and are subjected to next generation sequencing (NGS) (see Materials and Methods for the complete library preparation procedure).

Each sample was treated either with cycloheximide alone or with harringtonine and cycloheximide together (Fig. 3.1). Both drugs are translation inhibitors. Cycloheximide is a translation elongation inhibitor that interacts with actively translating ribosomes and stabilizes them on mRNA^{2,3}. The cycloheximide-treated samples were used to obtain insight into the mRNAs actively translated in the infected cells. Harringtonine, on the other hand, is an antibiotic that specifically interacts with the 60S ribosomal subunit prior to 80S ribosome assembly and selectively arrests these on translation initiation sites⁴. The harringtonine-treated libraries were used to map annotated and novel translation initiation sites.

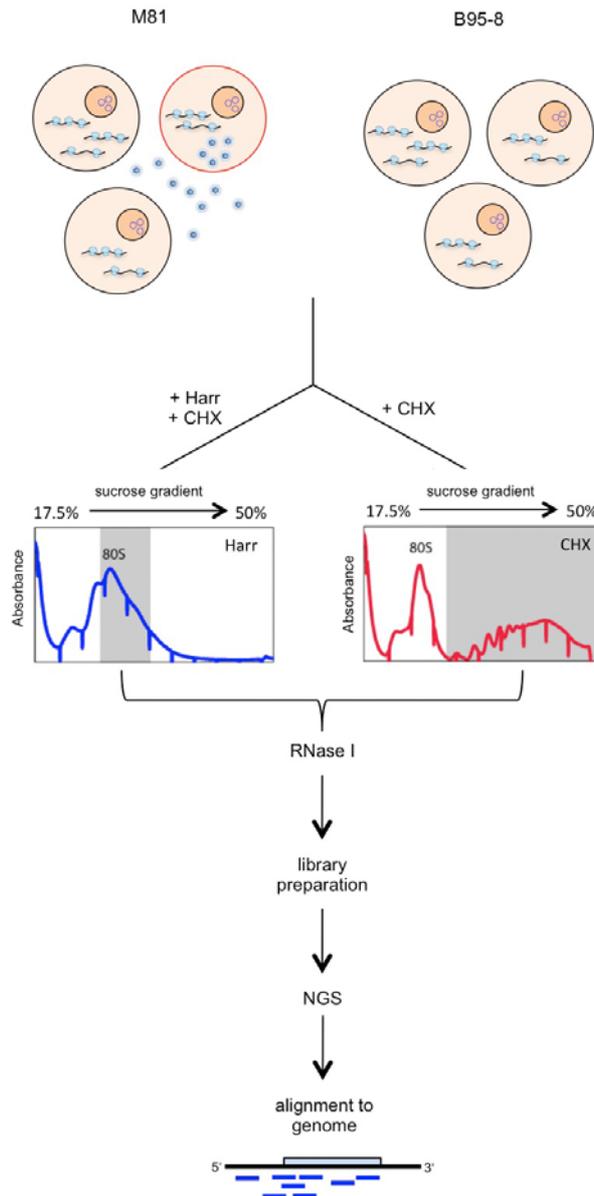


Figure 3.1: Schematic overview over the library generation procedure. M81 and B95-8-infected LCLs were treated with harringtonine (Harr) or cycloheximide (CHX) and lysed. The lysates were then loaded onto linear sucrose gradients and ultracentrifuged to separate mRNAs according to their ribosome loading. The gradients were then fractionated with a gradient fractionator and for harringtonine-treated lysates the fractions containing mRNAs loaded with one ribosome were collected (highlighted in grey on the left). For cycloheximide-treated lysates the fractions containing polysomal mRNAs were collected (highlighted in grey on the right). These fractions were then treated with RNase I to digest mRNA segments not covered by ribosomes. The remaining ribosome-protected fragments were then purified and converted into a cDNA library for next generation sequencing (NGS). The sequenced fragments were then aligned to the human and viral genomes.

Sucrose density gradient ultracentrifugation was used to separate polysomes from monosomes, ribosomal subunits and free RNAs (Fig. 3.1 and 3.2.a). Subsequent fractionation of the gradients enabled isolation of different ribosomal populations with

their associated mRNAs. During fractionation, the continuous UV absorbance profile of the single fractions was recorded at 254 nm.

For the cycloheximide-treated samples, the polyribosomal fractions were collected for library generation (Fig. 3.1). Treatment with harringtonine caused a shift in the distribution of ribosome-bound mRNAs. The polyribosomal fraction was depleted and a strong peak was observed in the 80S monosome-containing fraction (Fig. 3.1). This was due to the selective inhibition of initiating ribosomes. This fraction was then collected for library preparation. A summary of the sequencing reads obtained from each library can be viewed in Appendix I.

The lytic replication level of LCLs was determined in parallel on the days of profiling by immunofluorescence staining for the lytic cycle proteins BZLF1 and gp350. The quantification of cells positive for these proteins is given in Fig. 3.2b. BZLF1 is an immediate early lytic protein required to initiate lytic replication in cells. It is a marker for the early stages of lytic replication. The gp350 glycoprotein is the viral entry receptor and is expressed during the late stages of lytic replication.

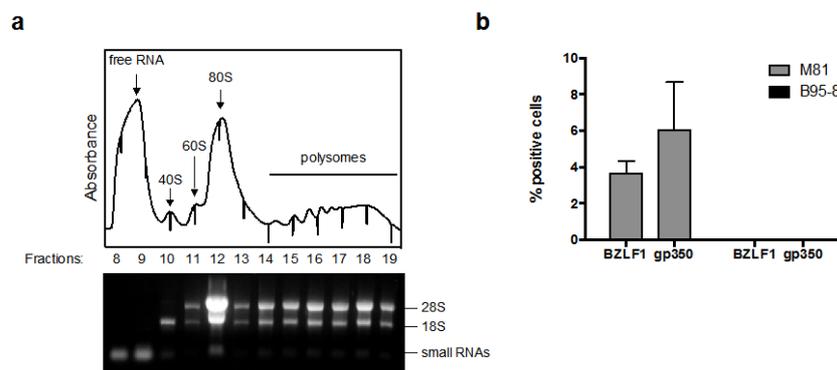


Figure 3.2: Polysome profile of LCLs and lytic replication levels in the different strains. a. The individual fractions from a polysome profile of B95-8-infected LCLs were collected, the RNA was purified and an agarose gel was run to visualize the 18S and 28S ribosomal RNAs present in the different fractions. Fractions 8 and 9 contain non-ribosome associated RNAs and small RNA species. Fraction 10 contains mainly the 40S ribosomal subunit. Fraction 11 contains mainly the 60S ribosomal subunit. Fractions 12 and 13 contain the translation-competent 80S ribosome and all subsequent fractions contain increasing numbers of translating 80S ribosomes. b. Lytic replication levels of infected cells used for ribosome profiling were determined on the day of sucrose density ultracentrifugation by immunofluorescence stainings for the lytic cycle proteins BZLF1 and gp350. The quantification of these stainings is shown as a bar plot. Quantifications of both replicates are summarized here. Error bars show the standard deviation between the two replicates.

As with conventional transcriptome data, abundance is measured during the sequencing reaction. The different reads are counted based on the number of times that particular sequence is read⁵. The different libraries contained varying amounts of reads that could be mapped to the viral genome (Appendix I). It ranged from 0.01 to 1% of total mapped reads. This is in line with Arvey et al. who reported that on average they could map 0.1 to 1% of total RNA sequencing-derived reads to the EBV genome⁶.

The reads derived from ribosome profiling range in size from 25 to 32 nucleotides. Ideally the majority of reads is in the size range of 27 to 30 nucleotides. This was the case for most of the LCL-derived libraries (Fig. 3.3 and Appendix II).

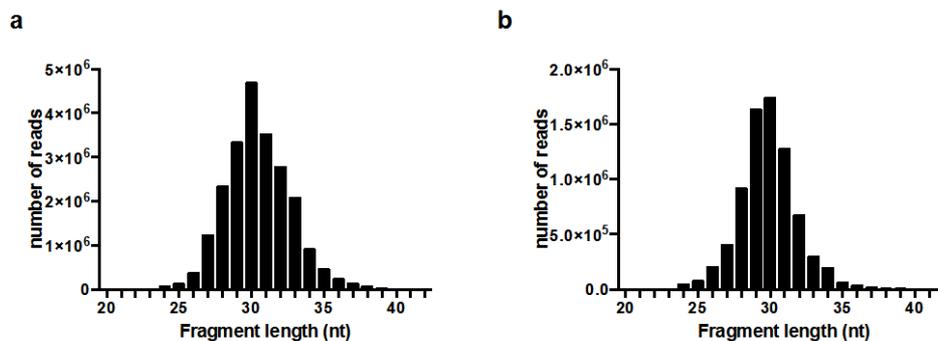


Figure 3.3: Fragment length distribution of ribosome footprints. Representative fragment length distribution plots of sequenced ribosome-protected fragments are shown a. The length distribution of reads from the cycloheximide-treated library generated from M81-infected LCLs is shown b. The length distribution of reads from the 5 min harringtonine-treated library generated from M81-infected LCLs is shown. For fragment length distribution plots of all libraries generated see Appendix II.

Following sequencing and mapping, the reads aligning to the cellular genome were used to compile metagene analyses of the data sets. This was done to further assess the quality of the generated libraries (Fig. 3.4). These revealed an increase in read density past the start codon in the cycloheximide-treated samples (Fig. 3.4a) and an accumulation of reads surrounding start codons upon harringtonine treatment (Fig. 3.4b).

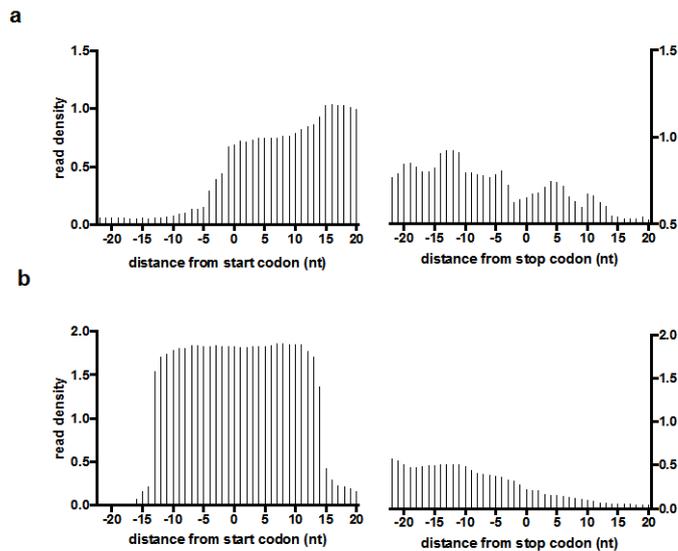


Figure 3.4: Metagene analysis. Relative RPF densities surrounding the start and stop codons of annotated and highly expressed coding cellular genes in representative libraries derived from M81-infected LCLs are shown a. A representative plot of relative read densities in the cycloheximide-treated library is shown. b. A representative plot of relative read densities in the harringtonine-treated library is shown.

The reads that aligned to the viral genome were sorted according to different categories of the viral gene locus (Fig. 3.5). The viral reads sequenced in the cycloheximide-treated libraries to a large proportion were mapped to the two EBV-encoded small RNAs (EBERs) (Fig. 3.5a). Between these two RNAs, EBER1 shows significantly higher coverage. Previous studies have shown that EBER1 is bound by the ribosomal protein L22 which is part of the 60S subunit⁷. The proportion of reads mapping to the EBER RNAs decreased in the harringtonine-treated libraries (Fig. 3.5b). This might be due to competitive interaction of harringtonine with the 60S subunit⁴.

A significantly smaller proportion of reads mapped to protein-coding genes (AUG start codon and coding DNA sequence (CDS)). In M81-infected LCLs, more reads aligned to coding regions of the transcriptome than in the B95-8-infected cells (Fig. 3.5a). The classification of the reads highlights the selective accumulation of ribosomes on translation initiation sites upon harringtonine treatment in the reads that aligned with the cellular and viral genes (Fig. 3.5b). Additionally, an increased fraction of reads mapped to the 5'leaders of transcripts in these libraries.

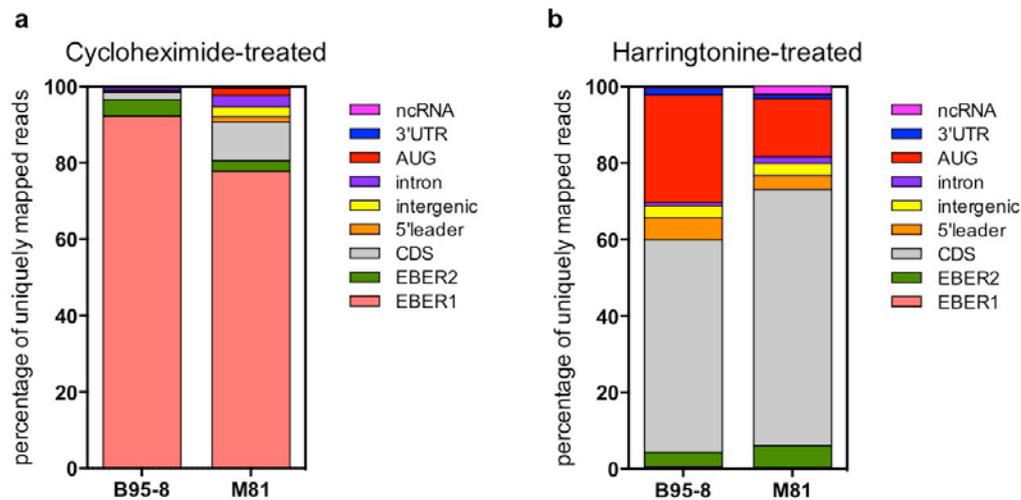


Figure 3.5: Read classification in the different ribosome profiling libraries. a. The stacked bar graphs show the classification of reads mapping to known viral genes from cycloheximide-treated libraries. Reads that intersected several classification features were sorted with following preference: AUG (start codon), CDS (coding DNA sequence), 5'leader, 3'untranslated region (3'UTR), non-coding RNA, intron, intergenic. The EBER RNAs were grouped into their own class due to their high abundance. b. Classification of reads obtained from the harringtonine-treated libraries as in a.

3.2. Differences in host gene translation between B95-8 and M81-infected LCLs

From the RPFs that were successfully aligned to the human genome, the read densities were calculated and expressed as reads per kilobase of coding DNA sequence normalized by the median over all transcripts. This quantification was used to infer differential translation between the two strains. A core analysis using the Ingenuity Pathway Analysis (IPA) software from Quiagen was performed. Only the cycloheximide-treated samples were used here. The top scoring canonical pathways and their representation in the two strains are shown in Fig. 3.6. IPA used the Fisher's exact test to assign significance through p-value calculation. Most of the top scoring pathways were more strongly enriched in the B95-8-infected samples. It is difficult to assess whether these differences truly stem from the differences between the two strains or whether this is due to the general broader expression of genes in B95-8 infected LCLs (Fig. 3.6a). Interestingly, the eIF2-signaling-associated pathway was more significantly enriched in the M81-infected samples. This pathway is directly linked to translation.

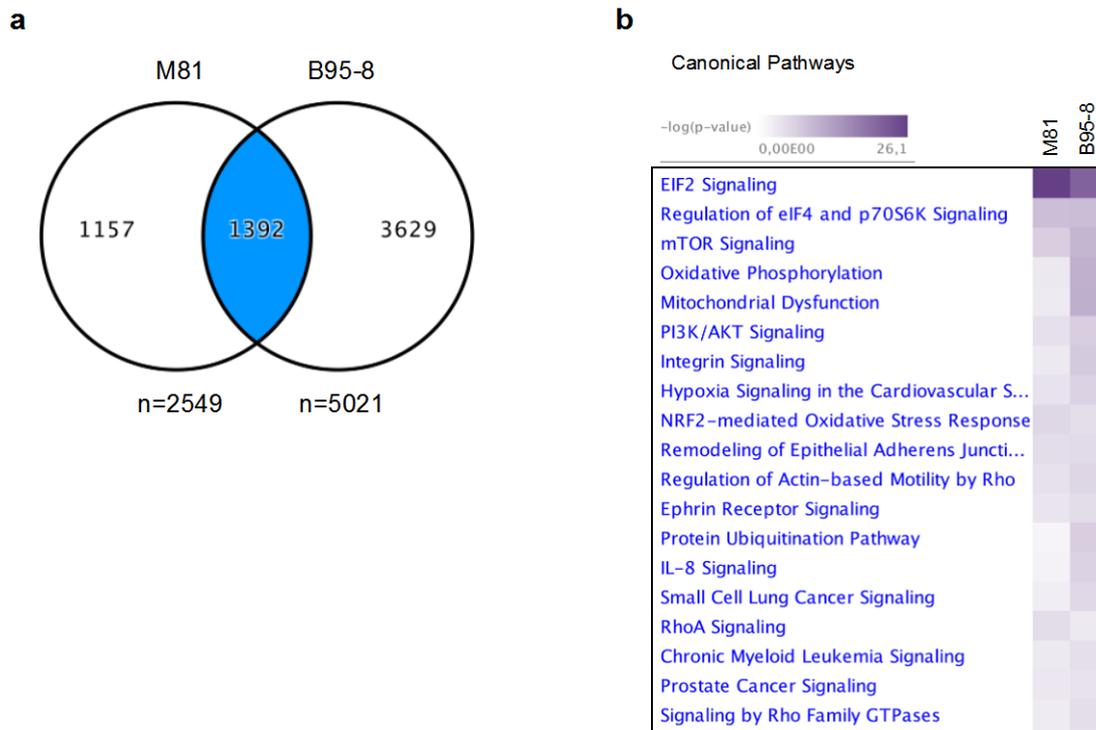


Figure 3.6: Ingenuity Pathway Analysis (IPA) of canonical pathways found to be most significantly enriched in genes covered by ribosome protected fragments (RPFs). a. Venn diagram summarizing the number of differentially translated genes between the two strains. The overlapping region between the two circles highlighted in blue indicates genes that were translated in both strains. b. The top canonical pathways that showed enrichment in genes covered by RPFs as determined by IPA analysis are shown. The heat map was generated by comparison of the cycloheximide-treated libraries from both strains. The p-value was calculated using Fisher’s exact test by the IPA software. The significance of the results is given as –log of the p-value.

A further functional analysis of the RPF-associated genes in the studied LCLs identified biological functions particularly represented in these cells. These included organismal injury and morphology and cancer-associated pathways (Fig. 3.7). Again, functions associated with translation such as protein synthesis and post-translational modifications ranked higher in M81-infected LCLs than in cells infected with B95-8.

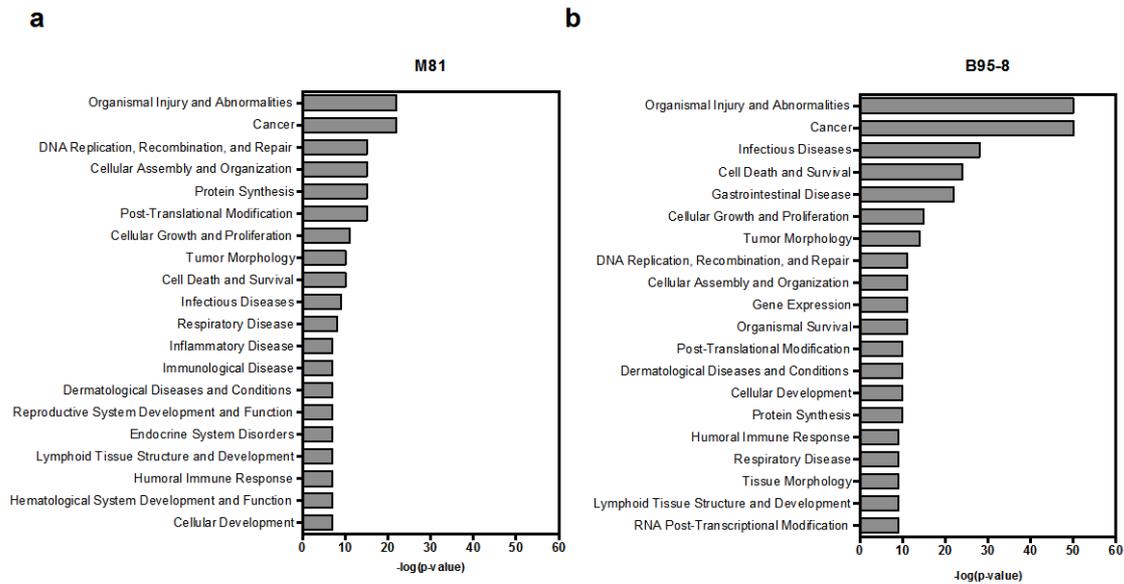


Figure 3.7: Diseases and biological functions associated with translated transcripts in M81 and B95-8-infected LCLs. The top 20 relevant pathways associated with different diseases and biological functions as determined by IPA are shown. The functions are sorted according to significance as determined by the p-value starting from the top with most significant to least significant. The $-\log p$ value is plotted as calculated by Fisher's exact test. A p-value of 0.05 was used as threshold for significance. a. Pathway ranking for M81-infected LCLs. b. Pathway ranking for B95-8-infected LCLs.

3.3. Viral mRNA translation in B95.8 and M81-infected LCLs

Following the functional analysis of host cell translation and the general classification of virus-derived RPFs, the individual EBV ORFs were analysed for ribosome coverage. A global overview over the read coverage for the cycloheximide-treated samples of the M81- and B95-8-infected LCLs is shown in Fig. 3.8. The M81-infected LCLs show almost complete coverage of the viral genome (Fig. 3.8a). The extent of bidirectional coverage over the whole EBV genome is lower in B95-8-infected LCLs and is mostly restricted to annotated viral genes (Fig. 3.8b).

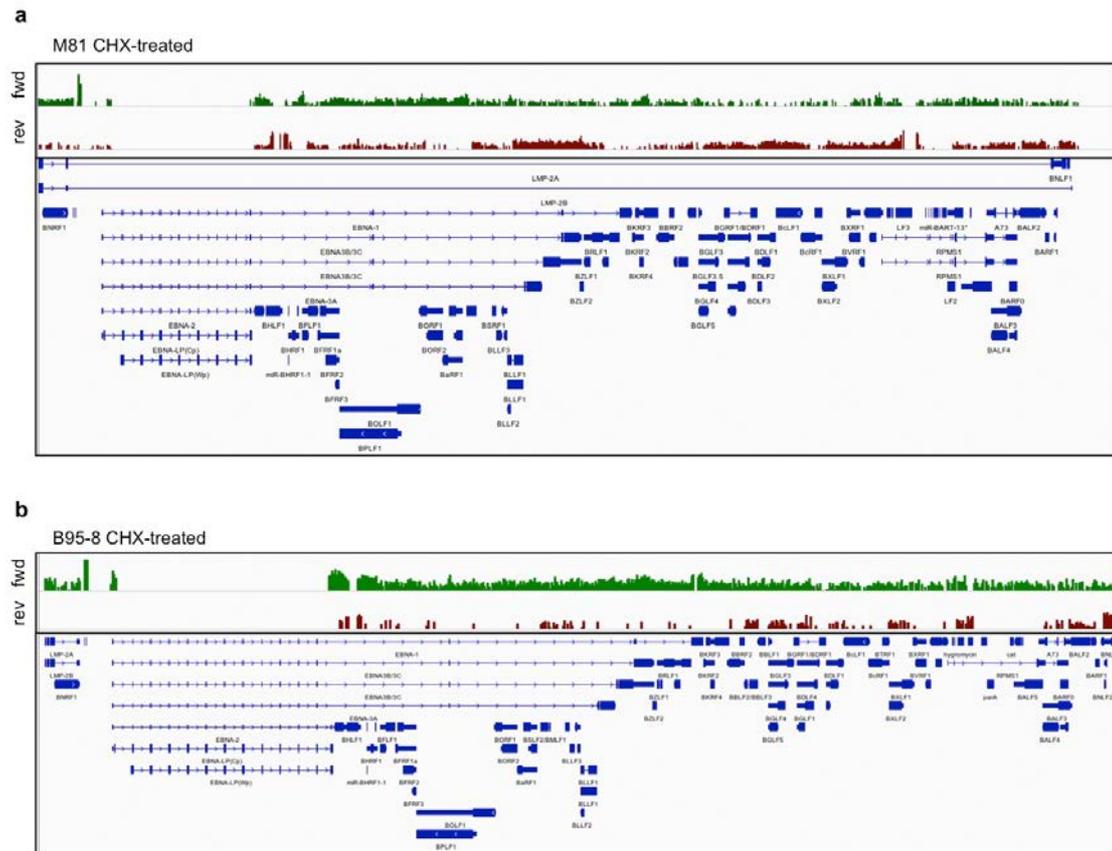


Figure 3.8: M81-infected LCLs show evidence for global bidirectional translation of the viral genome. a. Global overview over the M81 genome with coverage tracks of the cycloheximide-treated library separated according to strand direction of the mapped RPFs. The green track shows reads that are oriented in the rightward direction (defined as the forward (fwd) strand) and the red track shows reads that are oriented in the leftward direction (defined as the reverse (rev) strand). Below the tracks the gene annotation of the viral genome is shown. b. The B95-8 cycloheximide-treated library is shown with parameters as in a.

For detailed analysis of the individual ORFs the genes were classified according to the viral life cycle phase they were primarily associated with: (i) latent and (ii) lytic genes.

In the cycloheximide-treated libraries, the reads were additionally classified according to their localization within the respective mRNA transcript. Two categories were distinguished: reads that mapped within the 5'leader of the transcripts and reads that mapped within the protein-coding region (coding DNA sequence (CDS)). The cycloheximide-derived data was utilized to calculate the ratios between ribosome-protected footprints outside (within the 5'leader) versus inside a respective ORF (out:in ratio) (Fig. 3.9a). The method is derived from Chew et al.⁸. High ratio values indicated enrichment of footprints within the candidate ORF while low ratio values

meant a stronger accumulation of footprints within the 5'leaders of the analysed mRNA transcripts.

The harringtonine-derived data was used to calculate the ratios between footprints mapped to the 5'leader versus the footprints mapped to the annotated start codon (Fig. 3.9b). This ratio was indicative of the ribosome's efficiency to reach the start codon of the candidate gene.

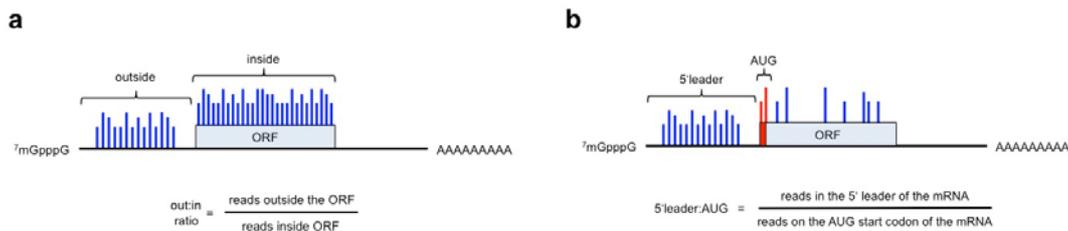


Figure 3.9: Data evaluation in the two library types. a. Ribosome footprints derived from cycloheximide-treated samples were used to calculate out:in ratios by dividing the number of normalized footprints mapped within the 5'leader of the mRNA transcript (out) by the number of reads mapped to the ORF (in). The nucleotide length of the respective feature was used to normalize the values. b. Ribosome footprints derived from harringtonine-treated libraries were used to calculate 5'leader:AUG ratios. Here the reads mapping to the 5'leader of a candidate gene were divided by the reads mapping to the annotated start codon (AUG) of the respective gene.

3.3.1. Latent genes

First, the ribosome coverage of EBV's latent genes was analysed. Seven of the ten latent genes encoded by EBV are expressed from one genomic region, termed the EBNA transcription unit (Fig. 3.10a). They are transcribed from one of two promoters, the C promoter (Cp) or W promoter (Wp). All seven genes exhibit unusually long 5'leaders. The other three genes (LMP1, LMP2A, LMP2B) are located close to or, in the case of the LMP2s, span the terminal repeats. LMP2A and LMP2B are protein isoforms stemming from the same genomic region. This makes definitive assignments of RPFs to either gene impossible and therefore these two proteins are summarized here as LMP2.

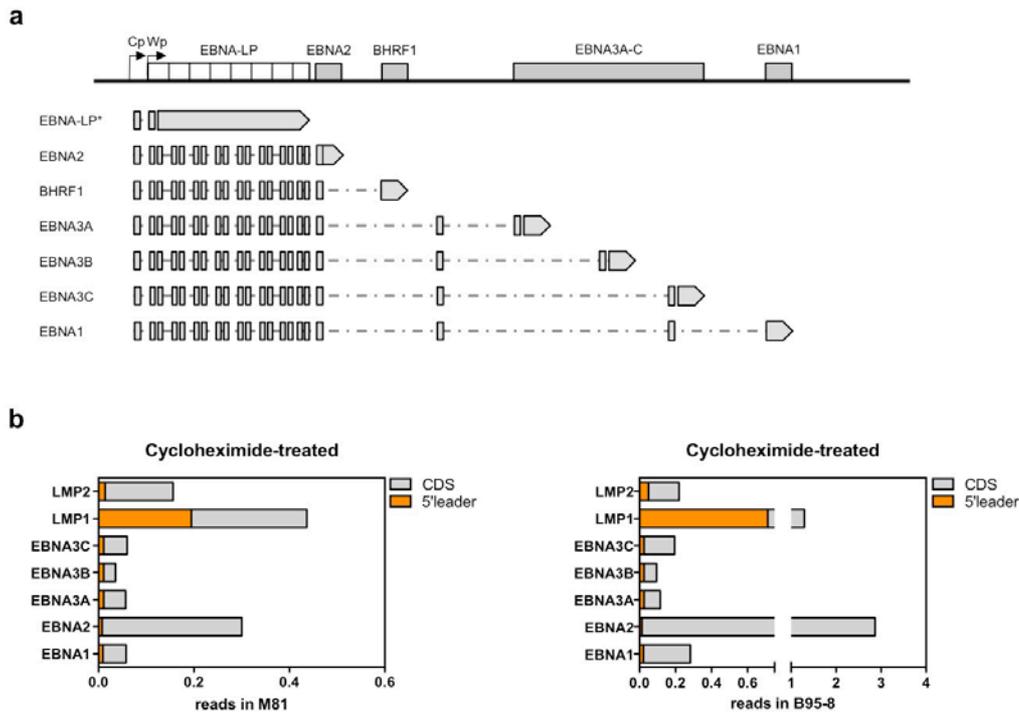


Figure 3.10: Ribosome coverage on latent transcripts. a. Overview over the EBNA transcription unit. The unit is under the control of two promoters (Cp and Wp). The expression of six EBNA genes and latent BHRF1 is directed from these two promoters. *EBNA-LP transcription is simplified in this figure and only the longest EBNA-LP isoform is shown. b. The read distribution on latent genes in the cycloheximide-treated samples is shown. Reads were classified according to their localization on mRNA transcripts: coding DNA sequence (CDS) and 5'leader of the mRNA transcript (5'leader). *Left:* Reads derived from M81-infected LCLs are plotted. *Right:* Reads derived from B95-8-infected LCLs are plotted. LMP2A and LMP2B transcripts are summarized as LMP2. EBNA-LP was omitted due to the repetitive nature of its exons and the difficulties of unambiguous mapping associated with it. The reads mapping to the 5'leader region shared by all Cp-derived genes were divided by seven (EBNA-LP, EBNA2, BHRF1, EBNA3A-C, EBNA1) and a proportion of these reads was ascribed to the total reads in the 5'leaders of each gene. Furthermore, the reads mapping to the U exon, which is shared by the EBNA3 family of genes and EBNA1, were divided by four and ascribed to the respective 5'leaders. For each gene the longest mRNA isoform was considered in the analysis.

The latent genes generally have an overall lower coverage in M81-infected LCLs compared to B95-8 (Fig. 3.10b). However, the overall pattern of latent protein translation is comparable between the two strains with EBNA2 and LMP1 being the two genes with the highest ribosomes coverage (Fig. 3.10). All latent genes show ribosome coverage within their 5'leaders in both strains. The out:in ratios of EBNA1, EBNA3A and EBNA3B were almost identical between the two strains (Fig. 3.11a). LMP1 had a higher out:in ratio in M81 compared to B95-8. This is also reflected in the 5'leader:AUG ratios indicating that more ribosomes accumulate in the LMP1 5'leader in M81 (Fig. 3.11b).

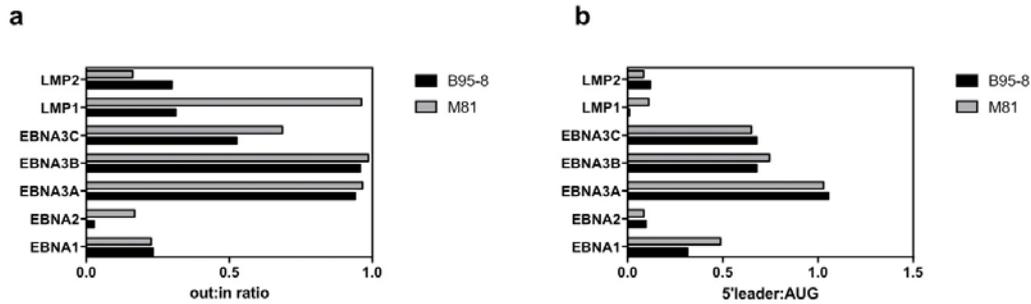


Figure 3.11: Ribosome coverage on latent transcripts. a. The ratios between reads mapping to the 5'leader vs. the reads mapping to the remaining open reading frame of the respective genes are shown. Cycloheximide-treated libraries were used for ratio calculation. b. Ratios of reads derived from the 5'leaders of the latent genes to the reads covering the annotated start codon are shown. Harringtonine-treated libraries were used for ratio calculation.

3.3.2. Lytic genes

In the M81-infected LCLs, all lytic genes annotated in the NCBI EBV genome annotation were found to be associated with ribosomes (Fig. 3.12). The extent of their association with the cellular translation machinery varied but because no transcriptome analysis was performed here, no assertion over the differential translation efficiency can be made.

Even though the B95-8 EBV strain has been described to be predominantly latent *in vitro*, several lytic cycle genes were found to associate with the cellular translation machinery (Fig.3.13). Some of these genes showed ribosome association rates that were comparable to those found in the M81-derived samples (Fig. 3.13). Furthermore, the BNLF2a gene exhibited stronger ribosome coverage in B95-8-infected LCLs than in M81. Though initially identified as a lytic cycle gene, BNLF2a has recently been detected at the mRNA level in latently infected cells⁹. The ribosome profiling data further supports this observation and indicates that not only is this gene transcribed during latency but it is also translated.

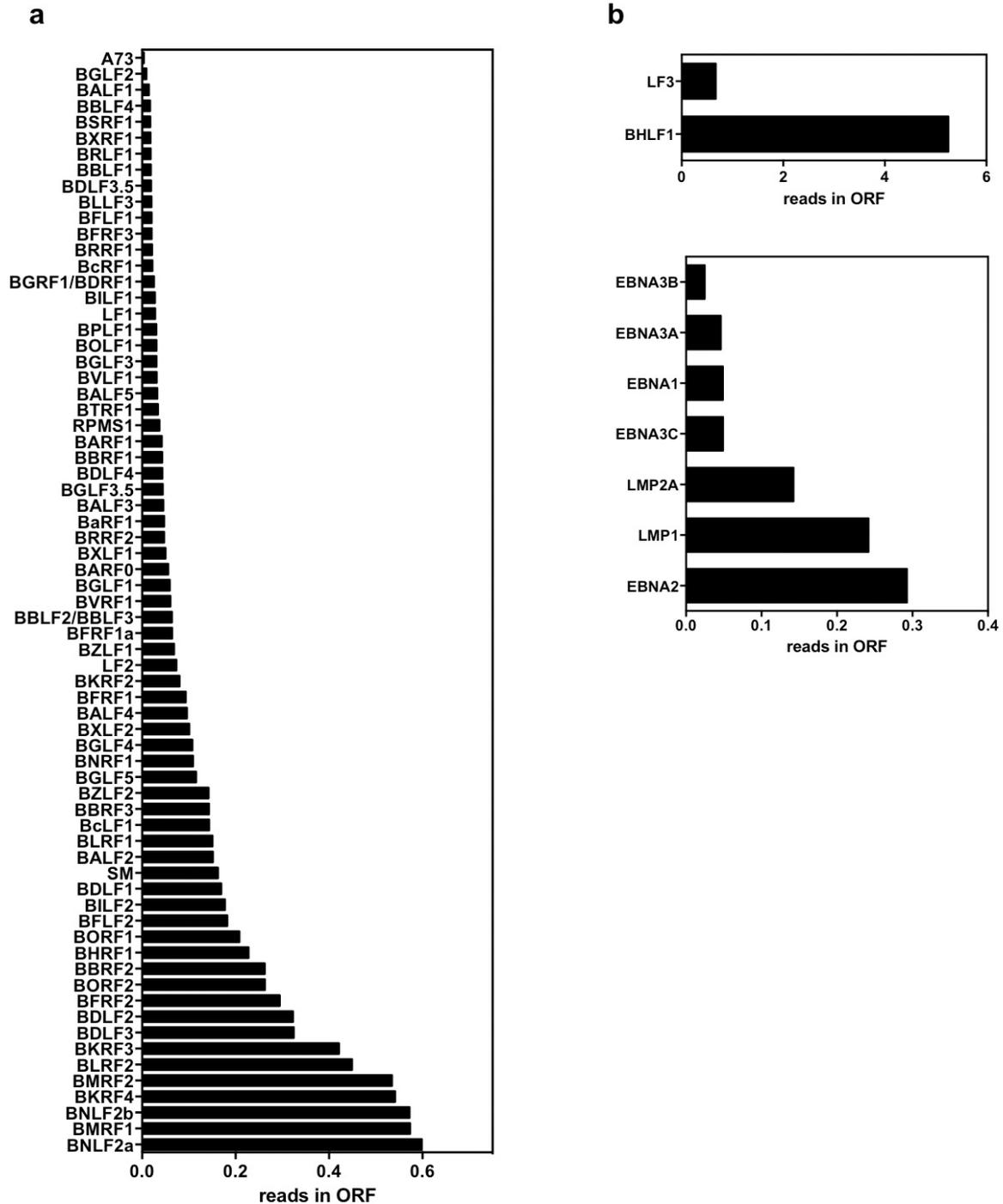


Figure 3.12: Lytic transcripts associate with ribosomes at varying degrees. a. The graph shows lytic genes sorted according to their association with ribosomes in M81-infected LCLs. The ribosome coverage of cycloheximide-treated samples is shown. The ORFs of the BdRF1 and BVRF2 genes overlap substantially as do the genes of BLLF1 and BLLF2. Therefore, these four genes could not be included in the analysis of the cycloheximide-treated samples as unambiguous assignment of reads to these genes is not possible. b. Read coverage of the lytic cycle genes BHLF1 and LF3 are shown separately due to their high abundance in lytically replicating cells. Below the ORF coverage of the latent genes is shown.

Finally, the BMRF1 gene was chosen to further validate the data obtained from ribosome profiling of B95-8-infected LCLs. Not only did this gene show clear ribosome footprints within its ORF in the cycloheximide-treated samples, but it also exhibited a very prominent peak on its start codon in the harringtonine-treated libraries (Fig. 3.14a). In order to confirm the presence of BMRF1 mRNA within B95-8-infected LCLs, a real-time quantitative PCR (RT-qPCR) analysis was performed on three independent LCLs established with the B95-8 strain. As a positive control, LCLs established from the same blood samples with the M81 strain were used. The negative controls were matched LCLs established with the M81 Δ ZR or the B95-8 Δ BMRF1 viruses. The M81 Δ ZR virus is a mutant derived from the M81 strain. Here the lytic cycle transactivators BZLF1 and BRLF1 are deleted and LCLs infected with this virus cannot initiate lytic replication. On the other hand the B95-8 Δ BMRF1 virus is a B95-8-derived mutant devoid of the BMRF1 gene. BMRF1 RNA was detected in B95-8-infected LCLs, albeit at approximately half the level found in M81-infected LCLs (Fig. 3.14b). Interestingly, BMRF1 RNA was also detected at low levels in M81 Δ ZR-infected LCLs even though the main transactivators of this gene: BZLF1 and BRLF1 were deleted in this strain. This indicates that BMRF1 on the one hand can be transcribed independently of BZLF1 and BRLF1 but also that it might be transcribed during latency at low levels as well.

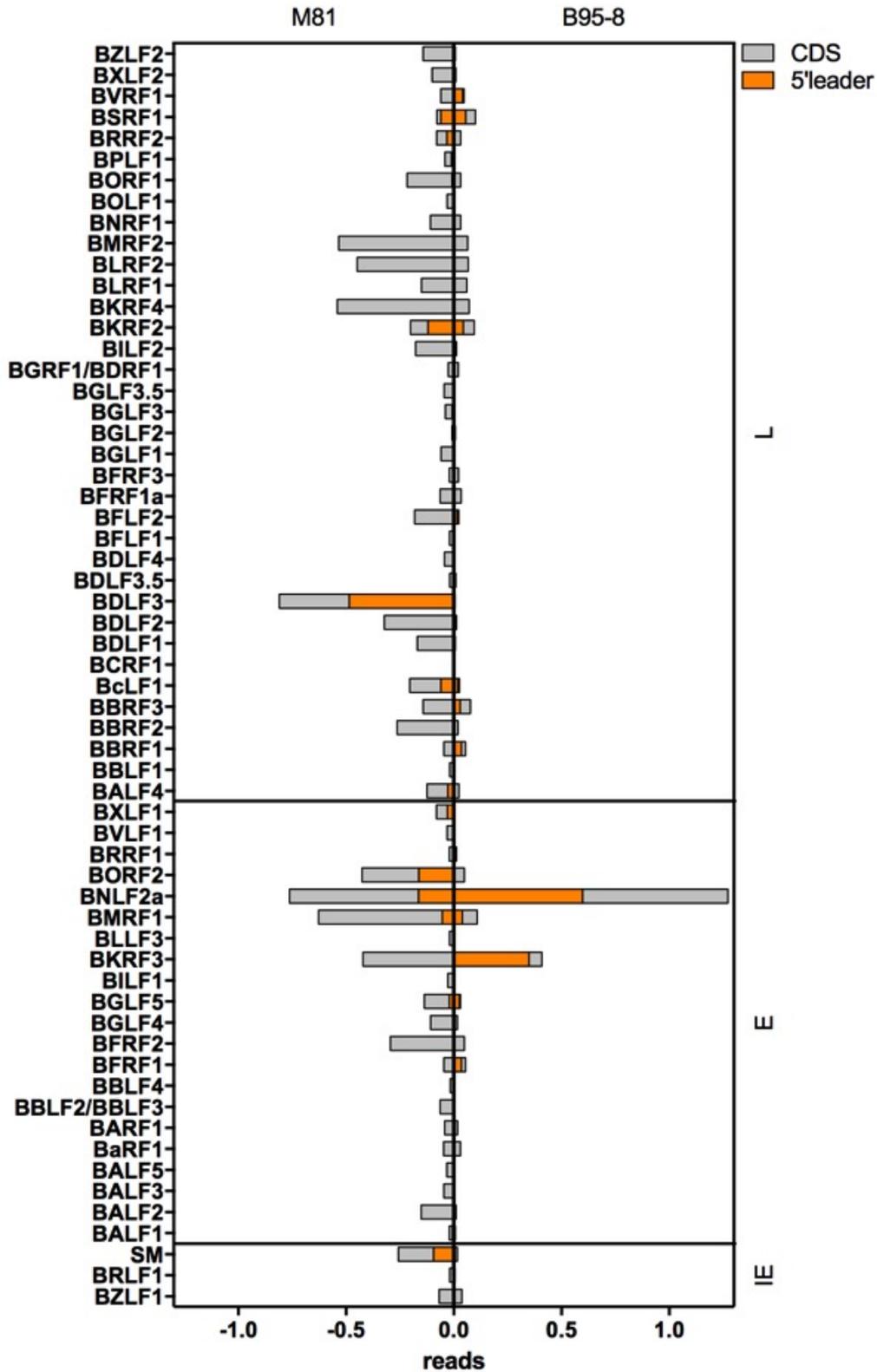


Figure 3.13: Comparison of ribosome coverage on lytic genes in B95-8 and M81-infected LCLs. a. The graph bar shows the ribosome coverage of lytic cycle genes in LCLs established with the B95-8 (positive values) or M81 strain (negative values). Depicted is the data from the cycloheximide-treated libraries. CDS: coding DNA sequence. The genes are grouped according to their temporal expression during reactivation: IE: immediate early genes. E: early genes. L: late genes.

The low levels of BMRF1 transcript in RT-qPCR analysis indicated that most probably only a low number of cells would express the protein. The protein product of BMRF1 was additionally analysed by immunofluorescence staining. This allowed us to specifically look at single cells in the culture and potentially identify BMRF1 expressing cells. Indeed, BMRF1 positive cells were identified in B95-8-infected cells albeit at a 40-fold lower level than in M81-infected LCLs (Fig. 3.14c).

These experiments confirmed that the ribosome coverage on lytic genes in B95-8 did not stem from artifacts or contaminations with M81 samples introduced during library preparation. Furthermore, they confirmed that the latency III gene expression program of B95-8-transformed is not strict but that a small number of cells can and do enter lytic replication.

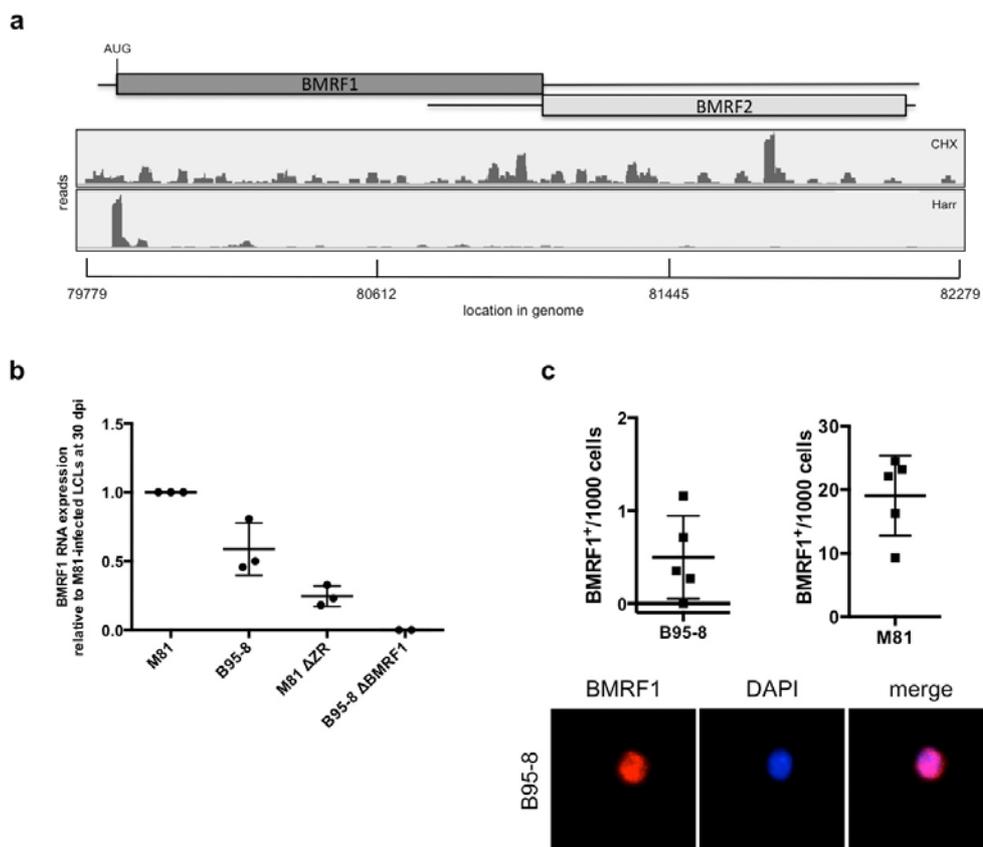


Figure 3.14: BMRF1 is expressed in B95-8-infected LCLs. a. Ribosome profiles around the BMRF1 gene in B95-8-infected LCLs. *Top panel:* cycloheximide-treated samples (CHX), *lower panel:* harringtonine-treated samples (Harr). b. The dot plot shows RT-qPCR measurements for BMRF1 transcript abundance in three independent LCLs generated with the viral strains indicated below. M81: positive control, lytically replicating strain. B95-8: predominantly latent strain. M81 Δ ZR: M81-derived mutant deleted in the BZLF1 (Z) and BRLF1 (R) genes and therefore incapable of lytic replication. B95-8 Δ BMRF1: B95-8-derived mutant deleted in the BMRF1 gene. Results are plotted relative to the expression level of M81-infected LCLs. c. Immunofluorescence staining for BMRF1 protein expression and quantification of positive cells. The dot plot shows the quantification for 5

independent LCLs established with the B95-8 or M81 strain. A representative image of a cell infected with the B95-8 strain expressing BMRF1 is shown below.

3.4. Ribosomes associate with non-coding RNAs

Next we analysed the ribosome association of viral non-coding RNAs. As already mentioned before, the EBERs show extensive association with the cellular translation apparatus (Fig. 3.5). The secondary structures of the EBER transcripts are shown in Fig. 3.15¹⁰. The majority of reads that align to EBER1 in all ribosome profiling libraries can be mapped to a region adopting stem-loop structures but not described to bind any cellular protein (Fig. 3.15a, highlighted in orange). For the EBER2 transcripts different regions of the transcripts are protected from RNase I digestion depending on which ribosome profiling library is analysed (Fig. 3.15b). In the cycloheximide-treated samples, the dominant peak observed in the libraries is at the 5'end of the transcript (Fig. 3.15b). It encompasses a part of the reported La antigen binding site. In harringtonine-treated samples the major peak is observed at the 3'end of the transcripts. This also encompasses part of the reported La antigen binding site but also includes a longer region of the transcript that is part of the stem-loop structure of EBER2.

We also see mapping of reads to viral and cellular small non-coding RNAs such as microRNAs and small nucleolar RNAs (data not shown). These observed peaks most likely do not reflect true translation but rather these transcripts are protected from nuclease digestion through their association with other RNA binding proteins such as the RNA-induced silencing complex into which miRNAs are assembled.

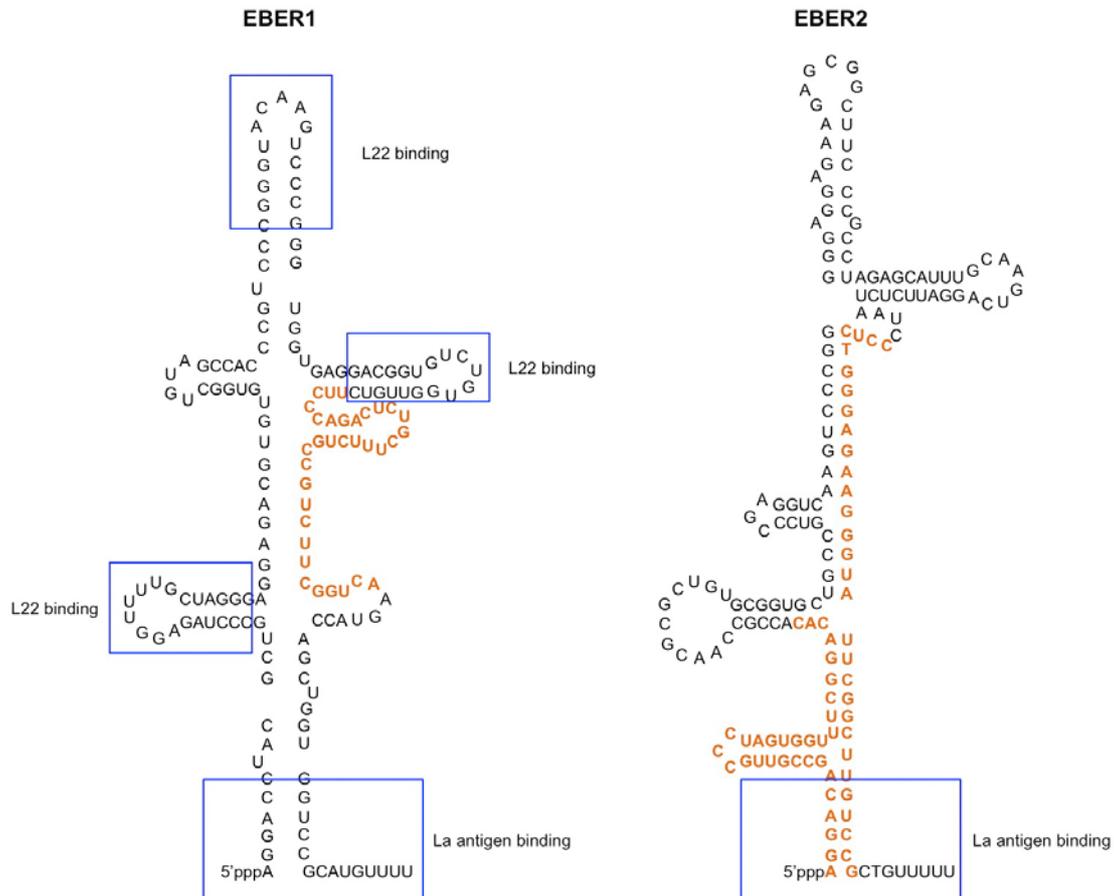


Figure 3.15: The EBER1 and EBER2 secondary structures reproduced from Iwakiri 2016¹¹. Schematic depictions of the EBER secondary structures are shown. The blue rectangles highlight EBER RNA structures previously shown to interact with cellular proteins. Highlighted in orange are the regions where the majority of reads aligning to the EBER transcripts map to. For EBER2 the left orange region is the dominant peak in the cycloheximide-treated libraries while the highlighted sequence on the right side of the stemloop structure is the predominant peak in the harringtonine-treated libraries (shown here for M81).

The BHLF1 gene which has recently been proposed to encode a long non-coding RNA contained high numbers of ribosome footprints¹² (Fig. 3.16a). The same was the case for its positional equivalent at the second oriLyt LF3 (data not shown). Most of the reads mapped to the 5' leaders of these transcripts (Fig. 3.16b). These genes consist of highly repetitive sequences in their putative coding region which makes the unambiguous mapping of reads in this region impossible. This was taken into account by quantifying the reads mapping within the tandem repeats just once. The ribosome association was observed in both strains for BHLF1 although the M81-derived samples exhibited significantly higher coverage (Fig. 3.16b/c). BHLF1 was initially described as a lytic gene and was only later discovered to also be expressed

in latent cells from a distinct latent promoter¹³⁻¹⁶. In the cycloheximide-treated B95-8-derived libraries, a significant proportion of reads mapped to the 5' leader of BHLF1 while in M81 a significant amount of reads were also identified within the ORF (Fig. 3.16b/c). This was also true for the LF3 gene in M81. The LF3 gene is deleted in the B95-8 genome.

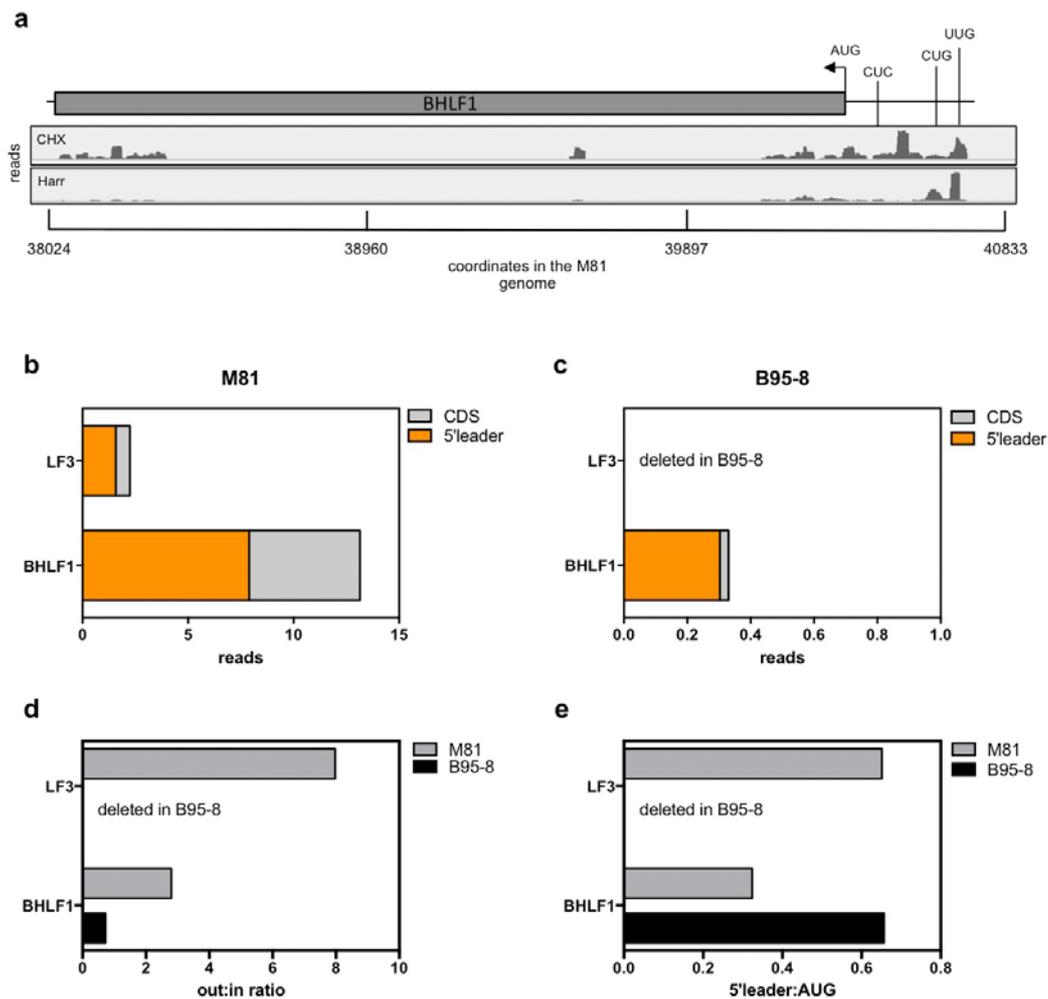


Figure 3.16: BHLF1 and LF3 transcripts show extensive association with ribosomes. a. Ribosome footprints mapping to the BHLF1 gene from M81-infected LCLs are shown. Both libraries are represented: cycloheximide-treated samples (CHX) and harringtonine-treated samples (Harr). The annotated translation initiation codon is depicted. The arrowhead indicates the direction of translation of the gene. b. Read distribution on the different segments of the BHLF1 and LF3 genes from the cycloheximide-treated M81-infected LCL samples. c. Read distribution on the different segments of the BHLF1 and LF3 genes from the cycloheximide-treated B95-8-infected LCL samples. d. The ratios between reads mapping to the 5' leader vs. the reads mapping to the remaining open reading frame of the respective genes are shown. Cycloheximide-treated libraries were used for ratio calculation. e. Ratios of reads derived from the 5' leaders of the latent genes to the reads covering the annotated start codon are shown. Harringtonine-treated libraries were used for ratio calculation.

The out:in ratios were high in both strains for BHLF1 (Fig. 3.16d). The same was true for LF3 in the M81-derived library. This further highlighted the strong accumulation of reads within the 5' leader of the transcripts, rather than in the protein-coding regions. On the other hand, the 5'leader:AUG ratio was comparable to the EBNA3 family of proteins (Fig. 3.11b) which are known to generate protein product. When analysing the distribution of BHLF1 transcripts within the polysomal fractions of M81-infected LCL lysates, the mRNA was found to preferentially localize to the lighter polysomal fractions devoid of translating ribosomes (Fig. 3.17).

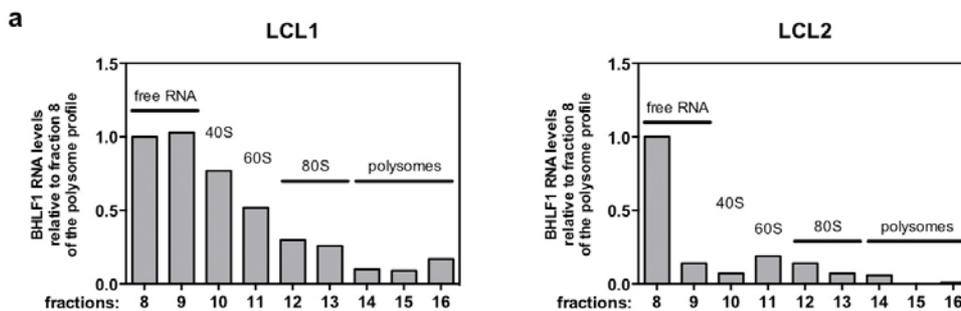


Figure 3.17: BHLF1 RNA preferentially sediments in the lighter polysomal fractions. a. Quantification of BHLF1 transcripts determined by RT-qPCR within the different fractions of a polysome profile. M81-infected LCLs were used for fractionation and two independent LCLs are shown. The signal intensity is given relative to the signals measured in fraction 8. The GAPDH gene was used for normalization of the measured Ct values.

3.5. Putative EBV proteins are associated with ribosomes

The ribosome profiling approach further enabled the analysis of putative EBV proteins where the definitive presence of a protein product remains to be demonstrated. These transcripts are derived from the EBV BART locus (Fig. 3.18a). This locus encodes a family of highly alternatively spliced and polyadenylated transcripts. The first four BART introns of this locus are processed into miRNAs. The leftward directed transcripts (RPMS1, RPMS1A and A73) have been subject of debate for a long time. Recombinant protein is produced in *in vitro* experiments but none of the putative protein products have been detected in natural infection^{17,18}.

The rightward directed LF1 gene has so far only been identified on the RNA level from RNA sequencing experiments during lytic replication of induced Akata cells. For LF2, the function of the protein has been established and is included here as reference for a protein-coding transcript expressed from this locus¹⁹⁻²¹.

The LF2 gene shows the strongest ribosome coverage in its coding region among these four genes (Fig. 3.18b). It also shows the lowest out:in ratio (Fig. 3.18c). The 5'leader:AUG ratio is also low (Fig. 3.18d). This indicates that the LF2 transcripts have more ribosomes within the coding ORF than in the 5'leader. In comparison, the LF1 gene shows high ribosome coverage within its 5'leader in the cycloheximide-treated samples (Fig. 3.18b). This leads to a high out:in ratio (Fig. 3.18c). Interestingly, this changes in the harringtonine-treated samples. Here, the 5'leader:AUG ratios are the lowest among the BART locus-derived genes (Fig. 3.18d). A closer analysis of the harringtonine-treated samples also shows a clear accumulation of reads on its start codon (data not shown), indicating that this gene might be translated.

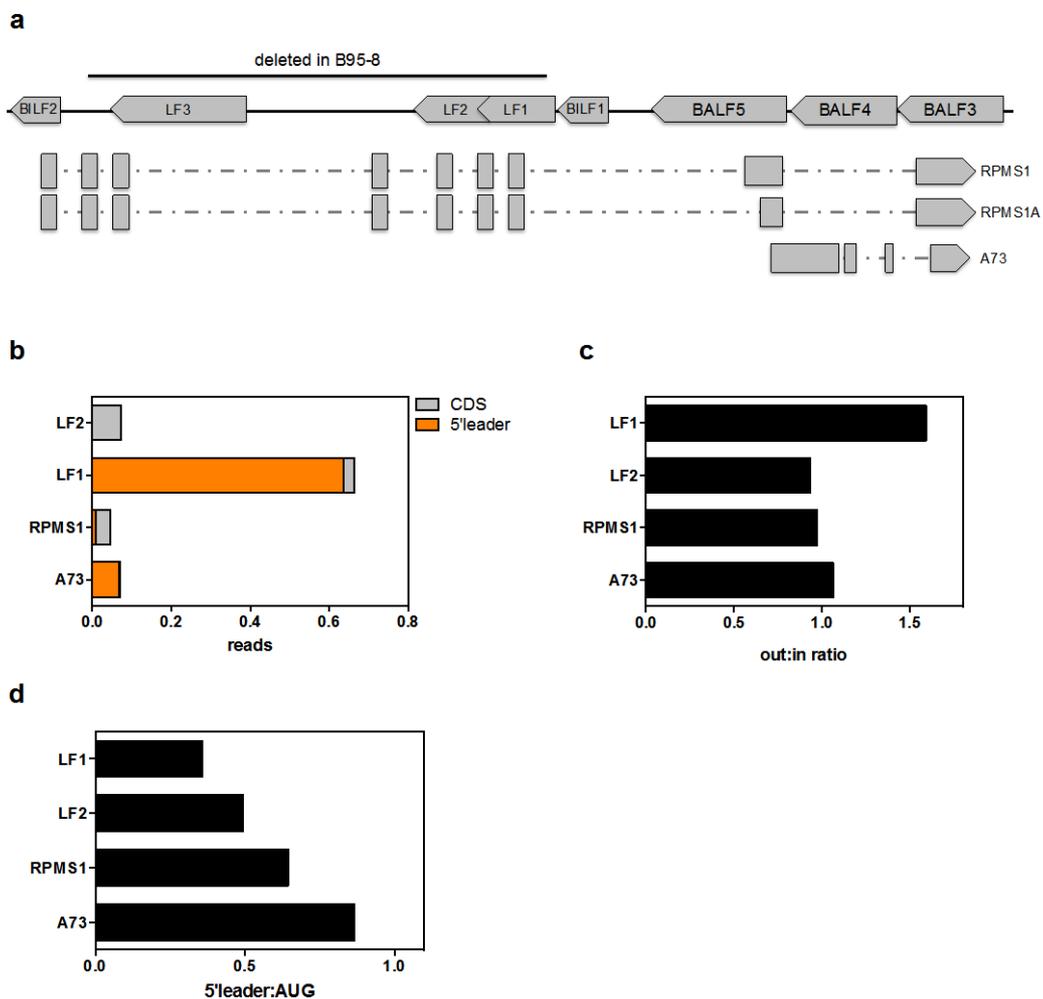


Figure 3.18: Ribosome profiling of BART transcripts. a. A schematic overview over the EBV BART region with the encoded genes is shown. The deletion in the B95-8 strain is indicated. Arrowheads indicate transcription direction of the genes. Image is not to scale. b. Read counts and distribution within the annotated ORFs of transcripts derived from EBV's BART locus. Only the results of M81-infected LCLs are shown as a large part of this locus is deleted in B95-8. c. Out:in ratios calculated from cycloheximide-treated samples. d. Ratios of reads mapping to the 5'leader versus the annotated AUG start codon of the respective genes are plotted. Harringtonine-treated samples were used for calculation of the ratios.

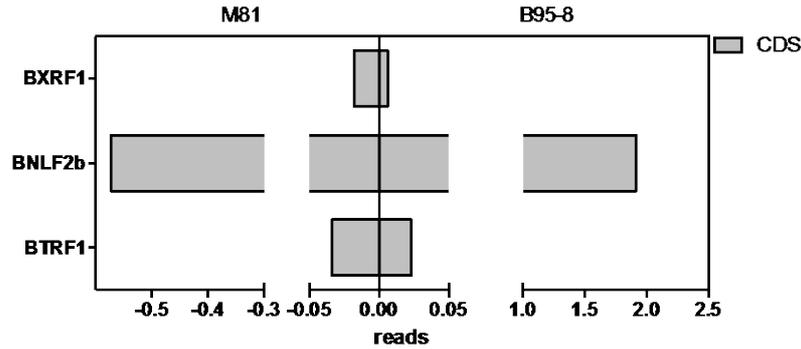


Figure 3.19: Ribosome coverage of poorly characterized EBV transcripts. Read distribution of poorly characterized EBV lytic transcripts. Data is derived from cycloheximide-treated libraries. BLLF2 was not included in the analysis due to extensive overlap with the BLLF1 gene. Only coding DNA sequence (CDS) coverage is shown as the 5' leaders of BXRf1 and BNLF2b overlap with other viral coding genes.

Furthermore, three EBV putative proteins, so far only described at the RNA level and with unknown functions, were also analysed (Fig. 3.19). All three genes are described as lytic genes and are known to be induced upon lytic reactivation but are detected in both strains in the ribosome profiling libraries. Of these three genes, BNLF2b shows strongest ribosome coverage in both strains. The other two genes are present at significantly lower levels. Interestingly, while BXRf1 and BTRF1 show higher coverage in M81-derived libraries as would be expected for lytic cycle genes, the BNLF2b gene shows a 3-fold higher per nucleotide ribosome coverage in B95-8. To assess the relevance of this observation, RT-qPCR measurements were performed for BNLF2a and BNLF2b transcripts in LCLs transformed by either of the two viruses (Fig. 3.20). BNLF2a is a lytic gene involved in immune evasion and has recently been reported to be expressed also in latently infected cells⁹. Indeed, both genes were detected in B95-8-infected LCLs at significantly lower levels than in M81 (Fig. 3.20). However, the ribosome profiling data indicates that BNLF2b is more efficiently translated in the B95-8 strain.

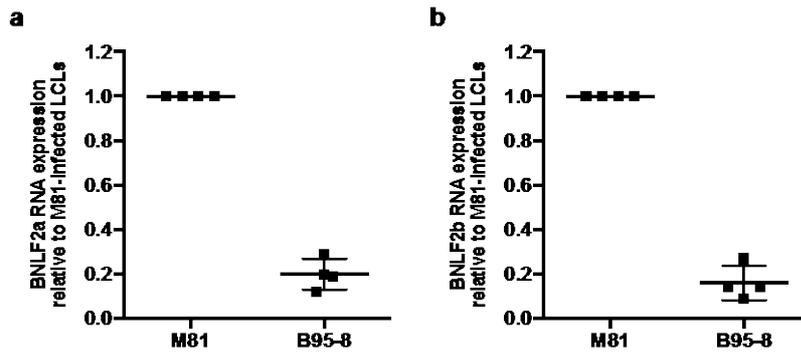


Figure 3.20: BNLF2 mRNA transcripts are expressed in B95-8-transformed LCLs. a. RT-qPCR analysis for BNLF2a transcripts in total RNA purified from M81- and B95-8-transformed LCLs. Results are plotted relative to the expression level of M81-infected LCLs. b. RT-qPCR analysis for BNLF2b transcripts in total RNA purified from M81- and B95-8-transformed LCLs. Results are plotted relative to the expression level of M81-infected LCLs.

3.6. Ribosome coverage of newly identified EBV-derived polyA⁺ RNA transcripts

Several recently published mRNA sequencing experiments of EBV-infected LCLs induced into lytic replication by B cell receptor crosslinking have identified about 300 novel virus-derived transcripts²²⁻²⁴. In these studies, the EBNA2 and EBNA3 family of proteins have been found to give rise to antisense transcripts. Of these, the protein encoded by the mRNA antisense to the encoded EBNA3C transcript has also been detected in independent mass spectrometry experiments²⁵.

The cycloheximide-treated samples, derived from the cells transformed by the lytically replicating strain M81, show extensive ribosome coverage antisense to the EBNA3C gene (Fig. 3.21a). This is not observed in the B95-8-transformed LCLs (Fig. 3.21b). On the other hand, the harringtonine-treated samples do not indicate translation initiation at the predicted ORF start²⁴ (data not shown). Antisense ribosome coverage is also observed for EBNA3A and EBNA3B in M81-infected LCLs (data not shown). However, all EBNA3 transcripts show RPFs located upstream and downstream of the putative antisense-ORFs as well, indicating that further proteins are encoded here.

There is no significant ribosome coverage of the antisense EBNA2 transcripts in M81 (Fig 3.22a). This is in line with O'Grady et al.'s prediction that the EBNA2 antisense transcript is most likely non-coding²⁴.

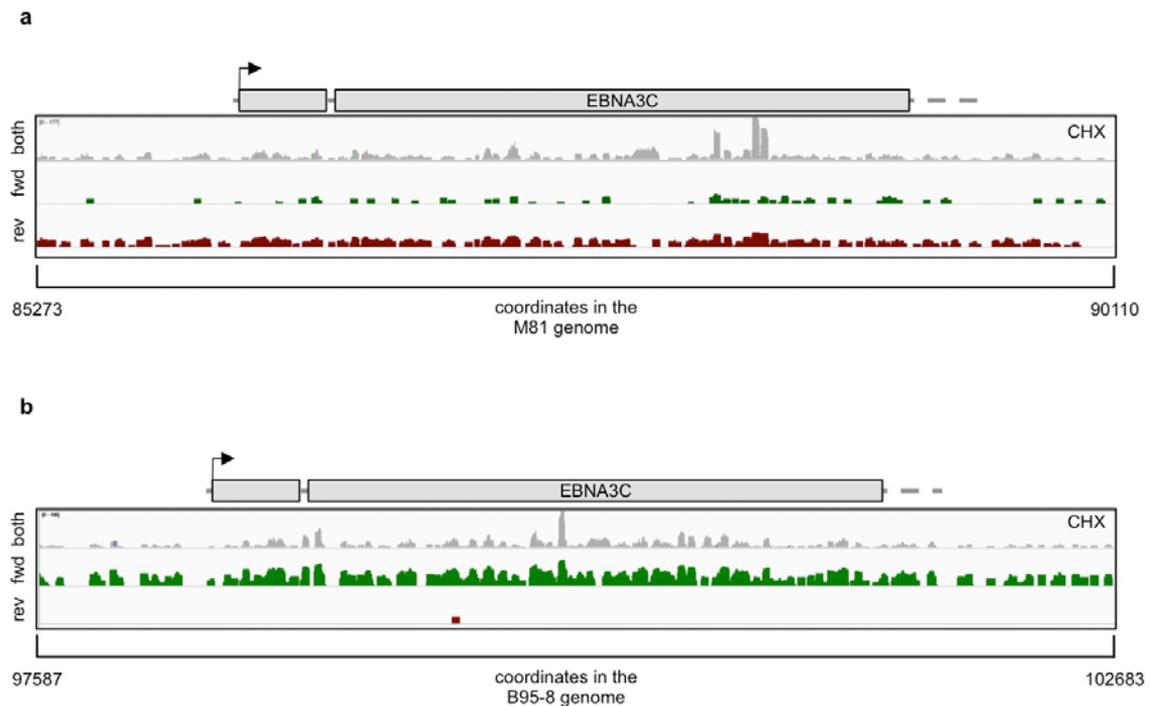


Figure 3.21: Antisense-EBNA3C shows ribosome coverage in the lytically replicating strain M81. a. Overview over the EBNA3C genomic region in the M81 strain. The grey track shows the combined reads mapping to this region. *Below:* The coverage tracks of the cycloheximide-treated library are separated according to strand direction of the mapped RPFs. The green track shows reads that are oriented in the rightward direction (defined as the forward (fwd) strand) and the red track shows reads that are oriented in the leftward direction (defined as the reverse (rev) strand). b. The B95-8 cycloheximide-treated library is shown with parameters as in a.

Many of the novel genes described by O'Grady *et al.* are alternatively spliced isoforms of previously annotated EBV genes²⁴. Due to the short fragment size of RPFs the ribosome profiling data cannot distinguish between the different protein isoforms of these genes.

Even though the cycloheximide-treated library of M81 shows extensive bidirectional ribosome coverage almost over the entire EBV genome (Fig. 3.8) the harringtonine-treated libraries do not show translation initiation sites that are present in all biological replicates.

Furthermore, EBNA1 has previously been described to encode an alternate protein within its ORF in a nested open reading frame^{26,27}. Here also, the ribosome profiling data could not confirm the translation of this ORF (data not shown).

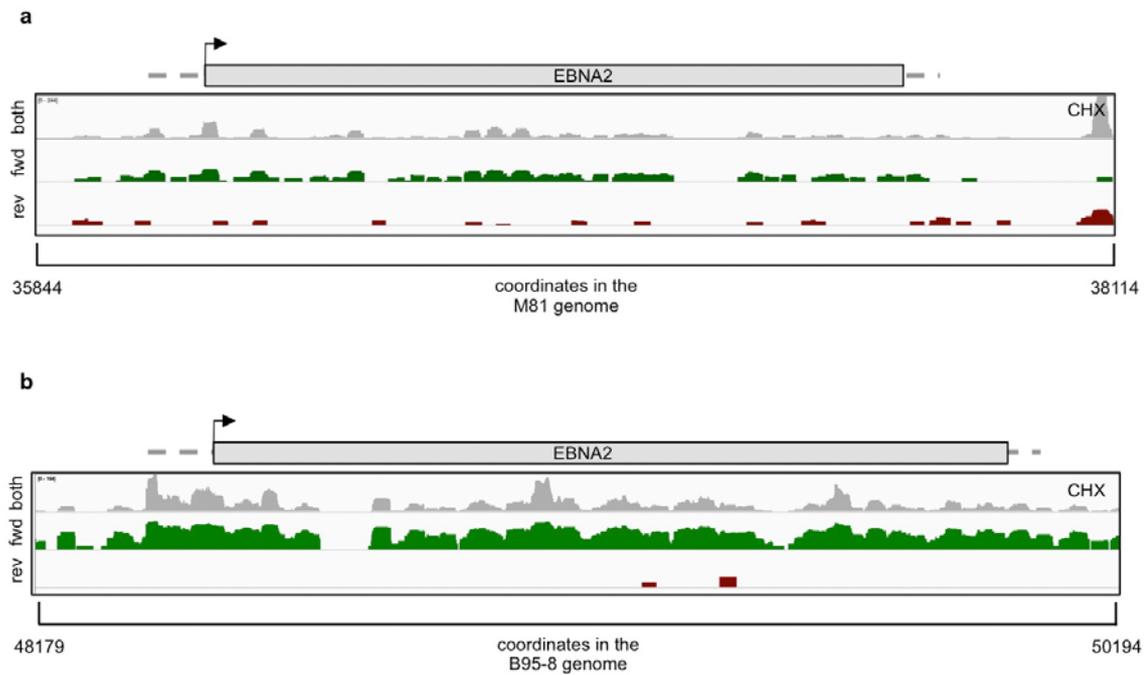


Figure 3.22: Antisense-EBNA2 is not significantly associated with ribosomes in M81. a. Overview over the EBNA2 genomic region in the M81 strain. The grey track shows the combined reads mapping to this region. *Below:* The coverage tracks of the cycloheximide-treated library are separated according to strand direction of the mapped RPFs. The green track shows reads that are oriented in the rightward direction (defined as the forward (fwd) strand) and the red track shows reads that are oriented in the leftward direction (defined as the reverse (rev) strand). b. The B95-8 cycloheximide-treated library is shown with parameters as in a.

3.7. Novel small open reading frames are encoded in the EBV genome

Further analysis of the harringtonine-treated libraries lead to the identification of 28 novel small open reading frames. Three of these were classified as small open reading frames (Table 1). The 25 remaining small ORFs were classified as upstream open reading frames as they were found to be encoded within 5'leaders of viral genes (Table 2).

Table 1: Identification of small ORFs in the EBV genome

| proximal annotated gene | strand | length of the encoded peptide [aa] | gene coordinates in M81 | start codon |
|-------------------------|--------|------------------------------------|-------------------------|-------------|
| BPLF1 | R | 34 | 52694-52798 | AUG |
| LMP1 | L | 18 | 169549-169495 | AUG |
| LMP1 | L | 37 | 169150-169040 | AUG |

aa: amino acids

Moreover, previously described uORFs of several human genes could be confirmed in the data set (see Appendix III). UORFs are short open reading frames localized in the 5' leaders of mRNA transcripts preceding the main ORF (mORF). They are thought to mainly function as repressors of downstream translation although in some cases uORFs have been shown to mediate reinitiation of translation at the downstream encoded mORF^{28,29}.

Ribosome profiling identified several of these regulatory elements in the EBV genome using the harringtonine-derived data set. Not all identified uORFs were present in both strains. An overview over their localization within the EBV genome in the different strains is given in Fig. 3.23 and their characteristics are summarized in Table 2.

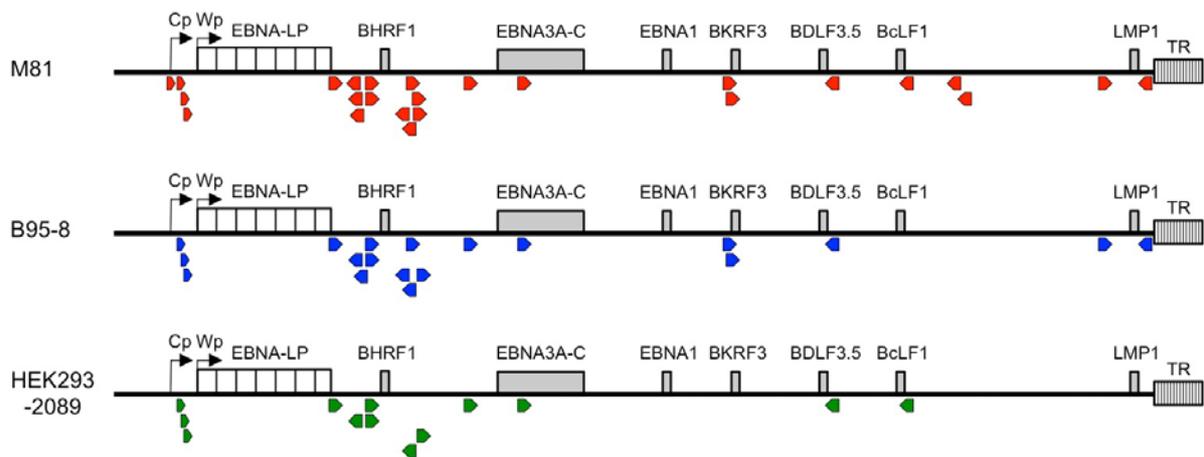


Figure 3.23: EBV encodes upstream open reading frames. A schematic depiction of the distribution of novel small open reading frames across the EBV genome in the different samples. Figure is not to scale.

All latent genes were found to encode at least one uORF in their 5' leader and several of the lytic genes were also found to utilize this type of gene expression regulation. The majority of the uORFs do not contain AUG as an initiation codon and among the non-canonical initiation codons CUG predominates (Table 2).

Table 2: Identification of uORFs in the EBV genome

| proposed nomenclature | length of the encoded peptide [aa] | gene coordinates in M81 | potentially regulated genes | start codon |
|-----------------------|------------------------------------|-------------------------|-------------------------------|-------------|
| Cp uORF1 | 1 | 11250-11255 | EBNA-LP & other EBNA's (Cp) | AUG |
| Cp uORF2 | 2 | 11339-11347 | EBNA-LP & other EBNA's (Cp) | AUG |
| Cp uORF3 | 1 | 11405-11410 | EBNA-LP & other EBNA's (Cp) | AUG |
| Cp uORF4 | 1 | 11451-11456 | EBNA-LP & other EBNA's (Cp) | AUG |
| BHLF1 uORF1 | 35 | 40471-40364 | BHLF1 | CUC |
| BHLF1 uORF2 | 20 | 40628-40566 | BHLF1 | CUG |
| BHLF1 uORF3 | 7 | 40693-40670 | BHLF1 | AUG |
| Y2uORF | 7 | 35675-35698 | EBNA2; BHRF1; EBNA1; EBNA3A-C | AUG |
| BHRF1 uORF1 | 14 | 41769-41813 | BHRF1 | CUG |
| BHRF1 uORF2 | 74 | 41837-42025 | BHRF1 | CUG |
| BFLF2 uORF1 | 12 | 44949-44911 | BFLF2 | CUG |
| BFLF2 uORF2 | 30 | 44996-44904 | BFLF2 | AUG |
| BFRF1 uORF | 10 | 46809-46841 | BFRF1 | UUG |
| BFRF3 uORF1 | 25 | 49357-49434 | BFRF3 | ACG |
| BFRF3 uORF2 | 41 | 49421-49546 | BFRF3 | UUG |
| UuORF | 6 | 55454-55474 | EBNA1; EBNA3A-C | AUG |
| BORF2 uORF | 17 | 64245-64298 | BORF2 | GUG |
| BKRF3 uORF1 | 45 | 97998-98135 | BKRF3 | CUG |
| BKRF3 uORF2 | 27 | 98052-98135 | BKRF3 | CUG |
| BDLF3.5 uORF | 8 | 117094-117068 | BDLF3.5 | AUG |
| BcLF1 uORF | 11 | 125380-125345 | BcLF1 | CUC |
| BXLF2 uORF1 | 24 | 130956-130882 | BXLF2 | UUG |
| BXLF2 uORF2 | 30 | 130913-130821 | BXLF2 | AUG |
| LMP2A uORF | 20 | 166026-166088 | LMP2A | UUG |
| LMP1 uORF | 15 | 169011-168964 | LMP1 | CUG |

uORF: upstream open reading frame

aa: amino acids

In order to further determine the relevance of the identified uORFs the sequence conservation level was determined by comparing the sequences of the uORFs identified by ribosome profiling with 122 sequenced EBV strains published at the time of analysis (Fig 3.24a). The LMP1 uORF conservation level is shown as an example in Fig. 3.24b. The sequenced strains exhibiting polymorphisms are listed below the consensus sequence. Apart from the BHLF1 uORFs the strongest sequence deviation from the consensus sequence was observed in the LMP1 uORF. The conservation level of the uORFs is generally high, indicating functional relevance.

and Y2 exons takes place, followed by splicing to the unique BHRF1 or EBNA genes. The BHRF1 protein is a viral bcl-2 homologue with anti-apoptotic activity just like its cellular counterpart³¹. It was primarily described as a lytic gene transcribed from its own promoter³². Later analyses identified its expression also in latent cells sharing a common 5' leader with the EBNA genes³³. Several uORFs identified in the ribosome profiling data are encoded within the 5' leader of these genes (Fig. 3.25). When analysing the uORF characteristics we also took a closer look at the Kozak consensus sequences surrounding the start codons of the uORFs (Fig. 3.26).

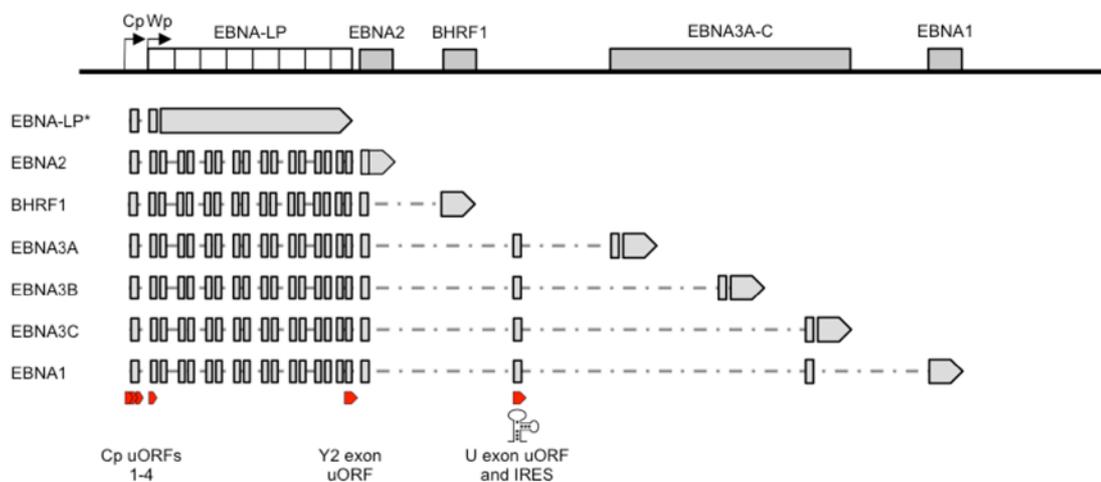


Figure 3.25: The EBNA transcription unit and the localization of the newly identified uORFs. Depicted is an overview over the EBNA transcription unit. The EBNA transcripts which encompass EBNA-LP, EBNA2, EBNA3A-C and EBNA1 as well as the latent form of BHRF1 are expressed from two promoters: Cp and Wp. *EBNA-LP transcription is simplified in this figure and only the longest EBNA-LP isoform is shown.

We found that all four of the identified short uORFs in the Cp-initiated EBNA transcripts are embedded within a weak Kozak sequence. Even though their ribosome coverage is quite pronounced in our data, it is unlikely that they will have a profound inhibitory effect on the translation of the downstream encoded EBNA genes because of their close proximity to the 5' cap and the distance to the start codons of the respective main ORFs (Fig. 3.25). The most likely gene these elements could actively regulate are the Cp-initiated EBNA-LP transcripts as the relative distance between the uORFs' termination codons to the mORF's start codon is the shortest. Interestingly, the Cp-initiated EBNA-LP start codon is surrounded by a strong Kozak sequence indicative of efficient translation initiation (**GAGCaugGC**) while the Wp-

initiated EBNA-LP transcripts that are not regulated by a long 5' leader sequence containing uORFs is surrounded by an intermediate Kozak sequence (ACAAaugGG).

| uORF | Kozak sequence strength |
|--------------|-------------------------|
| Cp uORF1 | - |
| Cp uORF2 | - |
| Cp uORF3 | - |
| Cp uORF4 | - |
| BHLF1 uORF1 | - |
| BHLF1 uORF2 | + |
| BHLF1 uORF3 | - |
| Y2 uORF | + |
| BHRF1 uORF1 | - |
| BHRF1 uORF2 | + |
| BFLF2 uORF1 | - |
| BFLF2 uORF2 | - |
| BFRF1 uORF | - |
| BFRF3 uORF1 | + |
| BFRF3 uORF2 | + |
| U uORF | + |
| BORF2 uORF | + |
| BKRF3 uORF1 | + |
| BKRF3 uORF2 | + |
| BDLF3.5 uORF | + |
| BcLF1 uORF | + |
| BXLF2 uORF1 | - |
| BXLF2 uORF2 | - |
| LMP2A uORF | - |
| LMP1 uORF | - |

| | |
|----|-----------------------------|
| - | weak Kozak sequence |
| + | intermediate Kozak sequence |
| ++ | strong Kozak sequence |

Figure 3.26: Kozak consensus sequence strength of the identified viral uORFs. Summarized is the strength of the Kozak consensus sequence surrounding the uORF start codons.

The Y2 uORF is surrounded by a strong Kozak sequence and it is in closer proximity to the start codons of genes it might regulate. The shortness of the uORF, which is only 7 amino acids long indicates that it may not completely abolish translation of the downstream encoded mORFs. The Kozak consensus sequences of the downstream encoded genes range from strong (EBNA3C) to weak (EBNA1). The distance between the Y2uORF termination codon and the start codon of the BHRF1 gene is short and indicates a strong regulatory function of the uORF for this protein. The BHRF1 ORF is present on several different mRNA isoforms. The latent BHRF1 ORF-containing transcripts are initiated from the Cp and Wp. The lytic BHRF1 transcripts

initiate from a more proximal lytic promoter. This endows the ORF with a different 5' leader sequence that harbours two alternate uORFs where one of them can further give rise to two different peptide products depending on the splicing of the BHRF1 5' leader. To complicate things further, BHRF1 has recently been described to be produced from a further mRNA isoform expressed during lytic reactivation²⁴. This transcript is produced from alternative splicing of W1-W1 exons instead of the conventional W1-W2 repeat exon splicing which endows the BHRF1 transcript with a shorter 5' leader.

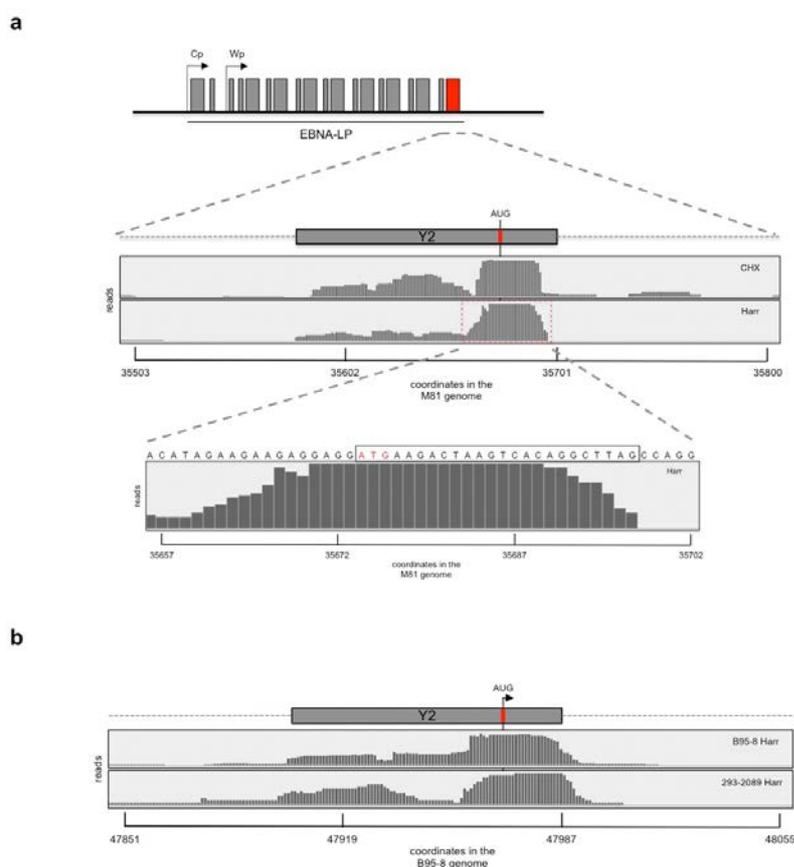


Figure 3.27: The Y2 exon encodes a 7 amino acid long uORF. a. *Upper panel:* A schematic overview over the Cp- and Wp-initiated EBNA-LP protein is shown with the W1 and W2 repeat exons depicted as grey rectangles. Highlighted in red is the Y2 exon which harbours the stop codon of the EBNA-LP reading frame and the Y2uORF in a shifted reading frame. *Middle panel:* The close-up figure shows the Y2 exon with the AUG start codon of the Y2uORF highlighted in red. The read coverage in representative M81-transformed ribosome profiling libraries is shown after cycloheximide treatment alone or with harringtonine. *Lower panel:* A close-up view of the red rectangle from the middle panel is shown displaying the ribosome coverage in detail for the Y2uORF. The nucleotide sequence of the uORF is boxed in black with the start codon highlighted in red. b. Representative ribosome coverage from the B95-8-transformed LCLs (upper panel) and HEK293-2089 cells (lower panel). RPF coverage is shown for harringtonine (Harr)-treated samples.

In addition to the short uORFs close to the 5' cap of the Cp-initiated EBNA transcripts and the downstream spliced Y2uORF, the EBNA1 and EBNA3 family mRNA transcripts further encompass the U exon further downstream in the EBV genome. This exon harbours an additional uORF that is placed upstream of the start codons of these respective genes. This UuORF is surrounded by a strong Kozak sequence and is similarly short as the Y2uORF (6 aa). It is further localized within an IRES sequence reported to direct EBNA1 translation³⁴ (Fig. 3.25 and 3.28). An IRES is a *cis*-acting RNA structure and offers the cell an alternative to cap-dependent translation initiation especially during stress conditions³⁵. A complex RNA secondary structure is formed that can directly bind the 40S subunit of the ribosome.

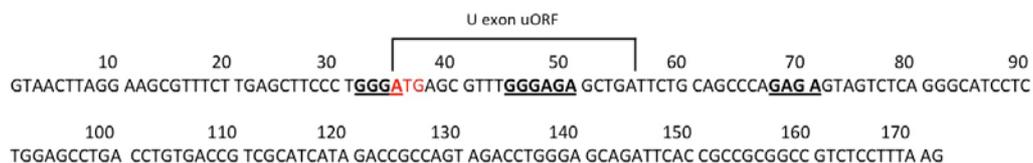


Figure 3.28: The U exon encodes an uORF within an IRES element. The nucleotide sequence of the U exon from the M81 strain is shown. The IRES consensus sequences identified by Isaksson et al.³⁴ are indicated in bold and underlined letters. The start codon of the uORF is highlighted in red.

3.7.2. Viral protein translation is regulated by upstream open reading frames

To verify that the small ORFs identified in the 5' leaders of several viral genes were indeed uORFs and that these did exert a regulatory role on the downstream encoded main ORF, six of the uORFs were studied in reporter assays. The 5' leader sequences were synthesized by Eurofins Genomics and were re-cloned upstream of the *Firefly* luciferase (FLuc) gene for reporter activity measurements (Fig. 3.29a). The Y2uORF was chosen as it showed strong ribosome accumulation in the M81 strain (Fig. 3.27a) and in the B95-8-derived libraries (Fig. 3.27b). The intergenic sequence between the uORF and the EBNA2 gene in the M81 virus background was cloned between the Y2uORF and the luciferase start codon.

The complete U exon comprising the UuORF and IRES sequence were cloned into the FLuc vector.

LMP1 contained a weak uORF that is very close to the annotated start codon (2 nt between uORF stop and LMP1 start). The LMP1 start codon of M81 is surrounded by a very strong Kozak sequence but nevertheless a distinct peak accumulates at the

start codon of the uORF in the harringtonine treated libraries. On the other hand, the 5'UTR of the B95-8 LMP1 gene contained the same uORF but less reads were visible in the harringtonine treated libraries on its start codon. Here, the majority of reads accumulated on the annotated start codon of LMP1. B95.8 contained 106x more reads on the LMP1 start codon compared to the uORF while in the M81 strain, there were only 2.1x more reads on the LMP1 start compared to the uORF (data not shown).

The BFLF2 ORF is, according to our analyses, regulated by two uORFs. Both uORF stop codons end close to the BFLF2 start codon. We chose to mutate the more proximal uORF here.

The BORF2 uORF is also surrounded by a strong Kozak sequence and exhibits close proximity to the translation start site of the mORF.

BKRF3 encodes two uORFs that extend beyond the start codon of the gene. A 5'leader was synthesized that had both uORFs deleted.

The different luciferase constructs were transfected into HEK293 cells alongside a *Renilla* luciferase (RLuc) encoding transfection control vector and luciferase activity measurements were performed 24h following transfection. Deletion of the Y2uORF start codon in the luciferase constructs lead to an approximately 2-fold increase in signal intensity arguing for a repressive function of this uORF (Fig. 3.29b). The UuORF with the IRES element included did not show any difference in luciferase assays. The deletion of the LMP1 uORF lead to a 1.5-fold increase in luciferase expression. The deletion of the BFLF2 uORF was comparable to the Y2uORF deletion with a 2-fold increase in FLuc expression. The uORFs of the other two lytic genes studied did not show an increase in FLuc expression upon deletion.

These results confirm that the identified uORFs, at least in part, act to repress the downstream encoded mORF.

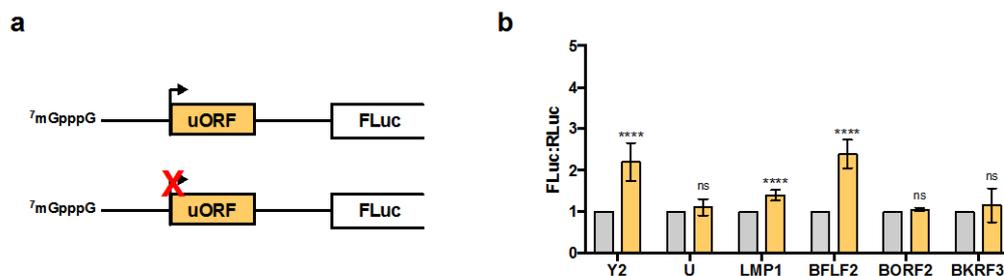


Figure 3.29: Mutational analysis of viral uORFs in dual luciferase reporter assays. a. Schematic depiction of the *Firefly* luciferase (FLuc) reporter constructs with different viral 5'leaders containing uORFs cloned upstream of

the luciferase gene. b. FLuc activity measurements normalized to co-transfected *Renilla* Luciferase (RLuc). Error bars show the standard deviation of at least six independent transfections. The paired-end t-test was used to calculate the p-value. Asterisks summarize the p-value: **** < 0.0001. ns: not significant.

3.8. Alternative translation initiation

As already highlighted above, ribosome profiling in the presence of harringtonine is a powerful tool to map translation initiation sites on a given transcript². Apart from the identification of small ORFs and regulatory uORFs in the viral genome the underlying data set also enabled the identification of alternative translation initiation sites (aTIS) within annotated viral genes (Table 3). The majority of identified aTISs were N-terminal truncations of annotated viral proteins. Typically they were initiated at non-canonical (non-AUG) start codons. This has been described to occur preferentially in these situations^{36,37}.

Table 3: Viral genes with alternative translation initiation sites

| Gene | start codon | gene coordinate in M81 | location relative to the annotated start codon [aa] |
|--------|-------------|------------------------|---|
| BFRF3 | CUG | 49428 | -14 |
| BLRF2 | AUG | 76939 | +17 |
| BFRF1a | AUG | 46637 | +51 |
| BKRF2 | CUG | 97998 | +93 |
| BKRF2 | CUG | 98052 | +111 |
| BKRF3 | CUG | 98417 | +101 |
| BKRF3 | CUG | 98570 | +151 |
| BALF1 | AUG | 164920 | +39 |

aa: amino acids

As an example, the BKRF3 gene is shown in Fig. 3.30. The ORF is preceded by two uORFs in its 5'leader that extend beyond the start codon of BKRF3. This has previously been described to promote the expression of N-terminally truncated protein isoforms³⁸⁻⁴⁰. This might also be the case for BKRF3 where pronounced translation initiation at two downstream CUG codons can be observed. However, in the luciferase assays described in the previous section no increase in FLuc activity was detected upon deletion of the two uORFs.

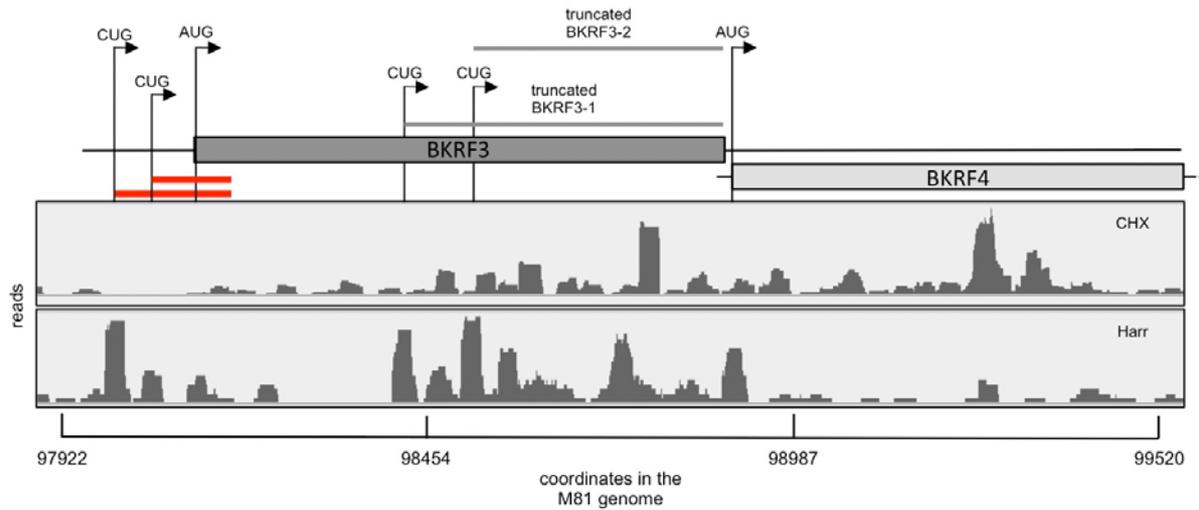


Figure 3.30: The BKR3 ORF is regulated by two uORFs and contains two internal alternative translation initiation sites. The schematic arrangement of the genomic region around BKR3 is shown with the two uORFs starting in the BKR3 5'leader highlighted in red. The two alternative translation initiation sites are indicated. The panel below shows the ribosome coverage of this region in the cycloheximide-treated library (CHX). The panel below that shows the ribosome coverage of this region from the 5 min harringtonine-treated library (Harr).

4. Discussion

This is the first study that analyses global translational processes in EBV-infected cells. This analysis on the nucleotide level has identified novel small open reading frames and alternative translation initiation sites. Moreover, the study is based on the physical association of ribosomes and thereby provides experimental data to identify these ribosome-associated regions rather than relying on *in silico* prediction methods.

The computational scoring of translated ORFs and initiation sites described so far perform badly when analysing ORFs with low ribosome coverage¹⁻⁵. The percentage of viral transcripts detected in LCLs was low compared to the cellular transcriptome. This results in significantly lower ribosome coverage of viral transcripts as well. This was particularly true for the majority of lytic transcripts analysed in the M81-transformed LCLs as the percentage of lytically replicating cells was below 10%. Therefore, the computational analyses published for ribosome profiling data were not applicable in our case. As a consequence, manual screening of the viral genome was performed. Partial manual annotation of novel ORFs has also been done for strongly lytically replicating viruses such as human cytomegalovirus (HCMV) and Kaposi's sarcoma herpesvirus (KSHV)^{5,6}.

A recent publication has shown that for the transcripts with low ribosome coverage, the translation inhibitor pateamine offers a more robust way of studying translation initiation sites by ribosome profiling compared to harringtonine⁷. This drug could be used as an alternative to harringtonine in future studies on EBV translation.

4.1 Ribosome profiling reveals differences in translation of cellular genes in two different EBV strains

B95-8-transformed LCLs showed a broader expression of cellular genes. This could be due to the host shut-off mechanism of lytically replicating cells mediated by the lytic cycle protein BGLF5 in M81-infected cells⁸. However, only 3-6% of M81-infected LCLs are positive for BZLF1 during lytic replication *in vitro*, making it unlikely that this small percentage of replicating cells would influence cellular protein translation so profoundly in the total library. Another cause could be the slower cell cycling kinetics observed in M81-infected LCLs compared to B95-8 (data not shown). B95-8-infected LCLs are also metabolically more active (data more shown). These two attributes of B95-8 could lead to a broader gene expression profile compared to M81.

Interestingly, genes that were associated with translation were more strongly represented in the M81-derived libraries. This leads to the assumption that translation is more directly influenced by EBV in the M81 strain. For example, the lytic cycle protein SM has been reported to alter translation and preferentially stimulate translation of its target genes⁹.

4.2 Ribosome profiling allows for the analysis of annotated viral proteins and the identification of novel viral genes

The global bidirectional association with ribosomes of almost the entire EBV genome in M81-infected LCLs is in line with O'Grady *et al.*'s report of global bidirectional transcription in lytically replicating cells^{10,11}. The transcripts they have identified in their transcriptome studies seem to be translated to a large part as well. This is at least observed in the cycloheximide-treated samples. The harringtonine data did not confirm any translation initiation on the newly identified transcripts that were consistent between biological replicates. The ribosome profiling data confirmed that antisense-transcripts generated from the EBNA3 family of genes are extensively associated with ribosomes in the M81-infected LCLs. Yet again, harringtonine data could not confirm initiation at the presumed start codons. The EBNA3A and EBNA3B antisense genes have been predicted by O'Grady *et al.* to be nuclear which they have also demonstrated by fluorescence *in situ* hybridization analysis¹⁰. The ribosome profiling data here suggests that these genes are actually translated. At the same time the global bidirectional ribosome coverage was selective, as was shown for the antisense EBNA2 gene. Here, both strains showed hardly any ribosome coverage supporting the claim that this region is not translated¹⁰.

Previous publications have used machine learning approaches to identify novel protein-coding genes using ribosome profiling⁴. Again, due to the low coverage of viral genes this approach is not reliable for this data set and it cannot be excluded that genes have been overlooked during the manual screening of the viral genome.

For future studies on translation of the newly identified genes, an EBV strain should be used that exhibits higher lytic replication levels. Unfortunately, EBV strains that spontaneously replicate *in vitro* are rare and M81 is a strain with the highest lytic replication levels observed in cell culture so far. An alternative would be to use a strain that can be induced into lytic replication to high levels, such as the Akata strain used by O'Grady *et al.*¹⁰. It is a more artificial system to study lytic replication but

would lead to a higher ribosome coverage of viral genes and allow for a more precise annotation of translated transcripts.

Surprisingly, the B95-8-infected LCLs showed evidence for the translation of several lytic cycle proteins even though the strain is known to be in latency III *in vitro*. BMRF1 expression in a few cells further showed that the proteins that were shown to be associated with ribosomes in the ribosome profiling experiments were actually translated and yielded protein product. BMRF1 is an early lytic gene and functions as the viral DNA polymerase processivity factor¹²⁻¹⁴. Its transcription is induced by BZLF1 and BRLF1¹⁵. The detection of BMRF1 RNA in RT-qPCR experiments with M81 Δ ZR-infected LCLs argues for an additional BZLF1 and BRLF1-independent expression of this gene.

BMRF1 has also been reported to act as a transcriptional regulator that enhances BALF2 promoter activity as well as activates BHLF1 transcription^{12-14,16,17}. Therefore, it is not unlikely that the protein is also expressed independently of its transactivator BZLF1.

4.3 Ribosomes associate with non-coding RNAs

The peaks observed on small non-coding RNAs most likely stem from protection of this RNA species from RNase digestion by associated proteins, such as the RNA-induced silencing complex. Interestingly, the peaks observed in the different libraries shift depending on which inhibitor is used. This is said to be indicative of true translation. Intriguingly, Laressergues et al. have described that miRNA precursors in plants encode short peptides that regulate subsequent miRNA processing¹⁸. Further validation is needed to determine whether the observed peaks are signs of true translation and whether EBV, and by extension mammalian cells, utilize this mechanism to regulate miRNA processing.

BHLF1 and LF3 show a high number of ribosome footprints. This is, on the one hand, in agreement with the previously reported high transcription level of these genes upon induction of lytic replication¹⁹⁻²¹. On the other hand, no definitive protein product has been detected in EBV-infected LCLs so far²²⁻²⁴.

The 5' leader of BHLF1 accumulates many RPFs which in part can be ascribed to the presence of three uORFs. Furthermore, the BHLF1 RNA is very GC-rich²⁵ and RNA hairpins form in GC-rich sequences that are inhibitory to translation. This is especially

the case when the GC-rich region is localized close to the 5'cap as is the case for BHLF1^{26,27}.

BHLF1 resembles the HCMV Beta2.7 RNA. This is a long RNA transcript that is generated from a repeat region within the HCMV genome similar to BHLF1^{25,28}. The Beta2.7 RNA transcript has sparked similar debate over its coding capacity. On the one hand, a putative ORF was identified in a laboratory strain of HCMV²⁹. However, sequencing of additional strains showed that this region was disrupted in four isolates analysed in a separate study, arguing against a peptide-dependent role of this gene³⁰. The putative ORF of BHLF1 is also not present in all EBV strains³¹.

Recently, the association of the HCMV Beta2.7 RNA with ribosomes was demonstrated by ribosome profiling⁵. Here, the RNA was predicted to encode several small ORFs. In a follow-up study, the group demonstrated that T cell responses against some of the small ORFs presumed to be translated from this long RNA transcript could be detected in the serum of HCMV⁺ individuals³. However, earlier reports have demonstrated the inhibitory effect of the 5'leader sequence of the Beta2.7 RNA on its own translation^{29,32}. Interestingly, one study reported the inhibitory effect to be mediated by the complex structure of the 5'leader²⁹ while the other identified two uORFs that mediated the translational repression³². Taking all these studies into consideration, it seems that the Beta2.7 RNA of HCMV does not encode an essential protein, but that the transcript is associated with the host cell translation machinery and can be translated under certain conditions. The same might hold true for BHLF1.

Recent studies have categorized BHLF1 as a viral long non-coding RNA (lncRNA). Rennekamp et al. have shown that the deletion of the transcription start site of BHLF1 leads to a drop in lytic replication levels in these cells while the mere silencing of the start codon of the putative BHLF1 ORF does not show any effect²⁵. This is supported by the polysome profiling experiments performed in my work. The BHLF1 mRNA was preferentially localized in the lighter fractions of the sucrose gradient. This argues that BHLF1 is not translated in LCLs. As a lncRNA, BHLF1 is hypothesized to function in *cis* at the OriLyt by interacting with the DNA and forming a stable RNA-DNA helix also termed R-loop. This generates a local single-stranded DNA region that is utilized for lytic replication initiation²⁵.

Previous studies on cellular lncRNAs have suggested their association with ribosomes confers a cellular mechanism to regulate their stability, subcellular

localization and degradation by triggering nonsense-mediated mRNA decay (NMD)³³⁻³⁵. NMD is frequently triggered by uORFs of which BHLF1 has three^{36,37}. Furthermore, Mendell *et al.* have shown that the knockout of the Rent1 protein which is essential for NMD function, leads to upregulation of specific mRNA transcripts that contain uORFs³⁸. Additionally, Smith *et al.* have shown that inhibition of NMD in cells increases the expression level of some lncRNAs that have been shown to be associated with ribosomes, arguing for the utilization of translation as a degradation signal for this RNA class³⁹. The BHLF1 protein product has been detected in LCLs chemically induced into lytic replication^{22,23}. Interestingly, the Western blots in both publications exhibit a ladder of protein products indicating extensive degradation of the protein. This suggests that if the BHLF1 transcript is indeed translated under certain settings, then the protein is highly unstable.

LF3 closely resembles BHLF1 and it is believed that the two genes have arisen out of gene duplication⁴⁰. It is the positional homologue of BHLF1 at the second OriLyt. For LF3 also, no consistent protein product was observed in Western blot but rather a protein ladder⁴¹.

4.4 Putative EBV proteins are associated with ribosomes

The LF1 and LF2 open reading frames have been described early on⁴². Using the Akata EBV strain, LF2 was identified as an early lytic gene of EBV⁴³. It has since been shown to fulfil different roles in the infected cell with different protein interactions being involved, such as the viral BRLF1 protein or the cellular IRF7⁴⁴⁻⁴⁷. LF1 shows some homology to LF2 and both genes are thought to have arisen from gene duplication of the viral BLLF3 gene⁴⁸. LF1 transcripts have been detected in different sequencing data and the gene is listed as a lytic gene. When expressed as a GST-fusion protein *in vitro*, it was identified as binding to SUMO residues⁴⁹.

As is common to many EBV lytic genes, LF1 has a short 5'leader sequence. Nevertheless, in the cycloheximide-treated library, many reads are localized to the 5'leader and only a low number of ribosomes are detected within the ORF. In the harringtonine-treated samples, this changes as is indicated in the low 5'leader:AUG ratio. The ratio is lower than for the translated LF2 gene. The cause for this drastic rearrangement of ribosomes upon harringtonine treatment is unknown. It could be that the LF1 ORF is not translated very efficiently and that this is reflected by the high number of reads detected in its 5'leader.

The BART-derived transcripts are another group of genes that have sparked debate over their translational status. BART transcripts are generated from a distinct genomic region in the viral genome and are characterized by extensive alternative splicing^{50,51}. In my analysis, only A73 and RPMS1 were included because these transcripts can be distinguished from one another. The BART transcripts have been shown to exhibit a predominantly though not exclusive nuclear localization⁵².

It is impossible to make assumptions about the translational state of these transcripts based on the ribosome profiling data alone. The coverage of these transcripts in all libraries is amongst the lowest of all viral genes. BART transcripts are not highly expressed in LCLs. On the other hand, BART transcripts have been reported to be expressed at very high levels in epithelial cells. A more precise statement over the translational status of these transcripts could perhaps be obtained from profiling of EBV-infected epithelial cells.

Additionally, BART miRNAs are encoded and processed from these transcripts⁵³. Due to a correlation between BART RNA expression and BART miRNA levels, it is believed that the BART transcripts are generated in order to produce the BART miRNAs rather than the putative proteins^{53,54}. The ribosome association of these transcripts could be explained with the same argument as for the lncRNAs: the association allows for the removal of these transcripts via the NMD pathway after miRNA processing. Small nucleolar RNAs (snoRNAs), for example, are processed from longer RNA transcripts. After processing of the snoRNAs, the transcripts are translated and this triggers their degradation through NMD³⁴. This allows the removal of the mature RNA after the biologically relevant snoRNAs have been excised from the introns of the message.

4.5 The EBV genome encodes small and upstream open reading frames

The ribosome profiling analysis of EBV-infected LCLs has identified three small open reading frames in the viral genome. ORFs are historically defined with a minimal length of 100 amino acids for functional protein generation⁵⁵. Small ORFs were long ignored and thought to be by-products and non-functional. But there is rising evidence that they too encode functional proteins and several small peptides have been described that fulfil significant biological roles⁵⁶⁻⁵⁸. The same could hold true for the newly identified viral small ORFs. Furthermore, previous ribosome profiling

studies on HCMV and KSHV have also identified small ORFs of unknown function in these viruses⁵.

EBV depends on the cellular translation machinery to translate its mRNAs and thus is also susceptible to the cell's regulatory networks. Translation enables almost immediate responses to extracellular and intracellular stimuli. Various regulatory elements function in *cis* on the affected mRNA. uORFs are such regulatory elements. Though described decades ago, these uORFs have in the last couple of years become widely appreciated as important fine-tuners of protein translation with relevance in several human diseases⁵⁹⁻⁶¹. In human cells, 49-64% of genes are predicted to be controlled by at least one uORF^{62,63}. They are especially clustered in oncogenes and transcription factors highlighting the importance for strict regulation of these genes⁶⁴. Viruses are also known to use uORFs for the regulation of translation^{5,6,65}.

Generally, uORFs are associated with a reduction in translation level of the mORF⁶². However, depending on the physiological state of the cell they can also enhance the translation of the downstream encoded mORF^{66,67}.

The observation that all of the viral latent genes harbour at least one uORF within their 5'leaders suggests an important role for these small ORFs in the translational control of these genes. Furthermore, the start codons of the uORFs were conserved among all EBV strains studied. This further highlights their importance. Latent genes are critical factors for the transformation of the infected B cell but can also be cytotoxic to cells⁶⁸. Therefore, latent gene expression needs to be tightly controlled.

The strongest differences in ribosome coverage of latent genes were in LMP1 and EBNA2. Both genes have, through ribosome profiling analysis, turned out to be regulated by uORFs. The higher 5'leader:AUG ratio for LMP1 in the M81 strain indicates that the gene is more tightly regulated in this strain. The ribosome coverage of the uORF is higher in M81-infected cells and luciferase assays have further shown that this uORF exerts a repressive role on the downstream encoded ORF.

The uORF-mediated regulation of EBNA2 is difficult to assess because the ribosome profiling data has shown that this gene is regulated by multiple uORFs. Additionally, EBNA2 shares its 5'leader sequence with other genes and it is impossible to determine how many of the RPFs detected in the 5'leader of this transcription unit are on transcripts that encode EBNA2. The Y2uORF which is one of the uORFs encoded upstream of the EBNA2 ORF has shown to have a repressive function in

luciferase reporter assays. However, not all cloned uORF sequences showed an effect on translation in the reporter assays.

The deletion of the UuORF within the IRES sequence did not show any effect on the expression level of the luciferase ORF. This uORF is different in its composition compared to the other viral uORFs identified. It is encoded within an IRES element⁶⁹. The lack of effect could be due to factors missing in the cells used for the reporter assays. That could be factors that are provided by the virus in natural infection. KSHV, for example, has been reported to control the expression of the vFLIP protein through an IRES. The IRES element is only functional in KSHV-infected but not in uninfected cells^{70,71}. Furthermore, it could be that a cellular factor is missing to make the UuORF functional. Mehta et al. have shown that the mere presence of uORFs does not mean that the mORF will be repressed in translation⁷². They studied the mRNA of the Her-2 gene which contains an uORF in its 5'leader⁷³ and showed that the uORF is non-functional in some cancer cells and that this overriding of uORF control is directed from a translational derepression element in the 3'UTR that can bind regulatory proteins⁷².

The next question to ask concerning the UuORF is: why is it embedded within an IRES element? In fact, the combination of an uORF within an IRES has been described in mammalian cells^{74,75}. The cat-1 gene, a stress-regulated arginine/lysine transporter, is translated through an IRES element that requires the translation of an uORF encoded within the IRES. Cellular stress leads to impairment of cap-dependent translation. IRES elements and uORFs have been described to ensure the translation of critical genes during stress responses. For the cat-1 ORF, the translation of the uORF remodels the structure of the RNA and renders the IRES functional. This enables the subsequent translation of the cat-1 ORF^{74,75}. Not only stress impairs cap-dependent translation. During mitosis, cap-dependent translation is also decreased. Here too, IRES-mediated translation is used to synthesize proteins critical for that cell cycle stage⁷⁶⁻⁷⁹.

One of the genes that contain the UuORF/IRES in its 5'leader is EBNA1. EBNA1 is essential for the propagation of the viral genome during mitosis by tethering the viral genome to chromosomes. Therefore, its expression needs to be guaranteed also during mitosis and this might be facilitated by the uORF/IRES in the U exon.

Furthermore, especially the uORFs encoded upstream of lytic genes might only be active during cellular stress. EBV has been reported to induce stress in cells through

expression of specific proteins⁸⁰. Also, the lytic cycle induction is known to exert stress on the infected cell⁸¹. This might explain why not all of the uORFs of the lytic cycle genes studied in the reporter assays showed an effect upon deletion.

The identified uORFs could also have a dual function depending on the physiological state of the cell. For example, Ebola virus uses an uORF to maintain the levels of L protein⁶⁵. Under optimal conditions in the cell the uORF represses translation of the downstream-encoded L protein. When cells are stressed and translation is therefore globally downregulated the uORF manages to uphold expression of L protein.

Another interesting aspect, is the biological relevance of the uORF-encoded peptides. Some of the identified uORFs showed strong conservation between EBV strains also within the coding region. UORFs have been reported to encode peptides that regulate the translation of the mORF. A well studied example is the uORF2 encoded in the 5'leader of the gpUL4 gene of HCMV^{82,83}. The function of this uORF depends on its amino acid sequence. The translated 12 aa long peptide remains associated with the ribosome at the termination codon as a peptidyl-tRNA complex.

Furthermore, uORFs can balance the translation between two different proteins on bicistronic RNAs. The C0 ORF of hepatitis B virus prevents translation of the core protein and promotes reinitiation of translation at the polymerase gene further downstream⁸⁴. The BKRF3 gene of EBV could be a good candidate for this type of uORF-mediated regulation. The BKRF3 gene encompasses alternative translation initiation and uORF-mediated regulation. According to the ribosome profiling data the 5'leader of BKRF3 encodes two uORFs that extend past the start codon of the mORF. Additionally the protein is predicted to contain two alternative translation initiation sites that start with CUG. BKRF3 is a DNA glycosylase protein⁸⁵. During lytic replication the protein is localized in the nucleus in a complex including BALF5, BMRF1 and BRLF1⁸⁶. If BKRF3 is deleted viral DNA replication is defective. A catalytic mutant can restore DNA replication indicating that the enzymatic activity is not essential for this function. Interestingly, the alternative translation initiation sites identified within the BKRF3 ORF either start within or further downstream of the catalytic domain. This indicates that BKRF3 isoforms could be produced in EBV-infected cells that lack enzymatic activity. The luciferase assays did not show an increase in activity upon deletion of these two uORFs. As mentioned in the previous section their functionality might depend on a factor of viral or cellular origin that was not present in the HEK293 cells used for the assays.

We find alternative translation initiation within several other viral genes as well. Most of these alternatively translated protein isoforms have a non-canonical start codon. Non-AUG alternative translation initiation was first discovered in viruses^{87,88}. Alternative translation initiation has proven to be an important factor for the generation of protein isoforms and protein diversity in cells^{4,89}. This leads to proteins differing in their N-terminal domains which in turn can alter their functions or change their subcellular localization^{90,91}. For example, PTEN has been identified to encode a long version of the canonical protein that is termed PTEN-long. This protein is initiated 519 bp upstream of the annotated AUG from a CUG start site. This adds 173 aa to the protein and makes it membrane-permeable. The long form is secreted and can enter into neighbouring cells thereby altering their behaviour⁹⁰.

One EBV N-terminal protein isoform has been described previously⁹². RAZ is transcribed from the same genomic locus as BZLF1. The N-terminal domain of RAZ is generated through a different transcription start site than BZLF1 and alternative splicing. This generates a protein that functions as an inhibitor of BZLF1.

4.6 EBV infection remodels cellular gene expression processes

EBV has been described to influence transcription in infected cells during lytic replication¹⁰. Multiple promoters are used to transcribe genes and extensive stop-codon readthrough is observed that also influences cellular genes. To what extent EBV also alters translation in infected cells is unknown. Several EBV proteins have been described to exert direct or indirect effects on translation and the associated proteins. LMP1, for example, can trigger the unfolded protein response (UPR)⁹³. This is a cellular stress response pathway. Furthermore, LMP1 has also been reported to cause phosphorylation of eIF2 α ⁹⁴. Both UPR and eIF2 α phosphorylation are known to act as global inhibitors of translation. In contrast, LMP1 has recently been described to increase eIF4E expression in NPC cell lines⁹⁵. LMP2A has been reported to modulate PI3K/Akt signaling in cells which too has direct consequences for translation⁹⁶. EBNA1 causes a redistribution of the ribosomal protein L4 from the cytoplasm to the nucleus⁹⁷. The EBER transcripts redistribute the ribosomal protein L22 and sequester it away from polysomes⁹⁸. Both L4 and L22 are components of the 60S ribosomal subunit. This is especially interesting when taking the hypothesis of “specialized ribosomes” into consideration⁴¹. The idea is that in mammalian cells, ribosomes are not all equal but as large molecular machines they can have unique

compositions that influence their activity. The heterogeneity in the ribosome can stem from the composition of the individual ribosome components ranging from variation in rRNA sequences to the association of different ribosome-associated factors. Post-translational modifications can also lead to the generation of specialized ribosomes that have preferences for certain mRNAs.

4.7 Future studies

The underlying study has looked at translational profiles of EBV⁺ LCLs of two different strains. The relevance of the new insights gained from ribosome profiling still needs to be further defined and validated.

Inclusion of uninfected primary B cells in ribosome profiling experiments is needed to further analyse the viral influence on translation. Furthermore, EBV is known to infect other cell types as well. Ribosome profiling experiments including infected epithelial cells could give us further insight into translation in another cell setting. It is already known that viral gene expression is different in different cell types. It would be interesting to know to what extent this applies to translation and the novel regulatory elements. Moreover, uORF-dependence has been shown to be highly dynamic in HCMV⁵. Translational profiling of lytically replicating cells at different timepoints after lytic cycle initiation would further our understanding into translation and translation-mediated regulation of protein expression in EBV-infected cells.

The approach of manual annotation of novel genes in combination with low lytic replication in the samples studied, most likely underestimates the novel ORFs identified. Arias et al. who have combined manual annotation with computational analysis have contributed to a 45% increase in KSHV genome annotation due to small peptide identification⁶. It is possible that LCLs with higher lytic replication levels will allow for even more small proteins to be identified in the EBV genome.

5. Materials

5.1 Eukaryotic cell lines and culture medium

HEK293 are human embryonic kidney cells generated by transformation with adenovirus¹ and were obtained from American Type Culture Collection (ATCC) (www.atcc.org) (ATCC: CRL-1573).

HEK293 producer cell lines for the generation of virus stocks were generated from HEK293 cells and have been described previously^{2,3}.

Primary B cells were isolated from anonymous human blood samples purchased from the blood bank of the University of Heidelberg. No consent from the ethics committee is required. The precise CD19⁺ B cell isolation protocol is described in the methods section.

Culture medium for all cell lines was RPMI-1650 medium (Invitrogen) supplemented with 10% fetal bovine serum (FBS) (Biochrom).

5.2 Bacterial strains, culture medium and antibiotics

DH5 α This E.coli strain was used for routine cloning procedures and high copy number plasmid amplification.

LB medium: 10g Tryptone (AppliChem), 5g Yeast extract (Sigma-Aldrich), 10g NaCl (Sigma-Aldrich) in 1l H₂O

LB Agar: 32g LB-Agar (Invitrogen) supplemented with 5g NaCl (1% NaCl)

Antibiotics used in this study

| Antibiotic | company | concentration |
|------------|---------|---------------|
| ampicillin | Serva | 100 /ml |

5.3 Plasmids

Vector plasmids

pGL4.5[Fluc/CMV/Amp] Expression plasmid with the *Firefly* luciferase gene placed under the control of the strong HCMV promoter (ampicillin resistance)

Expression and control plasmids

| | |
|----------------------|---|
| pCDNA3.1(+) | Expression plasmid with a HCMV cloned in front of the multiple cloning site (Invitrogen) (ampicillin resistance) |
| pGL4.5[Fluc/CMV/Amp] | Expression plasmid with the <i>Firefly</i> luciferase gene placed under the control of the strong HCMV promoter (ampicillin resistance) |
| pRL | Expression plasmid with <i>Renilla</i> luciferase placed under the control of a strong SV40 promoter (ampicillin resistance) |
| p509 | Expression plasmid for BZLF1 under the HCMV promoter (ampicillin resistance) |
| pRA | Expression plasmid for gp110 under the HCMV promoter (ampicillin resistance) |

5.4 5' transcript leader sequences (TLS) synthesized for reporter construct experiments

| # | TLS sequence of viral gene | Sequence (5'-3') |
|------|----------------------------|---|
| 1423 | LMP1 uORF | <u>AAGCTTGGCAGACCCCGCAAATCCCCCGGGCCTCCATCCCAGAAACAC</u> GCGTTGCTCTCTCGTAGGCGGCCTACATAAGCCTCTGTCACTGCTCTGTCA GCTTCTTTCTCAGTTGCCTTGCTCCTGCCACACTACCCTGACCATGGAAG <u>ATGCCAAAAACATTAAGAAGGGCCC</u> |
| 1431 | ΔLMP1 uORF | <u>AAGCTTGGCAGACCCCGCAAATCCCCCGGGCCTCCATCCCAGAAACAC</u> GCGTTGCTCTCTCGTAGGCGGCCTACATAAGCCTCTGTCACTGCTCCGTCA GCTTCTTTCTCAGTTGCCTTGCTCCTGCCACACTACCCTGACCATGGAAG <u>ATGCCAAAAACATTAAGAAGGGCCC</u> |
| 1424 | U uORF | <u>AAGCTTTGAGCCACCCACAGTAACCACCCAGCGCCAATCTGTCTACATAGA</u> AGAAGAGGAGGATGAAGACTAAGTCACAGGCTTAGCCAGGTAACCTTAGGAA GCGTTTCTTGAGCTTCCCTGGG ATGAGCGTTTGGGAGAGCTGA TTCTGCAG CCCAGAGAGTAGTCTCAGGGCATCCTCTGGAGCCTGACCTGTGACCGTCG CATCATAGACCGCCAGTAGACCTGGGAGCAGATTACCGCCGCGGCCGTCTC TCCTTTAAGACAAA ATGGAAGATGCCAAAAACATTAAGAAGGGCCC |
| 1425 | ΔU uORF | <u>AAGCTTTGAGCCACCCACAGTAACCACCCAGCGCCAATCTGTCTACATAGA</u> AGAAGAGGAGGATGAAGACTAAGTCACAGGCTTAGCCAGGTAACCTTAGGAA GCGTTTCTTGAGCTTCCCTGGG ACGAGCGTTTGGGAGAGCTGA TTCTGCA GCCAGAGAGTAGTCTCAGGGCATCCTCTGGAGCCTGACCTGTGACCGTC GCATCATAGACCGCCAGTAGACCTGGGAGCAGATTACCGCCGCGGCCGTCTC CTCCTTTAAGACAAA ATGGAAGATGCCAAAAACATTAAGAAGGGCCC |

| | | |
|------|--------------|--|
| 1426 | Y2 uORF | <u>AAGCTT</u> CCTTAGAGAGTGGCTGCTACGCATGAGAGCCAGCTTTGAGCCA CCCACAGTAACCACCCAGCGCCAATCTGTCTACATAGAAGAAGAGGAGGA TGAAGACTAAGTCACAGGCTTAG CCAGTTCCTCTTAATTACATTTGTGCC AGATCTTGTAGAGCAAG ATG GAAGATGCCAAAAACATTAAGAAGGGCCC |
| 1427 | ΔY2 uORF | <u>AAGCTT</u> CCTTAGAGAGTGGCTGCTACGCATGAGAGCCAGCTTTGAGCCA CCCACAGTAACCACCCAGCGCCAATCTGTCTACATAGAAGAAGAGGAGGA <u>CGAAGACTAAGTCACAGGCTTAG</u> CCAGTTCCTCTTAATTACATTTGTGCC AGATCTTGTAGAGCAAG ATG GAAGATGCCAAAAACATTAAGAAGGGCCC |
| 1432 | BFLF2 uORF | <u>AAGCTT</u> ATTTAAATCCACACAAGTGGCCAGAGTGGGCAAAACAATCCTCGT GG ATG TCACTAAGGAACTGGACGTGGTCTCGCATCCACGGCCTTGAC CTGGTACAGTCCTATCAA ACTTCCAGGTCTACGTGTGAAAAGTAAACCG ATG GAAGATGCCAAAAACATTAAGAAGGGCCC |
| 1433 | ΔBFLF2 uORF | <u>AAGCTT</u> ATTTAAATCCACACAAGTGGCCAGAGTGGGCAAAACAATCCTCGT GG ACG TCACTAAGGAACTGGACGTGGTCTCGCATCCACGGCCTTGAC CGGTACAGTCCTATCAA ACTTCCAGGTCTACGTGTGAAAAGTAAACCG ATG GAAGATGCCAAAAACATTAAGAAGGGCCC |
| 1434 | BKRF3 TLS | <u>AAGCTT</u> AGTCTGGCCTCCTTAAATTCACCTAAGAATGGGAGCAACCAGCT GGTCATCAGCCGCTGCGCAAACGGACTCAACGTGGTCTCCTTCTTTATCT CCATC CTGAAGCGAAGCAGCTCCGCCCTCACGAGCCATCTCCGTGAGTT GTTAACCA CCCTGGAGTCTCTTTACGGTTCATTCTCAGTGGAAGACCTGTT TGGTGCCA ACTTAAACAGATACGC ATG GAAGATGCCAAAAACATTAAGAA GGGCCC |
| 1435 | ΔBKRF3 uORFs | <u>AAGCTT</u> AGTCTGGCCTCCTTAAATTCACCTAAGAATGGGAGCAACCAGCT GGTCATCAGCCGCTGCGCAAACGGACTCAACGTGGTCTCCTTCTTTATCT CCATC CCGAAGCGAAGCAGCTCCGCCCTCACGAGCCATCTCCGTGAGTT GTTAACCA CC CCG GAGTCTCTTTACGGTTCATTCTCAGTGGAAGACCTGTT TGGTGCCA ACTTAAACAGATACGC ATG GAAGATGCCAAAAACATTAAGAA GGGCCC |
| 1436 | BORF2 uORF | <u>AAGCTT</u> ACATATAGAGGAGCTAACCTTCGGGGCGGTTGCCTGTCTGGGGA CATTTAGTGCTACTGACGGTTGGAGGAGGTCTGCCTTCAATTACCGTGGC TCTAGCCTCCCGTG G TGGAGATTGACAGCTTTTATTCCAACGTCTCTGAC TGGGAGGTGATTCTCTAG ACTTAACGGGAGGAAACAGGAGGAGGAGGGG GACAAGAGCACAAAAGTGGTTCAGTGGACACCCACCACACAGC ATG GAA GATGCCAAAAACATTAAGAAGGGCCC |
| 1437 | ΔBORF uORF | <u>AAGCTT</u> ACATATAGAGGAGCTAACCTTCGGGGCGGTTGCCTGTCTGGGGA CATTTAGTGCTACTGACGGTTGGAGGAGGTCTGCCTTCAATTACCGTGGC TCTAGCCTCCCGTG G CGGAGATTGACAGCTTTTATTCCAACGTCTCTGA CTGGGAGGTGATTCTCTAG ACTTAACGGGAGGAAACAGGAGGAGGAGGGG GGACAAGAGCACAAAAGTGGTTCAGTGGACACCCACCACACAGC ATG GAA AGATGCCAAAAACATTAAGAAGGGCCC |

Column 1 shows the plasmid number designated to the constructs after recloning into the 1272 Firefly luciferase expression plasmid. Column 2 lists the gene the uORF-containing 5' leader was derived from. Column 3 lists the sequences synthesized. The uORF sequence is highlighted in red, the mutation introduced to silence the uORF is

underlined in the mutant variant. The ATG of the luciferase gene is set in bold type. Each sequence is flanked by enzyme restriction sites (underlined sequences: 5': HindIII and 3': Bsp120I). TLS: transcript leader.

5.5 Recombinant EBV clones

| Clone | name | EBV strain | Description |
|-------|----------------------|------------|---|
| B240 | M81 wt | M81 | recombinant M81 wild-type virus ³ |
| B697 | M81 Δ ZR | M81 | M81 with a deletion of the BZLF1 and BRLF1 gene |
| 2089 | B95-8 wt | B95-8 | recombinant B95-8 wild-type virus ² |
| B072 | B95-8 Δ BMRF1 | B95-8 | recombinant B95-8 with a deletion of the BMRF1 gene |

5.6 Oligonucleotides

All oligonucleotides were synthesized by Eurofins Genomics.

Oligos used for ribosome profiling

| Oligo function | Sequence (5'-3') |
|----------------------------------|--|
| Upper size marker | 5'-AUGUACACGGAGUCGAGCUCAACCCGCAACGCGA-(Phos)-3' |
| Lower size marker | 5'-AUGUACACG (Phos) ACCCAACGCGA |
| Reverse transcription primer | 5'-(Phos)-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAG ATCTCGGTGGTCGC-(SpC18)-CAC TCA-(SpC18)-TTCAGACGTGT GCTCTTCCGATCTATTGATGGTGCCTACAG-3' |
| Biotinylated subtraction oligos* | 5'-GGGGGGATGCGTGCATTTATCAGATCA-3' 5'-TTGGTGA CTCTAGATAACCTCGGGCCGATCGCACG-3' 5'-GAGCCGCCTGGATACCGCAGCTAGGAATAATGGAAT-3' |

| | |
|--|---|
| | <p>5'-TCGTGGGGGGCCCAAGTCCTTCTGATCGAGGCC-3'</p> <p>5'-GCACTCGCCGAATCCCGGGGCGAGGGAGCGA-3'</p> <p>5'-GGGGCCGGGCGCCCTCCCACGGCGCG-3'</p> <p>5'-GGGGCCGGGCCACCCCTCCCACGGCGCG-3'</p> <p>5'CCCAGTGCGCCCGGGCGTCGTCGCGCCGTCGGGTCCCGGG-3'</p> <p>5'-TCCGCCGAGGGCGCACCACGGCCCGTCTCGCC-3'</p> <p>5'-AGGGGCTCTCGCTTCTGGCGCCAAGCGT-3'</p> <p>5'-GAGCCTCGGTTGGCCCCGGATAGCCGGTCCCCGT-3'</p> <p>5'-GAGCCTCGGTTGGCCTCGGATAGCCGGTCCCCCGC-3'</p> <p>5'-TCGCTGCGATCTATTGAAAGTCAGCCCTCGACACA-3'</p> <p>5'-TCCTCCCGGGGCTACGCCTGTCTGAGCGTCGCT-3'</p> |
| forward library PCR primers | 5'-AATGATACGGCGACCACCGAGATCTACAC-3' |
| Indexed reverse primers (from NEBNext) | <p>Index #2: 5'-CAAGCAGAAGACGGCATAACGAGATACATCGGTGACTGGAGTT CAGACGTGTGCTCTTCCGATC-s-T-3'</p> <p>Index #3: 5'-CAAGCAGAAGACGGCATAACGAGATGCCTAAGTGACTGGAGTTC AGACGTGTGCTCTTCCGATC-s-T-3'</p> <p>Index #4: 5'-CAAGCAGAAGACGGCATAACGAGATTGGTCAGTGACTGGAGTTC AGACGTGTGCTCTTCCGATC-s-T-3'</p> <p>Index #5: 5'-CAAGCAGAAGACGGCATAACGAGATCACTGTGTGACTGGAGTTC AGACGTGTGCTCTTCCGATC-s-T-3'</p> <p>Index #6: 5'-CAAGCAGAAGACGGCATAACGAGATATTGGCGTGACTGGAGTT CAGACGTGTGCTCTTCCGATC-s-T-3'</p> |

| | |
|--|--|
| | Index #7: 5'-CAAGCAGAAGACGGCATAACGAGATGATCTGGTGA CTGGAGTT CAGACGTGTGCTCTTCCGATC-s-T-3' |
|--|--|

*All biotinylated subtraction oligos were 5' modified with a biotin-triethyleneglycol (TEG) tag.

(Phos) indicates phosphorylation sites. (SpC18) indicates a hexa-ethyleneglycol spacer.

-s- indicates a phosphorothioate bond

Oligos used for Taqman qPCR*

| Oligo # | target gene | Sequence (5'-3') |
|----------------|----------------------|----------------------------------|
| 54 | EBV polymerase fwd | 5'-AGTCCTTCTTGGCTAGTCTGTTGAC-3' |
| 55 | EBV polymerase rev | 5'-CTTTGGCGCGGATCCTC-3' |
| 96 | EBV polymerase probe | Fam-CATCAAGAAGCTGCTGGCGGCC-Tamra |

fwd: forward primer. rev: reverse primer. Fam: fluorescein amidite. Tamra: Tetramethylrhodamine Azide.

*Human GAPDH was used as the endogenous control for the RT-qPCR experiments. The primer and probe mix for GAPDH was purchased ready-to-use from Applied Biosystems

Oligos used for SYBR Green qPCR

| Oligo # | target gene | Sequence (5'-3') |
|----------------|---------------------|-----------------------------|
| 1923 | BMRF1 fwd | 5'-CACCATGCTGGTGGTAGATG-3' |
| 1924 | BMRF1 rev | 5'-GAGATGAGTCTGGGCATGGT-3' |
| 2300 | BHLF1 fwd | 5'-ACCTACATGTCAACCGCCTC-3' |
| 2301 | BHLF1 rev | 5'-GTGTATTGCTCTCGTTGCCA-3' |
| 2559 | BNLF2a fwd | 5'-GTGCTTTGCTAGAGCAGCAGT-3' |
| 2560 | BNLF2a rev | 5'-TTAGTCTGCTGACGTCTGGGT-3' |
| 2561 | BNLF2b fwd (B95-8) | 5'-AGGACTGCATCCAACGCTT-3' |
| 2562 | BNLF2b rev (common) | 5'-CCACACCATCCCAATTCA-3' |
| 2563 | BNLF2b (M81) | 5'-AGGACTGCATCCAACGCTGC-3' |
| 1609 | GAPDH fwd | 5'-CAATGACCCCTTCATTGACC-3' |
| 1610 | GAPDH rev | 5'-TGGAAGATGGTGTGGGATT-3' |

fwd: forward primer. rev: reverse primer.

Oligo used for sequencing of luciferase constructs

Oligo # Sequence (5'-3')

1274 5'-CGCAAATGGGCGGTAGGCGTG-3'

5.7 Chemicals & Reagents

| Chemicals & reagents | company |
|--------------------------------------|----------------------|
| 1-butanol | AppliChem |
| 10 bp DNA ladder | Invitrogen |
| 1kb DNA ladder | Invitrogen |
| 1M Tris-HCl pH 8; RNase free | Ambion |
| 20% SDS solution | MP Biomedics |
| 3M sodium acetate pH 5.5, RNase-free | Ambion |
| Acetic Acid | Sigma-Aldrich |
| Agarose | Sigma-Aldrich |
| boric acid | Sigma-Aldrich |
| Bromphenol blue | Serva |
| CaCl ₂ | Invitrogen |
| Chloroform | Carl Roth |
| Complete Mini Protease Inhibitor | Roche |
| Cycloheximide | Sigma-Aldrich |
| D(+)-Saccharose | Roth |
| DAPI | Sigma-Aldrich |
| DMSO | Sigma-Aldrich |
| dNTP mix | Roche |
| EDTA | GERBU |
| EDTA, RNase-free | Invitrogen |
| Ethanol | Sigma-Aldrich |
| Ethidiumbromide | Carl Roth |
| Fetal Bovine Serum | Biochrom |
| Ficoll Plus | Amersham Biosciences |
| Formaldehyde | J.T. Baker |
| Formamide | Promega |
| Glycerol | VWR International |

| | |
|--|-------------------|
| Glycine | GERBU |
| GlycoBlue | Invitrogen |
| Harringtonine | LKT Laboratories |
| Heat-inactivated goat serum | Gibco |
| Heparin sodium salt | Sigma-Aldrich |
| Hygromycin B | Invitrogen |
| Isopropanol | Sigma-Aldrich |
| KH ₂ PO ₄ | Carl Roth |
| LB-Agar | Invitrogen |
| Metafectene | Biontex |
| Mini Protean TBE-Urea precast gels, 12 well, 15% | Bio-Rad |
| MnCl ₂ | Sigma-Aldrich |
| MOPS | Genaxxon |
| MyOne streptavidin C1 DynaBeads | Invitrogen |
| Na ₂ HPO ₄ | Sigma-Aldrich |
| NP-40 | Biochemika |
| 4% Paraformaldehyde (PFA) | AppliChem |
| PBS tablets for cell culture | Gibco |
| Phenol | Carl Roth |
| Potassium acetate (KAc) | Carl Roth |
| Random hexamers | Invitrogen |
| RbCl | Acros Organics |
| RNasin | Promega |
| RPMI-1640 (+)L-glutamine | Gibco |
| Sodium Acetate (NaAc) | Carl Roth |
| Sodium Chloride (NaCl) | Sigma-Aldrich |
| Sodium citrate | Sigma-Aldrich |
| sodium hydroxide (NaOH) | AppliChem |
| SYBR Gold | Invitrogen |
| TBE (10x), RNase-free | Thermo Scientific |
| Triton | AppliChem |
| TrizmaBase | Sigma-Aldrich |
| Trizol | Ambion |
| Trypan Blue | Sigma-Aldrich |
| Trypsin-EDTA (0.05%) | Invitrogen |
| Tryptone | AppliChem |

| | |
|------------------------------|---------------|
| Tween 20 | Sigma-Aldrich |
| Water, PCR-grade, RNase-free | Sigma-Aldrich |
| Yeast extract | Sigma-Aldrich |

5.8 Enzymes

All restriction enzymes were purchased from Thermo Fisher Scientific.

Other enzymes used in this study

| Enzyme | company |
|--------------------------------------|--------------------|
| Alkaline phosphatase | Roche |
| AMV Reverse transcriptase | Roche |
| Dnase I | Fermentas |
| Phusion High-fidelity DNA polymerase | ThermoScientific |
| Proteinase K | ThermoScientific |
| RNase A | Roche |
| Rnase I | Invitrogen |
| SuperScript III | Invitrogen |
| SYBR Green Master Mix | Applied Biosystems |
| T4 DNA ligase | Fermentas |
| T4 PNK | NEB |
| T4 RNA ligase 2, truncated | NEB |
| Taqman Universal Master Mix | Life technologies |

5.9 Antibodies

Antibodies used for immunofluorescence stainings

| Name | target | Species | Dilution | company |
|------|--|---------|----------|---------------------------|
| BZ.1 | BZLF1 | Mouse | 1:300 | Hybridoma supernatant |
| 72A1 | gp350/220 | Mouse | 1:250 | Hybridoma supernatant |
| R3 | BMRF1 | mouse | 1:1000 | Merck Millipore |
| Cy3 | α -mouse IgG conjugated with Cy3 | Goat | 1:300 | Jackson ImmunoResearch |

5.10 Commercial kits

| Kit | Company | Application |
|---------------------------------------|------------|------------------------------------|
| CircLigase | Epicentre | ribosome profiling |
| DETAChAbeAD C19 | Invitrogen | Isolation of human primary B cells |
| Dual Luciferase Reporter Assay System | Promega | Reporter assays |
| Dynabeads CD19 Pan B | Invitrogen | Isolation of human primary B cells |
| Jetstar 2.0 Plasmid Midiprep Kit | Genomed | Preparation of plasmid DNA |
| Universal miRNA cloning linker | NEB | ribosome profiling |

5.11 Buffers

| Buffer | Formula |
|-----------------------------------|--|
| Bind/wash buffer (2x) | 2M NaCl, 1mM EDTA, 5mM Tris (pH 7.5), 0.2% (v/v) Triton X-100 |
| Denaturing loading buffer (2x) | 98% (v/v) formamide, 10 mM EDTA, 300 µg/ml bromphenol blue |
| DNA gel extraction buffer | 300 mM NaCl, 10 mM Tris (pH 8), 1mM EDTA |
| DNA loading buffer | 0.25% bromophenol blue, 40% (w/v) sucrose, dissolved in H ₂ O |
| Lysis buffer mini prep | 1% SDS, 0.2M NaOH |
| MOPS (10x) | 0.2M MOPS, 20 mM NaAc, 10 mM EDTA |
| Mounting medium for IF | 90% glycerol in PBS |
| Neutralization buffer mini prep | 3M KaAc, 5M acetic acid, pH 5.5 |
| Nondenaturing loading buffer (6x) | 10mM Tris (pH 8), 1mM EDTA, 15% (w/v) Ficoll 400, 0,25% bromophenol blue |
| Paraformaldehyde (PFA) | 4 g paraformaldehyde in PBS, pH 7.4 |
| PBS (1x) | 137 mM NaCl, 2.7 mM KCl, 10 mM Na ₂ HPO ₄ , 2mM KH ₂ PO ₄ , pH 7.1-8.0 |
| PBS-Triton | 1x PBS (see above), 0.5% TritonX-100 |
| Polysome lysis buffer | 15 mM Tris-HCl, pH 7.4, 15 mM MgCl ₂ , 300mM NaCl, 1% Triton X-100, 0.1% β-mercaptoethanol, 200 U/ml RNasin, 1 tablet/10ml Complete Mini protease inhibitor |

| | |
|---------------------------|---|
| RNA gel extraction buffer | 300 mM sodium acetate, 1mM EDTA, 0.25% (w/v) SDS |
| RNA loading buffer | 50% formamide, 2.2 M formaldehyde, 1x MOPS, 1% Ficoll, 0.04% bromophenol blue |
| SSC (20x) | 3 M NaCl, 0.3 M Sodium Citrate; pH 7.0 |
| Staining buffer for IF | 10% Heat-inactivated goat serum (Gibco) in PBS |
| TAE (1x) | 40 mM Tris, 1 mM EDTA, 19 mM acetic acid |
| TBE (1x) | 89 mM Tris, 89 mM boric acid, 2 mM EDTA |
| TE | 10 mM Tris-HCl, 1 mM EDTA, pH 8.0 |
| TFB1 | 50 mM MnCl ₂ , 100 mM RbCl, 30 mM CaCl ₂ , 15% Glycerol, pH 5.8 with 0.2N acetic acid |
| TFB2 | 10 mM MOPS pH 7.0, 10 mM RbCl, 75 mM CaCl ₂ , 15% Glycerol |

IF: immunofluorescence

5.12 Consumables

| Consumable | Company |
|--|-------------------------|
| 0.5, 1.5 and 2 ml reaction tubes | Eppendorf |
| 15 and 50 ml Falcon tubes | TPP |
| Cell culture plates and flasks | TPP |
| Microscopy glass slides | Medco |
| Coverslips | Menzel |
| Syringe-driven sterile filter unit (0.45 µm) | Merck Millipore |
| Cryo.S Freezing tubes | Greiner Bio-One |
| Scalpel No 10 | Feather |
| 20 Gauge syringe-needle | TERUMO |
| Syringe (virus supernatant filtration) | TERUMO |
| Syringe (DNA gel extraction) | Braun |
| MicroAmp 96-well Rxn Plate | Applied Biosystems |
| Cuvettes semi-micro | Greiner-Bio |
| 96-well flat clear bottom plates | Corning Incorporated |
| 11x60 polyallomer tubes | Beckman Coulter |

5.13 Equipment

| Equipment | Company |
|---|----------------------|
| Beckman Coulter Optima LE-80K Ultracentrifuge with SW60 rotor | Beckman Coulter |
| Biometra standard power pack P25 | Biometra |
| BioPhotometer | Eppendorf |
| CO ₂ incubator, HERAcell 150 | Thermo Scientific |
| Electrophoresis MiniProtean | Bio-Rad |
| Hoefer miniVE (vertical electrophoresis system) | Amersham Biosciences |
| Leica DM5000B Microscope | Leica |
| Megafuge Heraeus 16R centrifuge | Thermo Scientific |
| Nanodrop 2000 | Thermo Scientific |
| Pico17 Heraeus centrifuge | Thermo Scientific |
| RC5C Centrifuge (Sorvall) | ThermoScientific |
| Roller RM5 | CAT |
| StepOnePlus RT-qPCR machine | Applied Biosystems |
| Teledyne Isco Foxy Jr. Gradient fractionator | Axel Semrau |
| Thermocycler PTC-200 | MJ Research |
| Thermomixer F1.5 | Eppendorf |
| UV gel documentation station | UVP |
| UV transilluminator (UV box) | Bio View |

5.14 Software

| Software | Version |
|-------------------------|------------------------------------|
| Graphpad | Prism Version 6.0c, March 21, 2013 |
| IGV | Version 2.3.25(29); 11/25/2013 |
| ImageJ | 1.47v |
| Leica Application Suite | 10 (X) |
| MacVector | Version 11.1.1, 2011 |

6. Methods

6.1. Eukaryotic experimental system

6.1.1. Culture conditions & maintenance

Mammalian cells were maintained in RPMI-1640 medium supplemented with 10% FBS. Primary B lymphocytes were cultured in RPMI-1640 medium supplemented with 20% FBS until lymphoblastoid cell lines (LCLs) grew out.

All cells were cultured in a 37°C incubator at 100% humidity and 5% CO₂ levels.

LCLs were split 1:5 or 1:10 twice a week. The frequency depended on their growth rate.

Adherently growing cell lines such as the HEK293 and HEK293 cell lines stably transfected with recombinant EBV (producer cells) were split before the culture plates reached confluency. The culture medium was removed, the cells were rinsed gently once with 1x PBS and incubated with 700 µl 0.05% Trypsin at 37°C for 1 min. The cells were then resuspended in RPMI/10%FBS. HEK293 cells were split 1:10. The producer cells were split 1:5.

6.1.2. Freezing and thawing cells

In order to freeze cells 1x10⁶ LCLs or a 10 cm culture dish of adherent cells were detached and spun down for 10 minutes at 1200 rpm. Following removal of the supernatant the cell pellet was resuspended in 900 µl of RPMI medium containing 20% FBS and transferred into a cryotube with pre-aliquotted 100 µl of DMSO. Cells were frozen in a -80°C freezer overnight and then transferred into liquid nitrogen-containing containers for long-term storage.

6.1.3. Induction of producer cells lines into virus production

Producer cell lines were induced to produce infectious virus by transfecting plasmids encoding the viral transactivator BZLF1 (encoded on the p509 plasmid) and viral glycoprotein gp110 (encoded on the pRA plasmid) using 3 µl of Metafectene transfection reagent/1µg of DNA according to manufacturer's instructions. The cells were first seeded onto 6 well plates at a cell density of 2.5x10⁵ cells/well. After allowing the HEK293 producer cells to attach to the bottom of the plate overnight the

cells were transfected with 0.5 µg of each plasmid in 200 µl RPMI medium not containing any supplements. The medium was changed 24 h after transfection and virus was harvested 72 h after medium change and filtered through a 0.45 µm filter.

6.1.4. Isolation of primary B lymphocytes

Human whole blood was centrifuged in a beaker with 1x PBS supplemented with 100 U/ml Heparin and underlaid with 50 ml Ficoll. The samples were spun at 1800 rpm for 30 min at room temperature and the upper Ficoll layer was carefully removed after centrifugation. The interphase was collected comprising approximately 30 to 40 ml and was divided into two 50 ml Falcons. The interphase solution containing the buffy coat was filled to 50 ml with 1x PBS. The samples were spun at 1600 rpm for 10 min at room temperature and the pellet was washed with 50 ml 1x PBS. The samples were spun down at 1400 rpm and the resulting pellet was washed with 50 ml 1xPBS. Again the samples were spun down, this time at 1200 rpm for 10 min at room temperature. The pellets from the two falcons were combined and resuspended in 20 ml RPMI supplemented with 1% FBS. The cells were counted at a 1:20 dilution. The peripheral blood mononuclear cells (PBMCs) were diluted to a concentration of 2×10^7 cells/ml in RPMI/1%FBS. Approximately 5% of PBMCs are estimated to constitute B cells. From the total number of PBMCs the B cell number was estimated and a ratio of 1:4 of B cells to CD19 Dynabeads was used for B cell isolation. The CD19 Pan B Dynabeads were transferred into 15 ml Falcons and washed three times with 5 ml RPMI/1% FBS by resuspending the beads, placing the Falcon tubes in a magnetic rack and removing the wash solution while taking care not to disrupt the beads on the side of the tube. Following washing the beads were mixed with the PBMCs and were incubated for 30 min at 4°C with gentle rolling. The beads were then washed again three times in 5 ml RPMI/1%FBS, again by using a magnetic rack. Finally, the beads were resuspended in 5 ml RPMI/1%FBS and the total number of cells was determined. For 1×10^6 cells 10 µl of CD19 Detachabeads were used to detach the isolated B cells from the CD19 Pan B Dynabeads. The samples were incubated for 45 min at room temperature by rolling. The samples were then placed in a magnetic rack and the supernatant was collected containing the B cells. The Falcons containing the Dynabeads were washed two more times with 5 ml RPMI/1%FBS and the supernatant was collected. The supernatants were pooled and spun for 5 min at 1200 rpm at room temperature. The resulting pellet was washed

once with 5 ml RPMI/1% FBS and resuspended in 5 ml RPMI supplemented with 20% FBS.

6.1.5. Infection of primary B lymphocytes and generation of LCLs

B cells were incubated with virus supernatants with a multiplicity of infection of 20 at room temperature for 2 h with gentle shaking. The cells were then spun down at 1200 rpm for 5 min and resuspended in 200 µl RPMI/20%FBS. Cells were cultured in U-bottom 96 well plates until LCLs grew out. They were then transferred into culture flask and were cultured in RPMI/10% FBS.

6.1.6. Dual luciferase reporter assays

For luciferase activity measurements HEK293 cells were seeded at a cell density of 1×10^4 cells/well on 24 well plates in 500 µl complete RPMI medium. After 24 *hours* 200 ng of plasmid with the different viral 5'leaders cloned in front of the *Firefly* luciferase gene were co-transfected with 200 ng of *Renilla* luciferase encoding plasmid and 1.5 µl Metafectene in a total volume of 50 µl of RPMI medium without any supplements. The Metafectene-DNA mixture was incubated at room temperature for 20 min before adding the mixture dropwise to the cells. Twenty-four hours later cells were washed once with 1x PBS and lysed in 100 µl Passive lysis buffer from the Dual-Luciferase Reporter Assay kit from Promega. Lysis was carried out at room temperature for 15 minutes with gentle shaking. Luciferase measurements were performed according to manufacturer's instructions. *Firefly* luciferase activities were normalized to co-transfected *Renilla* luciferase activity measured sequentially. This was done to correct for differences in transfection efficiency between samples.

6.2. Prokaryotic experimental system

6.2.1. Culture conditions

Bacteria were grown in LB medium with shaking or on LB Agar-containing plates to obtain single cell clones. Relative to the antibiotic resistance gene encoded on the plasmid, bacteria were culture in LB medium or LB agar supplemented with antibiotics. DH10α bacteria were grown at 37°C.

Glycerol stocks were prepared in 10% glycerol of bacteria cultured in exponential growth phase and frozen at -80°C for long-term storage.

6.2.2. Preparation of chemically competent bacteria

Overnight cultures of DH5 α were inoculated and grown at 37°C. The culture was diluted 1:50 the following day in 400 ml LB medium. The bacteria were grown until an optical density of OD_{600nm} = 0.4 was reached. The bacteria were then spun down at 4000 rpm for 15 min at 4°C and resuspended in 50 ml ice-cold TFB1 buffer and transferred into fresh tubes. The samples were then spun down at 3000 rpm for 10 min at 4°C. The pellet was resuspended in a final volume of 10 ml of ice-cold TFB2 buffer. 300 μ l aliquots were prepared and shock frozen on dry-ice/ethanol and were stored at -80°C until further use.

6.2.3. E.coli transformation by heat-shock

Chemically-competent DH10 α were thawed on ice, mixed with DNA and incubated on ice for 5 min. The bacteria were then incubated at 42°C for 2 min and transferred back to ice for 5 min. Finally 1 ml LB medium was added to the bacteria and they were cultured for 1 h at 37°C with shaking. After 1 h the bacteria were spun down at 4000 rpm for 10 min and the pellet was resuspended in 50 μ l LB medium and plated on LB agar plates with appropriate antibiotics. The plates were incubated over night at 37°C and single colonies were picked for further analysis.

6.3. Working with DNA

6.3.1. Cloning of overexpression plasmids

Synthesis of selected transcript leader sequences was performed by Eurofins Germany (<http://www.eurofins.de/de-de.aspx>). The sequences were delivered cloned into pEX-A2 vectors. The sequences were excised from these vectors with the restriction enzymes HindIII and Bsp120I. The sequences were then ligated into corresponding restriction sites in the pGL4.5[Fluc/CMV/Amp] vector placing the synthesised transcript leader sequences upstream of the Firefly luciferase gene. Verification was performed by restriction enzyme digestion and sequencing of the cloned region.

Cloning procedure in detail:

(1) Restriction enzyme digestion

10 µg of DNA was digested with approximately 30 units (U) of each restriction enzyme for 3 hours at 37°C in a final volume of 100 µl.

Following digestion 20 µl of 10x dephosphorylation buffer and 1U of alkaline phosphatase was added to the reaction tubes containing vector DNA and filled up to 200 µl. The samples were incubated for an additional hour at 37°C.

(2) Fragment and vector purification from agarose gels

After restriction enzyme digestion all samples were run on a 0.8% agarose gel and a gel extraction of vector DNA and insert fragments was performed.

See section 2.5.1 for details on the procedure.

(3) Ligation reaction

Ligation reactions were set up so that the molar ratio of vector DNA to insert DNA was approximately 1:3. Four units of T4 DNA ligase were used per reaction in a total volume of 10 µl. The ligation reactions were incubated for 1 hour at room temperature and were subsequently used to transform 50 µl of chemically-competent DH5α E.coli using the heat-shock method. Alkaline miniprep analysis was used to screen for correct clones.

6.3.2. Extraction of DNA with Phenol:Butanol

From agarose gels

Generally 0.8% agarose gels containing ethidiumbromide were run at 100 V for approximately an 1 hour. The desired DNA fragments were then cut from the gel using a UV box to visualize the DNA bands in the gel. A clean scalpel was then used to slice out the desired DNA fragments from the gel. The gel slice was then squeezed through a 20 gauge syringe-needle. To the gel slurry 1 gel volume of phenol was added and the samples were vortexed. The tubes containing the phenol-gel slurry were then places on dry ice for 5 min until the mixture was frozen and thawed in a 37°C water-bath. The freeze-thaw cycles were repeated twice more. The frozen gel-

slurry samples were subsequently centrifuged at 12000 rpm for 30 min at room temperature. The upper phase of the centrifuged samples was collected and mixed with 1 volume of 1-butanol. The samples were vortexed for at least one minute and centrifuged at 12000 rpm for 10 min at room temperature. The upper phase was discarded and the lower phase was collected and DNA was precipitated with 1/10 volume of sodium acetate and 2.5 volumes of ethanol. Precipitations were carried out for 1 hour in a -20°C freezer. Subsequently the samples were spun at 13000 rpm for 30 min. The DNA pellet was then washed with 1 ml of 70% ethanol followed by a 5 min centrifugation step at 13000 rpm. The DNA was then resuspended in 30 µl TE buffer. DNA purity and concentration were determined by Nanodrop.

In solution

DNA isolation from solutions were performed by adding 1 volume of phenol to the DNA solutions and vortexing for 1 min. The samples were then centrifuged for 10 min at room-temperature at 13000 rpm. The upper phase was collected and transferred to a fresh tube. 1 volume of 1-butanol was added, vortexed for 1 min and spun at 13000 rpm for 5 min at room-temperature. The upper phase was discarded and the lower phase into a fresh Eppendorf tube. The DNA was precipitated with 2.5 volumes of ethanol and 1/10 volume of 3M sodium acetate (pH 5.0). The precipitates were incubated for 20 min at -20°C and were subsequently spun at 13000 rpm for 20 min at room-temperature. The pellet was washed in 70% of ethanol to remove salts with 5 min centrifugation at 13000 rpm. The washed DNA pellet was resuspended in 1xTE buffer. DNA purity and concentration were determined by Nanodrop.

6.3.3. Plasmid DNA isolation (Mini & Midi preparation)

Miniprep

Single E.coli colonies were plated out on a ¼ of a LB agar plate with appropriate antibiotics and grown overnight at 37°C depending on the clones analysed. Bacteria were scraped off and resuspended in 200µl TE buffer containing concentration RNase A. Bacteria were lysed by addition of 200µl of lysis buffer (add components to materials), inversion of the samples and incubation for 5min at room temperature. Lysates were then neutralized with 200µl neutralization followed by 10min incubation on ice. The lysates were then cleared by centrifugation at 13 000rpm for 10min. The

supernatant was transferred into fresh tubes containing 500µl Isopropanol. Precipitations were carried out for 10min on ice and DNA was pelleted for 10min at 13 000rpm. The DNA pellet was washed with 1ml 70% Ethanol and were dissolved in TE followed by restriction enzyme analysis. Correct colonies were inoculated in 200ml of LB medium for midiprep analysis.

Midiprep

E.coli cell containing the desired plasmids were inoculated from glycerol-stock into 200 ml of LB medium supplemented with ampicillin (100 µg/ml). Bacteria were grown at 37°C overnight with gentle shaking. The following day bacteria were centrifuged at 4000 rpm for 15 min at room-temperature. The supernatant was discarded and the bacterial pellet resuspended in 4 ml of resuspension buffer from the Jetstar Midiprep kit. The resuspended cells were transferred to fresh tubes. 4 ml of lysis buffer were added and mixed with the samples by gentle inversion. Samples were incubated at room-temperature for 5 min. Lysis was neutralized by adding 4 ml of neutralization buffer 3. The samples were inverted 3-4 times and incubated on ice for 10 min. Lysates were then centrifuged for 10 min at 12000 rpm at room-temperature. The supernatant was filtered through a fluted filter and the Jetstart column was equilibrated with 10 ml equilibration buffer E4 in parallel. The filtered lysate was then loaded on the column. The column was allowed to empty by gravity flow and was washed twice with 10 ml wash buffer E5. The DNA was eluted from the column with 5 ml of elution buffer E5 into fresh Falcon tubes. The eluted DNA was mixed with 0.7 volumes of isopropanol and centrifuged for 30 min at 4°C at 4000 rpm. The supernatant was discarded. The DNA pellet was resuspended in 1 ml of 80% ethanol and transferred into a fresh 1.5 ml Eppendorf tube. The samples were centrifuged at 13000 rpm for 5 min at room-temperature. Ethanol was discarded and the pellet was resuspended in 400 µl TE buffer. DNA purity and concentration were determined by Nanodrop. Plasmid integrity was analysed by agarose gel electrophoresis.

6.3.4. Quantitative real-time PCR (qRT-PCR)

Taqman qRT-PCR for virus titer determination

To determine the virus titers from virus supernatants of induced producer cell lines 50µl of virus supernatant was transferred into a PCR tube and 1U of DNase I was

added. The samples were incubated in a thermocycler for 60min at 37°C followed by DNase inactivation for 10 min at 70°C. Afterwards 5 µl of DNase-treated supernatants was transferred into a fresh PCR tube and mixed with 5 µl of 1:100 diluted Proteinase K (10 mg/ml). The samples were again incubated in a thermocycler for 60 min at 50°C followed by Proteinase K inactivation for 20 min at 75°C. Finally the samples were diluted 1:10 in H₂O and 5µl were used for qPCR measurements.

qRT-PCR reaction mix:

12.5 µl Taqman 2x Universal Mastermix
2.5 µl EBV polymerase forward primer (10 µM)
2.5 µl EBV polymerase reverse primer (10 µM)
1 µl FAM-labeled EBV polymerase probe (20 µM)
1.5 µl H₂O

Cycling conditions were as follows:

| | | |
|---|------|--------|
| 1 | 50°C | 2 min |
| 2 | 95°C | 10 min |
| 3 | 95°C | 15 sec |
| 4 | 60°C | 1 min |

steps 3 and 4 were run for 40 cycles

SYBR Green qRT-PCR for gene expression analysis

SYBR Green qRT-PCR was used to determine RNA transcript abundance for viral genes. RNA was extracted from cells directly or from fractioned gradients from polysome profiling experiments (see RNA chapter for details on RNA purification). RNA was reverse transcribed into cDNA (see RNA chapter for detailed protocol). 5µl of cDNA were used for measurements.

qRT-PCR reaction mix:

10 µl SYBR Green master mix
0.4 µl forward primer (10 µM)
0.4 µl reverse primer (10 µM)
4.2 µl H₂O

Cycling conditions were as follows:

| | | |
|---|------|--------|
| 1 | 95°C | 10 min |
| 2 | 95°C | 15 sec |
| 3 | 60°C | 1 min |

steps 2 and 3 were repeated for 40 cycles

6.4. Working with RNA

6.4.1. RNA extraction from cell pellets

RNA extractions from cell pellets were performed with Trizol reagent. Cell pellets were resuspended in 1 ml Trizol reagent for cell lysis. 200 µl of chloroform were added/1 ml Trizol lysate. Samples were shaken vigorously for 30 sec and incubated at room-temperature for 3 min. Subsequently the samples were spun at 12000 rpm for 15 min at 4°C. In the mean time 500 µl of isopropanol were aliquotted to fresh 15 ml Eppendorf tubes. The upper phase of the samples (approximately 600 µl) was then collected and transferred to the isopropanol-containing tubes. Samples were mixed by inverting and incubated for 2 h at -80°C. In the mean time RNase-free water was heated to 95°C. After 2 h of precipitation at -80°C the samples were spun at 12000 rpm for 10 min at 4°C. Isopropanol was removed carefully and 75% ethanol prepared with RNase-free water was used to wash the pellet. The samples were centrifuged at 8000 rpm for 5 min at 4°C. The RNA pellet was resuspended in 50 µl pre-heated RNase-free water.

6.4.2. RNA extraction from sucrose solutions after polysome profiling

RNA extractions from polysome fractionation in sucrose solution were performed by mixing equal volumes of fraction with ready-to use phenol:chloroform:isoamyl alcohol (25:24:1) mix. Samples were shaken for 30 sec and spun at 12000 rpm for 15 min at 4°C. After centrifugation the upper phase was transferred into fresh 1.5 ml Eppendorf tubes. 500 µl of isopropanol was added and samples were mixed by inversion. Precipitations were carried out at -80°C for 2 h. In the mean time RNase-free water was heated to 95°C. After 2 h of precipitation at -80°C the samples were spun at 12000 rpm for 10 min at 4°C. Isopropanol was removed carefully and 75% ethanol prepared with RNase-free water was used to wash the pellet. The samples were

centrifuged at 8000 rpm for 5 min at 4°C. The RNA pellet was resuspended in 20 µl RNase-free water.

6.4.3. Reverse transcription of RNA

The cDNA synthesis was prepared from total RNA extracted from cell pellets or from RNA purified from fractionated gradients. 1 µg of RNA in a final volume of 14.4 µl was pipetted into a clean 0.5 ml PCR tube. The sample was heat denatured at 90°C for 3 min and transferred onto ice immediately after.

5.6 µl of master mix containing:

- 4 µl 5x RT buffer
- 0.4 µl 10 mM dNTPs
- 0.5 µl random hexamers
- 0.2 µl RNasin
- 0.5 µl AMV reverse transcriptase

was added. The mixture was incubated for 10 min at 25°C followed by a 1 h incubation at 42°C. Subsequently heat inactivation was performed at 90°C for 5 min. The cDNA was diluted 16-fold with nuclease-free water. 5 µl of diluted cDNA was used in RT-qPCR measurements.

6.5. Immunofluorescence analysis

Cells for immunofluorescence analysis were washed twice with PBS, dropped on microscopy slides and left to air dry. After drying cells were fixed depending on the antibody either with 4% PFA for 20 min at room temperature or in acetone for also for 20 min. PFA fixed cells were permeabilized in PBS with 0.5% Triton X-100 for 2 min. Incubation with the primary antibody was performed for 30 min at 37°C in a humidity chamber. Afterwards the slides were washed three times for 5 min in PBS. Incubation with the secondary Cy3-labeled antibody was carried out for another 30 min at 37°C in a humidity chamber. Again, slides were washed three times in PBS for 5 min. Nuclei were counterstained in DAPI solution for 1 min. Cover slips were mounted with 90% glycerol in PBS.

6.6. Polysome profiling

6.6.1. Preparation of linear sucrose gradients

To prepare linear sucrose gradients for polysome and ribosome profiling 5 different sucrose solutions were prepared in polysome lysis buffer: 50%, 41.9%, 33.8%, 25.6% and 17.5%.

To pour the gradients first 790 μ l of 50% sucrose solution was slowly pipetted into the bottom of a Beckman polyallomer tube. The tube was covered with aluminum foil and frozen for 20 min at -80°C . The frozen 50% sucrose was carefully overlaid with 790 μ l of 41.9% sucrose solution. Again the tubes were covered with aluminum foil and frozen for 20 min at -80°C . This procedure was continued until all the sucrose solutions were frozen on top of one another, ending with 17.5%.

Gradients were stored at -80°C until use. The night before polysome or ribosome profiling experiments the gradients were transferred into the 4°C room to thaw.

6.6.2. Preparation of cells for ultracentrifugation

Prior to profiling LCLs were expanded to large culture flasks. The cells were split 1:3 48 hours prior to profiling. On the day of the profiling experiment the LCLs were then diluted 6×10^5 cells/ml and left in the incubator for another 1.5 hours. A total of 3×10^7 cells were used per profiling experiment. For profiling LCLs were treated with 100 μ g/ml cycloheximide for 5 minutes to stall ribosomes on RNA. Cells were then centrifuged at 4°C for 5 minutes at 1500 rpm. Supernatant was removed and LCLs were lysed in 300 μ l polysome lysis buffer. Lysates were incubated for 10 minutes at 4°C on a roller. Afterwards nuclei were removed by centrifugation at 10000 rpm for 10 minutes at 4°C . The resulting supernatant was loaded onto a linear sucrose gradient. Ultracentrifugation was run at 4°C at 35000 rpm for 2.5 hours in an SW60Ti rotor.

6.6.3. Gradient fractionation

Following centrifugation the gradients were fractionated using a Teledyne Isco Foxy Jr. Gradient fractionator. The device separated the gradient into 12 fractions containing 400 μ l of sucrose gradient. RNA was purified from the single fractions using organic solvent extraction followed by isopropanol precipitation.

6.7. Ribosome profiling

6.7.1. Generation and purification of ribosome protected fragments (RPFs)

Ribosome profiles were generated according to a modified protocol from Ingolia et al⁴. LCLs were prepared according to the polysome profiling protocol (see section: Polysome profiling). In addition to treatment with cycloheximide samples that were prepared for open reading frame identification were incubated for 2 or 5 min with 2 µg/ml harringtonine at 37°C prior to addition of 100 µg/ml cycloheximide for another 5 min. The cells were then transferred into a falcon tube and centrifuged for 5 min at 4°C at 1200 rpm. The pellet was lysed with 300 µl ice-cold polysome lysis buffer and ultracentrifuged and fractionated as described in the polysome profiling protocol. Fractionated samples were treated with 640U RNase I per 1 OD A₂₆₀ for 15min on a roller at room temperature. RNA was purified with phenol:chloroform:isoamyl alcohol as described in section 6.4.2. with a slight modification: Precipitations were carried out at -80°C overnight and the recovered RNA was resuspended in 5 µl of 10mM RNase-free Tris (pH 8).

6.7.2. RNA size selection

RPFs were size-selected by polyacrylamide gel electrophoresis using a pre-cast 15% polyacrylamide TBE-urea gel. Samples were mixed with 5 µl of 2x denaturing loading buffer and were run with synthesized control oligonucleotides of 26 and 34 nt size and a 10 bp DNA ladder. The samples were denatured at 80°C for 90 sec. The gel was run for 65min at 200 V and was subsequently stained with 1:20000 dilution of SYBR Gold dye in 1x TBE buffer for 5 min. The region on the gel was excised that was demarcated by the two control oligonucleotides ranging from 26 to 34 nt. The gel slices were transferred into a fresh RNase-free tube. Additionally the marker oligonucleotides were also excised as internal control. To the gel slices 400µl RNA gel extraction buffer was added and the samples were frozen for 30min on dry ice. The samples were then left on a roller overnight at room temperature.

The following day the RNA was precipitated with 1.5 µl GlycoBlue and 500 µl isopropanol. The samples were placed into the -80°C freezer for 2h and were then spun at max speed for 30min at 4°C. Size-selected RNA was then resuspended in 10µl 10mM Tris (pH 8) and transferred to a fresh RNase-free tube.

6.7.3. Dephosphorylation & linker ligation

For the dephosphorylation reaction 33 μ l of RNase-free water was added to the 10 μ l of resuspended, size-selected RNA. The samples were then denatured for 90sec at 80°C and left to equilibrate to 37°C.

The following reaction was then set up (50 μ l final volume):

1x T4 PNK buffer
20U RNasin
10U T4 PNK

The reaction was incubated at 37°C for 1h and the enzyme was then heat-inactivated for 10 min at 70°C.

RNA was precipitated by addition of 39 μ l water, 1 μ l GlycoBlue, 10 μ l 3M sodium acetate and 150 μ l isopropanol. The samples were placed into the -80°C freezer for 2h and were then spun at max speed for 30min at 4°C. Dephosphorylated RNA was resuspended in 8.5 μ l 10mM Tris (pH 8) .

To the 8.5 μ l of dephosphorylated RNA 1.5 μ l of preadenylated linker from the Universal miRNA cloning linker kit were added and denatured at 80°C for 90sec and then let to cool to room temperature.

The following linker ligation reaction was set up (20 μ l final volume):

1x T4 Rnl2 buffer
15% (w/v) PEG 8000
20U RNasin
200 U T4 Rnl2(tr)

The reaction was incubated at room temperature for 2.5h. After linker ligation 338 μ l of RNase-free water were added to the reaction and precipitated with 40 μ l 3M sodium acetate, 1.5 μ l GlycoBlue and 500 μ l isopropanol. The samples were placed into the -80°C freezer for 2h and were then spun at max speed for 30 min at 4°C.

The ligation reactions were separated by polyacrylamide gel electrophoresis as described in the size-selection section of the protocol.

The ligation products were excised from the gel as demarcated by the ligated marker oligos. RNA was recovered again in 400 μ l RNA gel extraction buffer as described in the size-selection section of the protocol.

After isopropanol precipitation the ligated RNA was resuspended in 10µl 10mM Tris (pH 8).

6.7.4. Reverse transcription

To the ligated RNA 2 µl of reverse transcription primer (1.25 µM) was added and denatured for 2 min at 80°C. The samples were then immediately transferred on ice. A thermal cycler was pre-warmed to 48°C.

The following reverse transcription reaction was set up (20 µl final volume):

- 1x first strand buffer
- 0.5 mM dNTPs
- 5 mM DTT
- 20U RNasin
- 200U SuperScript III

The reaction mixture was incubated for 30min at 48°C. RNA was then hydrolysed by the addition of 2.2 µl of 1N NaOH and incubation for 20min at 98°C. The cDNA was then precipitated by addition of 20 µl 3M sodium acetate, 2 µl GlycoBlue, 156 µl water and 300 µl isopropanol. The samples were placed into the -80°C freezer for 2h and were then spun at max speed for 30min at 4°C and resuspended in 10 µl 10mM Tris (pH 8). Precipitated cDNA was further separated from unextended primer by polyacrylamide gel electrophoresis as described in the size-selection section of the protocol. As a control 10 µl of reverse transcription primer at 1.25 µM concentration was prepared and run with the samples. The reverse transcribed product was excised from the gel and transferred into a fresh tube. The cDNA was extracted with 400 µl DNA gel extraction buffer. The gel slice in the buffer was frozen on dry ice for 30 min and was then transferred onto a roller at room temperature overnight. The eluted cDNA was precipitated as described in the size-selection section of the protocol. Precipitated cDNA was resuspended in 15 µl 10mM Tris (pH 8) and transferred to a PCR tube.

6.7.5. Circularization

The circularization reaction was prepared as following (20 µl final volume):

1x CircLigase buffer
50mM ATP
2.5mM MnCl₂
100U CircLigase

The reaction mixture was incubated at 60°C for 1 h.

6.7.6. Depletion of rRNA

For rRNA depletion 5 µl of circularized cDNA was mixed with 1 µl of subtraction oligo mix, 1 µl 20x SSC and 3 µl water. The mix was denatured at 100°C for 90 sec and then an annealing reaction was setup at 0.1°C/s to 37°C. Once the thermocycler reached 37°C the samples were incubated for 15min.

In the mean time MyOne Streptavidin C1 DynaBeads were vortexed and 25µl of the beads was used per subtraction reaction. The beads were aliquotted to a fresh tube and washed three times using a magnetic rack in 1x bind/wash buffer. After the final wash the beads were resuspended in 10 µl/subtraction reaction in 2x bind/wash buffer. The beads were transferred to 37°C and the samples from the thermocycler were directly transferred to the beads. The mix was incubated for 15min at 37°C with vigorous shaking. After the incubation time the samples were directly transferred to the magnetic rack and beads were isolated for 1min at a time on the rack. 17.5 µl of eluate was recovered from the rRNA depletion procedure and transferred to a fresh tube. The depleted cDNA was recovered by precipitation with 2 µl GlycoBlue, 6 µl 5M NaCl, 74 µl water and 150 µl isopropanol. DNA was frozen at -80°C for 2 h and spun at 13000 rpm for 30 min at 4°C. The recovered cDNA was resuspended in 5 µl 10mM Tris (pH 8).

6.7.7. PCR amplification and barcoding

To the 5 µl of cDNA 95 µl of following master mix was added:

- 20 µl 5x Phusion HF buffer
- 2 µl 10 mM dNTPs
- 0.5 µl forward library primer
- 0.5 µl reverse indexed primer*
- 71 µl RNase-free water
- 1 µl Phusion polymerase (2U/µl)

**A different indexing primer was used for each sample.*

The 100 µl reaction was divided up into five 16.7 µl aliquots. PCR amplification was performed with varying cycle numbers. The five tubes containing the samples were placed into a thermal cycler and following cycling conditions were used:

| step | temperature | duration |
|-------------|--------------------|-----------------|
| 1 | 98°C | 30 sec |
| 2 | 98°C | 10 sec |
| 3 | 65°C | 10 sec |
| 4 | 72°C | 5 sec |

steps 2-4 were run for 18 cycles

Following cycling 3.3 µl of 6x nondenaturing loading dye was added to each tube and the samples were run on an 8% polyacrylamide nondenaturing gel. The gel was run for 40 min at 180 V. The gel was then stained in 1x SYBR Gold prepared in 1x TBE buffer. The bands were visualized by placing the gel on a UV box. A product band of approximately 175 nt size was excised from the gels. The gel slices were placed into 1.5 ml Eppendorf tubes and the PCR amplification products were recovered by adding 400 µl of DNA gel extraction buffer to each gel slice. The gel slice in the buffer was frozen on dry ice for 30 min and was then transferred onto a roller at room temperature overnight. The eluted cDNA was precipitated as described in the size-selection section of the protocol. Precipitated cDNA was resuspended in 15 µl 10mM Tris (pH 8) and transferred to a PCR tube.

6.7.8. Sequencing

The samples were sequenced at the Genomics and Proteomics Core Facility of the DKFZ on a HiSeq V3 sequencer from Illumina.

6.8. Bioinformatics

6.8.1. Read alignment & normalization

Adaptor sequences were trimmed with FastX clipper. STAR was used for the alignment. First, reads mapping to rRNA sequences were removed and then aligned to the human genome issue HG-19 as well as to the corresponding viral genomes B95-8 (accession number NC_007605.1) and M81 (accession number KF373730.1).

Read lengths of sequenced reads were determined with Fastq.

Reads were normalized in two ways depending on the read origin. For RPFs derived from cellular transcripts a scaling factor was calculated. The scaling factor is the ratio between the total number of reads in M81-infected LCLs to those in B95-8-infected LCLs. The absolute RPF counts were then multiplied by this scaling factor. RPF densities were calculated as reads per kilobase of coding DNA sequence.

This normalization method is not suitable for the low number of reads that mapped to the viral genome. Total count normalization described by Dillies *et al.*⁵ was used to compare RPF densities on viral transcripts instead. Here, absolute RPF counts were divided by the number of uniquely mapped reads that aligned to the viral genome in the respective libraries.

6.8.2. Metagene analysis

To analyse the aggregation of reads around start and stop codons (metagene analysis) the reads of 27 and 28 nt length were considered. Reads of this length were counted on transcripts and were normalized by transcript length. Transcripts with an average length of > 2 were considered for analysis. Average read depth of nucleotide positions 20 bases up- and downstream of start and stop codons was then calculated.

6.8.3. Identification of translation initiation sites

Harringtonine-treated libraries were used to identify translation initiation sites. The machine learning approach described by Stern-Ginossar *et al.*⁶ was not reliable enough due to the low ribosome coverage of the viral genome. The identification of novel small viral ORFs was carried out by manual screening of the alignments in IGV. Putative ORFs were scored as translated when at least two of the three harringtonine-treated libraries from a strain were positive for it. The start codon within a peak was designated as the 15th nucleotide of reads with a length of 27-28 nt and the 16th if the reads were longer.

Figures showing ribosome coverage were generated using the Sashimi Plot function in IGV.

6.8.4. Calculation of ratios

The out:in ratios were calculated by dividing the number of RPFs mapping within 5'leaders of a transcript (out) by the number of RPFs mapping to the coding region of the transcript (in). An arbitrary count of 1 was added to both out and in counts. Read coverage was length normalized by nucleotide length of the respective feature (5'leader and coding region).

5'leader:AUG ratios were calculated by dividing the number of RPF mapping within the 5'leaders of a transcript by the number of RPFs mapping to the start codon of a transcript (AUG). An arbitrary count of 1 was added to both out and in counts. Read coverage was length normalized by nucleotide length of the respective feature (5'leader and AUG triplet).

6.8.5. UORF conservation analysis

Genome sequences of different EBV strains were downloaded from the NCBI database and genomic regions of interest were aligned against each other using MacVector software.

6.8.6. IPA analysis

Ingenuity pathway analysis (IPA) software from Qiagen was used to assess differentially regulated genes and the associated pathways and biological functions in the two strains. For the analysis the cycloheximide-derived data sets were used.

Transcript translation levels were first calculated by quantifying the read number per kilobase of coding sequence and normalizing with the median calculated over all transcripts. These expression values were loaded into the IPA software for analysis.

6.9. Statistical analysis

Fisher's exact test was used to assess significance of results generated from IPA analysis. The Fisher's exact test was performed with the IPA software.

For all other analyses the unpaired t test was used to assess significance. Prism software was used for performing the unpaired t test. Where depicted, error bars represent standard deviations of the data.

6.10. Work distribution

I performed the generation of the ribosome profiles and all the associated wet-lab experiments. The pre-processing and alignment of reads from the sequenced libraries were done by Dr. Olaf Klinke. Quality control assessments were conducted by Dr. Agnes Hotz-Wagenblatt from the DKFZ Core Facility Genomics and Proteomics and by myself.

7. References

- 1 Epstein, M. A., Achong, B. G. & Barr, Y. M. Virus Particles in Cultured Lymphoblasts from Burkitt's Lymphoma. *Lancet* **1**, 702-703 (1964).
- 2 de-The, G. *et al.* Sero-epidemiology of the Epstein-Barr virus: preliminary analysis of an international study - a review. *IARC scientific publications*, 3-16 (1975).
- 3 Venkitaraman, A. R., Lenoir, G. M. & John, T. J. The seroepidemiology of infection due to Epstein-Barr virus in southern India. *Journal of medical virology* **15**, 11-16 (1985).
- 4 Niederman, J. C., Evans, A. S., Subrahmanyam, L. & McCollum, R. W. Prevalence, incidence and persistence of EB virus antibody in young adults. *The New England journal of medicine* **282**, 361-365, doi:10.1056/NEJM197002122820704 (1970).
- 5 Zur Hausen, H. The search for infectious causes of human cancers: where and why. *Virology* **392**, 1-10, doi:10.1016/j.virol.2009.06.001 (2009).
- 6 Kieff, E. & Rickinson, A. B. in *Fields Virology* (eds D.M. Knipe & P.M. Howley) 2603-2654 (Lippincott Williams & Wilkins, 2007).
- 7 Pavlova, S. *et al.* An Epstein-Barr virus mutant produces immunogenic defective particles devoid of viral DNA. *Journal of virology* **87**, 2011-2022, doi:10.1128/JVI.02533-12 (2013).
- 8 Penkert, R. R. & Kalejta, R. F. Tegument protein control of latent herpesvirus establishment and animation. *Herpesviridae* **2**, 3, doi:10.1186/2042-4280-2-3 (2011).
- 9 Cheung, A. & Kieff, E. Long internal direct repeat in Epstein-Barr virus DNA. *Journal of virology* **44**, 286-294 (1982).
- 10 Raab-Traub, N. & Flynn, K. The structure of the termini of the Epstein-Barr virus as a marker of clonal cellular proliferation. *Cell* **47**, 883-889 (1986).
- 11 Baer, R. *et al.* DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* **310**, 207-211 (1984).
- 12 Young, L. S. & Murray, P. G. Epstein-Barr virus and oncogenesis: from latent genes to tumours. *Oncogene* **22**, 5108-5121, doi:10.1038/sj.onc.1206556 (2003).

- 13 Lam, N. & Sugden, B. CD40 and its viral mimic, LMP1: similar means to different ends. *Cellular signalling* **15**, 9-16 (2003).
- 14 Caldwell, R. G., Wilson, J. B., Anderson, S. J. & Longnecker, R. Epstein-Barr virus LMP2A drives B cell development and survival in the absence of normal B cell receptor signals. *Immunity* **9**, 405-411 (1998).
- 15 Glickman, J. N., Howe, J. G. & Steitz, J. A. Structural analyses of EBER1 and EBER2 ribonucleoprotein particles present in Epstein-Barr virus-infected cells. *Journal of virology* **62**, 902-911 (1988).
- 16 Houmani, J. L., Davis, C. I. & Ruf, I. K. Growth-promoting properties of Epstein-Barr virus EBER-1 RNA correlate with ribosomal protein L22 binding. *Journal of virology* **83**, 9844-9853, doi:10.1128/JVI.01014-09 (2009).
- 17 Toczyski, D. P., Matera, A. G., Ward, D. C. & Steitz, J. A. The Epstein-Barr virus (EBV) small RNA EBER1 binds and relocalizes ribosomal protein L22 in EBV-infected human B lymphocytes. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 3463-3467 (1994).
- 18 Lee, N., Moss, W. N., Yario, T. A. & Steitz, J. A. EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. *Cell* **160**, 607-618, doi:10.1016/j.cell.2015.01.015 (2015).
- 19 Pfeffer, S. *et al.* Identification of virus-encoded microRNAs. *Science* **304**, 734-736, doi:10.1126/science.1096781 (2004).
- 20 Grundhoff, A., Sullivan, C. S. & Ganem, D. A combined computational and microarray-based approach identifies novel microRNAs encoded by human gamma-herpesviruses. *Rna* **12**, 733-750, doi:10.1261/rna.2326106 (2006).
- 21 Klinke, O., Feederle, R. & Delecluse, H. J. Genetics of Epstein-Barr virus microRNAs. *Seminars in cancer biology* **26**, 52-59, doi:10.1016/j.semcancer.2014.02.002 (2014).
- 22 Cai, X. *et al.* Epstein-Barr virus microRNAs are evolutionarily conserved and differentially expressed. *PLoS pathogens* **2**, e23, doi:10.1371/journal.ppat.0020023 (2006).
- 23 Feederle, R. *et al.* A viral microRNA cluster strongly potentiates the transforming properties of a human herpesvirus. *PLoS pathogens* **7**, e1001294, doi:10.1371/journal.ppat.1001294 (2011).

- 24 Feederle, R. *et al.* The members of an Epstein-Barr virus microRNA cluster cooperate to transform B lymphocytes. *Journal of virology* **85**, 9801-9810, doi:10.1128/JVI.05100-11 (2011).
- 25 Seto, E. *et al.* Micro RNAs of Epstein-Barr virus promote cell cycle progression and prevent apoptosis of primary human B cells. *PLoS pathogens* **6**, e1001063, doi:10.1371/journal.ppat.1001063 (2010).
- 26 Lin, X. *et al.* The Epstein-Barr Virus BART miRNA Cluster of the M81 Strain Modulates Multiple Functions in Primary B Cells. *PLoS pathogens* **11**, e1005344, doi:10.1371/journal.ppat.1005344 (2015).
- 27 Qiu, J. & Thorley-Lawson, D. A. EBV microRNA BART 18-5p targets MAP3K2 to facilitate persistence in vivo by inhibiting viral replication in B cells. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 11157-11162, doi:10.1073/pnas.1406136111 (2014).
- 28 Hutzinger, R. *et al.* Expression and processing of a small nucleolar RNA from the Epstein-Barr virus genome. *PLoS pathogens* **5**, e1000547, doi:10.1371/journal.ppat.1000547 (2009).
- 29 Moss, W. N. & Steitz, J. A. Genome-wide analyses of Epstein-Barr virus reveal conserved RNA structures and a novel stable intronic sequence RNA. *BMC genomics* **14**, 543, doi:10.1186/1471-2164-14-543 (2013).
- 30 Lin, Z. *et al.* Whole-genome sequencing of the Akata and Mutu Epstein-Barr virus strains. *Journal of virology* **87**, 1172-1182, doi:10.1128/JVI.02517-12 (2013).
- 31 O'Grady, T. *et al.* Global bidirectional transcription of the Epstein-Barr virus genome during reactivation. *Journal of virology* **88**, 1604-1616, doi:10.1128/JVI.02989-13 (2014).
- 32 Young, L. S., Arrand, J. R. & Murray, P. G. in *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis* (eds A. Arvin *et al.*) (2007).
- 33 O'Grady, T. *et al.* Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucleic acids research*, doi:10.1093/nar/gkw629 (2016).
- 34 Tugizov, S. M., Berline, J. W. & Palefsky, J. M. Epstein-Barr virus infection of polarized tongue and nasopharyngeal epithelial cells. *Nature medicine* **9**, 307-314, doi:10.1038/nm830 (2003).

- 35 Ruoslahti, E. RGD and other recognition sequences for integrins. *Annual review of cell and developmental biology* **12**, 697-715, doi:10.1146/annurev.cellbio.12.1.697 (1996).
- 36 Jones, J. F. *et al.* T-cell lymphomas containing Epstein-Barr viral DNA in patients with chronic Epstein-Barr virus infections. *The New England journal of medicine* **318**, 733-741, doi:10.1056/NEJM198803243181203 (1988).
- 37 Wensing, B. & Farrell, P. J. Regulation of cell growth and death by Epstein-Barr virus. *Microbes and infection / Institut Pasteur* **2**, 77-84 (2000).
- 38 Thompson, M. P. & Kurzrock, R. Epstein-Barr virus and cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **10**, 803-821 (2004).
- 39 Fingerroth, J. D. *et al.* Epstein-Barr virus receptor of human B lymphocytes is the C3d receptor CR2. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 4510-4514 (1984).
- 40 Nemerow, G. R., Wolfert, R., McNaughton, M. E. & Cooper, N. R. Identification and characterization of the Epstein-Barr virus receptor on human B lymphocytes and its relationship to the C3d complement receptor (CR2). *Journal of virology* **55**, 347-351 (1985).
- 41 Tanner, J., Weis, J., Fearon, D., Whang, Y. & Kieff, E. Epstein-Barr virus gp350/220 binding to the B lymphocyte C3d receptor mediates adsorption, capping, and endocytosis. *Cell* **50**, 203-213 (1987).
- 42 Ogembo, J. G. *et al.* Human complement receptor type 1/CD35 is an Epstein-Barr Virus receptor. *Cell reports* **3**, 371-385, doi:10.1016/j.celrep.2013.01.023 (2013).
- 43 Li, Q. *et al.* Epstein-Barr virus uses HLA class II as a cofactor for infection of B lymphocytes. *Journal of virology* **71**, 4657-4662 (1997).
- 44 Sixbey, J. W. *et al.* Replication of Epstein-Barr virus in human epithelial cells infected in vitro. *Nature* **306**, 480-483 (1983).
- 45 Shannon-Lowe, C. & Rowe, M. Epstein-Barr virus infection of polarized epithelial cells via the basolateral surface by memory B cell-mediated transfer infection. *PLoS pathogens* **7**, e1001338, doi:10.1371/journal.ppat.1001338 (2011).

- 46 Birkenbach, M., Tong, X., Bradbury, L. E., Tedder, T. F. & Kieff, E. Characterization of an Epstein-Barr virus receptor on human epithelial cells. *The Journal of experimental medicine* **176**, 1405-1414 (1992).
- 47 Yoshiyama, H., Imai, S., Shimizu, N. & Takada, K. Epstein-Barr virus infection of human gastric carcinoma cells: implication of the existence of a new virus receptor different from CD21. *Journal of virology* **71**, 5688-5691 (1997).
- 48 Molesworth, S. J., Lake, C. M., Borza, C. M., Turk, S. M. & Hutt-Fletcher, L. M. Epstein-Barr virus gH is essential for penetration of B cells but also plays a role in attachment of virus to epithelial cells. *Journal of virology* **74**, 6324-6332 (2000).
- 49 Wang, H. B. *et al.* Neuropilin 1 is an entry factor that promotes EBV infection of nasopharyngeal epithelial cells. *Nature communications* **6**, 6240, doi:10.1038/ncomms7240 (2015).
- 50 Shannon-Lowe, C. D., Neuhierl, B., Baldwin, G., Rickinson, A. B. & Delecluse, H. J. Resting B cells as a transfer vehicle for Epstein-Barr virus infection of epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 7065-7070, doi:10.1073/pnas.0510512103 (2006).
- 51 Feederle, R. *et al.* Epstein-Barr virus B95.8 produced in 293 cells shows marked tropism for differentiated primary epithelial cells and reveals interindividual variation in susceptibility to viral infection. *International journal of cancer* **121**, 588-594, doi:10.1002/ijc.22727 (2007).
- 52 Herrmann, K., Frangou, P., Middeldorp, J. & Niedobitek, G. Epstein-Barr virus replication in tongue epithelial cells. *The Journal of general virology* **83**, 2995-2998, doi:10.1099/0022-1317-83-12-2995 (2002).
- 53 Temple, R. M. *et al.* Efficient replication of Epstein-Barr virus in stratified epithelium in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 16544-16549, doi:10.1073/pnas.1400818111 (2014).
- 54 Tsang, C. M. *et al.* Epstein-Barr virus infection in immortalized nasopharyngeal epithelial cells: regulation of infection and phenotypic characterization. *International journal of cancer* **127**, 1570-1583, doi:10.1002/ijc.25173 (2010).

- 55 Tsang, C. M. *et al.* Cyclin D1 overexpression supports stable EBV infection in nasopharyngeal epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E3473-3482, doi:10.1073/pnas.1202637109 (2012).
- 56 Hardwick, J. M., Lieberman, P. M. & Hayward, S. D. A new Epstein-Barr virus transactivator, R, induces expression of a cytoplasmic early antigen. *Journal of virology* **62**, 2274-2284 (1988).
- 57 Miller, G. The switch between latency and replication of Epstein-Barr virus. *The Journal of infectious diseases* **161**, 833-844 (1990).
- 58 Zimmermann, J. & Hammerschmidt, W. Structure and role of the terminal repeats of Epstein-Barr virus in processing and packaging of virion DNA. *Journal of virology* **69**, 3147-3155 (1995).
- 59 Pfuller, R. & Hammerschmidt, W. Plasmid-like replicative intermediates of the Epstein-Barr virus lytic origin of DNA replication. *Journal of virology* **70**, 3423-3431 (1996).
- 60 Amon, W. *et al.* Lytic cycle gene regulation of Epstein-Barr virus. *Journal of virology* **78**, 13460-13469, doi:10.1128/JVI.78.24.13460-13469.2004 (2004).
- 61 Summers, W. C. & Klein, G. Inhibition of Epstein-Barr virus DNA synthesis and late gene expression by phosphonoacetic acid. *Journal of virology* **18**, 151-155 (1976).
- 62 Farina, A. *et al.* BFRF1 of Epstein-Barr virus is essential for efficient primary viral envelopment and egress. *Journal of virology* **79**, 3703-3712, doi:10.1128/JVI.79.6.3703-3712.2005 (2005).
- 63 Torrisi, M. R. *et al.* Localization of Epstein-Barr virus envelope glycoproteins on the inner nuclear membrane of virus-producing cells. *Journal of virology* **63**, 828-832 (1989).
- 64 Laichalk, L. L. & Thorley-Lawson, D. A. Terminal differentiation into plasma cells initiates the replicative cycle of Epstein-Barr virus in vivo. *Journal of virology* **79**, 1296-1307, doi:10.1128/JVI.79.2.1296-1307.2005 (2005).
- 65 Anagnostopoulos, I., Hummel, M., Kreschel, C. & Stein, H. Morphology, immunophenotype, and distribution of latently and/or productively Epstein-Barr virus-infected cells in acute infectious mononucleosis: implications for the interindividual infection route of Epstein-Barr virus. *Blood* **85**, 744-750 (1995).

- 66 Niedobitek, G., Agathangelou, A., Steven, N. & Young, L. S. Epstein-Barr virus (EBV) in infectious mononucleosis: detection of the virus in tonsillar B lymphocytes but not in desquamated oropharyngeal epithelial cells. *Molecular pathology : MP* **53**, 37-42 (2000).
- 67 Kenney, S. C. & Mertz, J. E. Regulation of the latent-lytic switch in Epstein-Barr virus. *Seminars in cancer biology* **26**, 60-68, doi:10.1016/j.semcancer.2014.01.002 (2014).
- 68 Amon, W. & Farrell, P. J. Reactivation of Epstein-Barr virus from latency. *Reviews in medical virology* **15**, 149-156, doi:10.1002/rmv.456 (2005).
- 69 Tsurumi, T., Fujita, M. & Kudoh, A. Latent and lytic Epstein-Barr virus replication strategies. *Reviews in medical virology* **15**, 3-15, doi:10.1002/rmv.441 (2005).
- 70 Sinclair, A. J. & Farrell, P. J. Host cell requirements for efficient infection of quiescent primary B lymphocytes by Epstein-Barr virus. *Journal of virology* **69**, 5461-5468 (1995).
- 71 Bechtel, D., Kurth, J., Unkel, C. & Kuppers, R. Transformation of BCR-deficient germinal-center B cells by EBV supports a major role of the virus in the pathogenesis of Hodgkin and posttransplantation lymphomas. *Blood* **106**, 4345-4350, doi:10.1182/blood-2005-06-2342 (2005).
- 72 Jochum, S., Moosmann, A., Lang, S., Hammerschmidt, W. & Zeidler, R. The EBV immunoevasins vIL-10 and BNLF2a protect newly infected B cells from immune recognition and elimination. *PLoS pathogens* **8**, e1002704, doi:10.1371/journal.ppat.1002704 (2012).
- 73 Jochum, S., Ruiss, R., Moosmann, A., Hammerschmidt, W. & Zeidler, R. RNAs in Epstein-Barr virions control early steps of infection. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E1396-1404, doi:10.1073/pnas.1115906109 (2012).
- 74 Kalla, M. & Hammerschmidt, W. Human B cells on their route to latent infection--early but transient expression of lytic genes of Epstein-Barr virus. *European journal of cell biology* **91**, 65-69, doi:10.1016/j.ejcb.2011.01.014 (2012).
- 75 Woellmer, A. & Hammerschmidt, W. Epstein-Barr virus and host cell methylation: regulation of latency, replication and virus reactivation. *Current opinion in virology* **3**, 260-265, doi:10.1016/j.coviro.2013.03.005 (2013).

- 76 Alfieri, C., Birkenbach, M. & Kieff, E. Early events in Epstein-Barr virus infection of human B lymphocytes. *Virology* **181**, 595-608 (1991).
- 77 Woisetschlaeger, M., Yandava, C. N., Furmanski, L. A., Strominger, J. L. & Speck, S. H. Promoter switching in Epstein-Barr virus during the initial stages of infection of B lymphocytes. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 1725-1729 (1990).
- 78 Altmann, M. & Hammerschmidt, W. Epstein-Barr virus provides a new paradigm: a requirement for the immediate inhibition of apoptosis. *PLoS biology* **3**, e404, doi:10.1371/journal.pbio.0030404 (2005).
- 79 Kalla, M., Schmeinck, A., Bergbauer, M., Pich, D. & Hammerschmidt, W. AP-1 homolog BZLF1 of Epstein-Barr virus has two essential functions dependent on the epigenetic state of the viral genome. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 850-855, doi:10.1073/pnas.0911948107 (2010).
- 80 Wen, W. *et al.* Epstein-Barr virus BZLF1 gene, a switch from latency to lytic infection, is expressed as an immediate-early gene after primary infection of B lymphocytes. *Journal of virology* **81**, 1037-1042, doi:10.1128/JVI.01416-06 (2007).
- 81 Price, A. M. & Luftig, M. A. To be or not IIb: a multi-step process for Epstein-Barr virus latency establishment and consequences for B cell tumorigenesis. *PLoS pathogens* **11**, e1004656, doi:10.1371/journal.ppat.1004656 (2015).
- 82 Thorley-Lawson, D. A. & Mann, K. P. Early events in Epstein-Barr virus infection provide a model for B cell activation. *The Journal of experimental medicine* **162**, 45-59 (1985).
- 83 Tierney, R. *et al.* Epstein-Barr virus exploits BSAP/Pax5 to achieve the B-cell specificity of its growth-transforming program. *Journal of virology* **81**, 10092-10100, doi:10.1128/JVI.00358-07 (2007).
- 84 Shannon-Lowe, C. *et al.* Epstein-Barr virus-induced B-cell transformation: quantitating events from virus binding to cell outgrowth. *The Journal of general virology* **86**, 3009-3019, doi:10.1099/vir.0.81153-0 (2005).
- 85 Hoagland, R. J. The transmission of infectious mononucleosis. *The American journal of the medical sciences* **229**, 262-272 (1955).

- 86 Tugizov, S. M., Herrera, R. & Palefsky, J. M. Epstein-Barr virus transcytosis through polarized oral epithelial cells. *Journal of virology* **87**, 8179-8194, doi:10.1128/JVI.00443-13 (2013).
- 87 Thorley-Lawson, D. A. & Allday, M. J. The curious case of the tumour virus: 50 years of Burkitt's lymphoma. *Nature reviews. Microbiology* **6**, 913-924, doi:10.1038/nrmicro2015 (2008).
- 88 Thorley-Lawson, D. A. & Gross, A. Persistence of the Epstein-Barr virus and the origins of associated lymphomas. *The New England journal of medicine* **350**, 1328-1337, doi:10.1056/NEJMra032015 (2004).
- 89 Young, L. S. & Rickinson, A. B. Epstein-Barr virus: 40 years on. *Nature reviews. Cancer* **4**, 757-768, doi:10.1038/nrc1452 (2004).
- 90 Nikitin, P. A. *et al.* An ATM/Chk2-mediated DNA damage-responsive signaling pathway suppresses Epstein-Barr virus transformation of primary human B cells. *Cell host & microbe* **8**, 510-522, doi:10.1016/j.chom.2010.11.004 (2010).
- 91 Thorley-Lawson, D. A. & Strominger, J. L. Reversible inhibition by phosphonoacetic acid of human B lymphocyte transformation by Epstein-Barr virus. *Virology* **86**, 423-431 (1978).
- 92 Roughan, J. E. & Thorley-Lawson, D. A. The intersection of Epstein-Barr virus with the germinal center. *Journal of virology* **83**, 3968-3976, doi:10.1128/JVI.02609-08 (2009).
- 93 Thorley-Lawson, D. A. Epstein-Barr virus: exploiting the immune system. *Nature reviews. Immunology* **1**, 75-82, doi:10.1038/35095584 (2001).
- 94 Kirchmaier, A. L. & Sugden, B. Dominant-negative inhibitors of EBNA-1 of Epstein-Barr virus. *Journal of virology* **71**, 1766-1775 (1997).
- 95 Hochberg, D. R. & Thorley-Lawson, D. A. Quantitative detection of viral gene expression in populations of Epstein-Barr virus-infected cells in vivo. *Methods in molecular biology* **292**, 39-56 (2005).
- 96 Hochberg, D. *et al.* Acute infection with Epstein-Barr virus targets and overwhelms the peripheral memory B-cell compartment with resting, latently infected cells. *Journal of virology* **78**, 5194-5204 (2004).
- 97 Miller, G., Shope, T., Lisco, H., Stitt, D. & Lipman, M. Epstein-Barr virus: transformation, cytopathic changes, and viral antigens in squirrel monkey and marmoset leukocytes. *Proceedings of the National Academy of Sciences of the United States of America* **69**, 383-387 (1972).

- 98 Arrand, J. R. *et al.* Molecular cloning of the complete Epstein-Barr virus genome as a set of overlapping restriction endonuclease fragments. *Nucleic acids research* **9**, 2999-3014 (1981).
- 99 Raab-Traub, N., Dambaugh, T. & Kieff, E. DNA of Epstein-Barr virus VIII: B95-8, the previous prototype, is an unusual deletion derivative. *Cell* **22**, 257-267 (1980).
- 100 Edson, C. M. & Thorley-Lawson, D. A. Epstein-Barr virus membrane antigens: characterization, distribution, and strain differences. *Journal of virology* **39**, 172-184 (1981).
- 101 Skalsky, R. L. *et al.* The viral and cellular microRNA targetome in lymphoblastoid cell lines. *PLoS pathogens* **8**, e1002484, doi:10.1371/journal.ppat.1002484 (2012).
- 102 Palser, A. L. *et al.* Genome diversity of Epstein-Barr virus from multiple tumor types and normal infection. *Journal of virology* **89**, 5222-5237, doi:10.1128/JVI.03614-14 (2015).
- 103 Rickinson, A. B., Young, L. S. & Rowe, M. Influence of the Epstein-Barr virus nuclear antigen EBNA 2 on the growth phenotype of virus-transformed B cells. *Journal of virology* **61**, 1310-1317 (1987).
- 104 Aitken, C., Sengupta, S. K., Aedes, C., Moss, D. J. & Sculley, T. B. Heterogeneity within the Epstein-Barr virus nuclear antigen 2 gene in different strains of Epstein-Barr virus. *The Journal of general virology* **75 (Pt 1)**, 95-100, doi:10.1099/0022-1317-75-1-95 (1994).
- 105 Apolloni, A. & Sculley, T. B. Detection of A-type and B-type Epstein-Barr virus in throat washings and lymphocytes. *Virology* **202**, 978-981 (1994).
- 106 Kyaw, M. T. *et al.* Expression of B-type Epstein-Barr virus in HIV-infected patients and cardiac transplant recipients. *AIDS research and human retroviruses* **8**, 1869-1874, doi:10.1089/aid.1992.8.1869 (1992).
- 107 Sculley, T. B., Cross, S. M., Borrow, P. & Cooper, D. A. Prevalence of antibodies to Epstein-Barr virus nuclear antigen 2B in persons infected with the human immunodeficiency virus. *The Journal of infectious diseases* **158**, 186-192 (1988).
- 108 Young, L. S. *et al.* New type B isolates of Epstein-Barr virus from Burkitt's lymphoma and from normal individuals in endemic areas. *The Journal of*

- general virology* **68 (Pt 11)**, 2853-2862, doi:10.1099/0022-1317-68-11-2853 (1987).
- 109 Tsai, M. H. *et al.* Spontaneous lytic replication and epitheliotropism define an Epstein-Barr virus strain found in carcinomas. *Cell reports* **5**, 458-470, doi:10.1016/j.celrep.2013.09.012 (2013).
- 110 Desgranges, C. *et al.* In vitro transforming activity of EBV. I-Establishment and properties of two EBV strains (M81 and M72) produced by immortalized Callithrix jacchus lymphocytes. *Biomedicine / [publiee pour l'A.A.I.C.I.G.]* **25**, 349-352 (1976).
- 111 Desgranges, C., Lavoue, M. F., Patet, J. & de-The, G. In vitro transforming activity of Epstein-Barr virus (EBV). II. Differences between M81 and B95-8 EBV strains. *Biomedicine / [publiee pour l'A.A.I.C.I.G.]* **30**, 102-108 (1979).
- 112 Rickinson, A. B. Co-infections, inflammation and oncogenesis: future directions for EBV research. *Seminars in cancer biology* **26**, 99-115, doi:10.1016/j.semcancer.2014.04.004 (2014).
- 113 Arvey, A. *et al.* An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. *Cell host & microbe* **12**, 233-245, doi:10.1016/j.chom.2012.06.008 (2012).
- 114 Evans, A. S., Niederman, J. C. & McCollum, R. W. Seroepidemiologic studies of infectious mononucleosis with EB virus. *The New England journal of medicine* **279**, 1121-1127, doi:10.1056/NEJM196811212792101 (1968).
- 115 Balfour, H. H., Jr. *et al.* Behavioral, virologic, and immunologic factors associated with acquisition and severity of primary Epstein-Barr virus infection in university students. *The Journal of infectious diseases* **207**, 80-88, doi:10.1093/infdis/jis646 (2013).
- 116 Balfour, H. H., Jr. *et al.* Age-specific prevalence of Epstein-Barr virus infection among individuals aged 6-19 years in the United States and factors affecting its acquisition. *The Journal of infectious diseases* **208**, 1286-1293, doi:10.1093/infdis/jit321 (2013).
- 117 Balfour, H. H., Jr. & Verghese, P. Primary Epstein-Barr virus infection: impact of age at acquisition, coinfection, and viral load. *The Journal of infectious diseases* **207**, 1787-1789, doi:10.1093/infdis/jit096 (2013).

- 118 Morris, M. C. *et al.* Sero-epidemiological patterns of Epstein-Barr and herpes simplex (HSV-1 and HSV-2) viruses in England and Wales. *Journal of medical virology* **67**, 522-527, doi:10.1002/jmv.10132 (2002).
- 119 Takeuchi, K. *et al.* Prevalence of Epstein-Barr virus in Japan: trends and future prediction. *Pathology international* **56**, 112-116, doi:10.1111/j.1440-1827.2006.01936.x (2006).
- 120 Condon, L. M. *et al.* Age-specific prevalence of Epstein-Barr virus infection among Minnesota children: effects of race/ethnicity and family environment. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **59**, 501-508, doi:10.1093/cid/ciu342 (2014).
- 121 Hjalgrim, H. *et al.* Infectious mononucleosis, childhood social environment, and risk of Hodgkin lymphoma. *Cancer research* **67**, 2382-2388, doi:10.1158/0008-5472.CAN-06-3566 (2007).
- 122 Krabbe, S., Hesse, J. & Uldall, P. Primary Epstein-Barr virus infection in early childhood. *Archives of disease in childhood* **56**, 49-52 (1981).
- 123 Hadinoto, V. *et al.* On the dynamics of acute EBV infection and the pathogenesis of infectious mononucleosis. *Blood* **111**, 1420-1427, doi:10.1182/blood-2007-06-093278 (2008).
- 124 Brady, G., MacArthur, G. J. & Farrell, P. J. Epstein-Barr virus and Burkitt lymphoma. *Journal of clinical pathology* **60**, 1397-1402, doi:10.1136/jcp.2007.047977 (2007).
- 125 Njie, R. *et al.* The effects of acute malaria on Epstein-Barr virus (EBV) load and EBV-specific T cell immunity in Gambian children. *The Journal of infectious diseases* **199**, 31-38, doi:10.1086/594373 (2009).
- 126 Klein, G. Specific chromosomal translocations and the genesis of B-cell-derived tumors in mice and men. *Cell* **32**, 311-315 (1983).
- 127 Leder, P. Translocations among antibody genes in human cancer. *IARC scientific publications*, 341-357 (1985).
- 128 Manolov, G. & Manolova, Y. Marker band in one chromosome 14 from Burkitt lymphomas. *Nature* **237**, 33-34 (1972).
- 129 Kovalchuk, A. L. *et al.* Burkitt lymphoma in the mouse. *The Journal of experimental medicine* **192**, 1183-1190 (2000).

- 130 Neri, A. *et al.* Epstein-Barr virus infection precedes clonal expansion in Burkitt's and acquired immunodeficiency syndrome-associated lymphoma. *Blood* **77**, 1092-1095 (1991).
- 131 Pathmanathan, R. *et al.* Undifferentiated, nonkeratinizing, and squamous cell carcinoma of the nasopharynx. Variants of Epstein-Barr virus-infected neoplasia. *The American journal of pathology* **146**, 1355-1367 (1995).
- 132 Raab-Traub, N. *et al.* The differentiated form of nasopharyngeal carcinoma contains Epstein-Barr virus DNA. *International journal of cancer* **39**, 25-29 (1987).
- 133 Levine, P. H. *et al.* The reliability of IgA antibody to Epstein-Barr virus (EBV) capsid antigen as a test for the diagnosis of nasopharyngeal carcinoma (NPC). *Cancer detection and prevention* **4**, 307-312 (1981).
- 134 Wara, W. M., Wara, D. W., Phillips, T. L. & Ammann, A. J. Elevated IGA in carcinoma of the nasopharynx. *Cancer* **35**, 1313-1315 (1975).
- 135 Zeng, Y. *et al.* Prospective studies on nasopharyngeal carcinoma in Epstein-Barr virus IgA/VCA antibody-positive persons in Wuzhou City, China. *International journal of cancer* **36**, 545-547 (1985).
- 136 Cao, S. M. *et al.* Fluctuations of Epstein-Barr virus serological antibodies and risk for nasopharyngeal carcinoma: a prospective screening study with a 20-year follow-up. *PloS one* **6**, e19100, doi:10.1371/journal.pone.0019100 (2011).
- 137 Chien, Y. C. *et al.* Serologic markers of Epstein-Barr virus infection and nasopharyngeal carcinoma in Taiwanese men. *The New England journal of medicine* **345**, 1877-1882, doi:10.1056/NEJMoa011610 (2001).
- 138 Ning, J. P., Yu, M. C., Wang, Q. S. & Henderson, B. E. Consumption of salted fish and other risk factors for nasopharyngeal carcinoma (NPC) in Tianjin, a low-risk region for NPC in the People's Republic of China. *Journal of the National Cancer Institute* **82**, 291-296 (1990).
- 139 Yu, M. C., Huang, T. B. & Henderson, B. E. Diet and nasopharyngeal carcinoma: a case-control study in Guangzhou, China. *International journal of cancer* **43**, 1077-1082 (1989).
- 140 Zou, X. N., Lu, S. H. & Liu, B. Volatile N-nitrosamines and their precursors in Chinese salted fish--a possible etiological factor for NPC in China. *International journal of cancer* **59**, 155-158 (1994).

- 141 Li, K., Lin, G. Z., Shen, J. C. & Zhou, Q. Time trends of nasopharyngeal carcinoma in urban Guangzhou over a 12-year period (2000-2011): declines in both incidence and mortality. *Asian Pacific journal of cancer prevention : APJCP* **15**, 9899-9903 (2014).
- 142 Hildesheim, A. *et al.* Herbal medicine use, Epstein-Barr virus, and risk of nasopharyngeal carcinoma. *Cancer research* **52**, 3048-3051 (1992).
- 143 Chen, C. J. *et al.* Multiple risk factors of nasopharyngeal carcinoma: Epstein-Barr virus, malarial infection, cigarette smoking and familial tendency. *Anticancer research* **10**, 547-553 (1990).
- 144 Jia, W. H. *et al.* Complex segregation analysis of nasopharyngeal carcinoma in Guangdong, China: evidence for a multifactorial mode of inheritance (complex segregation analysis of NPC in China). *European journal of human genetics : EJHG* **13**, 248-252, doi:10.1038/sj.ejhg.5201305 (2005).
- 145 Chang, E. T. & Adami, H. O. The enigmatic epidemiology of nasopharyngeal carcinoma. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **15**, 1765-1777, doi:10.1158/1055-9965.EPI-06-0353 (2006).
- 146 Hildesheim, A. & Wang, C. P. Genetic predisposition factors and nasopharyngeal carcinoma risk: a review of epidemiological association studies, 2000-2011: Rosetta Stone for NPC: genetics, viral infection, and other environmental factors. *Seminars in cancer biology* **22**, 107-116, doi:10.1016/j.semcancer.2012.01.007 (2012).
- 147 Hershey, J. W., Sonenberg, N. & Mathews, M. B. Principles of translational control: an overview. *Cold Spring Harbor perspectives in biology* **4**, doi:10.1101/cshperspect.a011528 (2012).
- 148 Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731-745, doi:10.1016/j.cell.2009.01.042 (2009).
- 149 Kozak, M. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *Journal of molecular biology* **196**, 947-950 (1987).
- 150 Kozak, M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *The Journal of biological chemistry* **266**, 19867-19870 (1991).

- 151 Kozak, M. Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *The EMBO journal* **16**, 2482-2492, doi:10.1093/emboj/16.9.2482 (1997).
- 152 Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic acids research* **15**, 8125-8148 (1987).
- 153 Kozak, M. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* **299**, 1-34 (2002).
- 154 Gerashchenko, M. V., Lobanov, A. V. & Gladyshev, V. N. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 17394-17399, doi:10.1073/pnas.1120799109 (2012).
- 155 Leprivier, G. *et al.* The eEF2 kinase confers resistance to nutrient deprivation by blocking translation elongation. *Cell* **153**, 1064-1079, doi:10.1016/j.cell.2013.04.055 (2013).
- 156 Shalgi, R. *et al.* Widespread regulation of translation by elongation pausing in heat shock. *Molecular cell* **49**, 439-452, doi:10.1016/j.molcel.2012.11.028 (2013).
- 157 Walsh, D. & Mohr, I. Viral subversion of the host protein synthesis machinery. *Nature reviews. Microbiology* **9**, 860-875, doi:10.1038/nrmicro2655 (2011).
- 158 Jang, S. K. *et al.* A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *Journal of virology* **62**, 2636-2643 (1988).
- 159 Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320-325, doi:10.1038/334320a0 (1988).
- 160 Tsukiyama-Kohara, K., Iizuka, N., Kohara, M. & Nomoto, A. Internal ribosome entry site within hepatitis C virus RNA. *Journal of virology* **66**, 1476-1483 (1992).
- 161 Isaksson, A., Berggren, M. & Ricksten, A. Epstein-Barr virus U leader exon contains an internal ribosome entry site. *Oncogene* **22**, 572-581, doi:10.1038/sj.onc.1206149 (2003).

- 162 Macejak, D. G. & Sarnow, P. Internal initiation of translation mediated by the 5' leader of a cellular mRNA. *Nature* **353**, 90-94, doi:10.1038/353090a0 (1991).
- 163 Stein, I. *et al.* Translation of vascular endothelial growth factor mRNA by internal ribosome entry: implications for translation under hypoxia. *Molecular and cellular biology* **18**, 3112-3119 (1998).
- 164 Nanbru, C. *et al.* Alternative translation of the proto-oncogene c-myc by an internal ribosome entry site. *The Journal of biological chemistry* **272**, 32061-32066 (1997).
- 165 Stoneley, M., Paulin, F. E., Le Quesne, J. P., Chappell, S. A. & Willis, A. E. C-Myc 5' untranslated region contains an internal ribosome entry segment. *Oncogene* **16**, 423-428, doi:10.1038/sj.onc.1201763 (1998).
- 166 Johannes, G. & Sarnow, P. Cap-independent polysomal association of natural mRNAs encoding c-myc, BiP, and eIF4G conferred by internal ribosome entry sites. *Rna* **4**, 1500-1513 (1998).
- 167 Spriggs, K. A., Stoneley, M., Bushell, M. & Willis, A. E. Re-programming of translation following cell stress allows IRES-mediated translation to predominate. *Biology of the cell / under the auspices of the European Cell Biology Organization* **100**, 27-38, doi:10.1042/BC20070098 (2008).
- 168 Kozak, M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 2850-2854 (1986).
- 169 Sehgal, A., Briggs, J., Rinehart-Kim, J., Basso, J. & Bos, T. J. The chicken c-Jun 5' untranslated region directs translation by internal initiation. *Oncogene* **19**, 2836-2845, doi:10.1038/sj.onc.1203601 (2000).
- 170 Chua, J. J. *et al.* Synthesis of two SAPAP3 isoforms from a single mRNA is mediated via alternative translational initiation. *Scientific reports* **2**, 484, doi:10.1038/srep00484 (2012).
- 171 Meijer, H. A. & Thomas, A. A. Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. *The Biochemical journal* **367**, 1-11, doi:10.1042/BJ20011706 (2002).
- 172 Derrigo, M., Cestelli, A., Savettieri, G. & Di Liegro, I. RNA-protein interactions in the control of stability and localization of messenger RNA (review). *International journal of molecular medicine* **5**, 111-123 (2000).

- 173 Moore, M. J. From birth to death: the complex lives of eukaryotic mRNAs. *Science* **309**, 1514-1518, doi:10.1126/science.1111443 (2005).
- 174 Wilkie, G. S., Dickson, K. S. & Gray, N. K. Regulation of mRNA translation by 5'- and 3'-UTR-binding factors. *Trends in biochemical sciences* **28**, 182-188, doi:10.1016/S0968-0004(03)00051-3 (2003).
- 175 Kozak, M. Bifunctional messenger RNAs in eukaryotes. *Cell* **47**, 481-483 (1986).
- 176 Barbosa, C., Peixeiro, I. & Romao, L. Gene expression regulation by upstream open reading frames and human disease. *PLoS genetics* **9**, e1003529, doi:10.1371/journal.pgen.1003529 (2013).
- 177 Wethmar, K., Smink, J. J. & Leutz, A. Upstream open reading frames: molecular switches in (patho)physiology. *BioEssays : news and reviews in molecular, cellular and developmental biology* **32**, 885-893, doi:10.1002/bies.201000037 (2010).
- 178 Arribere, J. A. & Gilbert, W. V. Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome research* **23**, 977-987, doi:10.1101/gr.150342.112 (2013).
- 179 Pelechano, V., Wei, W. & Steinmetz, L. M. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**, 127-131, doi:10.1038/nature12121 (2013).
- 180 Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7507-7512, doi:10.1073/pnas.0810916106 (2009).
- 181 Somers, J., Poyry, T. & Willis, A. E. A perspective on mammalian upstream open reading frame function. *The international journal of biochemistry & cell biology* **45**, 1690-1700, doi:10.1016/j.biocel.2013.04.020 (2013).
- 182 Wethmar, K. The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley interdisciplinary reviews. RNA* **5**, 765-778, doi:10.1002/wrna.1245 (2014).
- 183 Hanfrey, C. *et al.* A dual upstream open reading frame-based autoregulatory circuit controlling polyamine-responsive translation. *The Journal of biological chemistry* **280**, 39229-39237, doi:10.1074/jbc.M509340200 (2005).

- 184 Mize, G. J. & Morris, D. R. A mammalian sequence-dependent upstream open reading frame mediates polyamine-regulated translation in yeast. *Rna* **7**, 374-381 (2001).
- 185 Mize, G. J., Ruan, H., Low, J. J. & Morris, D. R. The inhibitory upstream open reading frame from mammalian S-adenosylmethionine decarboxylase mRNA has a strict sequence specificity in critical positions. *The Journal of biological chemistry* **273**, 32500-32505 (1998).
- 186 Amrani, N., Sachs, M. S. & Jacobson, A. Early nonsense: mRNA decay solves a translational problem. *Nature reviews. Molecular cell biology* **7**, 415-425, doi:10.1038/nrm1942 (2006).
- 187 Peccarelli, M. & Kebaara, B. W. Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. *Eukaryotic cell* **13**, 1126-1135, doi:10.1128/EC.00090-14 (2014).
- 188 Vattam, K. M. & Wek, R. C. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 11269-11274, doi:10.1073/pnas.0400541101 (2004).
- 189 Lu, P. D., Harding, H. P. & Ron, D. Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *The Journal of cell biology* **167**, 27-33, doi:10.1083/jcb.200408003 (2004).
- 190 Kozak, M. Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. *Molecular and cellular biology* **7**, 3438-3445 (1987).
- 191 Poyry, T. A., Kaminski, A. & Jackson, R. J. What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? *Genes & development* **18**, 62-75, doi:10.1101/gad.276504 (2004).
- 192 Kozak, M. Constraints on reinitiation of translation in mammals. *Nucleic acids research* **29**, 5226-5232 (2001).
- 193 Luukkonen, B. G., Tan, W. & Schwartz, S. Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. *Journal of virology* **69**, 4086-4094 (1995).

- 194 Child, S. J., Miller, M. K. & Geballe, A. P. Translational control by an upstream open reading frame in the HER-2/neu transcript. *The Journal of biological chemistry* **274**, 24335-24341 (1999).
- 195 Wang, L. & Wessler, S. R. Inefficient reinitiation is responsible for upstream open reading frame-mediated translational repression of the maize R gene. *The Plant cell* **10**, 1733-1746 (1998).
- 196 Grant, C. M. & Hinnebusch, A. G. Effect of sequence context at stop codons on efficiency of reinitiation in GCN4 translational control. *Molecular and cellular biology* **14**, 606-618 (1994).
- 197 Lincoln, A. J., Monczak, Y., Williams, S. C. & Johnson, P. F. Inhibition of CCAAT/enhancer-binding protein alpha and beta translation by upstream open reading frames. *The Journal of biological chemistry* **273**, 9552-9560 (1998).
- 198 Grant, C. M., Miller, P. F. & Hinnebusch, A. G. Sequences 5' of the first upstream open reading frame in GCN4 mRNA are required for efficient translational reinitiation. *Nucleic acids research* **23**, 3980-3988 (1995).
- 199 Arias, C. *et al.* KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS pathogens* **10**, e1003847, doi:10.1371/journal.ppat.1003847 (2014).
- 200 Stern-Ginossar, N. *et al.* Decoding human cytomegalovirus. *Science* **338**, 1088-1093, doi:10.1126/science.1227919 (2012).
- 201 Irigoyen, N. *et al.* High-Resolution Analysis of Coronavirus Gene Expression by RNA Sequencing and Ribosome Profiling. *PLoS pathogens* **12**, e1005473, doi:10.1371/journal.ppat.1005473 (2016).
- 202 Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature protocols* **7**, 1534-1550, doi:10.1038/nprot.2012.086 (2012).
- 203 Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218-223, doi:10.1126/science.1168978 (2009).

- 204 Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789-802, doi:10.1016/j.cell.2011.10.002 (2011).
- 205 Rutkowski, A. J. *et al.* Widespread disruption of host transcription termination in HSV-1 infection. *Nature communications* **6**, 7126, doi:10.1038/ncomms8126 (2015).

8. List of Figures

Chapter 1: Introduction

| | |
|--|----|
| Figure 1.1. EBV particle structure. | 6 |
| Figure 1.2. The linear U _s /U _L organization of the EBV genome. | 7 |
| Figure 1.3. Organization of the viral genome according to BamHI restriction enzyme digestion. | 8 |
| Figure 1.4. EBV-encoded miRNAs. | 9 |
| Figure 1.5. EBV latency states during EBV infection. | 15 |
| Figure 1.6. Post-transcriptional gene expression control. | 19 |
| Figure 1.7. Cap-dependent translation initiation. | 20 |
| Figure 1.8. The Kozak consensus sequence. | 21 |
| Figure 1.9. <i>Cis</i> -acting regulatory elements in the 5'leaders of mammalian mRNAs. | 23 |
| Figure 1.10. Upstream open reading frames regulate translation. | 24 |
| Figure 1.11. Structural characteristics influencing the functionality of uORFs. | 25 |
| Figure 1.12: An overview over the ribosome profiling method. | 27 |

Chapter 3: Results

| | |
|--|----|
| Figure 3.1: Schematic overview over the library generation procedure. | 30 |
| Figure 3.2: Polysome profile of LCLs and lytic replication levels in the different strains. | 31 |
| Figure 3.3: Fragment length distribution of ribosome footprints. | 32 |
| Figure 3.4: Metagene analysis. | 33 |
| Figure 3.5: Read classification in the different ribosome profiling libraries. | 34 |
| Figure 3.6: Ingenuity Pathway Analysis (IPA) of canonical pathways found to be most significantly enriched in genes covered by ribosome protected fragments (RPFs). | 35 |

| | |
|---|----|
| Figure 3.7: Diseases and biological functions associated with translated transcripts in M81 and B95-8-infected LCLs. | 36 |
| Figure 3.8: M81-infected LCLs show evidence for global bidirectional translation of the viral genome. | 37 |
| Figure 3.9: Data evaluation in the two library types. | 38 |
| Figure 3.10: Ribosome coverage on latent transcripts. | 39 |
| Figure 3.11: Ribosome coverage on latent transcripts. | 40 |
| Figure 3.12: Lytic transcripts associate with ribosomes at varying degrees. | 41 |
| Figure 3.13: Comparison of ribosome coverage on lytic genes in B95-8 and M81-infected LCLs. | 43 |
| Figure 3.14: BMRF1 is expressed in B95-8-infected LCLs. | 44 |
| Figure 3.15: The EBER1 and EBER2 secondary structures reproduced from Iwakiri 2016 ¹¹ . | 46 |
| Figure 3.16: BHLF1 and LF3 transcripts show extensive association with ribosomes. | 47 |
| Figure 3.17: BHLF1 RNA preferentially sediments in the lighter polysomal fractions. | 48 |
| Figure 3.18: Ribosome profiling of BART transcripts. | 49 |
| Figure 3.19: Ribosome coverage of poorly characterized EBV transcripts. | 50 |
| Figure 3.20: BNLF2 mRNA transcripts are expressed in B95-8-transformed LCLs | 51 |
| Figure 3.21: Antisense-EBNA3C shows ribosome coverage in the lytically replicating strain M81. | 52 |
| Figure 3.22: Antisense-EBNA2 is not significantly associated with ribosomes in M81. | 53 |
| Figure 3.23: EBV encodes upstream open reading frames. | 54 |
| Figure 3.24: Conservation level of putative uORFs and their surrounding Kozak consensus sequence. | 56 |
| Figure 3.25: The EBNA transcription unit and the localization of the newly identified uORFs. | 57 |
| Figure 3.26: Kozak consensus sequence strength of the identified viral uORFs. | 58 |

| | |
|---|----|
| Figure 3.27: The Y2 exon encodes a 7 amino acid long uORF. | 59 |
| Figure 3.28: The U exon encodes an uORF within an IRES element. | 60 |
| Figure 3.29: Mutational analysis of viral uORFs in dual luciferase reporter assays. | 61 |
| Figure 3.30: The BKRF3 ORF is regulated by two uORFs and contains two internal alternative translation initiation sites. | 63 |

9. List of Tables

| | |
|---|----|
| Table 1: Identification of small ORFs in the EBV genome | 53 |
| Table 2: Identification of uORFs in the EBV genome | 55 |
| Table 3: Viral genes with alternative translation initiation sites | 62 |

10. Appendix

10.1. Appendix I: Summary of the sequencing reads obtained from each sample

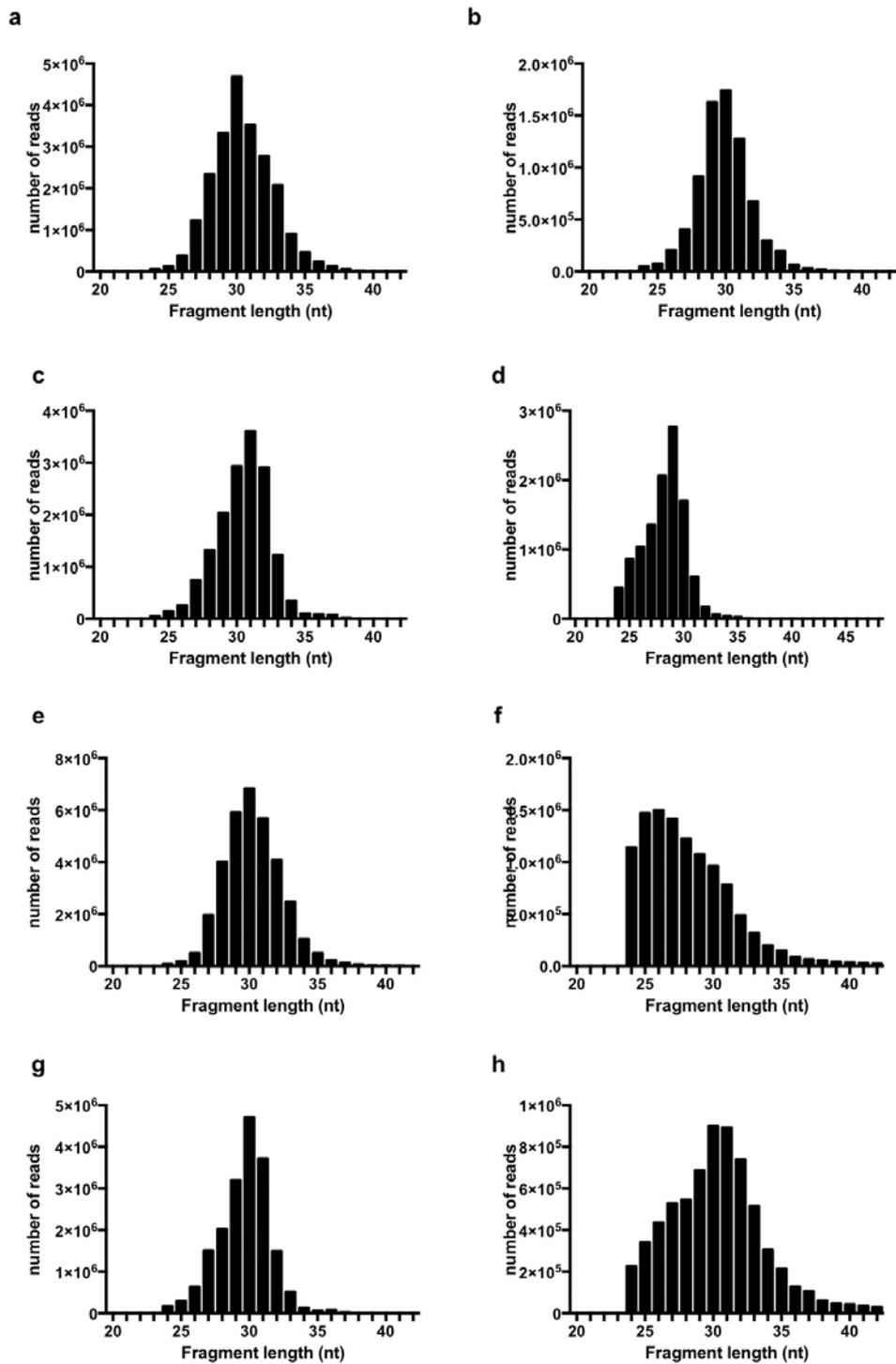
Appendix I: Summary of the sequencing reads obtained from each sample

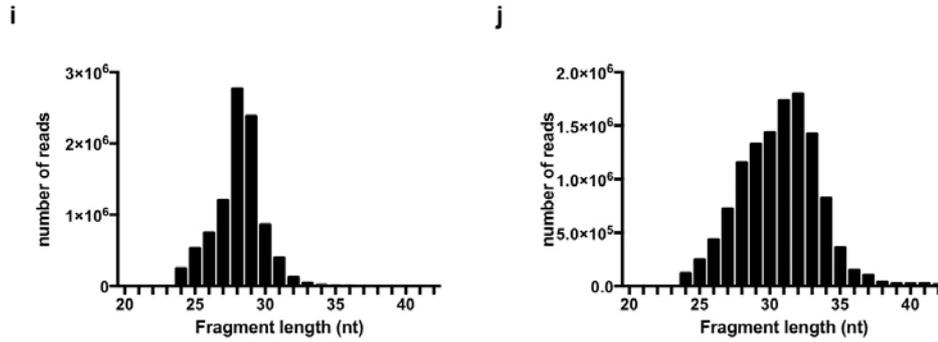
The first column identifies the sequenced sample, the second column shows the total number of reads sequenced in each sample, the third column lists all reads remaining following linker removal, the fourth column shows the number of reads remaining in each sample after rRNA sequence removal, the fifth column lists the number of reads that aligned to our human and viral genomes, the sixth column lists the reads that mapped to the viral genome and the last column indicates the number of these reads that were mapped to the EBER genes. Harr: harringtonine; CHX: cycloheximide.

| sample | sequenced reads | reads following clipping | reads following rRNA removal | mapped reads | total viral reads | EBER reads |
|--------------------------------|-----------------|--------------------------|------------------------------|--------------|-------------------|------------|
| B95-8 2 min Harr treated 01 | 45,666,734 | 39,400,972 | 15,383,128 | 8,235,491 | 15.050 | 7.930 |
| B95-8 5 min Harr treated 01 | 49,161,723 | 45,998,693 | 7,392,648 | 3,861,791 | 5.298 | 212 |
| B95-8 2 min Harr treated 02 | 17,476,623 | 11,038,224 | 4,170,313 | 2,439,254 | 21.287 | 19.593 |
| B95-8 CHX treated 01 | 49,874,563 | 44,682,658 | 29,039,482 | 15,514,135 | 475.700 | 458.764 |
| B95-8 CHX treated 02 | 16,216,980 | 9,944,246 | 8,131,636 | 6,013,662 | 107.380 | 99.059 |
| M81 2 min Harr treated 01 | 58,997,731 | 56,141,014 | 8,481,210 | 5,810,660 | 8.550 | 1.340 |
| M81 5 min Harr treated 01 | 35,032,225 | 33,753,242 | 6,445,370 | 3,406,665 | 3.992 | 244 |
| M81 2 min Harr treated 02 | 27,108,396 | 24,591,719 | 11,160,336 | 5,224,299 | 18.100 | 3.916 |
| M81 CHX treated 01 | 33,592,695 | 30,042,652 | 16,881,387 | 9,945,378 | 393.252 | 313.907 |
| 293-2089 2 min Harr treated 01 | 18,763,156 | 12,997,976 | 6,231,563 | 3,246,322 | 258.258 | 19.945 |

!

10.2. Appendix II: Fragment length distribution of sequenced ribosome profiling samples





Appendix II: Fragment length distribution of all sequenced ribosome profiling libraries. a. M81-infected LCLs, cycloheximide treated. b. M81-infected LCLs, 5 min harringtonine treated. c. M81-infected LCLs, 2 min harringtonine treated, first round. d. M81-infected LCLs, 2 min harringtonine treated. e. B95-8-infected LCLs, cycloheximide treated, first library. f. B95-8-infected LCLs, cycloheximide treated, second library. g. B95-8-infected LCLs 2 min harringtonine treated, first library. h. B95-8-infected LCLs, 2 min harringtonine treated, second library. i. B95-8-infected LCLs, 5 min harringtonine treated. j. HEK293-2089 producer cells lines induced into lytic replication and treated with harringtonine for 5 min.

10.3. Appendix III: A subset of upstream open reading frames identified in host cell genes.

Appendix III: upstream open reading frames identified in host cell genes.

This table summarizes some cellular uORFs identified in the ribosome profiling libraries.

| Reads | Gene | Peptide Sequence |
|-------|---------|--|
| 52 | ADAM10 | LKWSEREVLRRFSCQGRSRLPVEAPDQAPSASPSGSMCCC |
| 17 | ADIPOR1 | M |
| 5 | ADRB2 | MRLPGVRSRPAEPRRGSAR |
| 151 | AMD1 | MAGDIS |
| 837 | ATF4 | MAY |
| 688 | ATF4 | M |
| 29 | ATF4 | MALLTAFSSSVAVTDKDTFELSTFLDSSKAPQHDRNELPEQRGVGGGLDVPLRPVGFGG |
| 349 | ATF5 | LGSQSLTTVSPPLPA |
| 317 | ATF5 | MAL |
| 255 | ATF5 | MESSTFALVPVFAHLSILQSLVPAAGAASPVAISAQHLCYSHVTPGDPGAGAGQGAPAS |
| 105 | BIRC2 | LKNLPIPIILSPCSNKSHYGDLETL |
| 54 | BIRC2 | MEISKLYKGI |
| 23 | BIRC2 | MTDNYS |
| 11 | BIRC2 | MGRRASGAPGLIRAERAVSPCRRR |
| 372 | BIRC3 | MET |
| 164 | CDKN1B | VARRGLCLLAPRAVAGLPRGVRAA |
| 130 | CDKN1B | LGH |
| 628 | CHOP | MLKMSGWQRQSQNQSWNLRRECSRRKCFIHHT |
| 18 | CHRNA5 | MRPEPARSC |
| 59 | CYP1B1 | MGIDATHRPPVSISTL |
| 67 | DAP | LSSLAHARPAREPRRRLGPAEAPPRHVAFASRRETRD |
| 12 | DBP | MISRDPLELHVSAPRTSNPRRS |
| 25 | DGCR8 | MKRAMWPA |
| 14 | DGCR8 | MAAAVGR |
| 11 | DGCR8 | MKTDSLRSRQSLKLSAL |
| 75 | DICER1 | ME |
| 41 | DICER1 | MQFRQEQRDLKH |
| 558 | GADD34 | MQDAARPRARLSPRHLRQPEIL |
| 411 | GADD34 | MNALASLTVRTCDFWQTEPALLPPG |
| 564 | HDLBP | LAG |
| 23 | HTT | MDGRSGSAFTCGPEPHSLPRC |
| 54 | IFRD1 | MYRFRSQLFTGISAAATAHSYPRRFSTLLAEDSPLSRPPHRTSKKCSSIG |
| 123 | MDM2 | MEGEAAEPEGRPRPL |
| 20 | MDM2 | MIPEAQGVVLPAP |
| 69 | MKKS | MSLRNLWRDYKVLVVMVPLVGLIHLGWYRIKSSPVFQIPKNDIPEQDSLGLSNLQKSQIQGK |
| 21 | MKKS | MKNTSWIRKNWLLVAGISFIGVHLGTYFLQRSQSAKQSVKQSQSKQKSIEE |
| 61 | PAPOLA | MPRAAAAAVAGGK |
| 40 | PAPOLA | MLGRRGRRRSA |
| 5407 | TUBB | LRPAEKKNYLFSCPIHLRRAKKNFNHEGNRAHPGWSVWQPDRQCQLGGDQ |
| 24 | TUSC2 | MVVRTAV |