Dissertation zur Erlangung des Doktorgrades
der Fakultät für Biologie der Ludwig-Maximilians-Universität München

The role of the EBNA2-EBF1 complex in EBV driven B cell transformation

Sophie Beer
München, 2021

Erstgutachterin:	Prof. Dr. Bettina Kempkes
Zweitgutachter:	Prof. Dr. Wolfgang Enard
Tag der Abgabe:	01.03.2021
Tag der mündlichen Prüfung:	09.09.2021

Parts of this thesis have been published in:

"EBF1 binds to EBNA2 and promotes the assembly of EBNA2 chromatin complexes in B cells", Glaser, LV, Rieger, S, Thumann, S, **Beer, S**, Kuklik-Roos, C, Martin, DE, Maier, KC, Harth-Hertle, ML, Grüning, B, Backofen, R, Krebs, S, Kempkes, B; PLoS Pathogens (2017); 10.1371/journal.ppat.1006664

Content

Abstract

The Epstein Barr virus (EBV) infection of B cells is associated with various malignancies, e.g. Burkitt's lymphoma, Hodgkin's lymphoma, lymphoproliferative diseases. Upon infection, EBV induces a complex viral and cellular gene expression program causing the transformation of naïve resting B cells to lymphoblastoid cell lines (LCLs) in vitro. These are continuously proliferating cultures and used to study molecular mechanisms of the transformation process. EBV nuclear antigen (EBNA) 2 is one of the first genes expressed after EBV infection and a key regulator of the transformation process. It preferentially binds to B cell specific enhancers and promotors via cellular adaptor proteins. CBF1 is the main DNA anchor but the knowledge about the contribution of B cell specific transcription factors (TFs) to EBNA2's activity is limited. The early B cell factor (EBF) 1, a key TF during B cell development, has been identified as an EBNA2 co-factor. By identifying the α1-helix in the EBNA2 N-terminal dimerization (END) domain as a crucial region for the interaction with EBF1, I was able to generate mutant EBV lacking this region (EBVΔα1) in order to study the role of the EBNA2-EBF1 complexes during the transformation of infected B cells. Analysis of cellular processes and RNA expression revealed that EBVΔα1 was impaired in reprogramming cellular gene expression leading to a cell cycle arrest at the early S phase and impaired metabolism. Nevertheless, it was possible to establish long-term LCLΔα1 cultures on CD40 ligand expressing feeder cells. The analysis of primary EBVΔα1 infected B cells and LCLΔα1 revealed reduced expression of MYC and LMP1, which are important EBNA2 target genes and required for optimal cell proliferation and survival. Moreover, reverse genetics identified 3 classes of chromatin binding sites 1) EBF1 independent EBNA2 binding, 2) EBF1 dependent EBNA2 binding, 3) sites at which EBF1 and EBNA2 required the complex formation in order to bind. These data indicated that a functional EBNA2-EBF1 complex is required for optimal B cell transformation upon EBV infection. Subsequently, two CRISPR/Cas9 based genome editing approaches were explored to establish conditional EBF1 LCLs to further delineate the contribution of EBF1 to the EBNA2 induced transformation process of EBV infected B cells.

Abbreviations

7-AAD	7-aminoactinomycin D
aa	amino acid
AIDS	acquired immune deficiency syndrome
ALL	acute lymphoblastic leukemia
BAC	bacterial artificial chromosome
BCR	B cell receptor
bHLH	basic HLH
bp	base pair
BrdU	Bromodeoxyuridine
CBF	C promotor binding factor
CCN	cyclin
CD	cluster of differentiation
CDK	cyclin dependent kinase
CDKN	cyclin dependent kinase inhibitor
cDNA	complentary DNA
ChIP	chromatin immunoprecipitation
CLP	common lymphoid progenitor
Co-IP	co-immunoprecipitation
Cp	C promotor
cpm	counts per million
CRISPR	clustered regularly interspaced short palindromic repeats
CTD	C-terminal transactivation domain
D1/2/3	donor 1/2/3
DBD	DNA binding domain
DGE	differential gene expression
DIM	dimerization domain
DMSO	dimethyl sulfoxide
DNA	deoxyribonucelic acid
dpi	days post infection
dpt	day post transfection
DSB	double strand break
dsDNA	double stranded DNA
EBER	EBV encoded small RNAs
EBF	early B cell factor
EBNA	EBV nucelar antigen
EBV	Epstein-Barr virus
END	EBNA2 N-terminal dimerization
ER	endoplasmatic reticulum
ER-LBD	ligand binding domain of the estrogen receptor
ESE	EBV super-enhancer
FACS	fluorescent activated cell sorting
FC	fold change
FDR	false discovery rate
FSC-A	forward scatter-area
FSC-H	forward scatter-height
GC	germinal center
GFP	green fluorescent protein
gp	glycoprotein
gRNA	guide RNA
GSEA	gene set enrichment analysis
GST	gluthation S transferase
H_2O	water
HA	hemagglutinin
HDR	homology directed repair
HEK	human embryonic kidney
HLH	helix loop helix
HMGU	Helmholtz Zentrum München für Gesundheit und Umwelt
HSC	hematopoietic stem cell
Ig	immunoglobulin
IL	interleukin
IM	infectious mononucleosis
IPT	immunoglobulin, plexins, TF-like
IRF	interferon regulatory factor
Kan	kanamycin
kb	kilo base
LCL	lymphoblastoid cell lines
LMP	latent membrane protein
LMPP	lymphoid multipotent progenitor
LMU	Ludwigs Maximilians University
MA	mean-average
MOI	multiplicity of infection
MPP	multipotent progenitor
mRNA	messenger RNA
mt	mutant
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NCID	Notch intracellular domain
NES	normalized enrichment score
NGFR	nerve growht factor receptor
NHEJ	non-homologous end joining
NK	natural killer
NLS	nuclear localization signal
OMP	olfactory marker protein
PCA	principle component analysis
PCR	polymerase chain reaction
PI	propidium iodide
PTLD	post-transplant lymphoproliferative disease
qPCR	quantitative PCR
RNA	ribonucleic acid
RNP	ribonucleoprotein
RT	reverse transcription
SSC-A	side scatter-area
ssDNA	single stranded DNA
T7E1	T7 endonuclease 1
TAD	transactivation domain
TF	transcription factor
tRNA	transfer RNA
UMI	unique molecular identifier
UPR	unfolded protein response
WB	Western blot
wt	wild type
Zn	zinc

List of figures

List of tables

1 Introduction

EBV was first described as herpesvirus-like particles found in cell cultures derived from Burkitt's lymphoma from sub-Sahara African patients by the team of Epstein, Barr and Achong in 1964 (Epstein et al., 1964). Soon after, the structure of the virus could be identified using electron microscopy (Hummeler et al., 1966) and its ability to infect and transform B cells has been verified (Henle et al., 1967; Pope et al., 1968). Thus, EBV has been identified as the first human oncogenic virus. In 2018, an estimated 210,000 new cancer cases worldwide were attributable to EBV infections (de Martel et al., 2020) and EBV is associated with multiple malignancies, including Burkitt's lymphoma, Hodgkin's disease, post-transplant lymphoproliferative disease (PTLD), gastric carcinoma, NK/T cell lymphoma, nasopharyngeal carcinoma and smooth muscle carcinoma (Taylor et al., 2015). EBV establishes a life-long latent infection in the resting memory B cell compartment in the human host and an estimated 90 % of the world population carry the virus. If the primary infection is taking place during childhood, it is usually asymptomatic. However, an infection during adolescence can cause infectious mononucleosis (IM), an acute but severe disease with symptoms like fever, swelling of the lymph nodes and fatigue (Taylor et al., 2015).

Like other human herpes viruses, EBV's double stranded DNA is packed into an icosahedral nucleocapsid that is surrounded by a tegument layer and the lipid envelope. EBV belongs to the γ-herpes virus subfamily and exhibits B cell tropism (Liu and Zhou, 2007). The genome is approximately 170 kb in size and EBV encodes for 80-100 proteins. In vitro, the growth transforming potential of EBV causes continuous proliferation of infected B cells and the establishment of lymphoblastoid cell lines (LCL).

1.1 EBV's life cycle

EBV spreads via saliva, enters the body through the epithelial layer of the Waldeyer's ring in the pharynx, and infects resting B cells residing in the lymphoid tissue. The attachment of the virus to the cell surface is mediated by binding of the viral glycoprotein gp350 to the cellular surface molecule CD21, which is followed by endocytosis of the virus (Nemerow et al., 1987;

Tanner et al., 1987). Once the viral genome has entered the nucleus, it rapidly acquires nucleosomes and repressive chromatin marks to control viral gene expression. The genome exists now as an episome in the nucleus (Woellmer and Hammerschmidt, 2013). Since EBV preferentially infects naïve, resting B cells, it has to activate and drive the differentiation of infected B cells to eventually become memory B cells in vivo. This process is called transformation and requires the regulation of viral and cellular gene expression. During the early phase of infection, EBV induced transformation is initiated by the expression of the viral latency program III, which is characteristic for expression of the viral proteins EBNA1, 2, 3A/B/C, LMP1, 2/AB and the non-coding small RNAs EBER 1/2 (Thorley-Lawson, 2015). Once latency III is activated, massive proliferation of the infected B cells is taking place causing an immune response by the host and clearance of the majority of infected B cells by cytotoxic T cells (CTLs, Figure 1). Nevertheless, some infected B cells can escape the clearance by CTLs. The cellular compartment in which EBV establishes life-long latency is the memory B cell compartment. How EBV manages to enter this compartment is still in debate but there is strong evidence that infected B cells are undergoing the germinal center (GC) reaction before differentiating to memory B cells. During this process, the viral gene expression is gradually shut off resulting in the latency II program. This program is characterized by the expression of EBNA1, LMP1, 2A/B and EBERs (Thorley-Lawson, 2015). Viral LMP1 mimics the endogenous CD40 signaling (Lam and Sugden, 2003) and viral LMP2A mimics the BCR signaling (Dykstra et al., 2001; Fruehling and Longnecker, 1997; Fruehling et al., 1998). Both signaling pathways are important for the proliferation and survival of infected B cells. Once, B cells have gone through the GC reaction, they enter the periphery as memory B cells and express latency 0/I (Figure 1). The expression of EBNA1 and EBERs (latency I) can be detected in dividing memory B cells whereas resting memory B cells only express EBERs (latency 0).

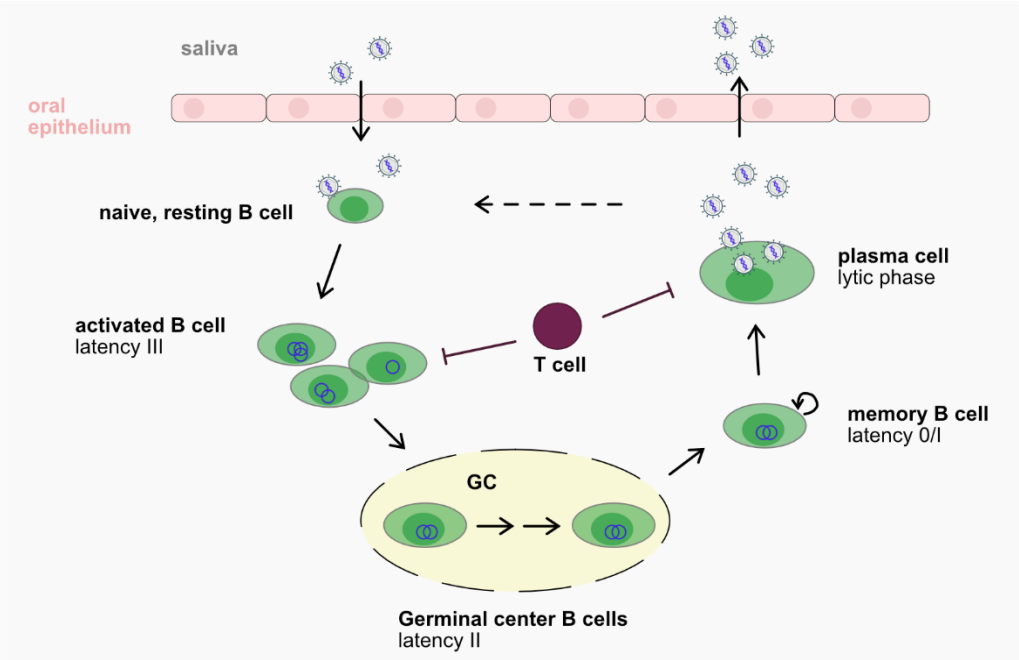


Figure 1 The life cycle of EBV. EBV infects naïve, resting B cells in the lymphoid tissue of the Waldeyer's ring. Due to the expression of the latency III program (EBNAs, LMPs, EBERs), EBV transforms the infected B cells to activated, hyperproliferating cells. Cytotoxic T cells (CTLs) recognize EBV antigens and eliminate the majority of infected, hyperproliferating B cells. Some of the B cells can escape the clearance by CTLs and migrate to the germinal center (GC), where they express the latency II program (EBNA1, LMPs, EBERs) and differentiate to memory B cells. Emigrating to the periphery, EBV infected memory B cells express the latency 0/I program (no latent genes, EBERs/EBNA1, EBERs) to escape immune surveillance by T cells and to establish life-long latent infection. The induction of the lytic phase can be triggered by differentiation of infected memory B cells to plasma cells. Newly produced EBV virions can infect naïve, resting B cells within the host or they are shed into the saliva and infect new hosts (adopted from Taylor et al., 2015; Thorley-Lawson, 2015).

The switch from the latent to the lytic phase is triggered by differentiation to plasma cells or external stimuli but the exact mechanism is not fully understood (Mckenzie and El-Guindy, 2015). However, the expression of viral BZLF1 and BRLF1 initiates the lytic phase of the EBV life cycle. BZLF1 is the master regulator of the lytic phase and a sequence-specific transcriptional activator (Grogan et al., 1987). The expression of BRLF1 is activated by BZLF1 (Bhende et al., 2005) and both activate the expression of early lytic genes. The lytic replication of the viral genome is independent of the cellular replication machinery. After all viral lytic genes are expressed the packaging of the viral DNA and assembly of virions can take place leading to shedding of the virus (Mckenzie and El-Guindy, 2015).

In vitro, EBV infection also activates B cells leading to continuous proliferation but, in contrast to an in vivo infection, the differentiation of infected B cells is blocked at this stage. This results

in the generation of continuously proliferating lymphoblastoid cell lines (LCL), which exhibit the latency III expression program. LCLs are used as a model system to study cellular and molecular processes during B cell transformation but are also a model for lymphoproliferative diseases like post-transplant lymphoproliferative disease (PTLD). PTLDs also display the latency III program, can arise in immunocompromised patients and the hyperproliferation of EBV positive cells is driven by EBNA2 (Taylor et al., 2015).

1.2 B cell transformation upon EBV infection

The concerted action of viral latent genes is crucial for EBV to induce B cell transformation and to establish and maintain latency in infected B cells. The regulation of viral latent gene expression is well controlled and provokes a complex network of cellular gene expression and the regulation of cellular processes, e.g. cell cycle, metabolism, signaling pathways.

1.2.1 The role of viral latent genes during B cell transformation

EBNA2 (further discussed in chapter 1.3) and EBNA-LP are the first viral latent genes expressed after EBV infection and are necessary for the cell cycle transition from G0 to G1 (Kempkes et al., 1995; Sinclair et al., 1994). EBNA2 plays a central role during the B cell transformation process as it activates the expression of all viral latent genes. EBNA-LP is a co-activator of EBNA2 but can also regulate gene expression independent of EBNA2 (Harada and Kieff, 1997; Kempkes and Ling, 2015; Szymula et al., 2018). Initially, the expression of EBNA2 is driven by the W promotor (Wp) of which several copies are present in the viral genome and most likely enhance the expression in the quiescent cell environment. The expression switches later to the C promotor (Cp), which is used in an activated cell environment (Woisetschlaeger et al., 1990). Another important step during the B cell transformation is the expression of MYC induced by EBNA2 (Kaiser et al., 1999). MYC further drives the cell cycle progression and induces cell growth and extensive hyperproliferation of infected B cells with mitosis occurring every 8 to 12 hours (Nikitin et al., 2010; Schuhmacher et al., 1999). Lastly, EBNA2 induces the expression of the EBNA3 family and the LMPs. EBNA3A and EBNA3C are viral oncogenes and necessary

for efficient B cell transformation (Hertle et al., 2009; Tomkinson et al., 1993). Several studies have demonstrated that EBNA3A and EBNA3C are involved in the modulation of the cell cycle and their action can promote cell cycle progression (Allday et al., 2015). Additionally, they can also epigenetically regulate gene expression (Harth-Hertle et al., 2013; Styles et al., 2017). EBNA3B is dispensable for B cell transformation but seems to be important for the control of tumor development (Allday et al., 2015; Chen et al., 2005). LMP1 is another viral oncogene (Kulwichit et al., 1998) and mimics the cellular CD40 signaling by associating with TNFR associated factors (TRAFs) and TNFR dead domains (TRADD). This initiates as singling cascade with the subsequent activation of NFkB, JNK/AP-1, Bcl2 proteins and cell survival and proliferation (Kieser and Sterz, 2015; Lam and Sugden, 2003; Smatti et al., 2018). Since LMP1 is constitutively active, infected B cells become independent of CD40 stimulated cell survival signals usually provided by T cells. LMP2A mimics tonic cellular BCR signaling by recruiting BCR associated proteins, like Lyn and Syk, which subsequently induce proliferation, survival and activation of B cells as well as migration and metastasis (Cen and Longnecker, 2015). The concerted action of LMP1 and LMP2A during B cell transformation provides infected B cells with important proliferation and survival signals that are independent of the cellular environment. The viral protein EBNA1 is important for the maintenance of latent infection because it tethers the viral episome to the host genome thereby ensuring proper replication and segregation of the viral genome by the host replication machinery during latent infection (Frappier, 2015).

1.2.2 Reprogramming of metabolic pathways during B cell transformation

Besides regulating and controlling the host cell gene expression, EBV needs to remodel metabolic pathways to efficiently transform infected B cells. The increased proliferation and cell growth upon EBV infection requires a high demand in energy, which is supplied by increased glycolysis. Cells infected with EBV display increased glycolytic activity as well as an increased uptake of glucose (McFadden et al., 2016; Mrozek-Gorska et al., 2019; Sommermann et al., 2011). The upregulation of gene expression related to glycolysis can be detected as early

as two days post infection (Wang et al., 2019a). Additionally, it has been shown that EBV and specifically EBNA2 induce the one-carbon (1C) metabolism in order to efficiently drive the transformation process (Wang et al., 2019a). 1C metabolism is a group of pathways that take place in mitochondria and the cytoplasm and are involved in nucleotide synthesis, amino acid homeostasis and redox regulation (Ducker and Rabinowitz, 2017). Another metabolic pathway that has been studied in the context of EBV infection is the mevalonate pathway (Wang et al., 2019b). The mevalonate pathway produces compounds used for the cholesterol synthesis on one hand but can also produce intermediate compounds, which are used for subsequent post-translational modification of proteins on the other hand. In the latter context, the study proposes that upregulation of the mevalonate pathway induces post-translational modification of proteins, which support trafficking of LMP1 and LMP2A. EBNA2 seems to play a key role also for the upregulation of the mevalonate pathway (Wang et al., 2019b).

The knowledge about how EBV remodels and reprograms the host metabolism is still scarce but findings of recent studies suggest that a broad range of metabolic pathways are affected. Understanding EBV driven reprogramming of B cells could help to develop new therapies against EBV induced lymphomas.

1.3 Epstein-Barr virus nuclear antigen (EBNA) 2

1.3.1 Functional regions of EBNA2

EBNA2 exerts its functions via protein-protein interactions and the activities can be mapped to specific regions (Figure 2A). The acidic C-terminal transactivation domain (C-TAD) can interact with components of the transcriptional machinery, e.g. TAF40, TFIIB, TFIIE via p100, RPA70 (Tong et al., 1995b, 1995a), thereby regulating transcription. Additionally, EBNA2 can interact with histone acetylases like p300/CBP and PCAF via the C-TAD to modulate chromatin (Wang et al., 2000) and might directly bind to EBNA-LP (Peng et al., 2004). The C-TAD (aa 448-471) is intrinsically unstructured but upon binding to Tfb1/p62, a subunit of TFIIH, it folds into a nine residue α-helix (Figure 2C) (Chabot et al. 2014). The second

transactivation domain is located at the N-terminus (N-TAD), which also serves as dimerization domain (DIM1). N-TAD, also called EBNA2 N-terminal dimerization (END) domain, folds into four β-sheets and one α1-helix. Dimerization of the END domain is mediated by two β-sheets from each monomer. The α1-helix and histidine 15 (H15) are required for the transcriptional activity of EBNA2 indicating that the END domain is providing functional activity to EBNA2 (Figure 2B) (Friberg et al. 2015). A second dimerization domain (DIM2) is separated by a poly proline (polyP) stretch from DIM1 and also mediates self-association of EBNA2 (Peng et al., 2004). EBNA2 amino acids 323 and 324, located within the adaptor region, interact with C promotor binding factor (CBF) 1, which provides EBNA2 chromatin access (Ling et al., 1993). The nuclear localization signal (NLS) flanks the C-TAD.

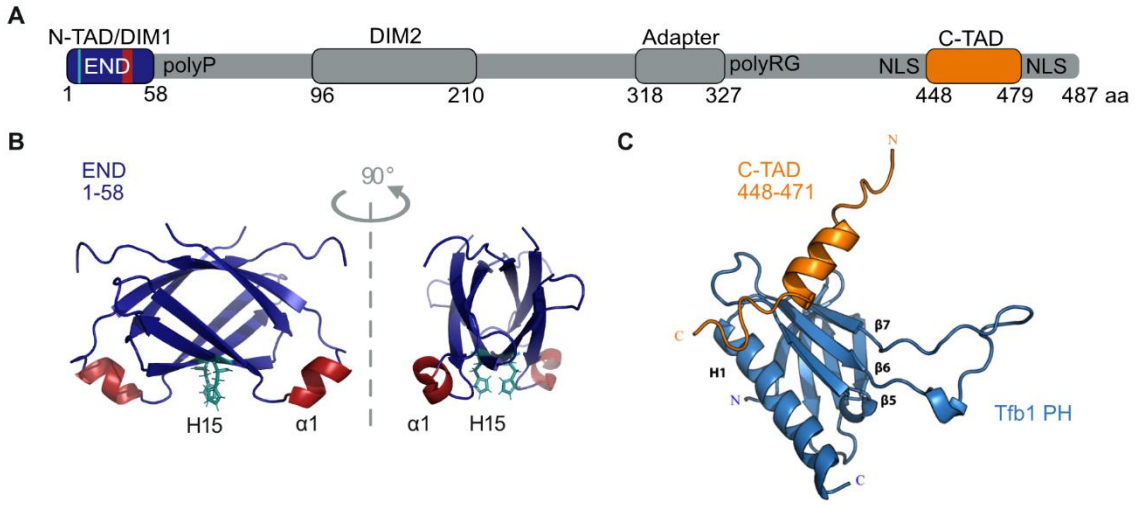


Figure 2 Functional regions of EBNA2 A) Schematic representation of functional domains of EBNA2. EBNA2 has an N- and C-terminal transactivation domain (N-/C-TAD), two dimerization domains (DIM1/2), and an adapter region, which binds to CBF1. The nuclear localization signal (NLS) flanks the C-TAD. The EBNA2 N-terminal dimerization (END) domain is formed by aa 1-58. (adapted from Friberg et al. 2015) B) Structure of the END domain as a homodimer. β-sheets and loops are colored in blue. The α1-helix is colored in red and H15 is colored in turquoise. Dimerization is mediated by β-sheets of each monomer. (adapted from Friberg et al. 2015) C) Structure of the EBNA2 C-TAD (orange) folds into a 9 residue α-helix upon binding to Tfb1 PH (blue). Without a binding partner, C-TAD is intrinsically unstructured. (from Chabot et al. 2014)

1.3.2 EBNA2 and CBF1 interaction

EBNA2 is a transcription factor (TF) that cannot bind directly to DNA but needs cellular anchor proteins. The main DNA anchor is CBF1 and was identified as a co-factor mediating EBNA2 binding to the viral Cp (Ling et al., 1993). CBF1, also known as recombination signal-binding protein for immunoglobulin kappa J region (RBPJ), is a ubiquitously expressed protein that is

the central molecule in Notch signaling. In the absence of Notch, CBF1 recruits repressor complexes, which prevent gene expression. In the presence of the Notch intra cellular domain (NICD), CBF1, NICD and mastermind-like (MAML) form a trimeric complex, attract additional coactivator and activate gene expression (Giaimo et al., 2021). By using CBF1 as an adaptor molecule, EBNA2 mimics the endogenous Notch signaling. Upon binding of EBNA2, the repressor complex is released and coactivator complexes are recruited, which initiates gene expression (Hsieh and Hayward, 1995). Although Notch and EBNA2 bind to the same region in CBF1, they partly bind to different amino acids within CBF1 (Fuchs et al., 2001) and binding of these factors seems to be mutually exclusive (Hsieh and Hayward, 1995). For the viral genome, several studies have demonstrated that EBNA2 regulated promotors (Cp, LMP1, LMP2A, LMP2B) contain at least one CBF1 binding motif and high affinity of CBF1 has been shown for the Cp and the LMP2A promotor (Kempkes and Ling, 2015). This highlights the importance of CBF1 for EBNA2 target gene regulation not just in the cellular but also in the viral genomic context. Furthermore, binding sites of other cellular factors flanking CBF1 motifs have been identified, which indicates that EBNA2 activity is regulated by additional factors (Kempkes and Ling, 2015). Within the host genome, EBNA2 and CBF1 primarily bind to intergenic regions and introns rather than promotor regions (Zhao et al., 2011).

1.3.3 Additional cellular co-factors for EBNA2 chromatin binding

Co-occupancy of CBF1 and EBNA2 binding sites supports CBF1 as the main DNA anchor. However, EBNA2 binding sites are also frequently co-occupied by other cellular TFs, e.g. PU.1, RUNX, IRF4, EBF1, BATF, ETS, NFκB (Zhao et al., 2011; Zhou et al., 2015). Although the role of these co-factors is not studied in detail, it could be demonstrated that PU.1 is important for the EBNA2 induced co-activation of the LMP1 promotor in reporter assays (Grossman et al., 1994; Johannsen et al., 1995; Laux et al., 1994; Sjoblom et al., 1995). However, the contribution of PU.1 to the activation of the LMP1 promotor in the viral genomic context is not well understood (Murata et al., 2016). Furthermore, the physical interaction of PU.1 and EBNA2 has been demonstrated only once (Yue et al., 2004).

Another potential B cell specific co-factor for EBNA2 is the early B cell factor (EBF) 1. EBNA2 binding sites are enriched for EBF1 and CBF1 motifs (Zhao et al., 2011) and EBNA2 and EBF1 can bind to the same chromatin sites (Lu et al., 2016). Furthermore, EBNA2 can induce new binding sites for EBF1 and CBF1, which is also linked to transcriptional activation (Lu et al., 2016). On the other hand, EBF1 is stabilizing CBF1 independent EBNA2 chromatin binding and CBF1 independent EBNA2 binding sites are enriched for EBF1 motifs (Glaser et al., 2017). Moreover, the protein complex formation of EBNA2 and EBF1 has been demonstrated and is most likely mediated by the END domain of EBNA2 (Glaser et al., 2017). Additionally, two EBF1 binding sites have been identified in the viral LMP1 promotor and overexpression of EBF1 can induce an increased LMP1 expression in LCL (Murata et al., 2016) further indicating a cooperative activity of EBNA2 and EBF1. Since EBNA2 activity is restricted to B cells and the transcription program initiated by EBNA2 is similar to the one of an activated B cells, EBF1 could be a co-factor that provides B cell specificity to EBNA2.

1.4 Early B cell factor (EBF) 1

Members of the transcription factor family Collier/Olf/EBF (COE) or EBF family were first identified as proteins binding to the olfactory marker protein-1 (OMP-1) promotor of olfactory neurons (Kudrycki et al., 1993) and to the B cell specific CD79a (mb-1) promotor in early B cells (Hagman et al., 1991). They were named Olf-1 and early B-cell factor (EBF) 1 accordingly. The family of these TFs is a well conserved group, which is found throughout the metazoan evolution. Homologues, for example, have been identified in *Drosophila melanogaster* called Collier (Crozatier et al., 1996) and in *Caenorhabditis elegans* (Wadsworth and Hedgecock, 1996). The mammalian EBF family consists of four members, EBF1, 2, 3 and 4, that are expressed in different tissues throughout the body and show a high sequence similarity. Within the hematopoietic system, EBF1 is only expressed in B lymphocytes and their progenitors but EBF1 expression can also be detected in other cell types like adipocytes (Griffin et al., 2013), skeletal muscle cells (Jin et al., 2014) and neurons (Kudrycki et al., 1993).

1.4.1 Structure of EBF1

The structure of EBF1 has been studied intensively and the EBF proteins consist of well-defined domains namely, the DNA binding domain (DBD), immunoglobulin, plexins, transcription factor-like/transcription factor immunoglobulin (IPT/TIG) domain, the helix loop helix (HLH) domain and the C-terminal transactivation domain (TAD, Figure 3A) (Siponen et al., 2010; Treiber et al., 2010a). EBF1 binds as a homodimer to the palindromic sequence 5'-TCCCNNGGGA-3', which is mediated by the DBD and the so called zinc knuckle (Travis et al., 1993; Treiber et al., 2010b). The zinc knuckle is a unique structure and Zn^{2+} stabilizes a loop that is formed by the amino acid motif H-X3-C-X2-C-X5-C (Figure 3A, purple colored, Zn^{2+} red sphere). Mutational studies of the histidine and cysteines revealed that DNA binding of the DBD is lost upon disruption of the zinc knuckle (Hagman et al., 1995). In addition to the zinc knuckle, two more interaction modules of EBF1 make contact in the minor and major groove of the DNA (Figure 3A, turquoise and red colored part of DBD). Downstream of the DBD is the IPT/TIG domain that displays an Ig-like fold (Figure 3A, colored green). The function of this domain is not fully understood but based on the structure it could be involved in protein-protein interactions (Treiber et al., 2010a). The nuclear localization signal (NLS) RRARR is located between DBD and IPT (Wang and Reed, 1993). The helix-loop-helix (HLH) domain identified by Treiber et al (Figure 3A, colored blue) consists of two helices. Siponen et al identified a third helix in EBF3 resulting in an atypical HLHLH motif (Figure 3B, Siponen *et al.*, 2010). Due to high sequence similarity between the EBF proteins, strong structural variations are unlikely and the third helix might be also present in EBF1. In contrast to typical bHLH domains of TFs, the HLH domain in EBF1 is not involved in DNA binding but in dimerization of the protein. Based on structural analyses, the third helix identified by Siponen et al. is not involved in dimerization but might be an interaction site for other proteins (Siponen et al., 2010). The TAD is intrinsically unstructured. Besides a transactivational function, it was also shown that the TAD supports binding of EBF1 to closed chromatin and regions with low

TF coverage (Boller et al., 2016a). An additional transactivation domain is located within the DBD indicated by transcriptional activity of truncated EBF1 (aa 18-251) (Hagman et al., 1995).

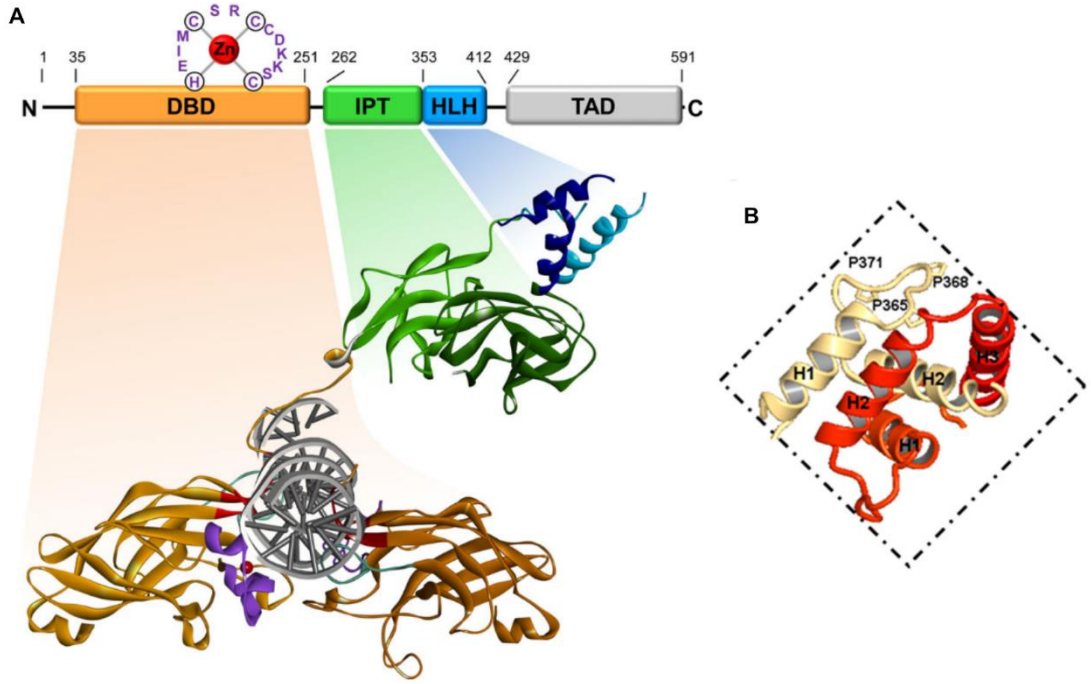


Figure 3 Structure of EBF1 homodimers A) EBF1 homodimer bound to DNA with the DNA binding domain (DBD) in orange, immunoglobulin/plexins/transcription factor-like domain (IPT) in green, helix-loop-helix domain (HLH) in blue and the C-terminal transactivation domain (TAD) in grey. The zinc-knuckle is located within the DBD and the areas of the DBD contacting the DNA are colored in red and turquoise (from Boller and Grosschedl, 2014); B) Structure of EBF3 helix-loop-helix-loop-helix (HLHLH) domain. One monomer is colored in beige and the second monomer is colored in red. The homodimer is formed by the helix 1 (H1) and helix 2 (H2) of each monomer. Helix 3 (H3) is not involved in dimerization. (from Siponen *et al.*, 2010)

1.4.2 EBF1 function during B-cell development

B lymphopoiesis is a process that involves a complex regulatory network of TFs, signaling pathways and epigenetic regulations. The following chapter will specifically focus on the role of EBF1 during B lymphopoiesis. Part of the development of highly specified cells is lineage specification and commitment during which pluripotent hematopoietic stem cells (HSC) lose their potential and develop into alternative lineages. EBF1 represents an important TF for B cell lineage specification, whereas Pax5 is an important B cell commitment factor. During B cell development, HSC differentiate into multipotent progenitors (MPP) that give rise to lymphoid multipotent progenitors (LMPP), which further differentiate into common lymphoid progenitors (CLP). Once cells have entered the B cell lineage, the rearrangement of

the Ig locus as well as the presence of the B cell receptor (BCR) and other surface markers define the developmental stages of B cells (Figure 4).

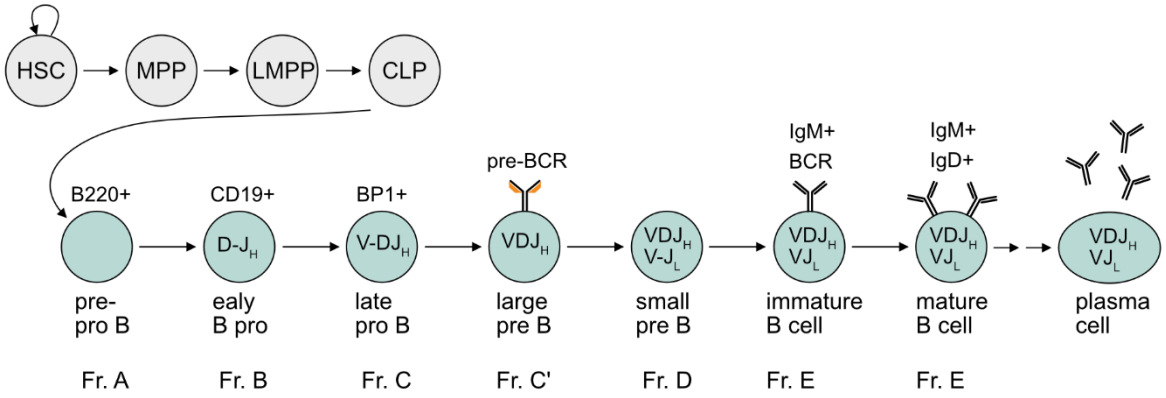


Figure 4 Schematic B cell development. Developmental stages of B lymphopoiesis are indicated below the cells (Hardy et al., 2007; Roessler and Grosschedl, 2006). Non-B lymphoid cells are colored in grey. B-lineage cells are colored in green. B cell developmental stages are defined by V(D)J rearrangement, (pre-)BCR and selected surface markers. B220 is expressed throughout B cell lineage. CD19 expression starts at the early pro B cell stage and is detected throughout. BP1 expression starts at the late pro B cell stage. Pre-BCR is expressed on large pre B cells and the functional BCR (IgM) is expressed on immature B cells. Mature B cells are IgD+ and IgM+. adapted from (Boller and Grosschedl, 2014; Lukin et al., 2008)

A prerequisite for the action of B cell specific TFs is the lineage-priming. This includes pre-marking of regulatory elements, like enhancers, with active or poised epigenetic marks in earlier progenitor stages (Mercer et al., 2011) as well as the transcriptional activation of B cell specific TFs. E2A, amongst other factors, is crucial for the B cell lineage-priming since, it can initiate EBF1 expression together with FoxO1. This leads to a positive feedback loop that enhances the expression of E2A, FoxO1 and EBF1 and as a consequence induces the expression of Pax5, a factor important for B cell commitment (Lin et al., 2010; Mansson et al., 2012; McManus et al., 2011; Roessler et al., 2007). Although the expression of EBF1 is initiated at the CLP stage (Dias et al., 2005; Roessler et al., 2007), functional activity is only detected at later stages. *EBF1* deficient mice display normal CLP development but a complete block of B cell development at the pre-pro B cell stage indicated by presence of germline μ and the lack of the Ig heavy chain rearrangement as well as the lack of B cell specific gene expression like CD79a, VpreB or λ5 (Lin and Grosschedl, 1995). Furthermore, the central role of EBF1 during B cell specification has been demonstrated in studies in which B cell development has been blocked at different progenitor stages, e.g. block at HSC, LMPP. The ectopic and enforced expression

of EBF1 could partially rescue the developmental block (Medina et al., 2004; Reynaud et al., 2008; Seet et al., 2004). The lineage-instructive potential of EBF1 is further supported by skewing progenitor cells of alternative lineages towards the B cell lineage upon enforced EBF1 expression (Pongubala et al., 2008; Zhang et al., 2003).

Important target genes of EBF1 are *CD79a/b* (Åkerblad et al., 1999; Feldhaus et al., 1992; Hagman et al., 1991), *VpreB* and *IglII* (λ5) (Sigvardsson et al., 1997). CD79a (Igα) and CD79b (Igβ) are involved in signal transduction of the pre-BCR and BCR (Hombach et al., 1990). *VpreB* and *IglII* encode for components of the surrogate light chain, VpreB and λ5, of the pre-BCR (Melchers et al., 1993). Importantly, EBF1 is regulating the expression of these genes in concert with other lineage specific TFs like E2A, PAX5 or Runx1 (Gisler and Sigvardsson, 2002; Maier et al., 2004; Sigvardsson et al., 2002). The cooperation of EBF1 with other TFs is also suggested by motif enrichment of hematopoietic TFs, e.g. NFkB, Stat1, Ets, within EBF1 bound regions (Treiber et al., 2010b). In addition, a global network including E2A, EBF1 and FoxO1 has been proposed to regulate B cell fate (Lin et al., 2010). Additional EBF1 regulated genes have been identified, which are involved in pre-BCR and Akt signaling, cell adhesion and migration (Treiber et al., 2010b). Besides activating the expression of B cell specific TFs, EBF1 can also repress gene expression, which is important for alternative lineage development (Banerjee et al., 2013; Pongubala et al., 2008; Thal et al., 2009).

Apart from regulating the gene expression necessary for the B cell lineage, another important function of EBF1 is the pioneering activity. EBF1's ability to modify the epigenetic signature of chromatin has been described for the CD79a promotor (Maier et al., 2004). Here, EBF1 increased the DNA accessibility and reduced methylation of CpG islands of the CD79a promotor suggesting that EBF1 is a lineage-specific pioneer factor (Hagman and Lukin, 2005). Later, a genome-wide approach has confirmed that EBF1 binding to chromatin can increase its accessibility and EBF1 binding leads to the acquisition of new epigenetic marks indicating that EBF1 can bind to closed chromatin. This activity has been localized to the C-terminal transactivation domain, which is also required for binding to chromatin with low TF

occupancy (Boller et al., 2016a). However, data suggest that EBF1 requires hematopoietic primed chromatin to act as a pioneer factor (Boller et al., 2018).

The role of EBF1 is well studied during the early development of B cells but little is known about the role during late development. Two studies have demonstrated that EBF1 seems to be required for the development of marginal zone B cells in spleen, B1 cells in the peritoneum and for the maintenance of germinal centers as the deletion of EBF1 in peripheral B cells resulted in the loss of these populations (Györy et al., 2012; Vilagos et al., 2012). Furthermore, deletion of EBF1 expression in peripheral B cells strongly impairs BCR signaling as detected by reduced phosphorylation of CD19, CD79a, Akt and reduced Ca^{2+} mobilization upon BCR stimulation (Györy et al., 2012). In summary, EBF1 plays a central role during early development of B cells as it activates important B cell specific genes and modifies chromatin. In mature B cells, EBF1 seems to be important for the maintenance of mature B cell populations and proper BCR signaling but the molecular mechanism is not known.

1.4.3 EBF1 expression in B cell derived malignancies

Genetic lesions or alterations of EBF1 expression have been detected in several B cell acute lymphoblastic leukemia (B-ALL) (Ramamoorthy et al., 2020; Somasundaram et al., 2015). A hallmark of B-ALL is the defect of differentiation of B lymphoid progenitor cells and dysregulation of EBF1 expression contributes to the B-ALL development. For example, mono- and bi-allelic deletions of EBF1 and other B cell specific TFs can contribute to the B-ALL phenotype in pediatric patients (Mullighan et al., 2007). Additionally, the repression of EBF1 expression by its antagonist ZNF423 has been associated with a contribution to B-ALL (Harder et al., 2013) and was also suggested for the EBF1 antagonist ZNF521 (Harder et al., 2013). The fusion of EBF1 to the receptor tyrosine kinase PDGFRB (EBF1- PDGFRB) in B-ALL samples rendered these cells independent of growth factors and increased proliferation of the cells (Roberts et al., 2012). Implications of EBF1 in B-cell malignancies derived from mature B cells have also been demonstrated. In diffuse large B cell lymphoma (DLBCL), EBF1 was fused in one case to the IGH locus, which was not associated with overexpression but might cause the

stimulation of anti-apoptotic genes (Bouamar et al., 2013). A hallmark of Hodgkin lymphoma is the loss of B cell specific transcription program in the tumorous Hodgkin and Reed Sternberg (HRS) cells. EBF1 is downregulated in these cells and re-expression of EBF1 has been demonstrated to induce a B cell specific transcription, e.g. expression of CD19, CD79a/b (Bohle et al., 2013). Another study in EBV derived B-cell lymphomas suggested that EBNA2 together with EBF1 might downregulate miR-34a expression and thereby upregulate PD-L1 (Anastasiadou et al., 2019), an important receptor causing immune evasion in the tumor environment (Alsaab et al., 2017). These data indicate that EBF1 expression might have tumor suppressive functions especially in B cell progenitor derived malignancies. In malignancies derived from mature B cell, EBF1 might have protective as well as tumor promoting functions.

1.5 Objective

EBNA2 is a key regulator of the transformation process induced after EBV infection in B cells as it initiates the expression of all viral latent genes and subsequently a complex network of cellular gene expression. Since EBNA2 cannot bind directly to DNA, it uses CBF1 as a DNA anchor. CBF1 is a ubiquitously expressed protein and central to Notch signaling. In contrast to CBF1, the activity of EBNA2 is restricted to B cells, which raises the question of additional B cell specific TFs supporting EBNA2 activity. The role of specific B cell TFs as co-factors of EBNA2 has not been studied extensively. On a global scale, studies have identified the co-occupation of EBNA2 chromatin binding sites with cellular TF as well as co-localization of binding motifs of EBNA2 and cellular TFs like EBF1, PU.1, Pax5, Ets. The Kempkes laboratory has recently identified EBF1 as co-factor of EBNA2. Both proteins are able to form protein complexes and EBF1 facilitates EBNA2 chromatin binding independent of CBF1. EBF1 plays a central role during B cell development since it activates the expression of B cell specific TFs. In addition, EBF1 displays pioneering activity as it can bind to closed chromatin and change the epigenetic landscape of chromatin in lymphocytes.

The goal of this thesis was to explore the specific contribution of EBF1 to the latency III transcription program initiated by EBNA2 after EBV infection of B cells. Since a protein

complex formation of EBNA2 and EBF1 has been demonstrated, the region within EBNA2 mediating the interaction with EBF1 had to be identified. Based on the results from the structural analysis, a specific EBV mutant lacking the interaction region within EBNA2 had to be generated and tested for the capacity to initiate short term activation and proliferation of infected primary human B cells as well as the ability to transform infected B cells into long-term proliferating cells. RNA sequencing was performed to analyze gene expression and to identify cellular processes affected by a loss of EBNA2-EBF1 interaction in EBV mutant infected B cells. Additionally, the contribution of the EBNA2-EBF1 interaction to chromatin binding of the factors was tested in long-term infected B cell cultures. In the last part, a conditional EBF1 LCL system should be established to further investigate the specific contribtuion of EBF1 to EBNA2's activity during B cell transformation.

2 Results

2.1 Identification of regions in EBNA2 and EBF1 critical for protein complex formation

Knowing the minimal binding region of proteins, which interact in a complex, allows us to study the function of such a protein complex by reverse genetics. By specifically manipulating the interaction sites, we can examine the effect of a disrupted protein complex on a cell and we can infer the biological function. In general, it is easier to generate and test structure-based mutants if the structure of the interacting proteins is known. The structure of N- and C-terminal parts of EBNA2 and the structure of EBF have been resolved in earlier studies (Friberg et al., 2015; Siponen et al., 2010; Treiber et al., 2010a), which facilitated the investigation of specific regions of the proteins.

2.1.1 The EBNA2 α1-helix is critical for the interaction with EBF1

To identify the region in EBNA2 that is important for the protein complex formation with EBF1, I focused on the END domain that has been characterized earlier (Friberg et al., 2015). This domain consists of the first 58 amino acids of EBNA2, shows transactivation activity and is one of two dimerization sites in EBNA2 (Figure 5A). Mutational studies of specific residues have shown that histidine 15 (H15) and the α1-helix are dispensable for the dimerization of the domain but are required for the biological activity of the N-terminal transactivation domain (Friberg et al., 2015). Additionally, H15 and the α1-helix are exposed on the surface of the END domain dimer (Figure 5B) and, therefore, could be potential interaction sites for other proteins.

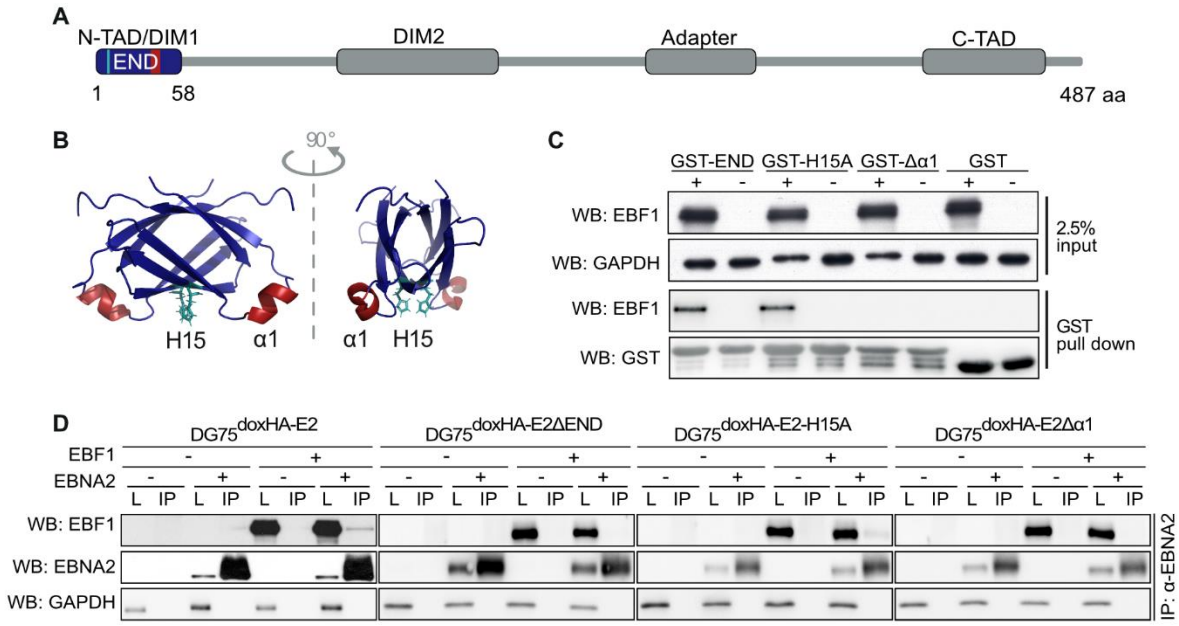


Figure 5 The EBNA2 α1-helix is critical for the interaction with EBF1 A) schematic representation of functional domains of EBNA2. EBNA2 has an N- and C-terminal transactivation domain (N-/C-TAD), two dimerization domains (DIM1/2), and an adapter region, which binds to CBF1. The EBNA2 N-terminal dimerization (END) domain is formed by aa 1-58. (adapted from Friberg et al. 2015) B) NMR structure of the END domain of EBNA2 with H15 (turquoise) and the α1-helix (aa 35-39, red). (adapted from Friberg et al. 2015) C) GST pull down with recombinant purified ENDwt (END), END-H15A (H15A) or ENDΔα1 (Δα1) coupled to GST. Whole cell lysate of DG75 cells transfected with an EBF1 expression plasmid (+) or empty vector control (-) was used to test complex formation of GST fusion proteins and EBF1. Western blot detection was performed with specific antibodies. GAPDH was used as a loading control. D) Co-IP with DG75doxHA-EBNA2 (E2), -EBNA2ΔEND (E2ΔEND), -EBNA2-H15A (E2-H15A) or –EBNA2Δα1 (E2Δα1) transfected with an EBF1 expression plasmid (+) or empty vector control (-). Expression of EBNA2 was induced by doxycycline. Co-IP was performed with EBNA2 specific antibodies. GAPDH was used as a loading control.

To test EBF1 binding, I performed GST-pull down assays with recombinant purified GST-ENDwt (GST-END), GST-END-H15A (GST-H15A), GST-ENDΔα1 (GST-Δα1) and GST only (Figure 5C). GST-END was sufficient to form protein complexes with EBF1 and the H15A mutation did not affect this complex formation. However, the deletion of the α1-helix led to a loss of interaction with EBF1. Next, the EBNA2 mutant lacking the END domain (E2ΔEND), EBNA2-H15A (E2-H15A) and EBNA2 lacking the α1-helix (E2Δα1) were co-expressed with EBF1 in DG75 cells to test the protein complex formation in a cellular background (Figure 5D). DG75 are human EBV negative Burkitt's lymphoma cells (Ben-bassats et al., 1977). Co-immunoprecipitations (Co-IP) with EBNA2 carrying the respective mutations confirmed the results obtained from the GST-pull down assays. The Co-IP with EBNA2Δα1 demonstrated that the α1-helix mediates the interaction with EBF1.

2.1.2 The C-TAD of EBF1 is dispensable for the interaction with EBNA2

Based on the protein structure, I generated different EBF1 domain mutants and tested them in Co-IP experiments to identify the minimal binding region in EBF1 (Figure 6A). The EBF1 mutant constructs were co-expressed with wild type EBNA2 in DG75 cells and the IP was performed with either an EBF1 specific antibody or a MYC-specific antibody since all EBF1 constructs carried a MYC epitope tag (Figure 6B). Wild type EBF1 (EBF1wt) and EBF1 lacking the transactivation domain (EBF1ΔTAD) could form protein complexes with EBNA2 demonstrating that the TAD is not involved in the interaction with EBNA2. Co-IP with all further EBF1 mutants did not show signals for EBNA2. This indicates that the minimal binding region in EBF1 might not localize to a specific domain as identified by the EBF1 x-ray structure but rather requires the cooperation of the DBD, IPT and HLH domain.

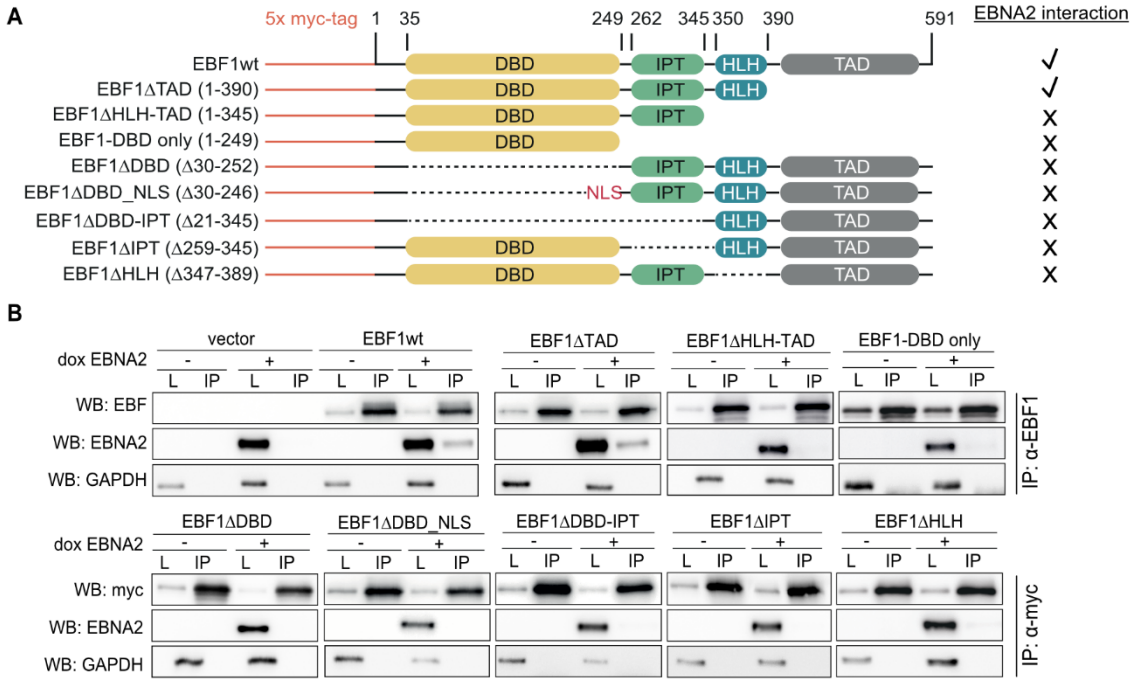


Figure 6 The EBF1 TAD is not involved in EBNA2/EBF1 complex formation. A) EBF1 domain deletion constructs used for Co-IP experiments with EBNA2. EBF1 constructs have a C-terminal 5x myc tag. B) DG75doxHA-EBNA2 cells were transfected with indicated EBF1 expression plasmids or empty vector control (-). EBNA2 expression was induced by doxycycline. Co-IP was performed with the indicated antibodies. GAPDH was used as a loading control. NLS- nuclear localization signal

2.2 In vitro EBV mutagenesis and generation of virus producing cells

Based on the identification of the EBNA2 α1-helix as the mediator of the interaction with EBF1, I was able to generate an EBV mutant that is unable to form EBNA2-EBF1 complexes in infected B cells. Generating EBV mutants and infectious viruses is a complex process that includes two major parts. During the first part, the EBV genome is manipulated in *E.coli* to induce the mutation of interest, while the second part includes the generation of virus producing HEK293 cells that carry the mutant EBV genome. Both parts will be explained in detail in the following two chapters.

2.2.1 EBV mutagenesis by recombineering in *E.coli*

EBV mutagenesis is based on a method called recombineering (recombination-mediated genetic engineering) and allows the mutagenesis in the heat sensitive *E.coli* strain SW105 (Wang et al., 2009). At 42 °C, SW105 express the proteins of the homologous recombination system of the λ bacteriophage. These proteins are exo, bet and gam and facilitate the insertion of specific DNA fragments into the mutation site (Warming et al., 2005). Mutating EBV in *E.coli* is possible because the EBV genome has been cloned into a plasmid carrying an F-factor, enhanced GFP (eGFP) and a puromycin resistance gene resulting in an EBV bacterial artificial chromosome (EBV-BAC) (Delecluse et al., 1998; Pich et al., 2019). In the first step of the recombineering procedure, a rpsL/Kan cassette is inserted into the site of interest by homologous recombination. Since the rpsL/Kan cassette disrupts the endogenous streptomycin resistance and provides a kanamycin resistance gene, clones with a successful recombination can be selected with kanamycin. In the second recombination step, a DNA fragment containing the desired mutation replaces the rpsL/Kan cassette, which restores the endogenous streptomycin resistance of SW105. Thus, clones with a successful recombination can be selected with streptomycin (Figure 7A).

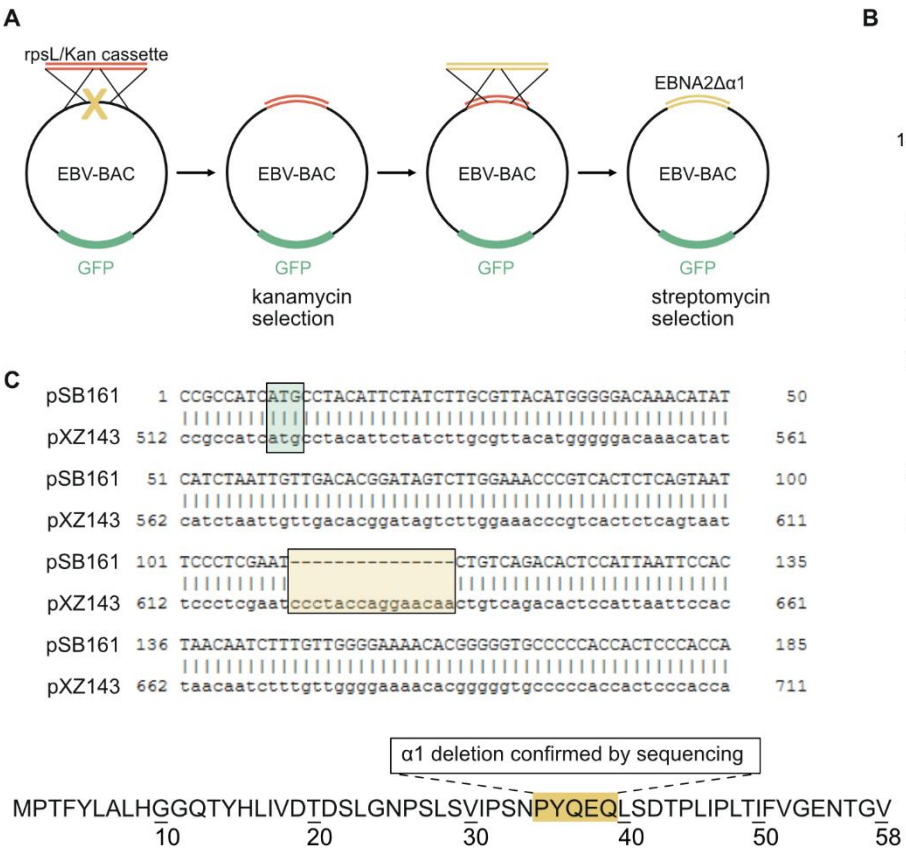


Figure 7 EBV mutagenesis by recombineering A) Schematic representation of the recombineering steps during EBV mutagenesis. The rpsL/Kan cassette is inserted into the mutation site and positive SW105 clones are selected by kanamycin. In a second recombineering step, a DNA fragment containing the mutation of interest replaces the rpsL/Kan cassette. Positive SW105 clones are selected by streptomycin and the BAC DNA is verified by restriction digest. B) KpnI restriction digest of EBV-HA-EBNA2 (pXZ143) and two clones of EBV-HA-EBNA2Δα1 (pSB161) after CsCl purification. C) Sequence alignment of sequenced EBV-HA-EBNA2 (pXZ143) and EBV-HA-EBNA2Δα1 (pSB161). The yellow box indicates the α1-helix nucleotides. The green box indicates the EBNA2 start codon. Amino acid sequence of the END domain of EBNA2 with the α1-helix highlighted in yellow.

The EBV genome used for mutagenesis in this thesis is based on p6008 genome and represents common field strains (Pich et al., 2019). The genomic sequence can be accessed under AJ507799 and NC_007605.1. My colleague, Xiang Zhang, added a C-terminal HA-tag to EBNA2 to be able to track EBNA2 in following experiments (pXZ143: EBV-HA-EBNA2). I used this newly generated EBV-BAC (pXZ143) to delete the α1-helix in EBNA2 (pSB161: EBV-HA-EBNA2Δα1). Once recombineering was completed and the BAC DNA was purified by a cesium chloride gradient, I performed a restriction digest to verify the DNA integrity (Figure 7B). No major deletions or aberrant alterations occurred during the recombineering procedure, which was indicated by the same patterns of DNA restriction fragments of the tested samples. To

verify the deletion of the α1-helix, the fragment carrying the Δα1 mutation was sequenced (Figure 7C). In the following chapters, virus generated from the BAC **EBV-HA-EBNA2 (pXZ143)** will be referred to as **EBVwt** and virus generated from the BAC **EBV-HA-EBNA2Δα1 (pSB161)** will be referred to as **EBVΔα1**.

2.2.2 Virus producing cells generate infectious cell culture supernatants

After having generated the EBV-BACs by recombineering, the next step was to establish HEK293 clones, which produced viral supernatants with a sufficient concentration of infectious particles. To this end, I transfected HEK293 cells with the purified EBV-BAC pXZ143 to produce EBVwt or pSB161 to produce EBVΔα1. As a predominantly latent virus, EBV does not produce virions spontaneously in HEK293 transfected cells. Therefore, single cell clones had to be raised and tested for their ability to produce infectious viral particles by inducing the lytic cycle. To induce the lytic cycle, expression plasmids of BZLF1 (p509) and BALF4 (p2670) were transfected into HEK293 clones. BZLF4 initiates the lytic cycle (Mckenzie and El-Guindy, 2015) and the gene product of BALF4, glycoprotein gp110, increases the infectivity of viral particles (Neuhierl et al., 2002). Since Raji cells can be easily infected with EBV, they were used as an indicator cell line to test the infectious supernatants produced by HEK293 cells. Therefore, Raji cells were incubated with HEK293 supernatants and the frequency of GFP positive cells was analyzed by FACS three days later (Figure 8A). GFP, encoded by the EBV-BAC, serves as a reporter gene that was used to monitor infected Raji cells. The viral titer of the supernatants obtained from HEK293 clones could be calculated based on the frequency of GFP positive Raji cells. EBVwt and EBVΔα1 viral supernatants were equally good in infecting Raji cells with EBVΔα1 infecting slightly more Raji cells than EBVwt (Figure 8A). In addition, the EBNA2 protein expression was similar for HEK293 producing EBVwt and HEK293 producing EBVΔα1 (Figure 8B). In summary, this demonstrated that infectious viral supernatants for EBVwt and EBVΔα1 could be generated, the supernatants were infectious and could be to study the function of the EBNA2-EBF1 protein complex in infected B cells.

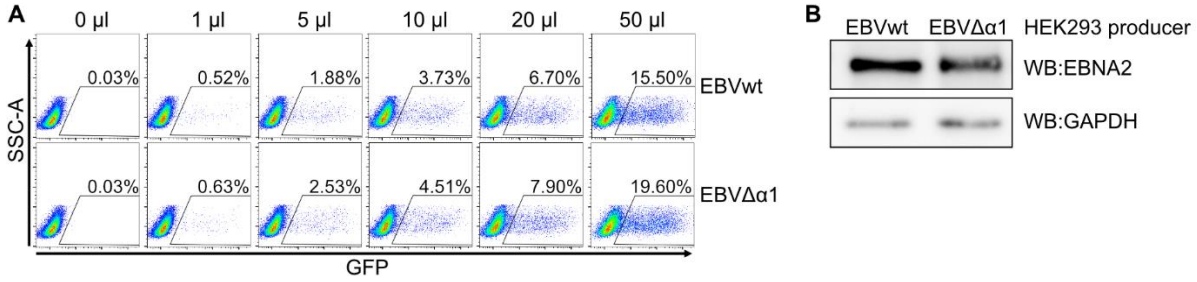


Figure 8 HEK293 clones can produce infectious EBVwt and EBVΔα1 particles. A) HEK293 supernatants containing EBVwt or EBVΔα1 virions were titrated on Raji cells and analyzed by FACS three days later. The cells were gated on lymphocytes and single cells. B) EBNA2 Western blot of whole cell lysates from EBVwt and EBVΔα1 HEK293 producer clones. GAPDH was used as a loading control.

2.3 Functional analysis of the EBNA2-EBF1 complex during in vitro transformation of EBV infected B cells

The newly produced viruses, EBVwt and EBVΔα1, were used to study the role of the EBNA2-EBF1 protein complex during the transformation process after B cell infection. For that purpose, B cells were isolated from human adenoid samples and infected with either EBVwt or EBVΔα1. I focused one part of the analysis on the early phase after infection, i.e. the first week after infection, which comprised the analysis of proliferation behavior, cell cycle and gene expression by RNA sequencing. The other part of the analysis focused on the latent phase of infection and included the establishment of lymphoblastoid cell lines (LCL) from EBVwt or EBVΔα1 infected B cells. I analyzed the protein and mRNA expression of specific target genes and studied the chromatin binding behavior of EBNA2 and EBF1 at specific binding sites in these LCLs.

2.3.1 EBVΔα1 infected B cells are arrested in the early S phase of the cell cycle

First, I examined the population expansion of primary B cells infected with EBVwt or EBVΔα1 during the first week after infection by an MTT assay (Figure 9A). The MTT assay is a colorimetric assay that measures the metabolic activity as an indicator of cell viability and proliferation. Viable cells metabolize MTT causing a change in color, which can be measured by a spectrometer. With an increasing number of viable and metabolically active cells in a culture, the purple color becomes more intense (Mosmann, 1983).

MTT conversion by EBVwt infected B cell cultures continuously increased over the course of 8 days post infection indicating an increase of total viable cells in the culture. In contrast, the MTT conversion by EBVΔα1 infected B cells remained at low levels during the first 8 days post infection indicating that this B cell population was viable but did not expand. Interestingly, the differences between EBVwt and EBVΔα1 infected B cells were first detected after day 2 post infection. Uninfected B cells served as a negative control and did not metabolize MTT (Figure 9A).

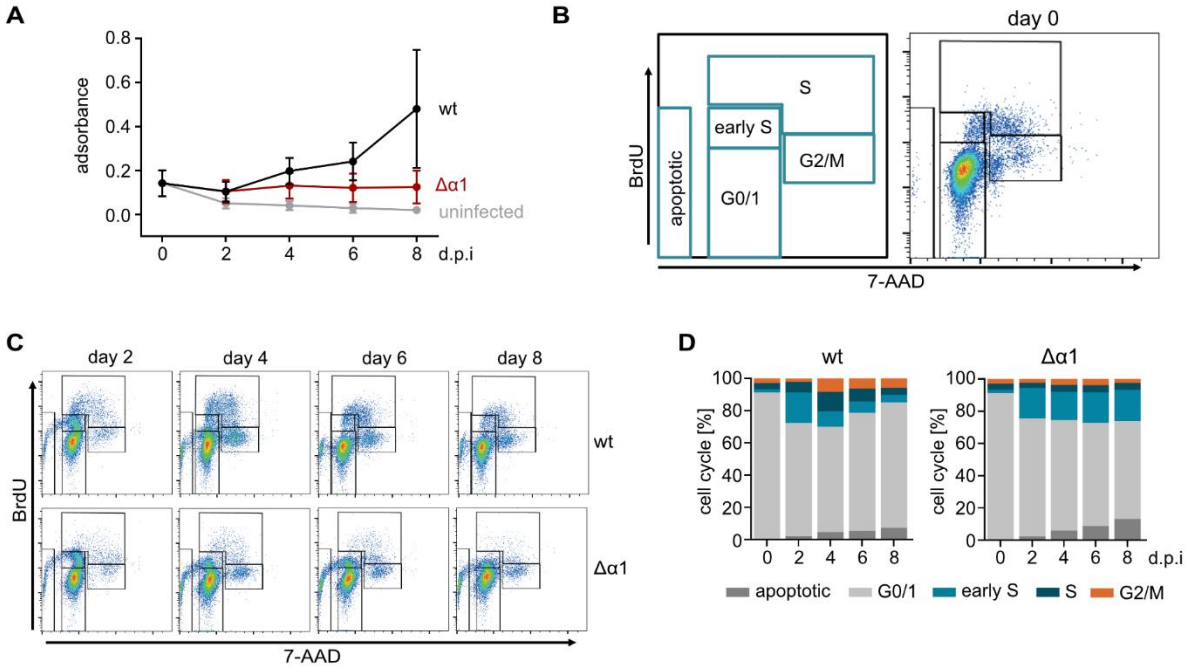


Figure 9 Primary B cells infected with EBVΔα1 are arrested in the early S-phase of the cell cycle. For the MTT assay and cell cycle analysis, primary B cells from 3 different adenoid samples were isolated and infected with either EBVwt or EBVΔα1. The assays were performed on day 0, representing uninfected B cells, and on day 2, 4, 6, 8 post infection. A) An MTT assay was performed with primary B cells infected with EBVwt, EBVΔα1 or uninfected B cells. The mean of 3 biological replicates is plotted. Error bars indicate SD. day 0 = uninfected B cells. B) Cell cycle analysis with Bromodeoxyuridine (BrdU) and 7-AAD, schematic gating of the cell cycle populations and one representative FACS plot of the day 0 uninfected sample, gated on lymphocytes and single cells. C) FACS plots of the BrdU assay of one representative experiment. B cells were infected with EBVwt (wt) or EBVΔα1 (Δα1) and the assay was performed on day 2, 4, 6 and 8 post infection. D) Summary of the cell cycle analysis of 3 biological replicates. The mean is plotted and cells were gated as indicated in B). day 0 = uninfected B cells

To understand the proliferation defect better, I analyzed the cell cycle of primary uninfected B cells (day 0) and EBVwt or EBVΔα1 infected B cells on day 2, 4, 6 and 8 post infection by bromodeoxyuridine (BrdU) incorporation. Cells actively replicating their genome, incorporate BrdU, a thymidine analog, into newly synthesized DNA. In combination with 7-AAD, staining of total DNA, the phases of the cell cycle can be distinguished. Cells in the G0/1 phase have low levels of BrdU incorporation as well as a low 7-AAD signal. Cells in the S-phase have increasing BrdU signals with increasing 7-AAD signals. My gating strategy allowed the discrimination between an early S and advanced S phase. Early S phase cells displayed lower BrdU and 7-AAD signals compared to cells in the advanced S phase. Cells in the G2/M phase display low BrdU signals and high 7-AAD signals. Apoptotic cells are defined by a low BrdU signal and a 7-AAD signal lower than G0/1 cells (Figure 9B).

The cell cycle analysis revealed that B cells infected with EBVwt entered the early S phase on day 2 post infection (Figure 9C, D, Table S 1). An increase of B cells in the G2/M phase was visible on day 4 post infection. From day 4 to day 8 post infection, the frequency of B cells in the early S, advanced S and G2/M decreased slightly and the population of apoptotic cells increased. This pattern of the cell cycle distribution was expected as is it was described before in another study (Mrozek-Gorska et al., 2019).

B cells infected with EBVΔα1 also entered the early S phase on day 2 post infection but, in contrast to EBVwt infected B cells, did not progress into the advanced S and G2/M phase (Figure 9C, D, Table S 2). There was no substantial increase of B cells in the G2/M or advanced S phase and the frequency of B cells in the early S phase stayed stable until day 8 post infection. The increase of apoptotic cells overtime was stronger compared to B cells infected with EBVwt. Taken together, the differences in the phenotypes between B cells infected with EBVwt or EBVΔα1 were already detectable in the first week after infection. The results demonstrate that B cells infected with EBVΔα1 could enter the cell cycle, however, these cells were arrested in the early S phase and did not progress with the cell cycle.

2.3.2 The cell cycle arrest of EBVΔα1 infected B cells is prevented by CD40L feeder cells

An early study has shown that primary human B cells can establish long-term cultures in vitro by stimulation of the CD40 receptor and in the presence of interleukine-4 (IL-4). CD40 stimulation alone induces proliferation but long-term survival of the cells is impaired in uninfected primary B cells (Banchereau et al., 1990). I wanted to know if it is possible to overcome the cell cycle arrest observed in EBVΔα1 infected B cells by co-culturing them with CD40 ligand (CD40L) expressing feeder cells. My colleague, Xiang Zhang, prepared the experiment as follows: primary B cells were infected with EBVwt, EBVΔα1 or stayed uninfected. One part of the cells was co-cultured with CD40L feeder cells and the other part was cultured without CD40L feeder cells. He then analyzed the cell cycle distribution by propidium iodide staining of nuclei and FACS analysis on day 0, 2, 4, 6, and 8 post infection. Apoptotic cells and cells of the G0/1, S and G2/M phases could be clearly identified in histograms (Figure 10A).

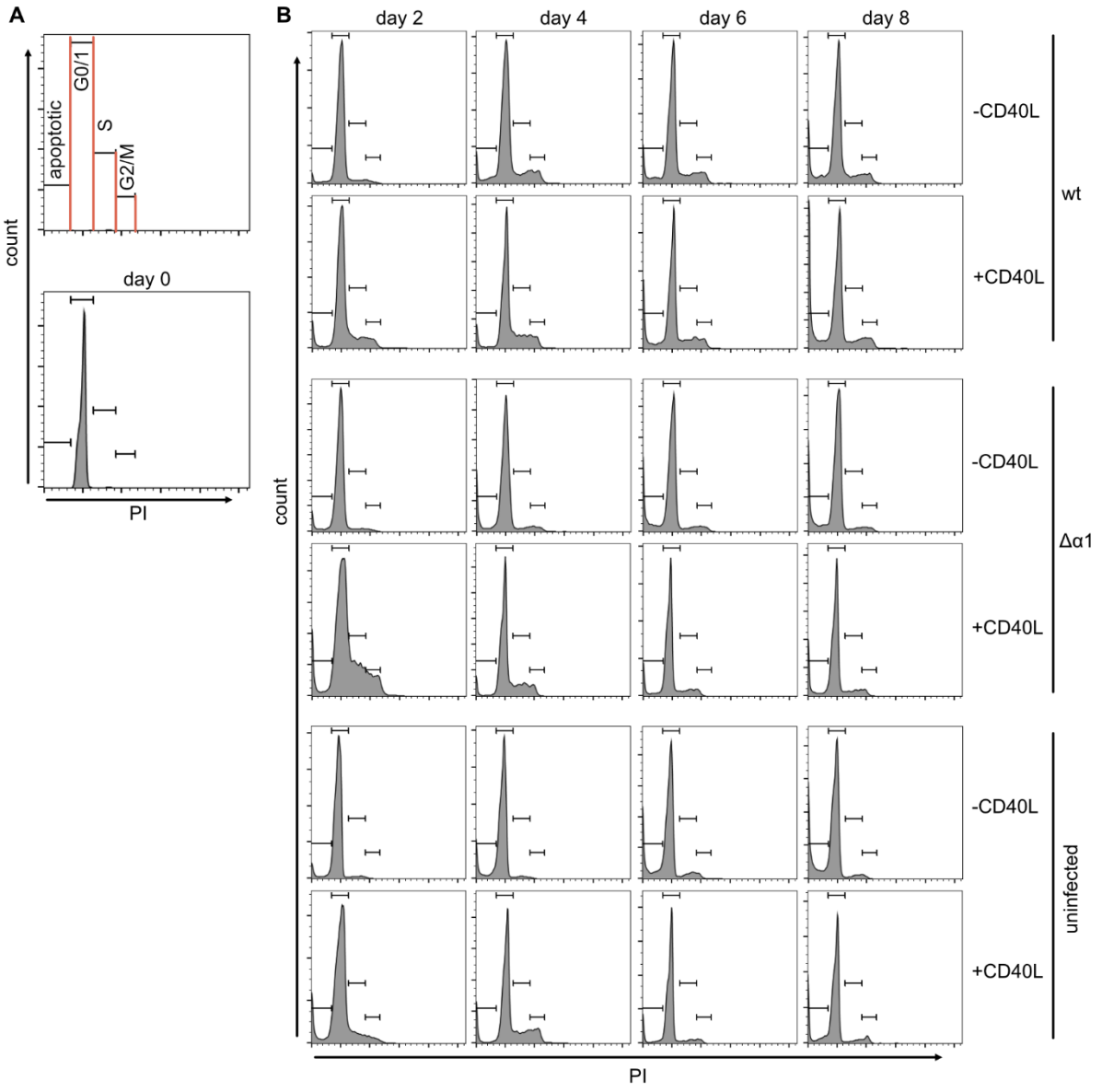

A
apoptotic
G0/1
S
G2/M
count
day 0
PI
B
day 2
day 4
day 6
day 8
-CD40L
+CD40L
wt
Δα1
uninfected
count
PI
C
uninfected
wt
Δα1
cell cycle [%]
- CD40L
+ CD40L
d.p.i.
apoptotic
G0/1
S
G2/M

Figure 10 The cell cycle arrest of EBVΔα1 infected B cells can be overcome by CD40L stimulation. A) Cell cycle analysis of propidium iodide (PI) stained nuclei, schematic gating of the cell cycle populations and one representative FACS plot for the day 0 uninfected sample. Cells were gated on lymphocytes and single cells. B) PI histograms of one representative experiment showing B cells infected with EBVwt (wt), EBVΔα1 (Δα1) or uninfected B cells on day 2, 4, 6 and 8 post infection. B cells were cultured without (-) CD40L expressing feeder cells or with (+) CD40L expressing feeder cells. C) Summary of the cell cycle analysis with PI of 3 biological replicates. The mean is plotted and cells were gated as indicated in A). day 0 = uninfected B cells

In contrast to EBV infected cells, only a small fraction of uninfected B cells cultured without CD40L feeder cells entered the S phase on day 2 with no substantial increase of the population until day 8. Co-cultivation of B cells with CD40L feeder had an overall positive effect on the cell cycle progression indicated by primary uninfected B cells entering the cell cycle on day 2 (Figure 10B, C, Table S 3). This result was expected since uninfected B cells can be cultured by CD40 stimulation (Banchereau et al., 1990). The fraction of B cells in the S and G2/M phase peaked at day 4 and decreased until day 8.

EBVwt infected B cells cultured without CD40L feeder cells display a similar cell cycle distribution as we have seen before (Figure 9D, Mrozek-Gorska et al. 2019). These cells entered the cell cycle on day 2 post infection with increasing populations of B cells in the S and G2/M phase and a peak on day 6 post infection. The supportive effect of CD40L feeder cell co-culture shifted this pattern to an earlier time point. Here, more EBVwt infected B cells entered the cell cycle on day 2 post infection. The fraction of B cells in the S and G2/M phase already peaked on day 4 post infection. Surprisingly, the population of apoptotic cells on day 8 post-infection was larger in the CD40 stimulated culture. Perhaps the high density of cells in this experimental setup caused a lack of nutrients in the culture (Figure 10C, D, Table S 4).

EBVΔα1 infected B cells without CD40L feeder cell co-culture displayed a similar cell cycle distribution as shown in Figure 9D. Only a small fraction of infected B cells entered the cell cycle. The population of EBVΔα1 infected B cells in the S and G2/M phase did not increase over time. Surprisingly, the positive effect of feeder cells was most pronounced for EBVΔα1 infected B cells. Already on day 2 post infection, a remarkably large population of S phase and G2/M cells was detectable. Additionally, those fractions were also higher compared to EBVwt infected B cells infected on day 2 post infection (Figure 10B, C, Table S 5).

Taken together, I have demonstrated that the proliferation defect observed in EBVΔα1 infected B cells could be overcome by co-culturing EBVΔα1 infected B cells with CD40L expressing feeder cells. Additionally, the positive effect of CD40L feeder cells on the cell cycle of B cells was strongest for EBVΔα1 infected B cells.

2.3.3 LCLΔα1 express less LMP1 and MYC

Since it was possible to establish long-term cultures of EBVΔα1 infected B cells on CD40L feeder cells, I was able to generate lymphoblastoid cell lines (LCL) from three different human adenoid samples (Figure 11B). For each adenoid sample, I established LCLs from B cells infected with EBVwt (LCLwt) and LCLs from B cells infected with EBVΔα1 (LCLΔα1). The deletion of the α1-helix in all three LCLΔα1 was confirmed by sequencing. All cell lines were maintained on CD40L feeder cells to ensure comparable culturing conditions between the wild type and mutant cell lines.

For studies on the protein and mRNA expression, all cell lines were cultivated in the absence of CD40L feeder cells for 10 day to ensure the elimination of exogenous survival and proliferation signals (Figure 11A). LCLΔα1 gradually stopped proliferating but remained viable over the time. For protein expression, whole cell lysate of GM12878, a lymphoblastoid cell line (Jiang et al., 2018), was included as a positive control for viral protein expression. Whole cell lysate of DG75 served as a negative control for viral protein expression (Figure 11B). The Western blot signals were quantified and normalized to GAPDH (Figure 11C, Figure S 1). The Western blot signals for LCLwt were set to 1 for each donor cell line pair. All LCLs, irrespective of the source of virus or the donor, expressed EBNA2, EBNA3A, EBNA3B and EBNA3C. The expression level of these proteins fluctuated between the cell lines but the expression of EBNA3A, B and C did not correlate with the co-expressed EBNA2 variant. Compared to LCLwt, viral LMP1 and cellular MYC protein expression was strongly reduced in all three LCLΔα1 whereas the EBF1 expression was increased (Figure 11B, C).

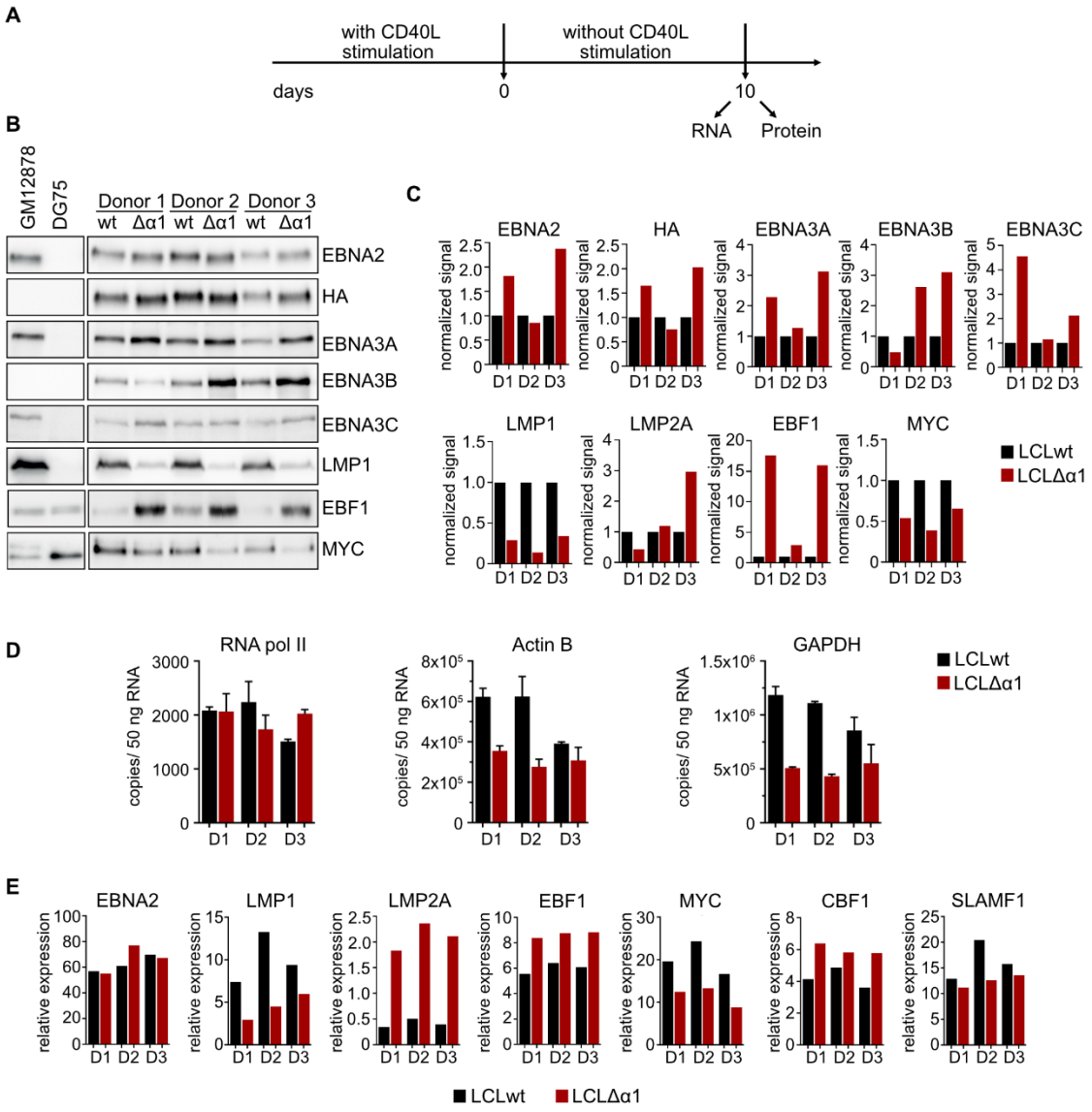


Figure 11 Reduced LMP1 and MYC expression in LCLΔα1 A) Experimental design for mRNA and protein harvest of LCLwt or LCLΔα1. The cell lines were cultured without CD40L feeder cells for ten days before mRNA and protein harvested. C) Protein expression of viral latent and cellular proteins in LCLwt or LCLΔα1 established from 3 donors. GM12878 served as a positive control for viral protein expression. DG75 served as a negative control for viral protein expression. E) Quantification of Western blot signals in B), normalized to GAPDH signals (see Figure S 1 for details) D) Absolute quantification of mRNA expression of RNA pol II, Actin B, GAPDH. Triplicates of each sample were analyzed. E) Relative expression of viral and cellular genes in LCLwt or LCLΔα1 analyzed by RT-qPCR. Expression levels were normalized to RNA pol II. Samples were analyzed in duplicates. The relative quantification was based on the ΔΔCt-method with correction for primer efficiency.

Before assessing the mRNA expression of specific genes, I had to establish reference genes for RT-qPCR data normalization. Genes of basal cell function are usually used as reference genes but their expression can differ depending on the applied experimental system. Ideally, reference genes for RT-qPCR analysis display stable mRNA expression under the tested

conditions. For my experiments, I analyzed the mRNA expression of RNA polymerase II (RNA pol II), Actin B and glyceraldehyde 3-phophase dehydrogenase (GAPDH). The mRNA expression of RNA pol II was very similar for all cell lines. In contrast, Actin B and GAPDH had significantly less mRNA expression in two LCLΔα1 (Figure 11D). Thus, I decided to use RNA pol II as a reference gene. The mRNA expression of EBNA2, LMP1, EBF1 and MYC reflected the results observed for the protein expression (Figure 11E). Viral LMP2A mRNA expression was strongly increased in all LCLΔα1. Cellular CBF1 mRNA expression was slightly increased in all LCLΔα1 whereas SLAMF1 mRNA expression was slightly reduced.

Since LMP1 expression is necessary for the maintenance of LCLs (Dirmeier et al., 2005), we were wondering if the proliferation defect observed in LCLΔα1 cultures without CD40L feeder cells might be caused by reduced LMP1 expression level. However, LMP1 overexpression in LCLΔα1 could not prevent the proliferation defect upon removal of CD40L feeder cells (data not shown).

2.3.4 The α1-helix assists in EBF1 dependent EBNA2 chromatin binding

It is well established that EBNA2 binds to chromatin by using CBF1 as an adaptor protein (Glaser et al., 2017; Hsieh and Hayward, 1995; Lu et al., 2016). However, the Kempkes laboratory has previously demonstrated that EBNA2 also requires EBF1 to bind to chromatin sites, which are independent of CBF1 (Glaser et al., 2017). Since I have generated a cellular system in which both interaction partners, EBNA2 and EBF1, are present, I was able to test chromatin binding behavior in the presence or absence of protein complex formation. In LCLwt, EBNA2 and EBF1 can form protein complexes, whereas in LCLΔα1 the complex formation is inhibited due to the lack of the α1-helix in EBNA2. Chromatin binding of EBNA2 and EBF1 was tested for specific promotor and enhancer regions by chromatin immunoprecipitation and subsequent qPCR (ChIP-qPCR). CD2 is a T cell specific gene and serves as a negative control (Figure 12C). Zhou et al recently identified two EBV super enhancers (ESE), which regulate MYC expression and bind many B cell specific TFs and EBV proteins (Zhou et al., 2015). Binding to ESE1 and ESE2 was tested because MYC is a direct

target of EBNA2 (Figure 12A)(Kaiser et al., 1999). HES1 is a canonical Notch target gene and also a target of EBNA2 in B cells (Figure 12B)(Iso et al., 2003; Maier et al., 2006). EBF1 has been demonstrated to bind to the CD79a promotor and to activate the expression (Figure 12D)(Hagman et al., 1991). EBNA2 binds to the viral C-promotor (Cp) during latency III and regulates the expression of EBNA proteins (Tempera and Lieberman, 2014) (Figure 12E). The LMP2A promotor contains two CBF1 binding sites forming the EBNA2 responsive element, through which EBNA2 can initiate LMP2A expression (Figure 12F) (Strobl et al., 1997). LMP1 expression is activated by EBNA2 and the promotor region contains an EBF1 binding site, which can enhance LMP1 expression (Figure 12G)(Fahraeus et al., 1990; Murata et al., 2016).

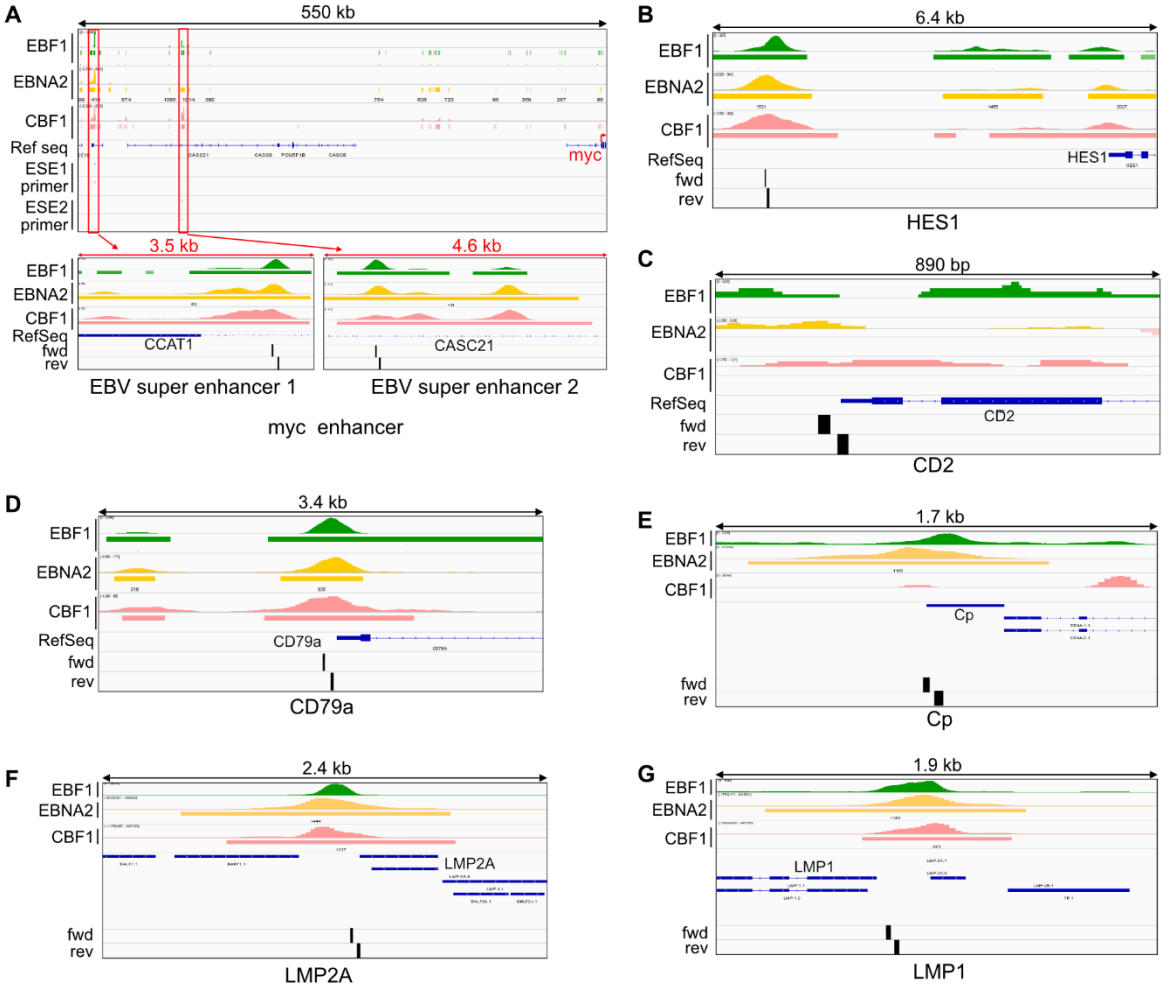


Figure 12 Genomic position of ChIP-qPCR primers for cellular and viral binding sites. EBF1 (green), EBNA2 (yellow), CBF1 (pink) ChIP signals and forward (fwd) and reverse (rev) ChIP-qPCR primer (black) position at cellular targets A) EBV super enhancer 1 (ESE1) and 2 (ESE2) of myc B) HES1 C) CD2 D) CD79a; and viral targets E) Cp F) LMP2A F) LMP1. In blue, are the RefSeq annotations for cellular genes and HHV-4 annotations for the viral genes. The EBF1 ChIP seq track was obtained from ENCODE. CBF1 and EBNA2 ChIP seq tracks were obtained from Zhao et al., 2011.

CD40L feeder cells were removed from both LCLwt and LCLΔα1 cultures one day before the chromatin immunoprecipitation (ChIP) was performed (Figure 13A). EBF1 did not bind to the HES1 and CD2 promotor in LCLwt and LCLΔα1 (Figure 13B). Binding of EBF1 to ESE1 and the CD79a promotor was not affected by the α1-helix deletion. However, EBF1 bound significantly less to ESE2 in LCLΔα1. The viral promotors of LMP2A, LMP1 and Cp bound EBF1 equally well in LCLwt and LCLΔα1 (Figure 13B). EBNA2 binding to the HES1 promotor was detectable in LCLwt and LCLΔα1 with slightly increased signals in LCLΔα1 (Figure 13C). As expected, EBNA2 binding to the CD2 promotor did not take place. EBNA2 bound significantly weaker to ESE1, ESE2, the CD79a and LMP1 promotor when the α1-helix was deleted but binding to the LMP2A and C-promotor was not affected by the EBNA2 mutation (Figure 13C)

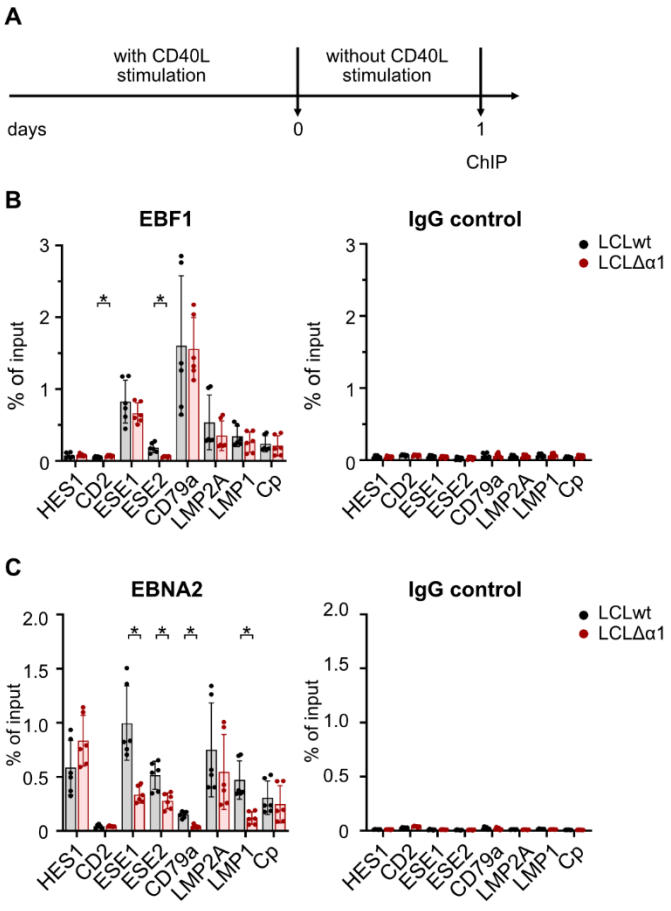


Figure 13 The α1-helix assists in EBF1 dependent EBNA2 chromatin binding. A) Experimental setup for ChIP-qPCR. LCLs were cultured without feeder cells for one day before the chromatin was harvested. B) ChIP-qPCR for EBF1 or IgG control or C) EBNA2 or IgG control in established LCLs at viral and cellular genomic regions. CD2 was used as a negative control. The mean of 3 biological replicates is plotted with error bars indicating SD. Statistical significance was calculated by Mann-Whitney or Welch's t test. Asterisks indicate $p < 0.05$.

These data indicate that there are different classes of EBNA2 chromatin binding sites. The first class (class 1) could be called 'EBF1 independent' binding sites, which do not require an interaction with EBF1. These sites would include HES1, LMP2A and Cp because EBNA2 and EBNA2Δα1 bound equally well irrespective of EBF1 chromatin binding or EBF1 interaction with EBNA2. The second class (class 2) of EBNA2 chromatin binding sites could be called 'EBF1 dependent'. These sites would include ESE1 and the CD79a and LMP1 promotor. At these sites, EBNA2 would require the interaction with EBF1 to bind to chromatin. Binding of EBF1 to the class 2 sites could be independent of EBNA2 indicated by similar EBF1 signals in the presence of EBNA2wt and EBNA2Δα1. A recent study has demonstrated that EBF1 binding to the LMP1 promotor and Cp is reduced in the absence of EBNA2 (Lu et al., 2016), which could not be confirmed in my experiments. The last class (class 3) of EBNA2 chromatin binding sites could be called 'EBNA2-EBF1 dependent sites' where a protein complex formation might be a prerequisite for both, EBNA2 and EBF1, to bind to these sites. An example of this group would be ESE2 (Figure 13B, C).

Altogether, the experiments demonstrate that the α1-helix of EBNA2 specifically contributes to EBNA2 chromatin binding at sites, which require EBF1 as a DNA anchor. EBF1 binding at most of the tested chromatin sites seemed to take place independently of EBNA2.

2.3.5 RNA sequencing during the first four days after infection

The results so far have demonstrated that EBVΔα1 is able to infect B cells, although, the transformation of the cells seemed to be impaired. EBNA2 is the first gene expressed after EBV infection and a key regulator of B cell transformation because it induces a cascade of viral and cellular gene expression. We investigated the gene expression early after infection to better understand the impact of the α1-helix deletion in EBNA2 during the transformation process. The results of these analyses will be presented in the following chapters.

2.3.5.1 Experimental design

To perform RNA sequencing, we collaborated with the group of Wolfgang Enard (LMU). They have developed a protocol to prepare cDNA libraries for single cell sequencing with increased sensitivity and efficiency (mcSCRBseq, Bagnoli et al. 2018). Applying this method to bulk RNA sequencing, as for our experiment, enabled us to use a relatively small number of cells, i.e. 10,000 cells/sample. This was of interest because the yield of primary B cells from adenoid samples was limited. The cDNA library was prepared from polyA RNA and sequencing of the cDNA libraries was performed at the Laboratory of Functional Genome Analysis (LAFUGA, LMU). All six adenoid samples were tested for EBV infection and only EBV negative adenoids were used for the experiment.

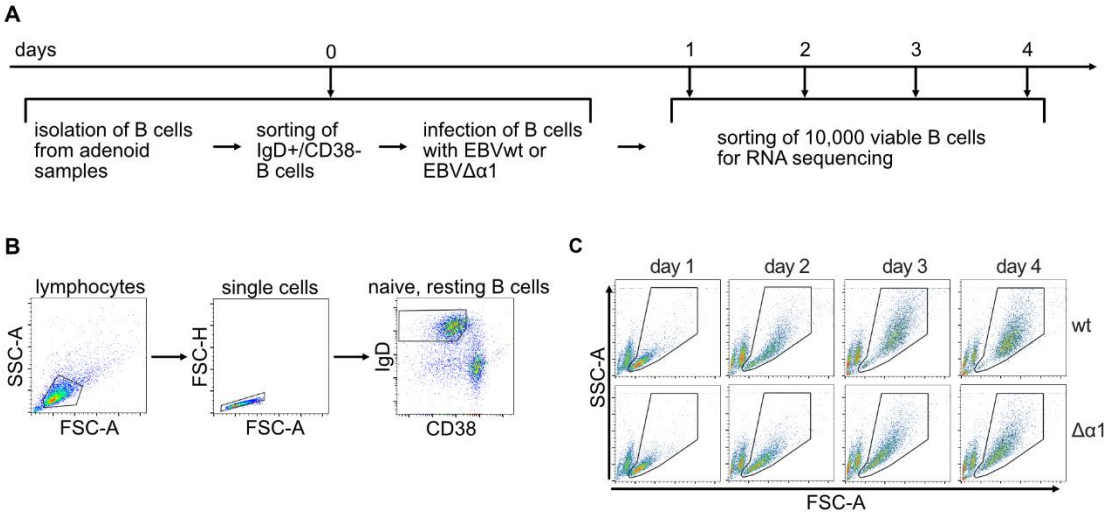


Figure 14 Sample preparation and collection for RNA sequencing A) Experimental design of sample preparation and collection for RNA sequencing. B cells were isolated from adenoid samples. Naïve, resting B cells (IgD+/CD38-) were sorted and infected with EBVwt or EBVΔα1. 10,000 cells of uninfected naïve, resting B cells were collected as the day 0 uninfected sample. On day 1, 2, 3, 4 after infection, 10,000 viable cells were sorted and collected for RNA sequencing. B) Gating strategy for the sort of naïve, resting B-cells. FACS plots of one representative experiment, n=6 C) Gating of EBVwt or EBVΔα1 infected, viable B-cells on day 1, 2, 3, 4 post infection. Cells within the gate were sorted and 10,000 cells were collected for RNA preparation. FACS plots of one representative experiment, n=6.

We decided to analyze the RNA expression during the first 4 days post infection because the difference between EBVwt and EBVΔα1 infected B cells appeared during this time, e.g. proliferation and cell cycle defect (Figure 9). On day 0, primary B cell were isolated from adenoid biopsies and stained for IgD and CD38 to sort naïve resting B cells (IgD+/CD38-; Figure 14A and B). 10,000 cells of sorted naïve resting B cells were saved as the uninfected sample and

stored at – 80°C. The rest of the cells was infected with either EBVwt or EBVΔα1. On day 1, 2, 3 and 4 post infection, 10,000 viable cells were sorted based on the side scatter (SSC) and forward scatter (FSC, Figure 14C). After the collection of all samples from 6 biological replicates, my collaboration partner, Lucas Wange (AG Enard, LMU), prepared the RNA and cDNA libraries and processed the raw sequencing data applying the zUMI pipeline (Parekh et al., 2018).

2.3.5.2 Quality control of sequenced libraries

After receiving the processed data from Lucas Wange, I performed a general quality control. The library size of the sequenced samples differed strongly (Figure 15A). The library size of EBVwt infected B cells on day 3 and 4 post infection was the largest whereas the library size of uninfected samples was the smallest. The increase in library size on day 3 and 4 might be due to more RNA used for the cDNA library preparation, although the RNA amount used for the cDNA library preparation was adjusted (Table 17). The number of detected cellular genes ranged from around 10,000 to around 25,000 genes. Samples with less than 10,000 genes were excluded from downstream analyses. This filter excluded one sample from the uninfected group (day 0) and one sample from EBVΔα1 infected B cells each on day 2 and 3 post infection (Figure 15B). Another quality control that was performed was the analysis the fraction of mitochondrial genes among the detected cellular genes. The expression of mitochondrial genes can differ between cell types but should be comparable between samples of the same type. High fractions of mitochondrial genes could indicate low sample quality, e.g. apoptotic or lytic cells within the sample can increase the number of detected mitochondrial genes (AlJanahi et al., 2018). The fraction of mitochondrial genes for my samples was between 4% and 8% indicating that my samples were of similar quality (Figure 15C). Regarding viral genes, only 4920 reads could be mapped to the viral genome of which 3095 were mapped to EBNA2. This, unfortunately, provided us with little information about viral gene expression and EBNA2 was the only gene with meaningful expression levels.

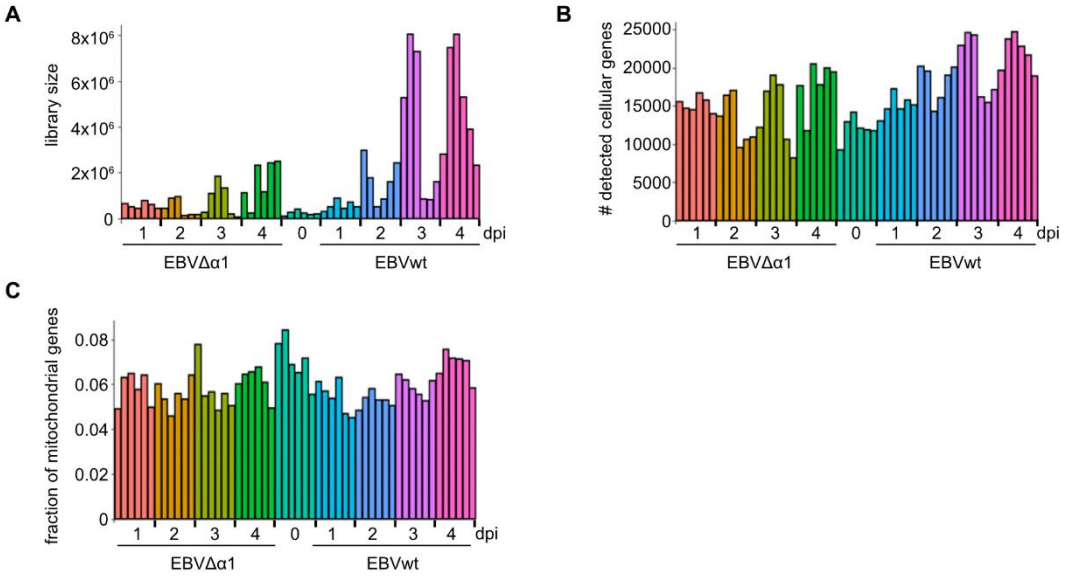


Figure 15 Quality control of RNA seq data; 6 biological replicates for each time point and infection condition were collected. A) Size of sequenced libraries for all collected samples. B) Number of detected cellular genes. Samples with less than 10,000 genes were excluded from downstream analyses. C) Fraction of mitochondrial genes detected among cellular genes.

2.3.5.3 Principal component analysis (PCA)

The principal component analysis (PCA) is a tool to reduce the dimensionality of a large dataset while preserving as much variability at the same time, which makes it easier to interpret a given dataset (Jolliffe and Cadima, 2016). A log-transformed expression matrix was used to perform the PCA. Samples with less than 10,000 cellular genes expressed (cpm >1) were excluded from this analysis (see Quality control of sequenced libraries). The first PCA included all time points and infection conditions (Figure 16A). Here, uninfected samples (day 0) separated from samples of EBVwt and EBVΔα1 infection. Since the variation between the time points of EBVwt and EBVΔα1 infected B cells was rather small, we excluded uninfected samples from the PCA. This analysis revealed a stronger variation between the time points and infection condition (Figure 16B). EBVwt and EBVΔα1 infected samples on day 1 post infection clustered together but, over time, EBVwt and EBVΔα1 infected samples separated. EBVwt infected samples on day 3 and 4 post infection clustered together indicating stronger similarity among these samples. In contrast, the clustering of EBVΔα1 infected samples on day 2, 3 and 4

post infection was rather dispersed indicating a higher heterogeneity among these samples (Figure 16B).

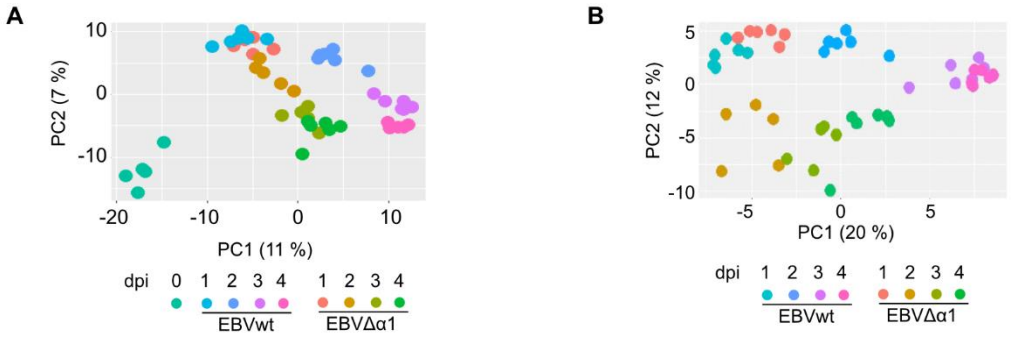


Figure 16 Principal component analysis of uninfected and infected B cells A) PCA including uninfected samples B) PCA with EBVwt and EBVΔα1 infected B cells; dpi – days post infection

2.3.5.4 Differential gene expression (DGE) analysis of protein coding genes

2.3.5.4.1 Gene expression of selected genes

Before looking at global changes of differential gene expression, I was interested in the gene expression of EBNA2, EBF1 and CBF1 (Figure 17). EBNA2wt and EBNA2Δα1 were both induced on day 1 post infection and decreased slightly until day 4. The expression of EBNA2Δα1was slightly but not significantly (FDR <0.1) increased compared to EBNA2wt (Figure 17). The EBF1 expression of EBVwt infected B cells decreased over time whereas the expression in EBVΔα1 infected B cells increased and was significantly (FDR <0.1) elevated on day 3 and 4 post infection. This result was in line with the observation of increased EBF1 protein and mRNA level in LCLΔα1 compared to LCLwt (Figure 11B, C, E). The expression of CBF1, the main EBNA2 DNA anchor, was similar in EBVwt and EBVΔα1 infected B cells (Figure 17).

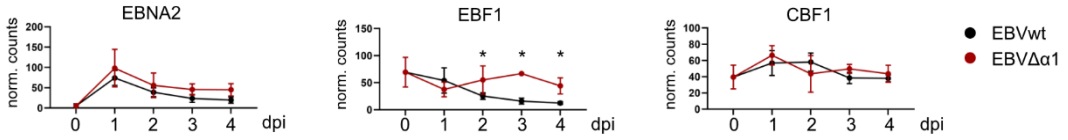


Figure 17 Gene expression of EBNA2, EBF1 and CBF1 in EBVwt (black) and EBVΔα1 (red) infected B cells. Normalized counts plotted over the course of 5 time points. Mean is plotted and the error bars indicate SD. Asterisks indicate FDR < 0.1 as calculated by DESeq2. dpi – days post infection

Since I observed a cell cycle arrest of EBVΔα1 infected B cells, I wanted to have a closer look at the gene expression of some cyclins (CCN), cyclin-dependent kinases (CDK) and additional genes important for the regulation of the cell cycle. The complex formation of cyclin D (D1, D2, D3) with CDK4 or CDK6 activates CDK4/6 and drives the cells through the G1 phase (Baker and Reddy, 2012). The expression of CDK4 was significantly reduced in EBVΔα1 infected B cells on day 2, 3 and 4 post infection compared to EBVwt infected B cells, whereas the expression of CDK6 was not different. The expression of cyclin D2 and 3 (CCND2/3) was similar until day 2 post infection but significantly increased in EBVwt infected B cells on day 3 and 4 post infection (Figure 18A). CDK2, cyclin A and cyclin E are G1/S target genes and expressed as a consequence of an active G1 phase. The association of CDK2 and cyclin E drives the G1/S transition. Cyclin A replaces cyclin E and forms a complex with CDK2 towards the end of the S phase thereby promoting the transition to the G2 phase (Ding et al., 2020). The overall expression of CDK2 was increased in EBVΔα1 infected B cells compared to EBVwt infected B cells but only significantly on day 1 post infection. The expression of cyclin A2 was significantly increased in EBVwt infected B cells on day 3 and 4 post infection. The expression of cyclin E2 was similar in both infection conditions (Figure 18B). The activation of CDK1 by cyclin A promotes the cell cycle progression through the G2 and transition into the M phase. The complex formation of CDK1 with cyclin B maintains the CDK1 activity and is important for the mitosis (Ding et al., 2020). The expression of both CDK1 and cyclin B1was significantly increased in EBVwt infected B cell on day 3 and 4 post infection compared to EBVΔα1 infected B cells (Figure 18C). p53 (TP53) is a potent tumor suppressor and its activation causes cell cycle arrest and apoptosis (Chen, 2016). In contrast to EBVwt infected B cells, the expression of p53 was strongly upregulated in EBVΔα1 infected B cells on day 2 post infection, decreased thereafter but displayed elevated levels compared to EBVwt infected B cells (Figure 18D). p21 (CDKN1A) is an inhibitor of CDKs and can cause a cell cycle arrest (Mansilla et al., 2020). The expression was significantly increased in EBVΔα1 infected B cells on day 1. On day 4 post infection, p21 was significantly reduced in the EBVΔα1 infection.

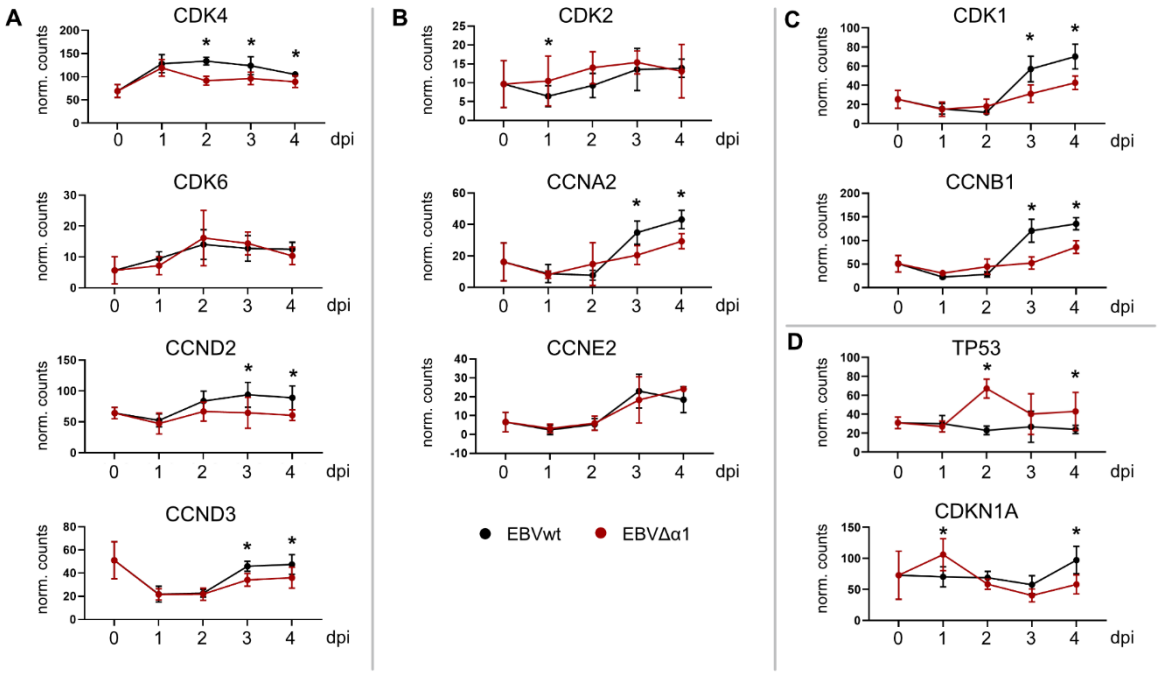


Figure 18 Expression of cell cycle associated genes; normalized expression of cell cycle genes of the A) G1 phase B) G1/S phase C) G2/M phase D) normalized expression of cell cycle inhibitors p53 (TP53) and p21 (CDKN1A); Mean is plotted and the error bars indicate SD. Asterisks indicate FDR < 0.1 as calculated by DESeq2. dpi – days post infection

2.3.5.4.2 Global changes in gene expression upon B cell infection

The differential gene expression (DGE) analysis was performed with the DESeq2 package in R. Only samples with more than 10,000 cellular genes were included (see Quality control of sequenced libraries). For the analysis, genes with a low expression level (cpm <10) were excluded from all samples. First, I compared the gene expression of EBVwt infected B cells on day 1, 2, 3 and 4 against the uninfected sample (day 0). Secondly, I compared the gene expression of EBVΔα1 infected B cells on day 1, 2, 3 and 4 post infection against the uninfected sample (day 0). And lastly, I analyzed the differential gene expression of EBVΔα1 and EBVwt infected B cells on day 1, 2, 3 and 4 post infection (Figure 19).

The DGE analysis of EBVwt infected B cells vs uninfected B cells revealed an increasing number of differentially expressed (DE) genes over the course of four days. The highest number of differentially expressed genes was detected on day 3 post infection. 1706 genes were upregulated of which 741 had at least a 2-fold increase in expression. 1955 genes were downregulated of which 1387 genes showed an expression level of at least 2-fold reduction.

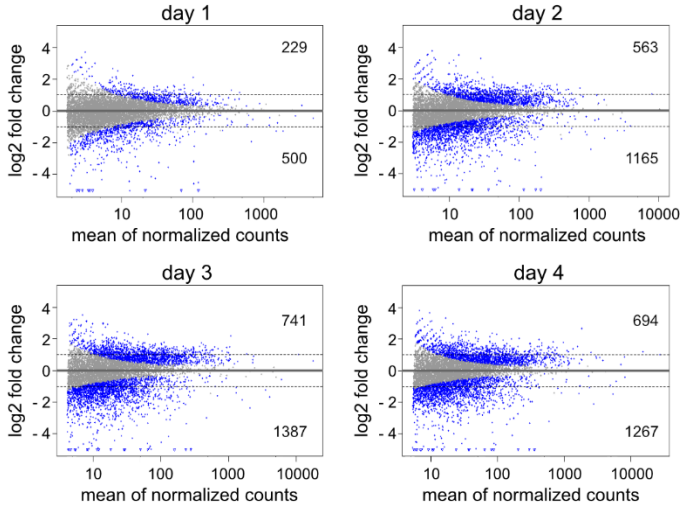
Interestingly, more genes were downregulated than upregulated each day after infection compared to the uninfected samples (Figure 19A, B). At this point of the studies, we cannot differentiate between active downregulation of genes and loss of transcripts, e.g. through cell division. Despite this, I will use the term "downregulation" to indicate a reduction in transcript abundance.

The DGE analysis of EBVΔα1 infected B cells on day 1, 2, 3 and 4 post infection vs uninfected cells also detected an increasing number of DE genes, which continually increased until day 4 post infection. Here, 594 genes were upregulated with 197 genes at least 2-fold and 897 genes were downregulated with 608 at least 2-fold (Figure 19C, D). On day 1 post infection, the number of DE genes in EBVwt and EBVΔα1 infected B cells was similar. On day 2, 3, and 4 post infection, EBVΔα1 infected B cells had dramatically less DE genes compared to EBVwt infected B cells (Figure 19B, D). This observation was in line with the PCA where the day 1 samples of EBVwt and EBVΔα1 infected B cells clustered together but separated the following days. Overall also in EBVΔα1 infected B cells, more genes were downregulated than upregulated each day after infection (Figure 19B).

The DGE analysis of EBVΔα1 infected vs EBVwt infected B cells on day 1, 2, 3, and 4 identified genes that were differentially expressed between the two infection conditions (Figure 19E, F). A log2FC > 0 indicates a stronger expression in EBVΔα1 infected B cells and a log2FC < 0 indicates a stronger expression in EBVwt infected B cells. On day 1 post infection, 96 genes were stronger expressed in EBVΔα1 infected B cells and 136 genes were stronger expressed in EBVwt infected B cells. The number of DE genes increased over time and peaked at day 4 post infection. 1500 genes were stronger expressed in EBVΔα1 infected B cells and 1513 genes were stronger expressed in EBVwt infected B cells (Figure 19E, F).

The DGE analysis demonstrated that EBVwt infection caused dramatic changes in gene regulation, which has also been shown before (Mrozek-Gorska et al., 2019). Infection of B cells with EBVΔα1 also caused changes in gene expression; however, significantly less genes were regulated, which could imply a reduced transformation efficiency of EBVΔα1.

A

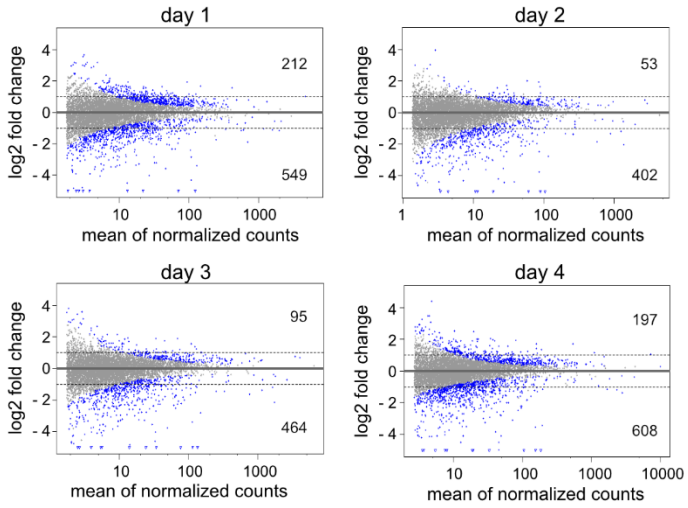


B

EBVwt infected vs uninfected

	log2FC > 0	log2FC < 0	log2FC > 1	log2FC < -1
day 1	588	742	229	500
day 2	1372	1690	563	1165
day 3	1708	1955	741	1387
day 4	1699	1916	694	1267

C

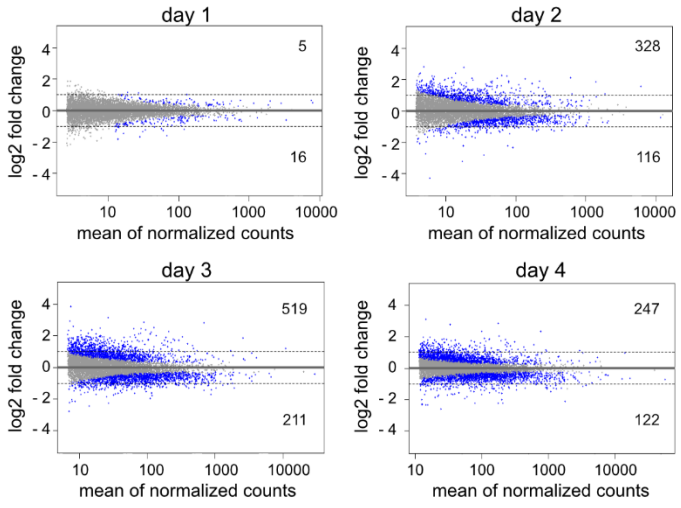


D

EBVΔα1 infected vs uninfected

	log2FC > 0	log2FC < 0	log2FC > 1	log2FC < -1
day 1	604	755	212	549
day 2	233	490	53	402
day 3	306	606	95	464
day 4	594	897	197	608

E



F

EBVΔα1 infected vs EBVwt infected

	log2FC > 0	log2FC < 0	log2FC > 1	log2FC < -1
day 1	95	136	5	16
day 2	785	794	328	116
day 3	1384	1387	519	211
day 4	1500	1513	247	122

Figure 19 Differential gene expression analysis of cellular protein coding genes A) Mean-average (MA) plot of DE genes of EBVwt infected samples on day 1, 2, 3 and 4 vs uninfected samples B) total number of DE genes identified in A and with an FDR<0.1 C) MA plot of DE genes of EBVΔα1 infected samples on day 1, 2, 3 and 4 vs uninfected samples D) total number of DE genes identified in C and with an FDR<0.1 E) MA plot of DE genes of EBVΔα1 vs EBVwt infected samples on day 1, 2, 3 and 4 F) total number of DE genes identified in E and with an FDR<0.1

2.3.5.5 Gene set enrichment analysis (GSEA)

Inspecting significantly regulated genes (fold change > 2, FDR <0.1) discovered in the DGE analysis was not very informative because many genes were regulated and they did not seem to share obvious biological functions. A powerful tool that helps to overcome this problem is the gene set enrichment analysis (GSEA) (Subramanian et al., 2005). It can identify genes sharing common biological processes, e.g. metabolic or signaling pathways, or common regulation. Furthermore, it assists in interpreting datasets, which, for example, do not display dramatic changes in gene expression. The DGE differential gene expression analysis of my dataset revealed that over the time of four days, the majority of DE genes displayed rather small changes (fold change < 2) in their expression level (Figure 19F).

For the GSEA, I performed three sets of analyses: i) EBVwt infected B cells (day 1, 2, 3, 4) vs uninfected (day 0), ii) EBVΔα1 infected B cells (day 1, 2, 3, 4) vs uninfected (day 0) and iii) EBVΔα1 infected B cells vs EBVwt infected B cells on day 1, 2, 3 and 4. I focused my analysis on the 'hallmark gene sets', which are genes sets with reduced redundancy and well-defined biological states or processes (Liberzon et al., 2015). I included all genes from the differential gene expression analysis that had an FDR <0.1 and 0 < log2FC >0. In the following subchapters, I will present the result of GSEA for selected gene sets with significant enrichment (GSEA FDR < 0.1). The complete list of significantly enriched gene sets can be found in the supplements (Table S 6, Table S 7, Table S 8).

2.3.5.5.1 Unfolded protein response (UPR)

The 'Unfolded protein response' is a gene set consisting of 113 genes, which are upregulated during a cellular stress response related to the endoplasmatic reticulum (ER) (https://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_UNFOLDED_PROTEIN_RESPONSE.html). In case of an accumulation

of un-/misfolded proteins in the ER due to a dramatic increase in protein biosynthesis, the UPR is induced to restore protein homeostasis in the cell (Johnston and McCormick, 2019).

GSEA identified an enrichment of this gene set on day 1 post infection. Genes of this set were upregulated in EBVwt (purple) and EBVΔα1 (turquoise) infected B cells (Figure 20A). There was no significant enrichment detected among DE genes of EBVΔα1 vs EBVwt infected B cells. When comparing the leading edge subset - genes contributing the most to the enrichment signal - the majority of the genes were present in both leading edge sub sets (Figure 20B). The size of the leading subsets was very similar with 40 genes in the EBVwt infection and 36 in the EBVΔα1 infection. EIF2S1, XBP1 and RRP9 are examples of the overlap of both leading edge subsets. *EIF2S1* encodes the α subunit of eIF2 that forms a heterotrimeric complex crucial for the initiation of protein translation (Johnston and McCormick, 2019). On day 1 post infection, the expression of this gene was induced in EBVwt and EBVΔα1 infected B cells. The expression level further increased in EBVwt infected B cells but stayed the same in EBVΔα1 infected B cells until day 4 post infection (Figure 20C). *XBP1* plays a central role during UPR. An isoform, generated by mRNA splicing, can acts as transcriptional activator whereas the unspliced form can act as a transcriptional repressor (Ron and Walter, 2007). The expression level of XBP1 in EBVwt and EBVΔα1 infected B cells was similar until day 2 post infection. On day 3 and 4, the expression was lower in EBVΔα1 infected B cells. RRP9 binds the U3 small nucleolar RNA resulting in a ribonucleoprotein complex involved in rRNA processing (Pluk et al., 1998). The expression of RRP9 peaked in EBVwt infected B cells on day 2 post infection. In EBVΔα1 infected B cells, the expression was initially induced on day 1 but did not change thereafter.

It was interesting to detect an enrichment of the UPR gene set on day 1 post infection, which might indicate that EBVwt as well as EBVΔα1 were able to induce gene expression resulting in increased protein synthesis. However, the expression level of the three examples also showed that there is differential gene expression in the following days post infection.

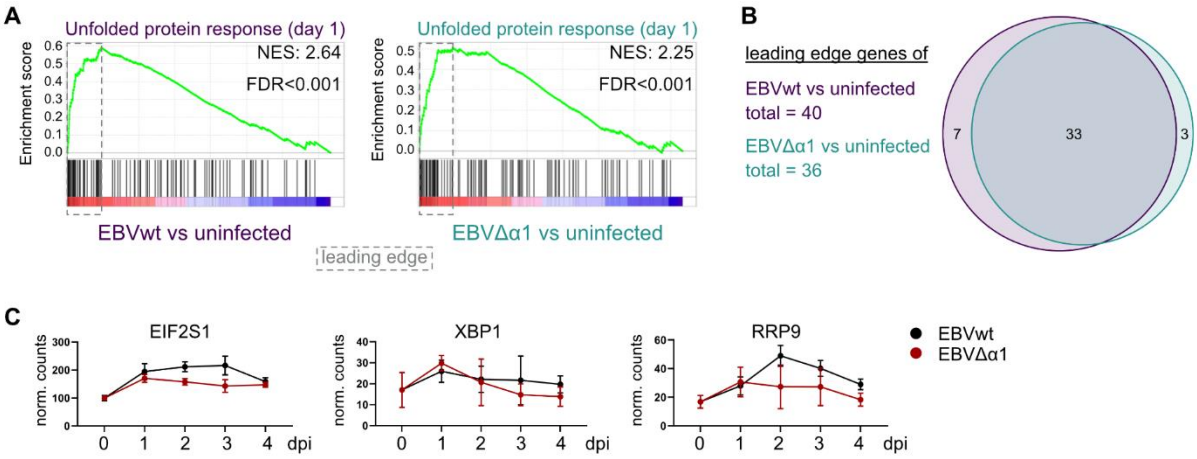


Figure 20 Enrichment of 'Unfolded protein response' on day 1 post infection. A) GSEA of EBVwt vs uninfected (purple) and EBVΔα1 vs uninfected (turquoise) identifies UPR as significantly enriched. B) Venn diagram of the leading edge subset identified in A. EBVwt vs uninfected (purple), EBVΔα1 vs uninfected (turquoise) C) Expression dynamics of XBP1, EIF2S1 and EIF2AK3 in EBVwt (black) and EBVΔα1 (red) infected B cells; All three genes were present in both leading edge subsets. day 0 – uninfected B cells, NES – normalized enrichment score

2.3.5.5.2 MYC targets Version 1 (V1)

On day 2 post infection, the gene set 'MYC targets V1' was significantly enriched. This is a set comprising 200 genes that are regulated by MYC (https://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_MYC_TARGETS_V1.html). Genes belonging to 'MYC targets V1' were upregulated in EBVwt (purple) and EBVΔα1 (turquoise) infected B cells. Differentially expressed genes between EBVΔα1 and EBVwt infected cells (yellow) belonging to 'MYC targets V1' were stronger expressed in EBVwt infected B cells (Figure 21A). Analyzing the overlap of the leading edge subset of each comparison showed that 65 genes were present in all three leading edge subsets. 30 genes were present in both EBVwt (purple) and EBVΔα1 (turquoise) infected B cells. 13 genes did not contribute to the leading edge of EBVwt (purple) and EBVΔα1 (turquoise) infected B cells but were exclusively present in the leading edge subset of the EBVΔα1 vs EBVwt comparison (yellow) subset (Figure 21B).

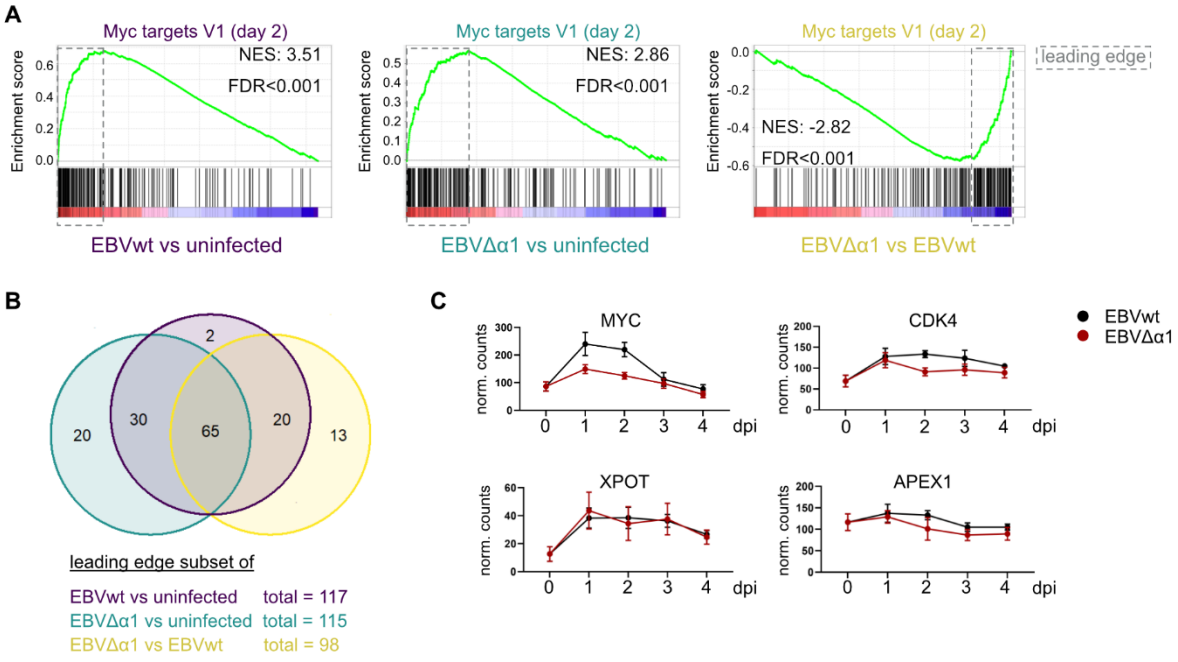


Figure 21 Enrichment of 'MYC targets V1' genes on day 2 post infection. A) GSEA of EBVwt vs uninfected (purple), EBVΔα1 vs uninfected (turquoise) and EBVΔα1 vs EBVwt (yellow) DE genes identified 'MYC targets V1' as a significantly enriched gene set. B) Venn diagram of the leading edge subsets identified in A. EBVwt vs uninfected (purple), EBVΔα1 vs uninfected (turquoise), EBVΔα1 vs EBVwt (yellow) C) Expression dynamics of MYC, CDK4 (present in all 3 leading edge subsets), XPOT (present in purple and turquoise leading edge subset) and APEX1 (only present in the leading edge subset of EBVΔα1 vs EBVwt, yellow); day 0 – uninfected B cells, NES- normalized enrichment score

MYC and CDK4 are examples of genes, which were present in all leading edge subsets. MYC is a direct target of EBNA2 and important for the transformation process of infected B cells (Kaiser et al., 1999). CDK4 is a target of MYC (Hermeking et al., 2000) and important for cell cycle progression through the G1 phase (Baker and Reddy, 2012). Both genes were induced upon infection; however, the induction of these genes was stronger in B cells infected with EBVwt (Figure 21C). XPOT is an example of genes present in both the EBVΔα1 (turquoise) and EBVwt (purple) leading edge subset. XPOT is a RAN-GTP-dependent receptor shuttling tRNA from the nucleus to the cytoplasm (Lin et al., 2019). As expected, the expression dynamics of XPOT was similar in EBVwt and EBVΔα1 infected B cells (Figure 21C). APEX1 is an example of genes, which were exclusively present in the leading edge subset of EBVΔα1 vs EBVwt (yellow). APEX1 is a DNA repair endonuclease with apurinic/apyrimidinic (AP) activity

(Whitaker and Freudenthal, 2018). The expression dynamics of APEX1 were similar in EBVwt and EBVΔα1 infected B cells with elevated levels in EBVwt infected B cells (Figure 21C).
This analysis demonstrated that both viruses were able to induce the expression of MYC and MYC target genes, which is important for the transformation process. However, the degree of induction was reduced for some genes in EBVΔα1 infected B cells, most importantly for MYC itself. The differential expression of MYC and MYC targets already on day 2 post-infection might contribute to the phenotypic differences observed in EBVΔα1 infected B cells.

2.3.5.5.3 E2F targets

The 'E2F targets' gene set comprises 200 genes, which encode cell cycle related targets of E2F transcription factors (https://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_E2F_TARGETS.html). On day 2 post infection, this gene set was enriched for DE genes of the EBVΔα1 vs EBVwt analysis (Figure 22A). At this time point, no significant enrichment was detected for the EBVwt vs uninfected comparison or EBVΔα1 vs uninfected comparison. On day 3 post infection, the 'E2F targets' gene set was significantly enriched for all three comparisons (Figure 22B). Genes of this gene set were upregulated in EBVwt (purple) and EBVΔα1 (turquoise) infected B cells. The direct comparison of EBVΔα1 with EBVwt infected B cells revealed a stronger expression of genes in EBVwt infected B cells (yellow). 43 genes contributing to the leading edge were present in all three subsets. 17 genes were present in the EBVwt (purple) and EBVΔα1 (turquoise) leading edge subset and 11 genes were exclusively present in the EBVΔα1 vs EBVwt (yellow) leading edge subset (Figure 22C).

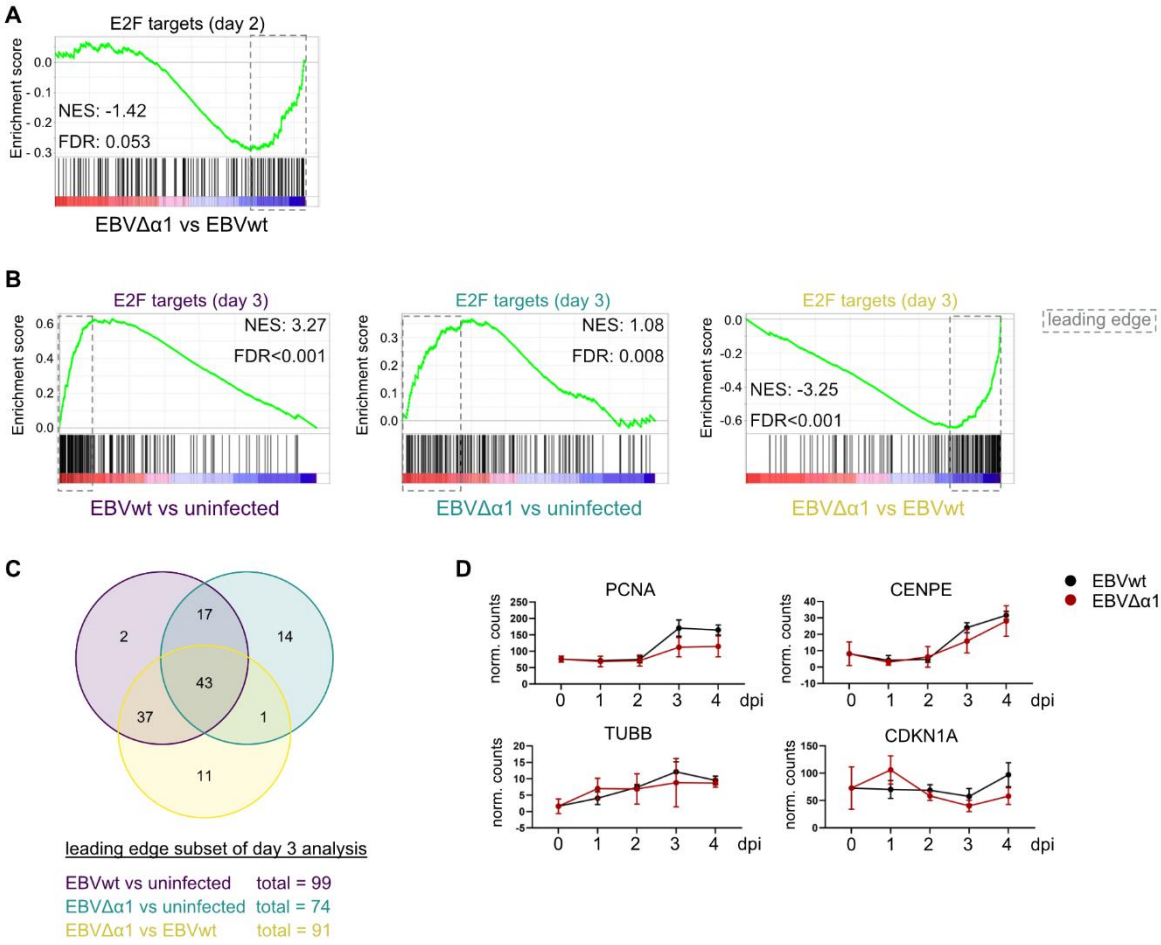


Figure 22 Enrichment of the 'E2F targets' gene set on day 2 and 3 post infection. A) GSEA on day 2 post infection of EBVΔα1 vs EBVwt identified 'E2F targets' as a significantly enriched gene set. B) GSEA on day 3 post infection of EBVwt vs uninfected (purple), EBVΔα1 vs uninfected (turquoise) and EBVΔα1 vs EBVwt (yellow) DE genes identified 'E2F targets' as a significantly enriched gene set. C) Venn diagram of the leading edge subsets identified in B; EBVwt vs uninfected (purple), EBVΔα1 vs uninfected (turquoise), EBVΔα1 vs EBVwt (yellow) D) Expression dynamics of PCNA, CENPE (present in all 3 leading edge subsets), TUBB (present in purple and turquoise leading edge subset) and CDKN1A (only present in the leading edge subset of yellow subset); day 0 – uninfected B cells, NES- normalized enrichments score

PCNA and CENPE are examples, which were present in all three leading edge subsets. PCNA is an important protein during DNA replication and repair (González-Magaña and Blanco, 2020). The expression of this gene was the same in EBVwt and EBVΔα1 infected B cells until day 2 post infection. On day 3 and 4, the expression was strongly increased in EBVwt infected B cells but only moderately increased in EBVΔα1 infected B cells (Figure 22D). CENPE is a centromere-associated protein that is driving the progression from the metaphase to the anaphase (Yen et al., 1991). The expression level of this gene only differed on day 3 post infection. EBVwt infected cells displayed a stronger expression compared to EBVΔα1 infected

B cells (Figure 22D). *TUBB* is an example of the leading edge subsets present in the EBVwt (purple) and EBVΔα1 (turquoise) subset. *TUBB* encodes for the tubulin beta protein, which is a major component of the cytoskeleton. The expression levels were very similar between the two infection conditions over time (Figure 22D). *CDKN1A* is an example of genes, which were only present in the EBVΔα1 vs EBVwt (yellow) leading edge subset. *CDKN1A* encodes for the protein p21, which can bind to cyclin-dependent kinases (CDK) and thereby inactivates them. The inactivation can trigger cell cycle arrest at the G1 or G2 phase (Mansilla et al., 2020). In the EBVΔα1 infection, the peak expression was reached at day 1 post infection and declined afterwards. In B cells infected with EBVwt, the expression was stable until day 3 post infection and increased until day 4 post infection (Figure 22D).

Identification of 'E2F targets' as a gene set enriched in both EBVwt and EBVΔα1 infected B cells indicated that the viruses were able to initiate the expression of genes that are associated with and important for the cell cycle. However, the analysis also demonstrated that on day 2 post infection genes of this gene set were already differentially expressed between the infection conditions, i.e. less expressed in EBVΔα1 infected B cells. The number of DE genes accumulated and the difference was even stronger on day 3 post infection.

2.3.5.5.4 G2M checkpoint

Based on the cell cycle analysis in chapter 2.3.2 and published studies (Mrozek-Gorska et al., 2019), I expected an upregulation genes that are important for the G2-M transition and cell division around day 3 to 4 post infection. In accordance to my expectation, the gene set 'G2M checkpoint' was significantly enriched on day 4 post infection in EBVwt vs uninfected (purple), EBVΔα1 vs uninfected (turquoise) and in EBVΔα1 vs EBVwt (yellow, Figure 23A). Importantly, the 'G2M checkpoint' gene set contains 200 genes, which are not only involved in the G2-M checkpoint but also important for other cell cycle checkpoints (https://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_G2M_CHECKPOINT.html). Nevertheless, genes belonging to this set were upregulated in EBVwt and EBVΔα1 infected B cells but many genes were stronger expressed in EBVwt infected B cells than EBVΔα1 infected B cells. 29 genes contributing to

the leading edge were present is all three subsets (Figure 23B). 14 genes were present in both the EBVwt (purple) and the EBVΔα1 (turquoise) subset and 16 genes were exclusively present in the EBVΔα1 vs EBVwt (yellow) subset.

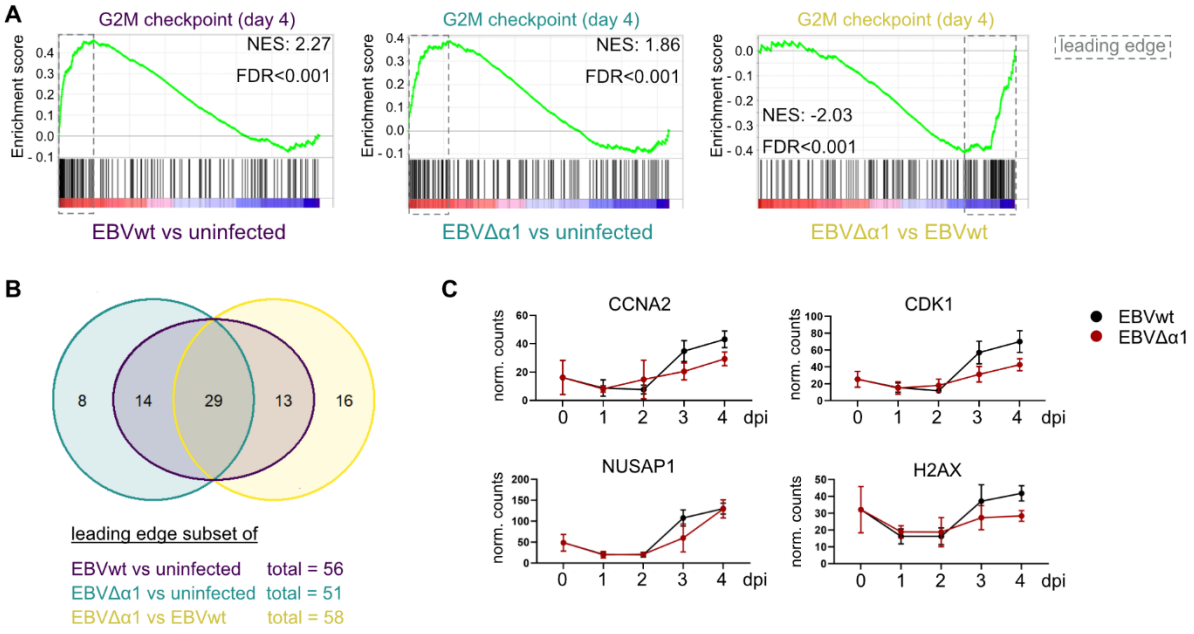


Figure 23 Enrichment of the 'G2M checkpoint' gene set on day 4 post infection. A) GSEA on day 4 post infection of EBVwt vs uninfected (purple), EBVΔα1 vs uninfected (turquoise) and EBVΔα1 vs EBVwt (yellow) DE genes identified 'G2M checkpoint' as a significantly enriched gene set. B) Venn diagram of the leading edge subsets identified in A; EBVwt vs uninfected (purple), EBVΔα1 vs uninfected (turquoise), EBVΔα1 vs EBVwt (yellow) C) Expression dynamics of CCNA2, CDK1 (present in all 3 leading edge subsets), NUSAP1 (present in purple and turquoise leading edge subset) and H2AX (only present in the leading edge subset of the yellow subset); day 0 – uninfected B cells, NES- normalized enrichment score

CCNA2 and CDK1 are examples of genes contributing to all three leading edge subsets. *CCNA2* encodes for Cyclin A2, which can bind to CDK2 and CDK1 and promotes the transition from the S to G2 phase (Ding et al., 2020). The expression of *CCNA2* decreased after the infection but increased after day 2 post infection. The increase was stronger for EBVwt infected B cells than EBVΔα1 infected B cells (Figure 23C). CDK1 is a cyclin-dependent kinase, which is important for the entry into the M phase of the cell cycle. The expression dynamics of CDK1 were very similar to *CCNA2*. After an initial decrease, the expression increased at day 3 and 4 post infection. The expression level in EBVwt infected B cells was higher compared to EBVΔα1 infected B cells (Figure 23C). NUSAP1 is an example of genes, which were present in the leading edge subset of both EBVwt (purple) and EBVΔα1 (turquoise) infected B cells. NUSAP1

is a microtubule-associated protein. Its expression peaks during the transition from the G2 to M phase (Raemaekers et al., 2003). In EBVwt and EBVΔα1 infected B cells, the expression was the same except on day 3 post infection. Here, the expression was lower in EBVΔα1 infected B cells (Figure 23C). H2AX is a histone variant of H2A and an indicator of DNA damage (Tarakanova et al., 2007). The expression dynamics were the same until day 2 post infection. On day 3 and 4, the expression was stronger in EBVwt infected B.

Since some of the genes in 'G2M checkpoint' are E2F targets, an enrichment of this gene set was expected. Furthermore, the result of this GSEA supports the observation of an early arrest of the cell cycle in primary B cells infected with EBVΔα1 (Figure 9) because genes of the 'G2M checkpoint' set were less expressed in EBVΔα1 infected B cells.

2.3.5.5.5 Glycolysis

Another interesting gene set that was enriched is the 'Glycolysis' set. This gene set contains genes, which are involved in the glycolysis and gluconeogenesis (https://www.gsea-msigdb.org/gsea/msigdb/cards/HALLMARK_GLYCOLYSIS.html). It was already significantly enriched in EBVwt infected B cells on day 2 post infection. An enrichment of this set was found among DE genes between EBVΔα1 vs EBVwt infected B cells, which demonstrates that genes of this set were stronger expressed in EBVwt infected B cells (Figure 24A). However, there was no significant enrichment for EBVΔα1 vs uninfected. On day 4 post infection, an enrichment of the 'Glycolysis' gene set was found in all three comparisons. There was an upregulation of genes in EBVwt (purple) and EBVΔα1 (turquoise) infected B cells. The stronger expression of genes within this gene set in EBVwt infected B cells was also present on day 4 post infection (yellow, Figure 24B). The majority of genes contributing to the leading edge of EBVΔα1 (turquoise) were also present in the EBVwt (purple) and EBVΔα1 vs EBVwt (yellow) leading edge subsets. 7 genes were present in both the EBVwt (purple) and EBVΔα1 (turquoise) leading edge subset and 15 genes were exclusively present in the EBVΔα1 vs EBVwt (yellow) subset (Figure 24C). *ENO1* and *GOT2* are examples of genes, which contributed to all three leading edge subsets. *ENO1* encodes for enolase 1, a protein catalyzing the conversion of 2-phosphoglycerate to

phosphoenolpyruvate. The expression of *ENO1* increased after EBV infection but was significantly elevated in EBVwt infected B cells (Figure 24D). GOT2 is a glutamate-aspartate translocator located in the membrane of mitochondria and is involved in amino acid metabolism (Borst, 2020). Over the course of four days, the expression was decreased in EBVΔα1 infected B cells (Figure 24D). G6PD is an example of genes, which were present in the leading edge subset of both EBVwt (purple) and EBVΔα1 (turquoise) infected B cells. *G6PD* encodes for glucose-6-phosphate dehydrogenase and catalyzes the first step in the pentose phosphate pathways (Ge et al., 2020). The expression of this gene was very similar between EBVwt and EBVΔα1 infected B cells (Figure 24D). *ALG1* is an example for genes, which were only present in the EBVΔα1 vs EBVwt (yellow) leading edge subset. ALG1 is an enzyme that is involved in the formation of lipid-linked oligosaccharides. The expression of this gene was induced in EBVwt infected B cells but was not induced in EBVΔα1 infected B cells (Figure 24D).

Similar to the previous gene sets, both viruses were able to induce the expression of genes which are necessary for the glycolysis and related pathways. The analysis also indicated that the induction of gene expression was more efficient in B cells infected with EBVwt. For example, the 'Glycolysis' gene set was already significantly enriched on day 2 post infection in the EBVwt whereas a significant enrichment in EBVΔα1 infected B cells was detected on day 4 post infection. This might indicate that genes important for glycolysis are expressed in EBVΔα1 infected B cells but the activation of expression might be delayed or the expression level are reduced compared to EBVwt infected B cells.

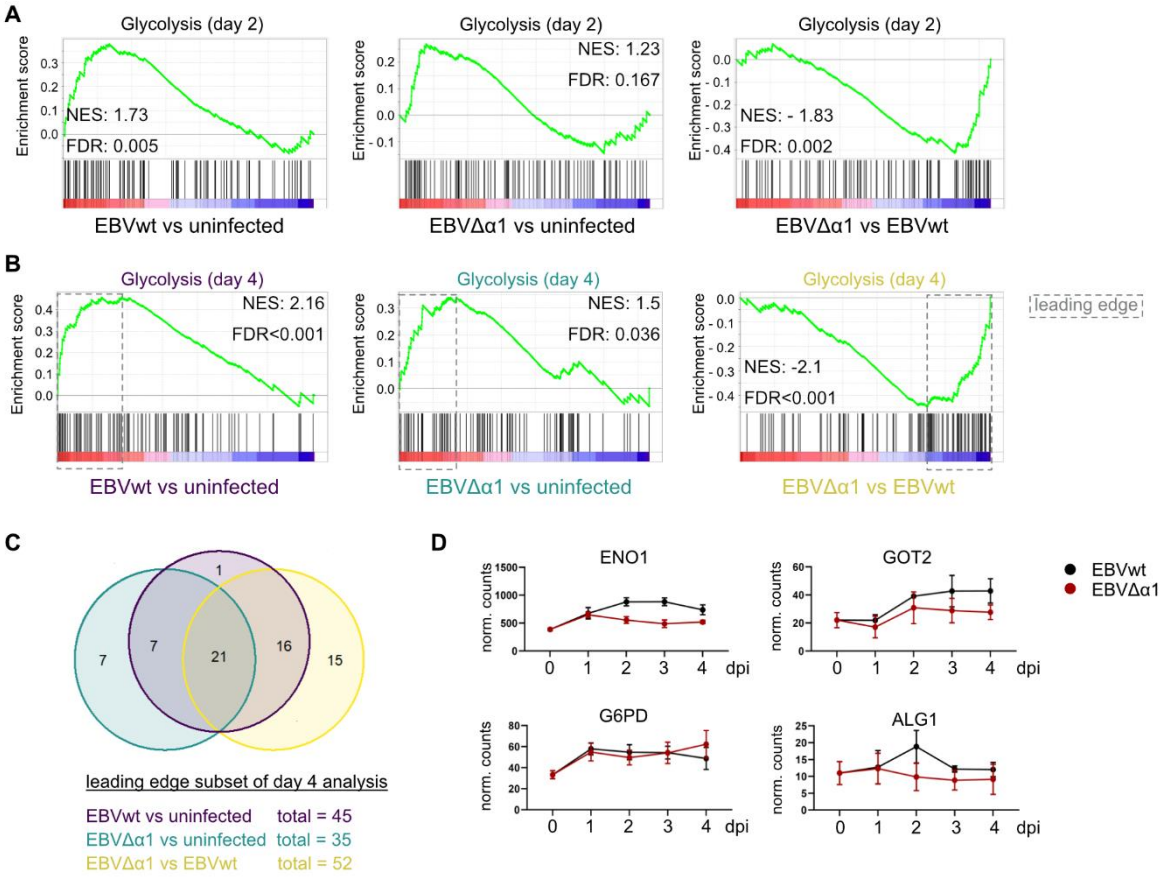


Figure 24 Enrichment of the 'Glycolysis' gene set on day 2 and 4 post infection. A) GSEA on day 2 post infection of EBVwt vs uninfected and EBVΔα1 vs EBVwt identified 'Glycolysis' as a significantly enriched gene set. 'Glycolysis' in the EBVΔα1 vs uninfected comparison was not significantly enriched. B) GSEA on day 4 post infection of EBVwt vs uninfected (purple), EBVΔα1 vs uninfected (turquoise) and EBVΔα1 vs EBVwt (yellow) DE genes identified 'Glycolysis' as a significantly enriched gene set. B) Venn diagram of the leading edge subsets identified in B; EBVwt vs uninfected (purple), EBVΔα1 vs uninfected (turquoise), EBVΔα1 vs EBVwt (yellow) C) Expression dynamics of ENO1, GOT2, (present in all 3 leading edge subsets), G6PD (present in purple and turquoise leading edge subset) and ALG1 (only present in the leading edge subset of the yellow subset); day 0 – uninfected B cells, NES- normalized enrichment score

Taken together, the RNA sequencing demonstrated that the infection of B cells with EBVwt and EBVΔα1 induced gene regulation including up- and downregulation. On day 1 post infection, B cells infected with either of the viruses displayed a similar phenotype and gene regulation, demonstrated by FACS (Figure 14), PCA (Figure 16) and differential gene expression analysis (Figure 19). The gene expression of EBNA2/EBNA2Δα1, EBF1 and CBF1 was similar for EBVwt and EBVΔα1 infected B cells or even elevated in the mutant infection. Furthermore, both viruses were able to induce the expression of genes of important pathways, e.g. glycolysis, however, the degree of gene expression was markedly reduced in EBVΔα1

infected B cells. The differences between EBVwt and EBVΔα1 infected B cells were already detectable at day 2 post infection and manifested thereafter. For example, FACS analysis showed a shift in cell size for EBVwt infected B cells (Figure 14); PCA revealed a separation of samples with EBVwt or EBVΔα1 infection (Figure 16) and dramatically less genes were regulated in EBVΔα1 infected B cells (Figure 19).

2.4 Exploring two approaches to generate conditional EBF1 cell systems

In the previous chapters, I have demonstrated that the EBNA2 α1-helix is important for the EBF1 interaction and that the deletion of the α1-helix has a global effect on the transformation of EBV infected B cells. To better understand the role of EBF1 during the transformation process, I wanted to establish lymphoblastoid cell lines (LCLs) with conditional EBF1 expression. Such a cell system would enable us to study the impact of an impaired EBNA2-EBF1 interaction in the presence of a fully functional EBNA2. Additionally, we would be able to systematically identify EBF1 target genes in LCLs, which are well studied in developing B cells but not in mature B cells. Furthermore, we could investigate mutual EBNA2 and EBF1 target gene regulation and chromatin binding of these factors in the presence and absence of EBF1. Studying the function of EBF1 in the context of an EBV infection requires a conditional system since EBF1 is important for the viability of LCLs (Lu et al., 2016).

I explored two approaches based on genome-editing by CRISPR/Cas9 to generate a conditional EBV cell system. The first approach aimed at controlling EBF1 transcription from a doxycycline inducible promotor system followed by the knock-out of endogenous EBF1. The second approach aimed at the expression of an estrogen-sensitive EBF1 fusion protein consisting of the ligand-binding domain of the estrogen receptor (ER-LBD) fused to the endogenous EBF1 open reading frame. In the cell system, EBF1 would be expressed at endogenous levels but the translocation to the nucleus and activity of EBF1 would be dependent on β-estradiol in the cell culture medium. The two approaches will be discussed in detail in the following chapters.

2.4.1 A conditional EBF1 cell system sensitive to doxycycline

The concept of the doxycycline sensitive system was to induce EBF1 expression in cells upon addition of doxycycline to the culture. Establishing such a cell system required a stable transfection of GM12878 (LCL) with a doxycycline inducible EBF1 expression plasmid followed by the knock-out of endogenous EBF1. The order of the steps was important because EBF1 seems to be important for the viability of LCLs (Lu et al., 2016). First, I cloned the EBF1 cDNA with a Flag-tag via a shuttle vector into the pRTR vector (Jackstadt et al., 2013). The pRTR vector has a tetracycline regulated bidirectional promotor that induces the expression of EBF1 and tNGFR/eGFP at the same time upon the addition of doxycycline to the cell culture medium (Figure 25A). tNGFR and eGFP are separated by an IRES sequence allowing independent translation of the proteins. tNGFR will be presented at the cell surface and can be used for cell sorting. eGFP is a fluorescent protein, which helps to monitor transfected cells and can be used for cell sorting as well. Additionally, the vector carries a truncated version of rat-CD2, which can be used for magnetic cell sorting. The viral protein EBNA1 binds to oriP, viral origin of replication, and the human genome, thus, tethering the plasmid to the host cell genome and ensuring episomal replication and proper segregation of the episome after mitosis (Bornkamm et al., 2005). The puromycin resistance gene, puromycin-N-acetyltransferase (Puro), allows the selection of transfected cells (Figure 25A). Another advantage of the pRTR vector is the control of gene expression level by titration of doxycycline (Figure 25B). EBF1 expression was induced in stably transfected GM12878 with increasing doxycycline concentrations resulting in increased EBF1 and Flag protein expression.

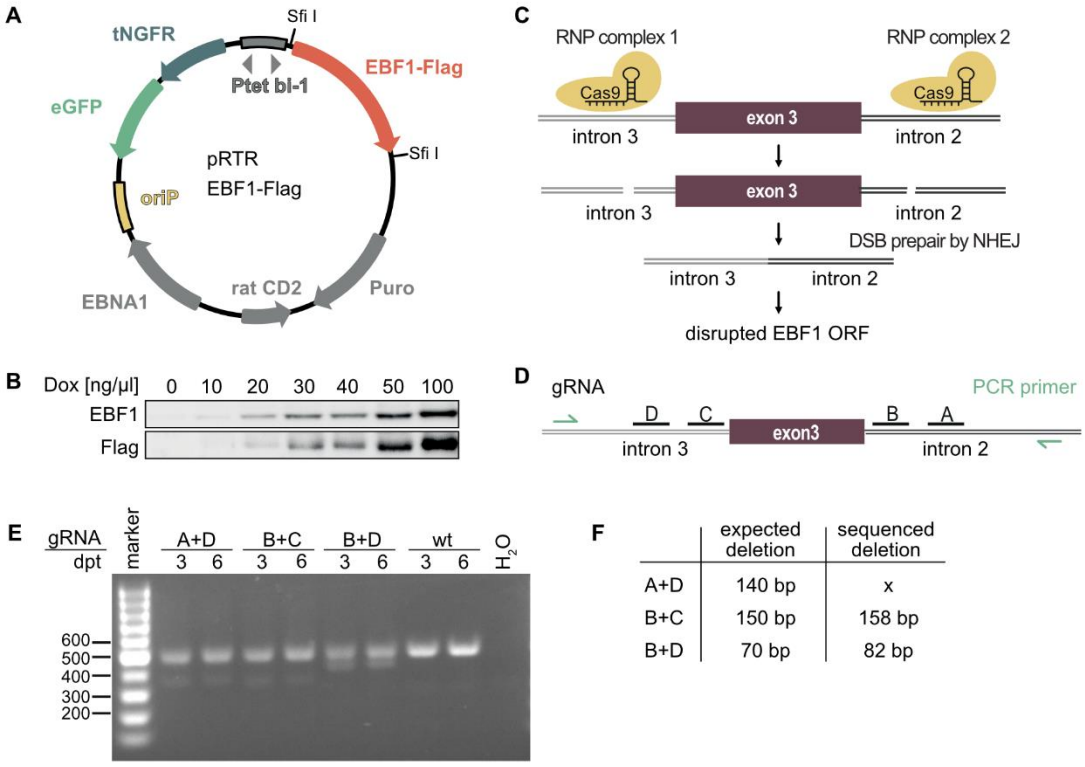


Figure 25 Strategy to create a conditional EBF1 cell system sensitive to doxycycline. A) pRTR plasmid carrying Flag-tagged EBF1 cDNA for stable transfection of LCL. The bidirectional doxycycline-regulated promotor Ptet bi-1 initiates the expression of EBF1 and tNGFR, eGFP simultaneously. tNGFR and eGFP are reporter genes that can be used to monitor transgene expression or for sorting of high expressing cells. EBNA1 is a viral protein that maintains the episomal replication of the vector in the nucleus. Truncated rat CD2 can be used for magnetic sorting of successfully transfected cells. oriP is the EBV episomal origin of replication which enables tethering of the episome to the host genome by EBNA1. B) Dose response of stably transfected LCLdoxEBF1-Flag to doxycycline. C) Schematic overview of the gene editing approach to knock out endogenous EBF1 in LCLdoxEBF1-Flag cell line. Two ribonucleoprotein (RNP) complexes induce two double strand breaks (DSB) before and after exon 3 of EBF1. The DSB will be repaired by non-homologous end joining (NHEJ), which leads to the deletion of exon 3 with a subsequent disruption of the EBF1 open reading frame. D) Relative localization of four gRNAs tested for gene editing. E) Agarose gel of the T7 endonuclease 1 assay of LCLdoxEBF1-Flag transfected with the indicated combinations of RNP complexes. Cells were harvested on day 3 and 6 post transfection. F) Expected and detected deletions in the EBF1 locus using different RNP complex combinations. dpt – days post transfection, wt – wildtype

I applied the CRISR/Cas9 system to knock-out endogenous EBF1 of stably transfected GM12878. Furthermore, I chose to use ribonucleoprotein (RNP) complexes in combination with transfection to deliver Cas9 with the guide RNA (gRNA) to the cells. Using two different RNP complexes for genome editing allows the deletion of a specific genomic region. We decided to target exon 3 of EBF1 because it was shown before to effectively disrupt the open reading frame of the EBF1 gene (Lin and Grosschedl, 1995). As outlined in Figure 25C, using RNP complexes that are targeting introns 2 and 3 of EBF1 will lead to two DNA double strand

breaks (DSB) that can be repaired by non-homologous end joining (NHEJ) causing a deletion of exon 3. I tested four different gRNA to find the most efficient gRNA combination. The relative locations of the gRNAs are indicated in Figure 25D with gRNA A and B targeting intron 3 and gRNA C and D targeting intron 2. The T7 endonuclease 1 (T7E1) assay can be used to assess the genome editing efficiency of RNPs. The T7E1 assay is based on the ability of the nuclease to cleave mismatched double stranded DNA. In a first step, the genomic region of interest is amplified by PCR from a bulk of cells that were transfected with RNP complexes. The resulting PCR product is purified and the DNA stands are re-annealed, which can lead to mismatched DNA duplexes. The T7E1 nuclease detects and cleaves the mismatched DNA duplexes. Subsequently, the resulting fragments can be analyzed by gel separation (Figure 25E). Indeed, the different RNP combinations resulted in editing of the EBF1 locus visible by a smaller and larger DNA fragment. Additionally, PCR product was sequenced to check for deletions. A deletion of 158 bp was detected for the RNP combination B+C and a deletion of 82 bp was detected for the RNP combination B+D (Figure S 2). Sequencing of cells transfected with the RNP combination A+D gave an ambiguous result for a deletion. To decide which RNP combination works best, capillary electrophoresis of the T7E1 cleaved PCR fragments could give detailed information about the genome editing efficiency. The next step in generating a doxycycline sensitive cell system would include single cell cloning and verification of the desired deletion. Positive clones could then be used to cultivate conditional EBF1 cell lines, which are doxycycline inducible.

2.4.2 A conditional EBF1 cell system sensitive to β-estradiol

2.4.2.1 The knock-in strategy

The second approach to generate a conditional EBF1 cell system was based on creating a fusion protein consisting of EBF1 and the ligand-binding domain of the estrogen receptor (ER-LBD). In this cell system, the expression of the EBF1 fusion gene (EBF1-ER) would be controlled by the endogenous promotor and the fusion protein expression levels most likely would resemble endogenous expression levels. In the absence of β-estradiol, the EBF1-ER protein would reside

in the cytoplasm. Upon β-estradiol addition to the cell culture medium, the EBF1-ER protein would translocate to the nucleus. In a previous study, an EBF1-ER fusion construct was used to study chromatin remodeling at the CD79a promotor (Gao et al., 2009). The authors transduced plasmacytoma cells with retroviral vectors encoding for EBF1-ER and regulated EBF1 activity by adding tamoxifen to the cell culture. James Hagman kindly provided the ER-LBD and linker sequence that was used in their study. My approach was a bit different because I tried to knock-in the ER-LBD into the endogenous EBF1 locus.

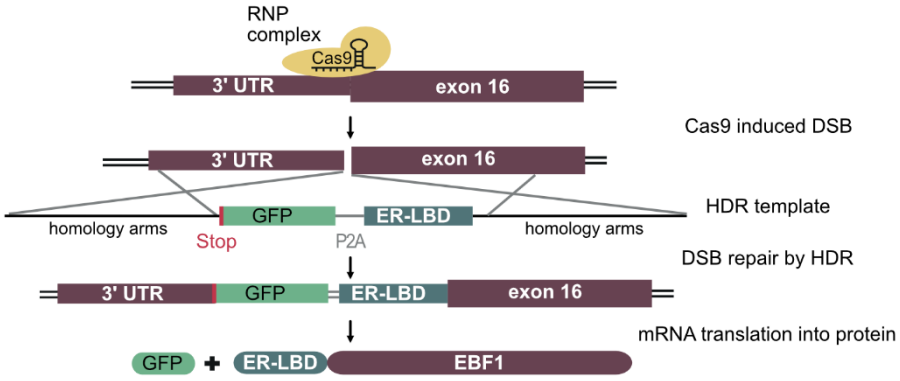


Figure 26 The knock-in approach to create an EBF1-ER-LBD fusion protein. A ribonucleoprotein (RNP) complex induces a double strand break (DSB), which is repaired by homology directed repair (HDR) if an HDR template is provided to the cell. Here, the template contains the ligand-binding domain of the estrogen receptor (ER-LBD) and GFP as well as flanking homology arms. After the successful knock-in of the HDR template just before the stop codon of EBF1, an EBF1-ER fusion protein and GFP protein will be expressed.

For the knock-in, RNP complexes were transfected into the cells where they induce DNA double strand breaks (DSB) just before the EBF1 stop codon. Providing the cells at the same time with a homology directed repair (HDR) template, the DSB should be repaired by HDR. The HDR template used in this approach included the ER-LBD sequence, a P2A and GFP sequence as well as left and right flanking homology arms. Both homology arms were around 400 bp in length (Figure 26, Figure 28A). After a successful recombination of the EBF1 locus, the transcribed mRNA would encode for EBF1-ER and GFP. Due to the P2A sequence, two proteins, EBF1-ER and GFP, would be translated from the same mRNA. GFP serves a reporter gene/protein that should be only expressed after a successful knock-in of the construct at the

correct genomic position and could be used to monitor and sort cells with the knock-in (Figure 26).

2.4.2.2 Selection of an efficient gRNA

I designed and tested the targeting efficiency of two gRNAs by applying the T7E1 assay and the TIDE web tool (Brinkman et al., 2014). Both gRNAs had overlapping sequences with a 3-nucleotide shift (Figure 27A). Genomic DNA was harvested 3 or 6 days post transfection of the RNP complexes. I tested two different time points post transfection to see if there are differences in the gene editing efficiencies. Since the T7E1 assay of 38rev treated samples on day 6 post transfection did not give a clear result on the agarose gel, I continued with the analysis of samples from day 3 post transfection. After performing the T7E1 assay, the intensity of cut fragments of 35rev treated samples on day 3 seemed to be less intense compared to the uncut fragment (Figure 27B). In contrast, the cut fragments of the 38rev treated samples on day 3 post transfection appeared to be more intense compared to the uncut fragment. This might indicate that the gRNA 38rev was more efficient in editing the target locus. An additional analysis tool providing information about the editing efficiency of gRNA is the web tool TIDE (Brinkman et al., 2014). TIDE needs the sequences of edited and unedited samples as well as the guide sequence of the gRNA to calculate the indel frequency induced by gRNAs. The indel frequency for 35rev was around 20 % and for 38rev around 60 % (Figure 27C). Having the result of both analysis tools – T7E1 assay and TIDE – I decided to continue with the gRNA 38rev to perform the knock-in of ER-LBD into the EBF1 locus.

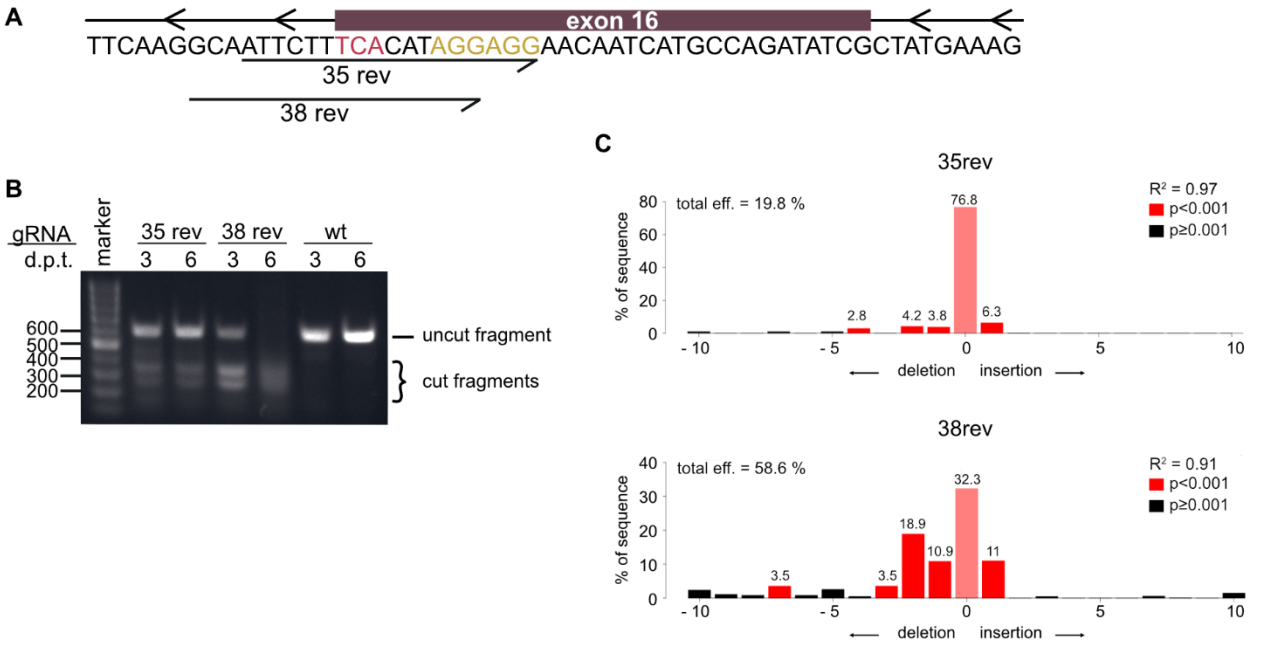


Figure 27 Selection of an efficient gRNA A) Position of two gRNAs, 35rev and 38rev, that were tested for the gene editing efficiency. Red letters indicate the stop codon of EBF1. Yellow letters indicate the PAM sequences of the two gRNAs. B) Agarose gel of the T7 endonuclease 1 assay of GM12878 transfected with the indicated RNP complexes. Cells were harvested on day 3 and 6 post transfection. d.p.t. – days post transfection, wt – wild type C) Indel frequency induced by gRNA 35rev or 38rev calculated by TIDE.

2.4.2.3 Knock-in of the ER-LBD into the EBF1 locus in Nalm6 and GM12878 cells

As mentioned earlier, a CRISPR/Cas9 directed knock-in of a DNA sequence requires a template that is used during the HDR. The company GenScript synthesized and cloned the required DNA sequence into the pUC57-mini vector (Figure 28A). The vector carried an ampicillin resistance gene (Amp), origin of replication as well as the HDR template sequence (homology arms – hom, ER-LBD, P2A, GFP). Since the pUC57-mini vector does not carry any promotor sequences, expression of genes from this vector should not be expected.

First, I performed a proof-of-concept experiment in Nalm6 cells because this cell line has been shown to have a high genome editing efficiency by homologous recombination (Adachi et al., 2006). Nalm6 cells are human pre-B cells derived from a patient with acute lymphoblastic leukemia (Hurwitz et al., 1979; Quentmeier et al., 2019). I transfected RNPs (38rev gRNA + Cas9) together with a linearized, nicked or circular vector containing the HDR template sequence and analyzed the frequency of GFP positive cells by FACS 3 days later. I gated on

lymphocytes, singlets, and live cells to exclude autofluorescence emitted by apoptotic cells (Figure 28B). Genome editing in combination with a linearized plasmid resulted in 1.23 % of GFP positive cells, the circular plasmid resulted in 2.27 % GFP positive cells and genome editing in combination with the nicked plasmid only resulted in 0.91 % of GFP positive cells. To confirm a knock-in into the EBF1 locus, I purified genomic DNA and performed a specific PCR with one primer located within the ER-LBD and the second primer outside of the left homology arm (Figure 28D). A successful knock-in would generate a PCR fragment of around 500 bp. Indeed, Nalm6 cells, which underwent genome editing with a linearized HDR template showed a significant fragment at around 500 bp for the PCR (Figure 28D). A faint signal was visible for cells provided with the circular HDR template. No signal was detectable for cells transfected with the nicked HDR template (nicked), with a GFP control vector (GFP), mock transfected cells or the negative PCR water control.

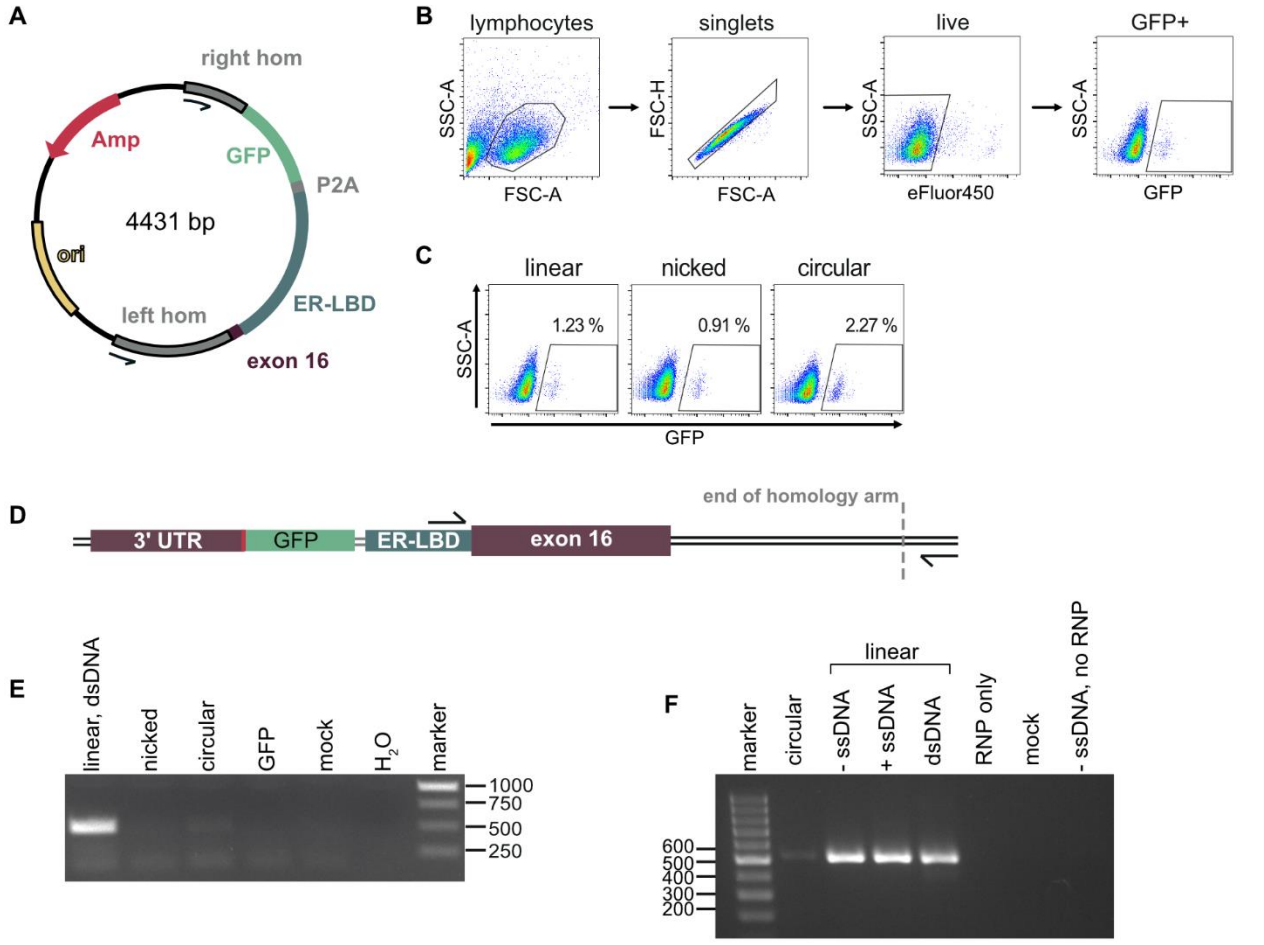


Figure 28 Confirmation of the ER-LBD knock-in in Nalm6 and GM12878 cells. A) The pUC57-mini plasmid carrying the HDR template with the left and right homology arms (right/left hom), ampicillin resistance (Amp) and the origin of replication (ori). Arrows indicate primer for ssDNA preparation. B) Gating strategy for the FACS analysis of transfected Nalm6 cells C) FACS analysis of Nalm6 cells transfected with RNPs and the HDR template. GFP+ cells indicate successful knock-in. HDR template was provided in a linear, nicked or circular plasmid version. D) Position of PCR primers to check for a knock-in of the HDR template. E) Agarose gel of a PCR to check for a knock-in in Nalm6 cells. The HDR template was provided as a linear, nicked or circular plasmid. A GFP expression plasmid was used a transfection control. Mock transfected cells only received the electric shock. H_2O is the negative control for the PCR. F) Agarose gel of a PCR to check for a knock-in in GM12878. The HDR template was provided as a circular plasmid, - or + strand of a single stranded DNA (-/+ ssDNA), double stranded DNA (dsDNA). RNP only, mock nucleofection and –ssDNA without RNP complexes served as negative controls.

The target cell line to establish the β-estrogen sensitive system was the lymphoblastoid cell line GM12878. It has been used in various studies and a lot of experimental data is publicly available. This would have been useful for comparable functional studies in the EBF1 conditional GM12878 cell system. Working with GM12878 cells turned out to be more difficult compared to Nalm6 cells. While Nalm6 cells with a knock-in of the ER-LBD construct were traceable by FACS (Figure 28C), this process was not traceable in GM12878 cells because of the

extremely low frequency of GFP positive cells. GFP served as a reporter to sort cells carrying the knock-in in order to generate single cell clones later on. Assuming that the frequency of GFP positive cells correlates with the frequency of cells carrying the knock-in, I tried to optimize the knock-in of the ER-LBD construct. Using single stranded (ss) DNA as an HDR template has been demonstrated to increase the knock-in efficiency and to decrease off-target effects (Roth et al., 2018). Therefore, I generated ssDNA by PCR amplifying the HDR template from the pUC57-mini plasmid with primers indicated in Figure 28A. The primer pairs used for the PCR carried a phosphate group at the 5'-end of either the forward or the reverse primer. A PCR run with these primer pairs leads to a fragment that has a phosphate group at the 5'-end of either the positive or negative strand. Incubating the PCR fragments with the λ exonuclease produces either –ssDNA or +ssDNA because the phosphate-tagged strand will be recognized and digested by the enzyme. Due to low yields, I did not validate with additional experiments if the HDR templates were really ssDNA but I will refer to them as –ssDNA or +ssDNA (Figure 28F). I tested different DNA confirmations to see which was most suited for the HDR in GM12878. I used the circular HDR template (Figure 28A), -ssDNA and +ssDNA and the linearized plasmid (dsDNA). As controls, I transfected RNPs alone, -ssDNA without RNPs and I mock transfected cells. 3 days post transfection, the frequency of GFP positive cell did not increase. Nevertheless, genomic DNA was purified and a PCR performed with the primers indicated in Figure 28D to check for a knock-in. A significant fragment at around 500 bp was detectable for cells provided with linearized HDR templates (Figure 28F). A faint signal was detectable for the template directed repair with the circular plasmid. The controls were all negative as expected. The PCR results demonstrated that the knock-in of ER-LBD was possible with all DNA configurations tested. I conclude that linearized DNA (-ssDNA, +ssDNA, dsDNA) was better for a homology directed repair because the PCR signals were stronger than the one for the circular plasmid. Judging by conventional endpoint PCR and agarose gel electrophoresis, there were no differences between the different linearized DNA configurations.

Based on the FACS and PCR data, it was possible to knock-in the ER-LBD into the EBF1 locus in Nalm6 cells and GM12878. For Nalm6 cells, it also seemed to be feasible to used GFP as a reporter gene to easily monitor cells with a successful homologous recombination. However, further analysis should be performed to verify that the GFP signal correlates with knock-in efficiency. Additionally, optimization is required to improve the knock-in efficiency in GM12878 cells.

3 Discussion

The viral transcription factor EBNA2 plays a key role in establishing and maintaining the latency III transcription program, which is critical for the phenotype and proliferation of EBV infected B cells. The aim of this thesis was to study the role of the EBNA2-EBF1 protein complex during the latency III transcription program in primary infected B cells and in long-term EBV infected B cell cultures. To this end, I first identified the α1-helix in the END domain of EBNA2 as the mediator of the interaction with EBF1. Subsequently, I generated a mutant EBV, which lacks the α1-helix in EBNA2, I infected primary B cells with the virus and analyzed cellular processes and RNA expression early after infection. In addition, I analyzed the protein and mRNA expression as well as chromatin binding of EBNA2/EBNA2Δα1 and EBF1 in long-term EBV infected B cell cultures. Lastly, I aimed to generate a conditional EBF1 LCL system to further study the contribution of EBF1 to the latency III phenotype. I explored two approaches, which both applied the CRISPR/Cas9 technology for genome editing.

3.1 Characterization of the interaction interface of the EBNA2-EBF1 complex

Information about the structure of full length EBNA2 is scarce but the structure of the END domain has been studied thoroughly and has been resolved in solution by NMR (Friberg et al., 2015). The END domain consists of amino acids 1-58 and is able to form homodimers. The function of different residues of the END domain have been studied and I focused on histidine (H15) and the α1-helix, which are important for the transactivation capacity of EBNA2. GST-pull down assays with the END domain and co-immunoprecipitation (Co-IP) experiments with full length EBNA2 both demonstrated that EBF1 complex formation is lost upon deletion of the α1-helix (Figure 5C, D). Thus, the α1-helix of the END domain is a crucial region for the EBNA2-EBF1 complex formation. Since Friberg et al already confirmed that ENDΔα1 and EBNA2Δα1 can form dimers, we can exclude that the loss of complex formation detected in the GST-pull down assays and Co-IP experiment was due to misfolding of the protein. In the same study, the deletion of the α1-helix within the END domain caused a loss of the transactivation capacity of EBNA2Δα1 (Friberg et al., 2015). Having identified EBF1 as a

binding partner to this region, it might play an important role for the function of EBNA2. Studies have demonstrated that EBNA2 and EBF1 frequently occupy the same chromatin regions and EBF1 motifs are enriched in EBNA2 binding sites (Lu et al., 2016; Zhao et al., 2011). EBF1 is known to access and to change the epigenetic signature, thus, enabling transcriptional activation (Boller et al., 2016a; Maier et al., 2004). Therefore, the transactivation capacity of the END domain might be conferred by the interaction with EBF1.

To identify the region within EBF1 that mediates the interaction with EBNA2, I generated several deletion mutants of EBF1 and tested their ability to interact with EBNA2 in Co-IP experiments (Figure 6). The strategy for the mutagenesis was based on the modular published structure of EBF1 that was resolved by two independent groups (Siponen et al., 2010; Treiber et al., 2010a). My Co-IP experiments demonstrated that the C-terminal transactivation domain (TAD) was dispensable for the interaction with EBNA2 but further information concerning the binding of EBNA2 to other regions in EBF1 could not be concluded. Up to date, not many interaction partners of EBF1 have been identified. An interaction between EBF1 and CNOT3, a subunit of the CCR4-NOT complex, has been shown in pro-B cells (Yang et al., 2016). CCR4-NOT is a multi-subunit complex, which is involved in various steps of RNA metabolism (Miller and Reese, 2012). Other interactions have been demonstrated for the transcription factor (TF) MEF2C and the dioxygenase TET2 by Co-IP experiments (Guilhamon et al., 2013; Kong et al., 2016). TET2 is involved in demethylation of DNA. EBF1 antagonists ZNF423 and ZNF521 also form protein complexes with EBF1 (Kang et al., 2012; Tsai and Reed, 1997). The interactions of EBF1 with CNOT3, TET2 or ZNF521 were identified on endogenous levels of both interaction partners, whereas the other interaction partners of EBF1 were identified by in vitro assays or upon overexpression of the proteins. Additionally, of the above mentioned studies, only the one identifying the interaction with CNOT3 mapped the specific interaction site in EBF1. This might indicate that an interaction of EBF1 with other proteins can be difficult to detect or to map. Interestingly, another study investigating the interaction of EBF1 and p300/CBP, has tried to map the interaction region within EBF1 (Zhao et al., 2003). Here,

p300/CBP was binding to the N- and C-terminal half of EBF1, indicating that binding of p300/CBP to EBF1 might not be restricted to specific domains of EBF1. In contrast to Zhao et al where p300/CBP could bind to either the N- or C-terminal half of EBF1, EBNA2 might require the presence of the DBD, IPT as well as HLH to interact with EBF1. The experiments performed in this thesis did not provide information about direct or indirect interaction of EBNA2 and EBF1. Affinity capture experiments with recombinant purified EBF1 and EBNA2 proteins were not informative due to high background signals.

In conclusion, the α1-helix of EBNA2 is crucial for the protein complex formation with EBF1. Biophysical methods like nuclear magnetic resonance spectroscopy could provide more information about binding regions within EBF1. Binding kinetics of the interaction partners could be determined by isothermal titration calorimetry. Additionally, interaction assays with purified recombinant EBNA2 and EBF1 would help to know if the interaction is direct or requires additional mediators like DNA or other proteins.

3.2 Functional analysis of the EBNA2-EBF1 complex in EBV driven B cell transformation

3.2.1 Generation of EBVΔα1

Recombineering based EBV mutagenesis allows the targeted mutation of EBV in *E.coli* facilitated by the application of EBV bacterial artificial chromosomes (EBV-BAC). These contain the EBV genome, an F-factor, a gene encoding for GFP and hygromycin/puromycin resistance (Delecluse et al., 1998). Transfecting the EBV-BACs into HEK293 cells and activating the virus production leads to shedding of the virus into cell culture medium, which can be used for infection studies. This approach enables researchers to study specific mutations of EBV in the context of the whole viral genome. Based on the identification of the α1-helix as a crucial region for the interaction with EBF1, I generated an EBV mutant that lacks this region (EBVΔα1, Figure 7). As demonstrated in Raji infection experiments, the deletion of the α1-helix did not affect EBV production by HEK293 producer cells and EBVwt and EBVΔα1

seemed to be equally good at infecting Raji cells (Figure 8). Both viruses were used in subsequent infection experiments to study the role and contribution of the EBNA2-EBF1 interaction to the latency III transcription program in primary infected B cells and long-term lymphoblastoid cell lines (LCL).

3.2.2 Proliferation defect and cell cycle arrest in EBVΔα1 infected B cells

With the newly generated virus, it was possible to perform infection experiments with primary B cells isolated from several adenoid samples. The analysis of these cells revealed that EBVΔα1 infected B cells exhibited impaired proliferation and a cell cycle arrest. The results of different experiments reinforce the observation and will be discussed in the following.

Cell cycle arrest detected by flow cytometric analysis

The expansion of cell populations can be measured by the MTT assay and is an indicator of cell viability and proliferation. The population of primary B cells infected with EBVwt expanded within 8 days post infection whereas EBVΔα1 infected B cells did not as demonstrated by MTT. The more detailed cell cycle analysis by flow cytometry of cells stained with BrdU and 7-AAD demonstrated that resting B cells infected with EBVwt entered the cell cycle between day 2 and 4 post infection as indicated by an increased population of cells in the advanced S and G2/M phase. This result confirmed a previous study that analyzed the cell cycle distribution upon EBV infection (Mrozek-Gorska et al., 2019). In contrast, EBVΔα1 infected B cells were arrested in the early S phase. The frequency of B cells in the early S phase did not change significantly at any time point of analysis. Additionally, there was no substantial increase of the advanced S and G2/M phase during the analysis. Thus, EBVΔα1 infected B cells were able to enter the cell cycle but were arrested between day 2 and 4 post infection (Figure 9).

Differential expression of cell cycle genes

During the G1 phase of the cell cycle, CDK4/6 and cyclin D family members form complexes, which activates CDK4/6 and as a consequence leads to the hyperphosphorylation of retinoblastoma (RB) protein. The hyperphosphorylation of RB triggers the release of E2F

transcription factors that subsequently activate the expression of target genes necessary for the G1/S as well as G2/M cell cycle progression (Baker and Reddy, 2012). Among these are cyclin E, A and B, CDK1 and 2 (Bracken et al., 2004). Upon EBV infection, an upregulation of cyclin D2, CDK4 and CDK6 proteins can be detected within 48 and 72 hours (Spender et al., 1999). In line with this, was the observation of increasing RNA expression levels of CDK4/6 and cyclin D2/3 in EBVwt and EBVΔα1 infected B cells between day 2 and 4 post infection (Figure 18). The temporal pattern of gene expression was also in accordance with the observed cell cycle entry identified by FACS analysis (Figure 9). However, the expression of CDK4 and cyclin D2/3 was significantly reduced in EBVΔα1 infected B cells compared to EBVwt infected B cells, which could cause decreased formation of CDK4/6-cyclin D complexes. Subsequently, this might lead to a reduced E2F activation and target gene expression necessary for the G1/S and G2/M progression. Indeed, the expression of cyclin A2, CDK1 and cyclin B1 was significantly reduced in EBVΔα1 infected B cells on day 3 and 4 post infection. Thus, a reduced CDK4/6-cyclin D complex formation might contribute to the cell cycle arrest at the early S phase in EBVΔα1 infected B cells.

Gene set enrichment analysis

In addition to the expression of selected cell cycle genes, the gene set enrichment analysis (GSEA) detected a significant (FDR <0.1) enrichment of the 'E2F targets' set (Figure 22) and the 'G2M checkpoint' set (Figure 23). For both gene sets, a reduced level of gene expression was detected in EBV Δα1 infected B cells although not all genes displayed reduced expression levels. As already mentioned, the activation of E2F targets genes is necessary for the cell cycle progression to the S phase and following phases (Baker and Reddy, 2012). An impaired activation of E2F target genes could, therefore, have a negative effect on the cell cycle progression. The 'G2M checkpoint' set contains genes that are involved in the cell cycle progression no only at the G2/M checkpoint but also at other cell cycle checkpoints (https://www.gsea-msigdb.org). Thus, a dysregulation of genes involved in different cell cycle checkpoints could further contribute to the cell cycle arrest in EBVΔα1 infected B cells. The

GSEA demonstrated that genes important for the cell cycle progression at later time points were induced in EBVΔα1 infected B cells but the induction of gene expression might not be sufficient to complete the cell cycle.

Taken together, these data demonstrate that EBVΔα1 infected B cells have a defective cell cycle resulting in the cell cycle arrest in the early S phase. The expression of genes important for the progression of the cell cycle is induced but not to the same level as in EBVwt infected B cells. To point out single events causing the cell cycle arrest, is difficult due to the complex regulation of the cell cycle. Nevertheless, I would like to discuss three observations, which could cause or contribute to the initiation of the cell cycle arrest.

Potential causes of the cell cycle arrest

An important cellular target gene of EBNA2 is **MYC**, which is induced within few hours after EBNA2 activation (Kaiser et al., 1999). RNA sequencing revealed that MYC expression is induced in EBVwt and EBVΔα1 infected B cells but the induction is severely reduced in the latter infection. Importantly, reduced MYC expression was probably not due to reduced EBNA2 expression since EBNA2 expression levels were similar in both infection conditions (Figure 17). Forced expression of MYC can induce the cell cycle entry of quiescent cells and the protein is involved in the regulation of the cell cycle at various stages (Bretones et al., 2015). Therefore, reduced MYC expression level most likely contributes or even causes a block in the cell cycle progression. In addition, CDK4 is a direct target gene of MYC (Hermeking et al., 2000). Thus, it is possible that reduced MYC expression could cause the drop in CDK4 transcript levels in EBVΔα1 infected B cells on day 2 post infection (Figure 18), which subsequently could lead to reduced expression of E2F target genes. Additionally, it has been demonstrated that strong expression of MYC renders the proliferation of EBV positive B cells independent of EBNA2 and LMP1 expression (Polack et al., 1996) adding to the importance of sufficient MYC expression for the proliferation of EBV infected B cells. Therefore, it is likely that the proliferation defect is caused by reduced MYC expression level in EBVΔα1 infected B cells.

In addition to MYC, there was a strong increase of **p53** in EBVΔα1 infected B cells on day 2. p53 is a TF that is activated upon stress stimuli, e.g. hypoxia, DNA damage, oncogene activation and can induce cell cycle arrest and apoptosis (Vousden and Lane, 2007). Activation of p53 can induce a downregulation of genes that would be necessary for the cell cycle progression, e.g. CDK2, CDK1 or cyclin A2 (Engeland, 2018). Thus, it is possible that EBVΔα1 infected B cells undergo a p53-mediated cell cycle arrest due to the strong upregulation of p53 on day 2. Since the peak expression of p53 was detected on day 2, it precedes the cell cycle arrest taking place between day 2 and 4 post infection. Hence, a p53-mediated cell cycle arrest seems plausible. An important mediator of p53-induced cell cycle arrest is **p21** (CDKN1A). Its expression is induced by p53 and p21 can inhibit CDK2 and CDK1, thereby arresting the cells at the G1 or G2 phase (Chen, 2016; Mansilla et al., 2020). However, the activation of p21 in EBVΔα1 infected B cells seems to be independent of p53. The peak expression of p21 was detected on day 1 post infection, thus, preceding the upregulation of p53 on day 2 post infection.

Unfortunately, RNA sequencing did not provide us with information about the expression of viral genes other than EBNA2. The expression and concerted action of viral latent genes is important for EBV to efficiently transform infected B cells, and their interaction and interference with regulators of the cell cycle drives the cell cycle progression (Yin et al., 2019). Since EBNA2 initiates the expression of all viral latent genes in latency III (Kempkes and Ling, 2015), it would be interesting to see if the expression of viral latent genes is dysregulated in EBVΔα1 infected B cells and if this might also contribute to the cell cycle arrest. Further analysis of the RNA sequencing data in combination with target gene validation experiments could provide more information and insight into the mechanism behind the cell cycle arrest.

3.2.3 Growth rescue of EBVΔα1 infected B cells and LCLΔα1 establishment

Interestingly, the cell cycle arrest of EBVΔα1 infected B cells was similar to observations of a study investigating the effects of LMP1 mutations on the transformation efficiency of EBV (Dirmeier et al., 2003, 2005). Here, B cells infected with specific LMP1 mutants could only establish long-term cultures in the presence of irradiated fibroblast feeder cells. Another study

demonstrated that long-term cultures of uninfected primary B cells can be generated by stimulation of the CD40 receptor and IL-4 supplementation. CD40 stimulation without IL-4 induces proliferation of uninfected B cells but long-term cultures do not survive (Banchereau et al., 1990). Based on these two studies, we decided to culture the EBVΔα1 infected B cells on CD40L expressing feeder cells hoping that the proliferation defect can be overcome. Indeed, EBVΔα1 infected B cells overcame the cell cycle block (Figure 10) and it was possible to generate long-term proliferating cultures. The mRNA and protein expression of established LCLs revealed that LCLΔα1 expressed EBNA2, EBNA3A, B and C equally well or even stronger compared to LCLwt. Interestingly, LMP1 and MYC expression level (mRNA and protein) were severely reduced in LCLΔα1. Since the removal of CD40L expressing feeder cells caused a gradual halt of proliferation of LCLΔα1, we assumed that this was due to the low expression of LMP1 and MYC. Although MYC is a direct target of EBNA2 (Kaiser et al., 1999), its expression is also regulated by the activity of LMP1 (Dirmeier et al., 2005). Therefore, we wanted to test if an overexpression of LMP1 can reverse the proliferation block. However, overexpression of LMP1 in LCLΔα1 cultured without CD40L feeder cells did not increase the proliferation. In addition, preliminary experiments (data not shown) indicate that LCLΔα1 require the CD40L stimulation. Thus, LMP1 overexpression is not sufficient to overcome the proliferation defect and additional stimuli are required for the long-term cultivation of LCLΔα1. Since LMP1 and CD40 can activate the same signaling pathways in B cells (Lam and Sugden, 2003), the mechanisms underlying the CD40L driven proliferation has to be analyzed in more detail in the future to better understand the proliferation block in LCLΔα1. Nevertheless, the established LCLwt and LCLΔα1 could be used for subsequent analysis of chromatin binding behavior.

3.2.4 Three classes of EBNA2 chromatin binding sites were identified

Since EBNA2 cannot bind directly to DNA, it requires the interaction with a cellular DNA anchor. CBF1 is the main DNA anchor but EBF1 has been demonstrated to facilitate EBNA2 binding at CBF1 independent binding sites (Glaser et al., 2017). Performing chromatin

immunoprecipitation assays followed by qPCR (ChIP-qPCR), we can assess if EBNA2 and EBF1 bind equally well to chromatin sites in the presence (LCLwt) and absence (LCLΔα1) of complex formation. I focused the investigation of the binding on specific chromatin regions, which are known to bind either of the proteins or both (Figure 12, Figure 13). Based on the ChIP-qPCR results, I was able to infer three classes of EBNA2 binding sites.

The first class was termed **'EBF1 independent'** EBNA2 binding sites and included the promotor of cellular HES1, viral LMP2A and the viral C promotor (Cp) (Figure 29). EBNA2wt and EBNA2Δα1 signals were equally strong at these binding sites. Although EBF1 binding was detected at the LMP2A promotor and Cp with similar signals in presence of EBNA2wt or EBNA2Δα1, it seems that EBNA2 does not require the interaction with EBF1 to bind to these chromatin regions. The LMP2A promotor and Cp contain high-affinity binding sites for CBF1 (Kempkes and Ling, 2015), which indicates that EBNA2 binding might be rather dependent on the interaction with CBF1 than EBF1. *HES1* is a classical Notch target gene but the expression is also activated by EBNA2 in LCLs (Kohlhof et al., 2009). In pre-pro B cells, EBF1 can also activate HES1 expression as demonstrated by gain-of-function experiments (Treiber et al., 2010b) but it is unknown if EBF1 can also regulate HES1 expression in LCLs. Although the HES1 site tested in this thesis contains an EBF1 motif (Lu et al., 2016), binding of EBF1 was not detected in LCLwt or LCLΔα1. This further supports the notion that EBNA2 binding to this chromatin region is independent of EBF1 and most likely depends on the CBF1 interaction.

The class 2 binding sites were termed **'EBF1 dependent'** EBNA2 binding sites and included the cellular EBV super enhancer 1 (ESE1) regulating MYC expression, the cellular CD79a promotor and the viral LMP1 promotor (Figure 29A). EBF1 binding at these sites was similar in the presence of EBNA2wt or EBNA2Δα1 while EBNA2Δα1 binding was significantly reduced (Figure 13). This indicates that EBNA2 requires the interaction of EBF1 to bind properly to these sites. ESE1 was defined as an enhancer strongly bound by EBNA2 and EBNA3 proteins as well as NFκB subunits and B cell specific TFs like EBF1, PU.1 or Pax5 (Zhou et al., 2015). The CD79a promotor is strongly regulated by EBF1 (Hagman et al., 1991) but whether EBNA2

is also involved in the regulation of CD79a is unclear. In LCLs, the LMP1 expression is regulated by EBNA2 and the promotor contains a CBF1 binding motif (Kempkes and Ling, 2015) but also an EBF1 binding motif (Murata et al., 2016). In addition, overexpression of EBF1 can increase the LMP1 expression in LCLs (Murata et al., 2016). These data support the idea that EBNA2 binding to these sites requires the interaction with EBF1. Since EBNA2Δα1 binding was not completely lost at these sites, it is possible that EBF1 enhances or stabilizes EBNA2 binding. It is also possible but cannot be inferred from the experiments that EBNA2 needs CBF1 as an additional factor for these binding sites. Two studies looking at the binding behavior of EBNA2 and EBF1 found reduced binding of both factors upon CBF1 knock-down or in CBF1 knock-out cell lines (Glaser et al., 2017; Lu et al., 2016). Therefore, it is possible that EBNA2, EBF1 and CBF1 form a trimeric complex and that EBF1 stabilizes EBNA2 binding at the class 2 sites. Importantly, a study detected significantly reduced EBF1 binding at the HES1, LMP1 and LMP2A sites in the absence of EBNA2wt concluding that EBF1 binding is dependent on the presence of EBNA2 at these sites (Lu et al., 2016). I could not confirm this finding. First, I did not detect any EBF1 signal at the HES1 site despite using the same primers but different cell lines. Secondly, EBF1 binding to the LMP1 and LMP2A promotors in my experiments was not significantly different in the presence of EBNA2wt or EBNA2Δα1. Thus, I conclude that EBF1 binds independently of EBNA2 to the class 2 sites (Figure 29B).

The third class of EBNA2 binding sites could be called **'EBNA2-EBF1 dependent'** sites (Figure 29C). The EBV super enhancer 2 (ESE2) controlling MYC expression could be an example of this class. EBF1 as well as EBNA2Δα1 signals were significantly reduced in LCLΔα1. This indicates that EBNA2 as well as EBF1 might require the protein complex formation in order to bind to this enhancer. Moreover, binding of EBNA2 and EBF1 to ESE2 could also be influence by additional factors since many cellular TFs and the EBNA3 proteins bind to this enhancer (Zhou et al., 2015).

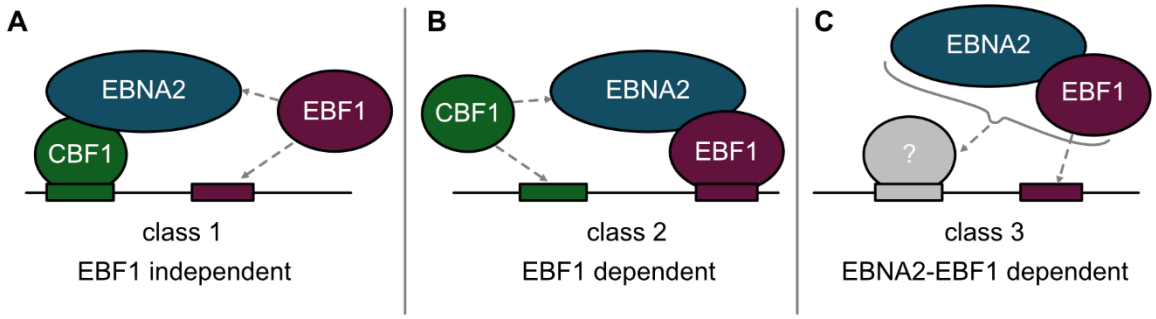


Figure 29 Potential classes of EBNA2 chromatin binding sites. A) EBF1 independent EBNA2 binding sites might require the interaction of EBNA2 with CBF1. EBF1 might also bind to chromatin at the same site but might not be required for chromatin binding of EBNA2 (indicated by dashed arrows). B) At EBF1 dependent EBNA2 binding sites, EBNA2 requires the EBNA2-EBF1 complex formation in order to bind. CBF1 might also bind to the complex but might be dispensable for stable chromatin binding by EBNA2 (indicated by dashed arrows). C) At EBNA2-EBF1 dependent sites, EBNA2 as well as EBF1 require protein complex formation in order to access chromatin. Additional cellular factors (grey) might enhance or support binding (indicated by gray arrows).

3.2.5 EBVΔα1 cannot efficiently reprogram B cell gene expression

Upon EBV infection, the gene expression of B cells is strongly regulated and reprogrammed in order to activate resting B cells and to induce the transformation process (Thorley-Lawson, 2015). The differential gene expression (DGE) analysis comparing EBVwt and EBVΔα1 infected B cells demonstrated that in both infection conditions a global gene regulation was induced but the extend of the regulation was reduced in EBVΔα1 infected B cells. Importantly, the global gene regulation was similar on day 1 post infection but differed strongly between EBVwt and EBVΔα1 infected B cells in the following days (Figure 19). The gene set enrichment analysis (GSEA) supported the findings from the DGE analysis and I will discuss five examples to highlight the gene regulation on different days post infection.

On day 1 post infection, an enrichment of the **'Unfolded protein response'** set was detected and genes of this set were upregulated in both EBVwt and EBVΔα1 infected B cells. The big overlap of genes contributing to the leading edge subset indicates that both viruses could induce the same genes equally well (Figure 20). The unfolded protein response (UPR) is a stress-response of the endoplasmatic reticulum (ER) to an accumulation of unfolded or misfolded proteins caused by an increase of protein translation. Induction of the UPR can pause protein translation but eventually aims to increase the folding capacity of the ER (Johnston and McCormick, 2019).

In the context of viral infection, the UPR is often induced upon lytic replication of viruses since it causes increased protein synthesis for virus production. XBP1 and eIF2α (EIF2S1) are, for example, mediators of the UPR. Splicing of the XBP1 mRNA generates an isoform that is able to induce the expression of chaperons amongst other genes. Phosphorylation of eIF2α, a subunit of the translation initiation factor 2, leads to reduced protein translation and subsequently less proteins entering the ER (Ron and Walter, 2007). If the enrichment of this gene set truly resembles an induced UPR by EBVwt and EBVΔα1 is not clear because the GSEA data is based on RNA expression. We do not know if the XBP1 RNA was spliced or if eIF2α was phosphorylated in the cells, which would be an indicator of UPR. Irrespective of the potential function of this gene set, we could consider the enrichment of the UPR gene set in both infections as an indicator of similar gene regulation on day 1 post infection.

The other gene sets I analyzed more closely were **'Myc target genes V1'** on day 2 post infection (Figure 21), **'E2F targets'** on day 3 post infection (Figure 22) and **'G2M checkpoint'** (Figure 23) and **'Glycolysis'** (Figure 24) on day 4 post infection. For all these gene sets, I observed the same pattern of gene regulation and enrichment. On the respective days, genes of each set were upregulated in both EBVwt and EBVΔα1 infected B cells but genes were stronger expressed in the EBVwt condition, which points towards an impaired gene regulation in EBVΔα1 infected B cells. The identification of the 'Myc target genes V1' as significantly enriched is in accordance with the fact that MYC expression was induced in EBVΔα1 infected B cells but the expression level was dramatically reduced already on day 1 post infection. Naturally, reduced MYC expression causes a reduced expression of direct and indirect MYC target genes, which can have dramatic secondary effects causing proliferation defects and cell cycle arrest. In line with this is also the significant enrichment of 'E2F targets' on day 3 and 'G2M checkpoint' on day 4 post infection with a reduced expression of genes in EBVΔα1 infected B cells. Genes of both sets are involved in the progression and control of the cell cycle. The reduced gene expression of these set probably contributes to the cell cycle arrest observed in EBVΔα1 infected B cells.

The identification of **'Glycolysis'** as a significantly enriched gene set is also interesting. It is known that EBV infection causes an upregulation of glycolytic genes and an increased energy demand indicated by an increased uptake of glucose (McFadden et al., 2016; Mrozek-Gorska et al., 2019; Sommermann et al., 2011). Additionally, the upregulation of these genes has been detected as early as 2 days post infection (Wang et al., 2019a). The GSEA for the 'Glycolysis' gene set in accordance with published data since an upregulation of genes was detected in EBVwt infected B cells on day 2 post infection. A significant enrichment for the EBVΔα1 vs uninfected comparison was only detectable on day 4 post infection supporting the idea that EBVΔα1 induces gene expression to a lesser extent or even at a later time point than EBVwt. Preliminary results of a cluster analysis also showed that genes involved in the non-coding and ribosomal RNA metabolism were strongly upregulated on day 1 post infection followed by a sudden drop of expression on day 2 in the EBVΔα1 infection. In contrast, the expression of these genes did not decrease after the initial induction in the EBVwt infection. Since ribosomal RNAs are important for the assembly of ribosomes and protein synthesis, an impaired expression might cause reduced protein synthesis, which could affect a growth in EBVΔα1 infected B cells.

Taken together, these data show nicely that both viruses can initiate gene regulation including up- and downregulation of genes. However, the gene regulation seems to be impaired in EBVΔα1 infected cells due to reduced expression level. Important cellular processes like the cell cycle and glycolysis are affected by this impairment and might contribute to the observed proliferation defect in primary infected B cells.

3.2.6 What is the contribution of EBF1 to EBNA2 target gene activation?

EBNA2 together with EBNA-LP are the first viral genes expressed after an EBV infection of B cells (Kempkes et al., 1995). The expression of both proteins is important for the cell cycle entry of resting B cells. Additionally, EBNA2 induces the expression of all viral latent genes in latency III and their concerted activity is required for an efficient transformation of B cells (Kempkes and Ling, 2015). Studies with EBNA2 negative EBV have demonstrated that these viruses

cannot transform B cells (Cohen et al., 1989; Hammerschmidt and Sugden, 1989). Therefore, a fully functional and active EBNA2 is necessary to initiate the transformation process.

In the present thesis, I could identify the α1-helix within the END domain as a crucial region for the interaction with EBF1. Deleting this region in EBV caused an impaired transformation of EBVΔα1 infected B cells identified by inefficient gene regulation that affects important cellular processes, e.g. cell cycle, growth, energy metabolism. EBF1 has been identified as an interaction partner and co-factor that enhances EBNA2 access to chromatin (Glaser et al., 2017). The molecular mechanism, however, underlying the contribution of EBF1 to the EBNA2 activity is not fully understood. Based on two specific examples and recent findings, I would like to suggest two potential modes of action.

Regulation of MYC expression

As already discussed, MYC is an important cellular target gene of EBNA2 (Kaiser et al., 1999). Activation of MYC leads to extensive cell growth and hyperproliferation of infected B cells. Two so called EBV super-enhancers (ESE) have been identified to regulate MYC expression and are bound by EBNA2, EBF1 and other cellular and viral proteins. They are located approximately 500 kb (ESE1) and 400 kb (ESE2) upstream of the transcriptional start site of MYC (Zhou et al., 2015). The ChIP experiments in this thesis revealed that EBNA2Δα1 binding was reduced at the ESE1 despite normal binding of EBF1. EBNA2Δα1 and EBF1 binding was reduced at the ESE2. These two results indicate that upon loss of EBNA2-EBF1 complex formation the binding of one or both partners to chromatin can be affected, which could ultimately cause inefficient gene activation. Reduced MYC expression was demonstrated early after infection but also in long-term LCLΔα1 cultures.

EBF1 is known to be able to access closed chromatin (Boller et al., 2016b) and change the epigenetic landscape to allow opening of chromatin and transcriptional activation (Li et al., 2018). It is doing that by interacting, for example, with Brg1, a subunit of the chromatin remodeling complex SWI/SNF (Gao et al., 2009). The CTD of EBF1 recruits Brg1 to chromatin and can induce liquid-liquid phase separation in vitro, which indicates that EBF1 forms

condensates. Thereby, it can facilitate chromatin remodeling of various regions at the same time by interacting with Brg1 (Wang et al., 2020). EBNA2 has also been demonstrated to interact with components of the SWI/SNF complex (Wu et al., 2000). It is possible that EBF1 recruits Brg1 to chromatin thus facilitating chromatin remodeling. Due to condensate formation, the EBV super-enhancers (ESE) are brought into close proximity to the MYC promotor. Consequently, EBNA2 would benefit from the chromatin remodeling abilities by interacting with EBF1, accessing open chromatin and thereby transactivating MYC expression. Additionally, EBF1 induced condensate formation might create a hub of permissive chromatin with additional associated activator proteins that subsequently facilitate EBNA2 induced transactivation of additional target genes. Conversely, inhibiting the EBNA2-EBF1 interaction by deleting α1-helix would then diminish EBNA2 access to open chromatin and MYC activation.

Regulation of LMP1 expression

Another example of EBNA2-EBF1 cooperative actions is the LMP1 promotor activation. During latency III, LMP1 expression is induced from the proximal ED-L1 promotor and regulated by EBNA2 (Murata et al., 2016). EBNA2 has been shown to associate with CBF1 and PU.1 to regulate LMP1 expression. Additionally, an EBF1 binding site in ED-L1 promotor has been identified and overexpression of EBF1 can cause an upregulation of LMP1 expression (Murata et al., 2016). The ChIP experiments performed in this thesis demonstrated that EBNA2Δα1 binding was reduced at the LMP1 promotor, whereas EBF1 binding was not affected by the loss of EBNA2 interaction. In contrast to the study by Murata et al, LMP1 expression was reduced in LCLΔα1 despite increased EBF1 expression. This together with the ChIP result might indicate that although EBF1 binds the promotor, EBNA2 requires the interaction with EBF1 to transactivate LMP1 expression. However, due to the loss of the EBNA2-EBF1 interaction, the access of EBNA2Δα1 to the promotor might be impaired resulting in reduced LMP1 expression in LCLΔα1. A recent study showed that EBF1 is enhancing the interaction of Pax5 with KTM2A, a catalytic subunit of the MLL complex that

is a histone methyltransferase complex associated with epigenetic remodeling and transcriptional activation (Bullerwell et al., 2021). Hypothetically, if EBF1 exhibits a similar function with EBNA2 and the MLL complex, it could allow epigenetic remodeling resulting in transcriptional activation of the LMP1 promotor. Nevertheless, the mechanism behind the interaction of Pax5 and MLL enhanced by EBF1 is not known.

Based on the experiments, we do not know whether the phenotype observed in EBVΔα1 infected B cells is only due to the loss of the interaction of EBNA2 and EBF1. It could be that additional cellular factors bind to the α1-helix in EBNA2 and thereby contribute to transcriptional regulation by EBNA2. Therefore, a conditional EBF1 cell system would be a valuable tool to further investigate the contribution of EBF1 to the characteristic features of EBVΔα1 infected B cells.

3.3 Exploring two approaches to generate conditional EBF1 cell lines

3.3.1 Two CRISPR/Cas9 based approaches can be used to generate conditional cell lines

In the last part of my thesis, I wanted to create conditional EBF1 LCLs, which would be useful to further investigate the contribution of EBF1 to the latency III transcription program and LCL phenotype. I wanted to establish the conditional EBF1 expression in the GM12878 cell line, which is an LCL used in various molecular biological studies (Jiang et al., 2018). To this end, I explored two approaches using the CRISPR/Cas9 technology for genome editing. The first approach aimed at expressing EBF1 from a doxycycline inducible expression plasmid stably transfected into GM12878. The stable transfection was followed by a CRISPR/Cas9 mediated knock-out of endogenous EBF1. In this system, the EBF1 expression is activated by doxycycline in the cell culture medium. The second approach aimed at the CRISPR/Cas9 based knock-in of the ligand binding domain of the estrogen receptor (ER-LBD) into the endogenous EBF1 locus. In this system, the EBF1 fusion gene is under the control of the endogenous promotor but the activity of the protein depends on the presence of β-estradiol in the cell culture medium. In

both cases, successful genome editing would be followed by single cell cloning to establish the respective cell lines.

I have demonstrated that both approaches are technically possible but both require additional optimization. We have established stably transfected GM12878 carrying a doxycycline inducible EBF1 expression plasmid and I was able to knockout exon 3 of EBF1 by using two gRNAs in combination with the Cas9 protein (Figure 25). Following the knock-out of exon 3, single cell clones would have to be generated and screened for bi-allelic deletion of exon 3 before generating cell lines. For the β-estradiol sensitive system, I successfully knocked-in the ER-LBD construct into the EBF1 locus of Nalm6 cells as a proof-of-concept. The knock-in of the construct also worked in GM12878 cells but a systematic optimization has to be carried out to increase the frequency of GFP positive cells after transfection of the RNPs and the HDR template (Figure 28). GFP expression is an indicator of successful insertion of the ER-LBD into the EBF1 locus and therefore correlates with the HDR efficiency. Additionally, using GFP as a reporter gene helps to enrich cells with a successful insertion and to screen for bi-allelic insertion. Optimizations that need to be carried out include the transfection conditions in order to increase the uptake of the HDR template and RNPs. Another important factor to increase the HDR efficiency is the optimization of the HDR template. For example, the length of the homology arms and as well as the DNA configuration can strongly affect the HDR efficiency (Chu et al., 2015; Roth et al., 2018). The CRISPR/Cas9 system has been studied extensively and additional optimizations for a successful knock-in could include the position of the gRNAs (Paquet et al., 2016), interference with the non-homologous end joining (NHEJ) repair pathway (DSB) (Liu et al., 2019) or modification of the Cas9 protein to increase the HDR efficiency (Ling et al., 2020).

3.3.2 The advantages and disadvantages of the two approaches

Both, the doxycycline and β-estradiol sensitive system, have their advantages and disadvantages and the choice for one depends on the different factors. The advantage of the first approach with doxycycline responsive plasmid is the relatively easy disruption of the open

reading frame (ORF) of EBF1 by knocking out exon 3. Here, the double strand break (DSB) induced by Cas9 is repaired by NHEJ, a pathway which is favored over the HDR by cells (Liu et al., 2019). Using two gRNAs to delete exon 3 of EBF1 can also increase the chance of a disrupted ORF. A cell system with the pRTR expression plasmid has the advantage that the level of expression of the gene of interest (GOI) can be regulated by adjusting the concentration of doxycycline in the medium (Bornkamm et al., 2005; Jackstadt et al., 2013). Therefore, the effect of different EBF1 concentrations in the cell can be examined if it is required for the experimental design. At the same time, it might be more difficult to adjust the expression to endogenous level because fine tuning with different doxycycline concentrations is not that sensitive. In my particular setting, the Tet on system of the pRTR is a disadvantage due to the importance of EBF1 for cell viability and the short half-life of doxycycline the cell culture, around 24 hours. In a Tet on system, binding of doxycycline to the tetracycline controlled transcriptional activator (tTA) enables binding of tTA to the promotor, which initiates the transcription of the GOI (Gossen et al., 1995). Since EBF1 is important for the viability of the cells (Lu et al., 2016), a continuous supply of doxycycline is required for a constant EBF1 expression. This makes the maintenance of the cell culture laborious and requires precise working to avoid fluctuating expression level. Therefore, a Tet off system, in which the expression of the GOI is repressed in the presence of doxycycline, would be better for the maintenance of the cultures. Additionally, we have observed a loss of inducibility of pRTR plasmid in other cell systems. However, since the plasmid also encodes for GFP and NGFR, enrichment of cells carrying the expression plasmid can be performed either by FACS or magnetic cell sorting and should not present a problem.

The biggest advantage of the β-estradiol-sensitive system is the expression of the EBF1 fusion gene at endogenous level. Here, the translocation of the fusion protein and activity can easily be regulated by the presence of β-estradiol in the cell culture medium. This facilitates the investigation of the function of EBF1 in an environment, which is closer to the biology of B cells as compared to the doxycycline-sensitive system. However, setting up such a conditional

EBF1 cell system can be quite difficult depending on the cell type and due to overall low HDR efficiency. Nevertheless, once the optimal conditions are established, monitoring of successful insertions is straightforward since it can be accomplished via GFP expression and by FACS analysis. In general, inserting ER-LBD into the genomic locus of EBF1 might negatively influence the function of the protein due to potential misfolding. The design of the fusion construct, however, was based on a study, which used an EBF1-ER fusion protein to investigate chromatin remodeling at the CD79a promotor (Gao et al., 2009). Chromatin remodeling is a fundamental function of EBF1 and, therefore, malfunctioning of the fusion protein can be excluded. The difference in their experimental approach is that the gene has been cloned onto a viral vector allowing the expression of the fusion protein by transducing the cells with the vector.

In conclusion, the choice of conditional cell systems depends on the experimental requirements, e.g. endogenous expression vs inducible expression, but also on the cell type as a knock-in can be difficult for some but fast and easy for other cell lines. Comparing the doxycycline- and β-estradiol systems, I think the EBF1-ER cell system is the more elegant. Although the establishment is a more complex process, the function of EBF1 can be analyzed in an environment, which is closer to the biology of B cells, and the maintenance of cell cultures is more feasible.

3.4 Outlook

As discussed, the deletion of the α1-helix, which is formed by 5 amino acids, had dramatic effects on the phenotype of EBVΔα1 infected B cells. There are strong evidences that the loss of EBNA2-EBF1 interaction, due to the deletion of the α1-helix, contributes to the phenotype of EBVΔα1 infected B cells, primary as well as LCLs. I observed dysregulated gene expression causing impaired transformation of B cells as well as altered chromatin binding of EBNA2 and EBF1. Following up the results I found during this thesis, a conditional EBF1 system would help to fully delineate the contribution of EBF1 and to better understand how EBNA2 uses EBF1 to transactivate target genes. In addition, EBF1 target genes are well defined in

developing B cells but little is known about target genes in mature B cells. A conditional EBF1 cell system would allow us to identify EBF1 target genes and mutual target genes with EBNA2 and to better understand how EBNA2 uses B cell specific TFs to initiate and maintain the latency III program. EBF1 has also been implicated in other malignancies like B cell acute lymphoblastic leukemia (B-ALL) (Somasundaram et al., 2015), Hodgkin's disease (Bohle et al., 2013) or gastric carcinoma (Xing et al., 2020). Since the approaches to establish conditional EBF1 cell lines presented here are not restricted to LCLs, they could also be applied to other cell lines in order to study the role of EBF1 in different malignancies of B cell and non-B cell origin.

4 Material and Methods

4.1 Antibodies

Table 1 Primary antibodies

Specificity	Host	Application	Supplier
α-EBF1 (C-8)	mouse, IgG2a	WB, ChIP	Santa Cruz, cat# sc-137065
α-EBF1	rabbit,	ChIP	Sigma Aldrich, cat# ABE1294
α-EBNA2 (R3)	rat, IgG2a	WB, IP, ChIP	HMGU Monoclonal Antibody Core Facility
α-EBNA2 (1E6)	rat, IgG2a	ChIP	HMGU Monoclonal Antibody Core Facility
α-EBNA3A (4A5-1111)	rat,	WB	HMGU Monoclonal Antibody Core Facility
α-EBNA3B (AN 6C9-1-1)	rat,	WB	HMGU Monoclonal Antibody Core Facility
α-EBNA3C (A10 P2-583)	rat,	WB	HMGU Monoclonal Antibody Core Facility
α- GST (6G9)	rat, IgG2a	WB, affinity capture, ChIP isotype control	HMGU Monoclonal Antibody Core Facility
α-Dog CD3	rat, IgG1	ChIP	HMGU Monoclonal Antibody Core Facility
α-HA (3F10)	rat, IgG1	WB, ChIP	HMGU Monoclonal Antibody Core Facility
α-GAPDH (0411)	mouse, IgG1	WB	Santa Cruz, cat# sc-47724
α-LMP1 (1G6)	rat,	WB	HMGU Monoclonal Antibody Core Facility
α-LMP2A (TP14 B6)	rat, IgG2a	WB	HMGU Monoclonal Antibody Core Facility
α-myc (9E10)	mouse,	WB	HMGU Monoclonal Antibody Core Facility
α-IgD-FITC	mouse, IgG2a	FACS sorting	BD Biosciences, USA
α-CD38-PE-Cy7	mose, IgG1κ	FACS sorting	eBioscience, cat #25-0389- 42
rabbit IgG		ChIP	Merck cat# PP64

Table 2 Secondary antibodies

Specificity	Host	Application	Supplier
α-mouse-IgGκ BP-HRP		WB	Santa Cruz (sc-516102)
α-rat-IgG-HRP	goat	WB	Santa Cruz (sc-2006)
α-mouse-IgG-HRP	horse	WB	Cell Signaling Technologies(7076S)

4.2 Bacteria

Table 3 Bacterial strains

E.coli strain	Genotype	Application
DH5α	*F-endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 Φ80dlacZΔM15 Δ(lacZYA-argF) U169, hsdR17(rK- mK+), λ–*	general cloning and plasmid storage
SW105	DY380 *(cro-bioA)*<>araC-PBAD *Flpe gal+ ΔgalK*	EBV recombination
Rosetta2	*F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE2 (CamR)*	protein expression
BL21	*F- ompT lon hsdSB(rB- mB-) gal dcm (DE3)*	protein expression

4.3 Cell lines

Table 4 General cell lines

Cell line	Description
GM12878 Cas9	human ENCODE tier I lymphoblastoid cell line with stable Cas9 expression (Jiang et al., 2018)
Raji	human EBV positive Burkitt's lymphoma cell line
HEK293	human epithelial embryonic kidney cell line, (Graham et al., 1977)
DG75	human EBV negative Burkitt's lymphoma cell line (Ben-bassats et al., 1977)
LL8	murine fibroblasts stably transfected with human CD40L (Wiesner et al., 2008a)

Table 5 Lymphoblastoid cell lines established from donors

Cell line	Donor	Virus
SB268 LCLwt	1	XZ143 EBV (6008) HA-EBNA2
SB301 LCL Δα1		SB161 EBV (6008) HA-EBNA2Δα1
SB346.1 LCLwt	6	XZ143 EBV (6008) HA-EBNA2
SB346.2 LCLΔα1		SB161 EBV (6008) HA-EBNA2Δα1
CKR654.1 LCLwt	8	XZ143 EBV (6008) HA-EBNA2
CKR654.3 LCLΔα1		SB161 EBV (6008) HA-EBNA2Δα1

Table 6 Conditional cell lines

Cell line	Internal designation	Description
DG75 doxHA-EBNA2	CKR128-34	human EBV negative Burkitt's lymphoma cell line carrying the doxycycline inducible pRTR-HA-EBNA2 plasmid
DG75 doxHA-EBNA2ΔEND	CKR505.11	human EBV negative Burkitt's lymphoma cell line carrying the doxycycline inducible pRTR-HA-EBNA2ΔEND plasmid
DG75 doxHA-EBNA2-H15A	CKR606.17	human EBV negative Burkitt's lymphoma cell line carrying the doxycycline inducible pRTR-HA-EBNA2-H15A plasmid
DG75 doxHA-EBNA2Δα1	CKR607.20	human EBV negative Burkitt's lymphoma cell line carrying the doxycycline inducible pRTR-HA-EBNA2Δα1 plasmid
GM12878 doxEBF1	CKR621.2	human ENCODE tier I lymphoblastoid cell line with stable Cas9 expression, stably transfected with pRTR-EBF1-Flag
LCLΔα1_doxHA-LMP1	CKR724.1	LCL generated from adenoid biopsy, infected with EBVΔα1, stably transfected with pRTS-HA-LMP1

4.4 Plasmids

Table 7 Plasmids used and established during this thesis

Plasmid	Internal designation	Application
pUC57-EBNA2Δα1_maxiEBV	pSB147	EBV mutagenesis
pCDNA3+	-	negative control, transient transfection for Co-IP assays
pcDNA3+-EBF1 myc his	pSB72	transient transfection for Co-IP assays
pET28-END	AZG1_GST-END	GST affinity capture assay
pET28-ENDΔ-H15A	AZG43_GST-END-H15A	GST affinity capture assay
pET28-ENDΔα1	AZGII_GST-ENDΔα1	GST affinity capture assay
pRTR-hEBF1-Flag	pSB90	stable doxycycline inducible cell line GM12878
pRTR-HA-wtEBNA2	pCKR74.2	stable doxycycline inducible cell line DG75
pRTR-HA-EBNA2ΔEND(aa 3-52)	pCKR492	stable doxycycline inducible cell line DG75
pcDNA3+_EBF1ΔDBD_rrarr	pSB171	transient transfection for Co-IP assays
pcDNA3+_EBF1ΔDBD	pSB128	transient transfection for Co-IP assays
pcDNA3+_EBF1ΔHLH-C	pSB124	transient transfection for Co-IP assays
pcDNA3+_EBF1ΔDBD-IPT	pSB125	transient transfection for Co-IP assays
pcDNA3+_EBF1_DBD	pSB127	transient transfection for Co-IP assays
pCDNA3+_EBF1ΔC	pSB72	transient transfection for Co-IP assays
pcDNA3+_wtEBF1	-	transient transfection for Co-IP assays
p6012.new	-	plasmid for rpsL/Kan cassette PCR
p509 BZLF1	-	virus production in HEK293 cells
p2670 BALF4	-	virus production in HEK293 cells
pGEX-4T2	-	GST only expression plasmid, GST affinity capture assay
pRTS-HA-LMP1	pCKR682.3	stable doxycycline inducible LCLΔα1

4.5 Primer

Table 8 PCR and sequencing primer

Name	Application	Sequence	Annealing T (°C)	Product [bp]
SB113 fwd	rpsL/Kan cassette	TTGACACGGATAGTCTTGGAAACCCGTCACTCTCAGTAATTCCCTCGAATGGCCTGGTGATGATGGCGGGATCG	63	
SB113 rev		GTGTTTTCCCCAACAAAGATTGTTAGTGGAATTAATGGAGTGTCTGACAGTCAGAAGAACTCGTCAAGAAGGCG		
SB162 fwd	sequencing E2 alpha1 helix	CACAGGCTTAGCCAGGTGA		689
SB307 rev		CCCGTGTTTTCCCCAACAAAG		
SB185.1 fwd	PCR of exon3 EBF1 (T7E1 assay)	TTTCTTTTGGGAGAGAGGGAGC	60	497
SB185.1 rev		CCCCAGGTAAACGCGAGA		
SB198 fwd	PCR of EBF1 exon 16 (T7E1 assay)	TGTTAAAGACACATACCCATGCTG	60	602
SB198 rev		TGTCCCTTGTATAGAGCTTTACGG		

SB213 fwd	PCR to check ER-LBD knock-in	ACAGAGGATAGGGAATGCCAAG	67	501
SB287 rev		GTCAGAGCACCTGCATCCAT		
SB244 fwd	PCR for ssDNA preparation	5'P-AAAAAGAGTGACAAAGTGAAAA	52	2577
SB225 rev		TTACACAGCTTTAAAAACATCATA		
SB225 fwd	PCR for ssDNA preparation	AAAAAGAGTGACAAAGTGAAAA	54	2577
SB224 rev		5'P-TTACACAGCTTTAAAAACATCATA		

Table 9 qPCR primer for cDNA quantification

Gene	Internal designation	Sequence	Annealing T [°C]	Product [bp]
myc	SAR40.3 fwd	CTCTCAACGACAGCAGCTC	63	210
	SAR40.3 rev	CCACAGAAACAACATCGATTTC		
EBF1	SAR77.1 fwd	AGACCAGCATGGTACCGAAT	63	235
	SAR77.1 rev	ACGACAGTCAATGTGGATGG		
CBF1	SAR94.14 fwd	AGTTACTGGCATGGCACTCC	63	234
	SAR94.14 rev	ATGATTGTCCAGGAAGCGCC		
SLAMF1	SAR127.1 fwd	GTATCAAGGTGCAGGTCCCG	63	375
	SAR127.1 rev	GGCAGTTGGGAAGCAAAGTG		
EBNA2	LG521 fwd	TCTGCTATGCGAATGCTTTG	63	255
	LG521 rev	CACCGTTAGTGTTGCAGGTG		
GAPDH	BS688 fwd	GAAGGTGAAGGTCGGAGTC	63	153
	LG90 rev	TGGGTGGAATCATATTGGAAC		
LMP2A	SB444 fwd	CTTGGGATTGCAACACGACG	63	171
	SB443 rev	GAAACACGAGGCGGCAATAG		
LMP1	SB422 fwd	GGAAGAAGGCTAGGAAGAAGG	63	155
	SB422 rev	CTGGAATTTGCACGGACAGG		
Actin B	ST249 fwd	GGCATCCTCACCCTGAAGTA	63	203
	ST249 rev	GGGGTGTTGAAGGTCTCAAA		
RNA pol II	SB397 fwd	GTCTGTGACGGAGGGTGG	63	218
	SB397 rev	AAACTTTCATTGTCTTCACCAGG		

Table 10 qPCR for ChIP quantification

Target		Sequence	Annealing T [°C]	Source
CD2 locus	SB396 fwd	GTAAAGAGAGGCACGTGGTTAAG	63	
	SB396 rev	GAACTGAAGTGAGACTGGTGGAG		
CD79a	SB394 fwd	TTCCAAAGGCTCAGCAGTGA	63	www.chipprimers.com
	SB394 rev	GGCCACAATTCCCCTGAGA		
Cp	SB362 fwd	GGCGGGAGAAGGAATAACG	63	(Lu et al., 2016)
	SB364 rev	CTTGAGCTCTCTTATTGGCTATAATCC		
ESE1 myc	SB429 fwd	AGTTCCTTCCCACACCCAC	63	
	SB429 rev	GCTTACGTTGCTCTGAGATTTAG		
ESE2 myc	SB430 fwd	TGAGGAGTGGTGAGTCAGC	63	
	SB430 rev	CCCAGGTGTATCTTAGAACAGG		
HES1	SB432 fwd	TGGCAGGGCAGCTACCA	63	(Lu et al., 2016)
	SB432 rev	GACTGACTTTCTCACACTCAGATTCC		

LMP1	SB361 fwd	ACGTCAGAGTAACGCGTGTTTC	63	(Lu et al., 2016)
	SB361 rev	GCAGACCCCGCAAATCC		
LMP2A	SB360 fwd	GCGCCGCGGTTTCAG	63	(Lu et al., 2016)
	SB360 rev	TTACGCCCCAGCAAGCTT		

Table 11 gRNAs for EBF1 targeted genome editing

gRNA name	Location	Sequence *PAM site*
A-130	Intron 2	AGCGAGTGGACCAACTTTCG *AGG*
B-UCSC	Intron 2	TTTGCGTCCTGATGCAACCC *AGG*
C-354	Intron 3	GGATGACTGACCTTCGCTGG *AGG*
D-278	Intron 3	GTGAGGCCATAGACCCGACC *CGG*
35rev	Exon 16	AGGCAATTCTTTCACATAGG *AGG*
38rev	Exon 16	TCAAGGCAATTCTTTCACAT *AGG*

4.6 Chemicals and Reagents

Table 12 Chemicals and reagents

Product	Supplier	Product	Supplier
α-thioglycerol	Sigman-Aldrich, Germany	L-Glutamine	Gibco, UK
30 % (w/v) acrylamide	Roth, Germany	MS2 RNA	Roche Diagnostics, Germany
Ampicillin	Sigma-Aldrich, USA	OptiMEM	Gibco, UK
APS	MP Biomedicals, Germany	PBS	Sigma-Aldrich, USA
Bradford reagent	Merck,	PEI	Sigma-Aldrich, USA
Ciprobay	Bayer Vital, Germany	Penicillin/Streptomycin	Gibco, UK
Complete Protease Inhibitior	Roche Diagnostics, Germany	PhosStop	Roche Diagnostics, Germany
Cyclosporin A	Sigma-Aldrich, USA	PageRuler Protein ladder	Thermo Scientific, USA
GeneRuler	Thermo Scientific, USA	Puromycin	Merck (Calbiochem) Germany
Doxycycline	Sigma-Aldrich, USA	RPMI 1640 medium	Gibco, UK
Ethidium bromide	Roth, Germany	Salmon sperm DNA	Invitrogen, #15633-019
Fetal calf serum	anprotec, Germany	Sodium-Selenite	Gibco
Ficoll	GE Healthcare (Amersham), UK	Streptomycin	Sigma-Aldrich
Glutathione Sepharose 4B beads	GE Healthcare (Amersham), UK	TEMED	GE Healthcare, UK
Kanamycin		Trypsin 0.05 %	Gibco

4.7 Enzymes and Kits

Table 13 Enzymes and kits

Product	Supplier
BrdU-APC Flow Kit	BD Pharmingen,
ECL	GE Healthcare (Amersham), UK

LightCycler 480 SYBR Green I Master	Roche Diagnostics, Germany
NucleoSpin Gel and PCR Clean-up	Macherey & Nagel, Germany
Phusion Polymerase	New England Biolabs, USA
Q5 Polymerase	New England Biolabs, USA
QIAmp DNA Mini Kit	Qiagen, Germany
QIAshredder	Qiagen, Germany
RNase-free DNase Set	Qiagen, Germany
RNeasy Mini Kit	Qiagen, Germany
SuperScript IV First-Strand cDNA Synthesis Reaction	ThermoFischer, USA
λ exonuclease	New England Biolabs, USA
T7 endonuclease 1	New England Biolabs, USA

4.8 Material and devices

Table 14 Material and devices

Product	Supplier
Ultracrimp 11,5 ml tubes	Sorvall #03987
FACS Fortessa	BD Biosciences, USA
FACS Aria II	BD Biosciences, USA
FACS Canto	BD Biosciences, USA
BioPhotometer D30	Eppendorf, Germany
Fusion Fx with integrated program	Vilber Lourmat, France
Bioruptor Standard sonicator	Diagenode, Belgium

4.9 Software packages and web tools

Table 15 Software packages and tools

Name	Supplier/Source	Application
Bio1D	Vilber Lourmat, France	Western blot quantification
CRISPOR	http://crispor.tefor.net; (Haeussler et al., 2016)	gRNA design
DESeq2 1.30.0	(Love et al., 2014) https://bioconductor.org/packages/release/bioc/html/DESeq2.html	Differential gene expression analysis
DIVA	Becton, Dickinson and Company; 2019.	acquisition of FACS data
FlowJo™ Software	Becton, Dickinson and Company; 2019.	post-acquisition FACS analysis
GraphPad Prism	GraphPad Software	data analysis
GSEA	(Mootha et al., 2003; Subramanian et al., 2005), https://www.gsea-msigdb.org	Gene set enrichment analysis
LightCycler480® v1.5.0	Roche	analysis of RT-qPCR data
R 4.0.4	https://www.r-project.org	programming language
R Studio 1.4.1103	RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.	Integrated development environment for R
TIDE	https://tide.deskgen.com	Cas9 mediated genome editing, calculation of indel frequency
UCSC genome browser	https://genome.ucsc.edu	gRNA design
VennDiagram 1.6.20	https://www.rdocumentation.org/packages/VennDiagram	generating Venn diagrams in R

4.10 Working with eukaryotic cells

4.10.1 Cultivation of suspension and adherent cells

All cells were cultured at 37 °C, 6 % CO_2. DG75, GM12878 cells and irradiated LL8 feeder cells were cultured in RMPI 1640 supplemented with 10 % FCS, 4 mM L-Glut, 100 U/ml penicillin, 100 µg/ml streptomycin. Cell lines carrying the doxycycline inducible pRTR plasmid were cultured with additional 1 µg/ml puromycin. LL8 feeder cells are murine fibroblasts L929 stably transfected with the human CD40L gene that binds to CD40 and induces an internal signaling cascade (Wiesner et al., 2008b).

Raji, HEK293 cells and LCLs were cultured in RPMI 1640 medium supplemented with 10 % FCS, 100 U/ml penicillin, 100 µg/ml streptomycin, 1 mM sodium pyruvate, 100 nM sodium selenite, and 0.43 % α-thioglycerols in 20 µM BCS. HEK293 virus producer cells were cultured with additional 1 µg/ml puromycin. LCLs infected with EBVΔα1 were cultured on LL8 feeder cells with LCL medium.

4.10.2 Long-term storage of cells

Cells for long-term storage were spun down, resuspended in 1 ml freezing medium (90% FCS, 10% DMSO) and transferred to freezing vials. The vials were stored in Mr Frosty freezing containers at – 80° C before being transferred to the liquid nitrogen tanks for long-term storage. To re-culture frozen cells, they were quickly thawed in a 37 °C water bath, washed with 20 ml medium and cultured with fresh medium. Selection medium was added the next day if it was required for the cell line.

4.10.3 Cell number calculation

An aliquot of the cell culture was mixed with Trypan blue 1:2 to 1:10 depending on the cell density. 10 µl of the suspension were pipetted onto a Neubauer counting chamber. Due to the Trypan blue, dead cells can be distinguished from live cells by a blue color. Live cells of four big squares were counted and the cell concentration was calculated accordingly:

$$\frac{cells}{ml} = \frac{\#\ of\ live\ cells \times dilution\ factor \times 10^3}{counted\ area\ (mm^2) \times chamber\ depth\ (mm)}$$

4.10.4 Electroporation of suspension cells

10^7 DG75 cells were harvested, washed with OptiMEM, centrifuged at 300 g for 10 min and resuspended in 250 µl OptiMEM with the respective amount of DNA. The cells were transferred to an electroporation cuvette (0.4 mm), electroporated with 250 V and 950 µF and transferred to a cell culture flask containing 10 ml transfection medium (RPMI 1640, 20 % FCS, 100 U/ml penicillin, 100 µg/ml streptomycin). The cells were cultured overnight and processed the next day according to the respective experiment.

4.10.5 Irradiation of CD40L expressing LL8 stimulator cells

Irradiated CD40L expressing LL8 feeder cells were provided by AG Moosmann. In brief, trypsinized cells were washed and resuspended in medium and irradiated with 180 Gy at the radiation facility. 10^6 cells were seeded per one full plate (96, 48, 24, 12, 6-well) and ready to use the next day.

4.10.6 Isolation of B cells

Primary B cells were isolated from adenoid samples, which we received from hospitals. Adenoids were rinsed in medium and carefully pushed through a 100 µm cell strainer using a 10 ml syringe plunger. Depending on the adenoid size, the cell suspension was filled up to 25 ml or 50 ml with medium (RPMI 1640, 10% FCS, 100 U/ml penicillin, 100µg/ml streptomycin, 4 mM L-Glutamine). T cell rosettes were formed by adding 0.5 ml of sheep blood per 25 ml of cell suspension. 25 ml of the cell suspension with sheep blood was carefully added onto 20 ml of Ficoll and centrifuged at 666 g, 40 min, at 10 °C, decelerated without breaks. The interface was transferred to a new 50 ml Falcon and washed three times with 40 - 50 ml PBS/2 mM EDTA at (i) 666 g (ii) 527 g (iii) 403 g, 10 min each, at 10 °C. Erythrocytes were lysed by resuspending the cell pellet in 5 ml red blood cell lysis buffer (154 mM NH_4Cl; 9.98 mM

$KHCO_3$; 0.127 mM ETDA, pH8) followed by centrifugation for 5 min at 403 g, 10 °C. Cells were resuspended in LCL medium.

4.10.7 Primary B cell infection and generation of LCL

10^6 primary B cells were infected with an MOI 0.1 with EBVwt or EBVΔα1 supernatants and cultured in LCL medium. For the first one to two weeks, 0.5 µg/ml Cyclosporin A was added to inhibit T cells and 10 µg/ml Ciprobay was added as an antibacterial agent. To enhance the growth after one week of infection, B cells infected with EBVΔα1 were co-culture with irradiated LL8 feeder cells. B cells infected with EBVwt were kept without a feeder layer. Infected B cells were expanded to generate LCL and frozen for long-term storage.

4.10.8 Sorting of B cells

4.10.8.1 Sorting of primary B cells

10^8 isolated primary B cells were washed once with FACS buffer (PBS, 2% FCS, 2 mM EDTA) and resuspended in 1 ml FACS buffer. The cells were stained with 80 µl α-IgD-FITC and 24 µl α-CD38-PE-Cy7 for 1 h in the dark at 4 °C and washed twice with FACS buffer. The stained cells were resuspended in 6 ml FACS buffer and filtered through a 35 µm filter to obtain a single cell suspension. Sorting was performed on a FACS Aria II device with a 70 µl nozzle and a sorting velocity of around 7000 events/second with the "Purity" sorting mask. The gating strategy was the following: (i) lymphocytes (ii) single cells (iii) IgD+ CD38-. Around $1.3 * 10^7$ naïve resting B cells were sorted and used for infection and daily sorting in preparation of RNA sequencing. The sorted cells were split and infected with either EBVwt or EBVΔα1 at an MOI 0.1 and cultured in LCL medium until harvest.

4.10.8.2 Sorting of infected B cells

Infected viable cells were harvested daily until day 4 post infection to prepare RNA sequencing samples. For that, infected cells were washed with FACS buffer and centrifuged at 500 g for 5 min. The cells were resuspended in 500 µl FACS buffer and filtered through a 35 µm strainer which was washed with additional 500 µl FACS buffer to avoid loss of cells. Daily sorting was

performed using a 100 µm nozzle, a velocity of around 1000 events/second with the "Purity" sorting mask. Sorted cells were collected in FCS and processed as described in section 4.13.9 (RNA sequencing).

4.10.9 Cell cycle analysis with Bromodeoxyuridine (BrdU)

Primary B cells were infected with an MOI 0.1 with wildtype EBV or EBVΔα1 and seeded with a density of 10^6 cells/ml to analyze them on day 0, 2, 4, 6 and 8 post infection. Prior to cell harvest, the infected cells were pulsed with a final concentration of 10 µM BrdU for 1 h at 37 °C. Afterwards, the cells were harvested, washed with 1 ml of FACS staining buffer (PBS, 2 % FCS, 2 mM EDTA) and centrifuged at 500 g for 5 min. The cells were fixed and permeabilized with 100 µl of BD Cytofix/Cytoperm for 25 min at RT and washed with 1 ml FACS buffer. The samples were stored in freezing medium (90% FCS, 10 % DMSO) at – 80°C until samples from all time points were collected. On the day of FACS analysis, freshly thawed samples were washed in FACS buffer, re-fixed with 100 µl BD Cytofix/Cytoperm for 10 min at RT and washed with 1 ml 1X BD Wash/Perm Buffer. Cell pellets were resuspended in 100 µl DNase (300 µg/ml in PBS final concentration) and incubated for 1 h in the incubator at 37 °C. Afterwards, the cells were washed in 1 ml 1X BD Perm/Wash Buffer. Incorporated BrdU was stained with 1 µl α-BrdU-APC in 50 µl BD Perm/Wash buffer for 20 min in the dark at RT. After washing with 1 ml 1X BD Perm/Wash buffer, cells were resuspended in 20 µl 7-AAD solution, incubated for 5 min and filled up to 200 to 300 µl with FACS staining buffer. The analysis was done with FACS Fortessa.

4.10.10 Cell cycle analysis with propidium iodide (PI)

Primary B cells were infected with an MOI 0.1 with wildtype EBV or EBVΔα1 and seeded with a density of 10^6 cells/ml on plates either containing LL8 feeder cells or on plates without LL8 feeder cells. Cells were harvested on day 0, 2, 4, 6, 8 post infection in preparation of the PI staining. 10^6 cells were harvested and washed once in staining buffer (PBS, 2 % FCS). The cell pellet was fixed by adding 1 ml of ice-cold 70 % ethanol dropwise and stored at -20 °C until all

samples from all time points were collected. On the day of cell cycle analysis, the fixed cell pellets were washed once in staining buffer. The cells were resuspended in 500 µl PI solution (10 µg/ml PI, 175 µg/ml RNaseA in PBS) and analyzed by flow cytometry.

4.10.11 Cell viability analysis with MTT Assay

Isolated primary B cells were infected with a MOI 0.1 with wildtype EBV and EBVΔα1. Uninfected B cells were included as a negative control. For each time point and condition, 8 wells with 10^5 cells/well were seeded in 100 µl medium. 10µl of MTT (5 mg/ml in PBS) was added to each well and incubated for 4 hours at 37 °C. Plates with MTT were stored at – 20 °C until measurement. On the day of analysis, formazan crystals were dissolved with 200 µl of 1 N HCl in isopropanol and the absorption was measured at 550 nm (measurement wave length) and 690 nm (reference wave length).

4.10.12 Establishing HEK293 producer cells

HEK293 cells were used to obtain virus producing cell lines. Pure, supercoiled BACmid DNA (see section 4.11.5.2 Large-scale BACmid preparation) was used for the transfection. One day before transfection, HEK293 cells were plated at a density of around 80% in 6-well plates with 3 ml medium. The medium was replaced with 1 ml OptiMEM the next day and the transfection mix was prepared as follows: Component A (4,2 µl PEI (1 mg/ml) + 500 µl OPtiMEM) and component B (1 µg maxiEBV DNA + 500 µl OptiMEM) were mixed and incubated for 15 min at RT. OptiMEM was removed from the wells and the transfection mix was carefully added to the cells. After four hours, 2 ml of medium was added and the cells were incubated overnight at 37 °C. The next day, HEK293 cells were trypsinized and 10, 30, 100, 300 µl of transfected cells were plated on to 14 cm dishes with 25 ml medium supplemented with 1 µg/ml puromycin. Transfected cells grew until colonies were visible. Before harvest, the clones were checked for GFP expression under a fluorescent microscope. Then the medium was removed and a small piece of autoclaved, Trypsin soaked filter paper was placed on the clones. This was transferred to a well of a 6-well plate containing 3 ml medium with 1 µg/ml puromycin. The

clones were cultured for around 2 weeks before they were screened for virus production. The cells were split in a 6-well plate one day before the screen. BZLF1 was transfected to induce the lytic cycle and BALF4 was transfected to increase the infectivity. The transfection mix was prepared as follows: component A (1 µg DNA (p509(BZLF1) + p2670(BALF4)) + 100 µl RPMI 1640) and component B (6 µl PEI + 100 µl RPMI 1640) were mixed and incubated for 15 min at RT. The medium was replaced with 2 ml medium without puromycin, the transfection mix was carefully added to the wells and the cells were incubated at 37 °C, 5 % CO_2 for 3 days. The supernatant was harvested and the virus titer measured with Raji cells. Promising clones were expanded followed by a large-scale virus production.

4.10.13 Large-scale virus production

The large-scale virus production was done with HEK293 producer cell lines that generate high virus titers. In brief, cells were seeded to a density of around 30 % on 14 cm dishes one day before transfection. The medium was replaced with 25 ml puromycin-free medium just before the transfection. Component A (12 µg DNA p509(BZLF1) + p2670(BALF4), 1.2 ml RPMI 1640) and component B (72 µl PEI (1 mg/ml), 1.2 ml RPMI 1640) of the transfection solution were mixed and incubated for 15 min at RT. The solution was added dropwise to the cells and incubated for 3 days at 37 °C, 5 % CO_2. The supernatant was harvested, centrifuged at 1200 rpm for 10 min and filtered through a sterile 1.2 µm filter. The virus titer of the supernatant was measured using Raji cells.

4.10.14 Determination of the virus titer

Raji cells were used to calculate the titer of virus supernatants. 10^5 Raji cells in 1.5 ml were infected with different volumes of virus supernatant in a 24-well plate. To screen for HEK293 producer clones, 100 µl of the supernatant was added to the Raji cells. To precisely calculate the titer of a large batch, 1, 2, 5, 10 and 20 µl of the virus supernatant were added to the cells and incubated for 3 days at 37 °C, 5 % CO_2. The frequency of GFP expressing Raji cells was

then analyzed on a FACS Canto II device. The following formula was applied to calculate the virus titer for the screen of HEK293 producer clones:

$$GRU = \left(\frac{\%\ GFP + cells}{100\%}\right)\left(\frac{1000\ \mu l}{x\ \mu l\ of\ supernatant}\right)10^5$$

To precisely calculate the virus titer of a large batch, the percentages of GFP+ Raji cells was plotted against the administered volumes of the supernatant. The formula of a linear regression curve was calculated for the linear part of the graph using GraphPad Prism. The titer defined as "Green Raji Units/ml" was calculated as follows with **a** and **b** being given by the linear regression formula. $titer\ [\frac{GRU}{ml}] = \frac{(a \times 1000\ \mu l + b)}{100} \times 10^5$

4.10.15 CRISPR/Cas9 mediated genome editing

Genome editing was achieved by nucleofecting Cas9 ribonucleoprotein (RNP) complexes into the cells. The Lonza 4D Nucleofector X Unit was used to perform the nucleofection experiments. Cas9 protein, crRNA and tracrRNA were purchased from Integrated DNA Technologies, Inc. To prepare gRNAs, crRNA and tracrRNA were mixed in a 1:1 molar ratio to obtain a 100 µM mix. The mix was heated to 95 °C for 5 min followed by 1-hour cooling on the bench top to anneal the RNAs. To obtain Cas9 RNP complexes, the Cas9 protein was mixed with the gRNA in a 62 µM:100 µM ratio in 50 µl sterile PBS and incubated for 10 min at RT. Aliquots were stored at -20 °C until further application. 10^6 cells were harvested and washed once with PBS at 300 g for 10 min. The cells were resuspended in 15 µl P3 buffer (Lonza) and mixed with 5 µl RNP complexes. In case of two different RNP complexes, 2.5 µl of each complex were used. For knock in experiments, the respective HDR template was added to the nucleofection mix. The mix was transferred to a pre-cooled 16-well strip cuvette and nucleofected with the respective program. 100 µl of pre-warmed RPMI medium was added to the well and incubated for 10 min at 37 °C, 5 % CO_2. The cells were transferred to a 48-well plate containing medium and kept in the incubator overnight. The cells were split the next day and either doxycycline (20 ng/ml) or β-estrogen (1 µM) was added to the cultures.

4.11 Working with bacteria

4.11.1 Transformation of bacteria

100 µl of chemically competent bacteria were thawed on ice for 10 min and incubated with 10 µl DNA for 30 min on ice. Bacteria were heat shocked for 2 min at 42 °C with subsequent incubation on ice for 2 min. 1 ml of LB (1% Bacto-Trypton; 0,5% Yeast-Extract; 1% NaCl; pH 7,4) was added to the bacteria and incubated for 1 h at 37 °C. After incubation, transformed bacteria were plated on selection LB-plates (LB medium with 1.5 % Bacto Agar) and incubated overnight at 37 °C.

4.11.2 Preparation of chemically competent bacteria

This is a standard protocol in our laboratory. 500 ml LB medium were inoculated with 5 ml overnight culture of the respective bacterial strain and incubated under vigorous shaking at 37 °C until an OD_{595} of 0.3-0.4 was reached. The culture was distributed into pre-cooled 50 ml tubes, incubated on ice for 10 min and pelleted at 1,600 g, 7 min, 4 °C. Each pellet was resuspended in 10 ml ice cold $CaCl_2$ solution (60 mM $CaCl_2$, 10 mM PIPES, 15% Glycerol, sterile filtrated (0.22 µm)), incubated for 30 min on ice and centrifuged at 1,100 g, 5 min, 4 °C. Each cell pellet was resuspended in 2 ml ice-cold $CaCl_2$ solution. 200 µl aliquots in pre-cooled 1.5 ml reaction tubes were snap frozen on dry ice and stored at -80 °C.

4.11.3 Plasmid preparation from bacterial cultures

4.11.3.1 Small-scale isolation of plasmids

Quick, small-scale plasmid DNA preparation was performed using NucleoSpin Plasmid (Macherey & Nagel) and according to manufacturer's instructions. In general, the preparation is based on an alkaline lysis. 4 ml of LB medium with antibiotics was inoculated with a single colony and incubated shaking overnight at 37 °C. Bacteria were harvested, lysed and cleared lysate was transferred to the silica membrane. Plasmid DNA was eluted by addition of elution buffer and centrifugation.

4.11.3.2 Large-scale isolation of plasmids

Large-scale plasmid DNA preparation was performed with the PureLink kit (Invitrogen) and is based on alkaline lysis. In short, 400 ml of LB with antibiotics was inoculated with 500 µl of bacterial culture and incubated shaking overnight at 37 °C. Bacterial stocks were stored in freezing medium (90% FCS, 10% DMSO) at -80 °C. Bacteria were harvested, lysed and cleared lysate transferred to the manufacturer's columns. Eluted DNA was precipitated in isopropanol and washed with 70 % ethanol. Dried DNA pellets were resuspended in TE buffer (10 mM Tris/ HCl, pH 8,0; 1 mM EDTA, pH 8,0). The DNA concentration was measured with the nanodrop Biophotometer (Eppendorf) and adjusted to the required concentration with TE buffer.

4.11.4 EBV mutagenesis

EBV mutagenesis is based on homologous recombination and dual antibiotic selection in SW105 bacteria. Since the EBV genome is very big, F-factor based BACmids containing the EBV genome are used for the mutagenesis. SW105 bacteria carrying the EBV BACmid with HA-tagged EBNA2 (pXZ143) were established by Xiang Zhang and used for further mutation of the EBNA2 gene.

Step I – Introducing rpsL/Kan cassette

Preparation of the rpsL/Kan cassette for bacterial transformation was as follows. The rpsL/Kan cassette with homologous sequences for mutagenesis target position was PCR amplified from the p6012.new plasmid with Phusion High Fidelity polymerase in a 50 µl reaction according to the manufacturer's instruction. Agarose gel-electrophoresis was performed with the PCR product and it was purified with NucleoSpin Gel and PCR Clean-up kit according to the manufacturer's instructions and an additional ethanol precipitation. To prepare electro-competent SW105 bacteria, the following procedure was performed. A 5 ml no-salt LB (12.5 µg/ml chloramphenicol (Cm)) overnight culture (32 °C) with a fresh colony of the BAC clone SW105 with pXZ143 was prepared. The next day, 40 ml of no-salt LB medium (12.5 µg/ml Cm) were inoculated with 1.5 ml of the overnight culture and incubated at 32 °C shaking until an OD600 of around 0.6. The culture was transferred to a 42 °C water bath and further incubated

for 15 min, shaking at 200 rpm to induce the expression of recombinant proteins. The reaction was stopped by a 20 min incubation on ice. The followings steps were all performed on ice. The culture was evenly transferred to two pre-cooled 50 ml Falcons and centrifuged at 3000 rpm, 15 min, 4 °C. The cells were washed twice with cold sterile, deionized water and centrifuged at 3000 rpm, 10 min, 4 °C. The pellet was resuspended in 1.5 ml cold, deionized water, transferred to pre-cooled microcentrifuge tubes and centrifuged at full speed, 15 min, 4 °C. The pellet was carefully resuspended in 3x pellet volume cold, sterile 10 % Glycerol. 40 µl of the recombination- and electroporation-competent cells were transformed with 100 ng of the purified PCR product in a 1 mm cuvette and with the GenePulser II (BioRad, 1700 V, 200Ω, 25 µF). Transformed bacteria were immediately incubated with 1 ml conventional LB medium at 32 °C shaking for 30-60 min. 5, 20, 100 µl of the cells were plated on conventional LB plates with 12.5 µg/ml Cm, 30 µg/ml Kanamycin and incubated at 32 °C overnight. Single colonies were tested for a functional rpsL/Kan insertion by replica plating on (i) 12.5 µg/ml Cm + 30 µg/ml Kan (Cm/Kan) and (ii) 12.5 µg/ml Cm + 1 mg/ml Str (Cm/Str) LB plates. Clones giving no or little background on Cm/Str-LB plates were selected for an overnight culture in Cm/Kan-LB medium at 32°C. DNA integrity of selected clones was checked by BamHI restriction digest of a small-scale BACmid preparation (see 4.11.3.1 Small-scale isolation of plasmids).

Step II – Replacing rpsL/Kan cassette with a desired sequence

In preparation of transforming and preplacing the rpsL/Kan cassette with the DNA of interest, 100 µg of pSB147 was digested with BsmBI in a 400 µl reaction and the resulting fragments separated by agarose gel-electrophoresis. The fragment of interest was cut out and purified with NucleoSpin Gel and PCR Clean-up kit according to the manufacturer's instructions with a subsequent ethanol precipitation. Preparation of electro-competent rpsL/Kan SW105 bacteria and transformation of these was carried out as described in "Step I". 5, 20 and 100 µl of transformed cells were plated on conventional Cm/Str-LB plates and incubated at 32 ˚C overnight. Several clones were picked and streaked on half a Cm/Str-LB plate, incubated at 32 °C overnight and checked for successful recombination by restriction digestion of a small-scale

BACmid preparation. Clones with a successful recombination were expanded to prepare a large-scale BACmid extraction with a subsequent purification in a cesium chloride- ethidium bromide gradient ($CsCl_2$-EtBr; see 4.11.5.2 Large-scale BACmid preparation)

4.11.5 BACmid preparation from bacterial cultures

4.11.5.1 Small-scale BACmid preparation

To extract BACmid DNA quickly in a small-scale, single clones were spread on half a LB plate with the respective antibiotic combination and incubated at 32 °C overnight. The next day, the grown bacteria were scraped off with an American tooth pick and resuspended in 200 µl resuspension buffer (50 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 100 µg/ml RNase A). 200 µl lysis buffer (4 ml H_2O, 500 µl 2M NaOH, 500 µl 10% SDS) were added and the mix was briefly shaken before a 5 min incubation on ice. 200 µl of neutralization buffer were added, resuspended well and incubated for 5 min on ice. Subsequently, the debris was pelleted at full speed, 15 min, RT, and the supernatant was transferred to a new tube. The DNA was precipitated with 350 µl isopropanol and centrifugation at full speed, 15 min, RT. The DNA pellet was washed with 80 % EtOH with a short vortexing step followed by centrifugation at full speed, 10 min, RT. The supernatant was completely removed and the pellet was resuspended in 20 µl TE buffer. The entire BACmid preparation was used for an enzymatic restriction reaction in 80 µl and analyzed on a 0.7 % agarose gel (1x TBE, 0.01 % (v/v) EtBr) running at 5-8 V/cm for around 16 hours. The DNA was visualized under UV light.

4.11.5.2 Large-scale BACmid preparation

It was important to obtain pure, supercoiled BACmid DNA in order to establish HEK293 clones, which produce high virus titers. To that extend BACmid DNA was extracted and purified on a $CsCl_2$-EtBr gradient. 400 ml LB medium containing 15 µg/ml chloramphenicol were inoculated with a fresh single colony and incubated at 32 °C shaking overnight. The bacteria were pelleted at 4000 rpm, 30 min, 4 °C. BACmid DNA was extracted using the NuceloBond Xtra BAC kit and according to the manufacturer's instructions. The DNA was

dissolved in 400 µl TE buffer overnight at 4 °C. The next day, 1.6 g $CsCl_2$ was added to the solution, carefully dissolved and transferred to a 11.5 ml Ultracrimp tube (Sorvall). The tube was filled up with 1.55 g/ml $CsCl_2$ solution and 200 µl 1% EtBr. The tube was sealed and ultracentrifuged at 35,000 rpm, 20 °C, 3 days. The lower band containing supercoiled DNA was extracted with a large gauge veterinary needle and transferred to 15 ml Falcon. EtBr was removed by isobutanol solvent extraction followed by a dialysis in 2 l TE at 4 °C overnight. The DNA concentration was measured by Qubit fluorimetric quantitation.

4.12 Protein Biochemistry

4.12.1 Whole cell protein lysate

10^7 cells were harvested, washed once with PBS and lysed in 100-200 µl NP-40 lysis buffer (1 % NP-40, 150 mM NaCl, 10 mM Tris-HCl pH7.4, EDTA pH8, 3 % Glycerol, 1x Complete, 1 PhosphoStop). The mix was incubated at 4 °C for one hour. Cell debris was pelleted at 16000 g, 4 °C for 15 min. Cleared lysate was transferred to a new tube and the protein concentration was measured with the Bradford assay. Protein lysates were stored at -80 °C.

4.12.2 Protein concentration determination

The protein concentration was determined with the Bradford assay. Standard dilutions with BSA amounts ranging from 0-12 µg was prepared in 1x ready-to-use Bradford reagent. 1 ml of 1x ready-to-use Bradford reagent was mixed with 1 or 2 µl of sample and incubated for 5 min at RT. The adsorption was measure at 590 nm. The standard dilutions and samples were measured twice and the average value was used for the calculation. Protein concentrations were calculated with the formula of the linear regression curve.

4.12.3 Purification of GST-tagged proteins

GST-tagged proteins were purified as described before(Glaser et al., 2017). *E.coli* strain BL21 was transformed with expression plasmids (see 4.11.1 Transformation of bacteria). Bacteria were cultured in 400 ml of LB medium containing antibiotics at 37˚C until an OD of 0.5–0.7

was reached. Expression of proteins was induced with 1 mM IPTG for 3 h at 30˚C. After induction, bacteria were harvested at 4000 g, 20 min, 4 °C and resuspended in 20 ml ice cold binding buffer (25 mM HEPES, pH 7.6, 0.1 mM EDTA, pH 8, 12.5 mM $MgCl_2$, 10% Glycerol, 0.1% NP-40, 100 mM KCl, 1 mM PMSF, 1 mM DTT) and lysed by sonication (10x 10 sec on/1 min off, 10 % amplitude). Lysates were cleared by centrifugation at 48,000 g at 4˚C for 20 min. Glutathione Sepharose 4B beads (GE Healthcare) were washed and resuspended in binding buffer to prepare a 50% slurry. To coat the beads with GST or GST fusion protein, 100 µl of the 50% slurry were incubated with 20 ml of cleared lysates for 1 hour, 4˚C, and washed 3 times with 20 ml binding buffer.

4.12.4 GST pull down assay with cell lysate

The GST pull down assay was performed as described before (Glaser et al., 2017). 10^7 DG75 cells were transfected with EBF1 expression plasmids or empty vector controls. 24 h after transfection, cells were harvested and lysed in 500 µl NP-40 lysis buffer (50 mM HEPES, pH 7.6, 5 mM EDTA, pH 8, 150 mM NaCl, 0.1% NP-40, 1 mM PMSF) followed by sonication. Cell lysates were centrifuged for 15 min at 16,000 g, 4˚C, and the protein concentration was measured by Bradford assay. To pull down EBF1, the supernatants were incubated with the GST or GST fusion protein coated beads for 3 h at 4˚C. Subsequently, beads were washed 5x with binding buffer and the protein complexes were eluted in 2x Lämmli buffer (4% SDS, 20% Glycerol, 120 mM Tris/HCl, pH 6.8, 5% β-Mercaptoethanol, Bromphenol-blue). Samples were analyzed by SDS-PAGE and Western Blot.

4.12.5 Co-Immunoprecipitation (Co-IP)

Per Co-IP, 2 µg of antibody was incubated with 90 µl 50% slurry of Protein G beads overnight at 4 °C followed by washing of the beads 3x with NP-40 lysis buffer. 10^7 cells per Co-IP were harvested, washed with 1 ml ice cold PBS and lysed in 520 µl NP-40 lysis buffer. The mix was incubated rotating at 4 °C for 1 hour. After pelleting the cell debris at 16,000 g for 15 min at 4 °C, the cleared supernatant was transferred to a new tube. An aliquot of 40 µl mixed with

10 µl of 5x Lämmli buffer was prepared as the input sample. The remaining supernatant was incubated with the antibody-coupled beads for 3 hours at 4 °C. The beads were washed 5x with 1 ml lysis buffer and centrifuged at 2000 g, 2 min, 4 °C. To elute the protein from the beads, one bead volume of 2x Lämmli buffer (4% SDS, 20% Glycerol, 120 mM Tris/HCl, pH 6.8, 5% β-Mercaptoethanol, Bromphenol-blue) was added, heated for 5 min at 95 °C and centrifuged at 6000 g for 2 min. The supernatant was transferred to a new tube. 5 µl of a 1:4 dilution for the input sample and 15 µl for the Co-IP sample were analyzed by SDS-PAGE and Western blot.

4.12.6 Chromatin Immunoprecipitation (ChIP)

ChIP-qPCR was performed in LCLs as previously described (Glaser et al., 2017) with minor changes. $2x10^7$ cells were harvested and washed twice in cold PBS. The pellet was resuspended in RPMI and the crosslink performed with 37% formaldehyde for 7 min. The reaction was stopped with the addition of glycine and incubated for 5 min shaking. The cells were pelleted at 500 g, 4 °C, 5 min, washed twice with cold PBS and 3x with ice-cold lysis buffer (10 mM Tris-HCl, pH7.5, 10 mM NaCl, 3 mM$MgCl_2$, 0.5 % NP-40, 1x protein inhibitor) at 300 g, 4 °C, 10 min. The nuclei were resuspended in sonication buffer (50 mM Tris-HCl, pH8, 5 mM EDTA, pH8, 0.5 % SDS, 0.5 % Triton X-100, 0.05 % Na-deoxcycholate, 1x proteinase inhibitor) and sonicated 4x with the Biorupter device (10 min, 30 sec on/off, medium). Cell debris was pelleted and the supernatant used for the the IP. 50 µl were saved as the input sample. 230 µl chromatin was mixed with 750 µl ChIP dilution buffer (12.5 mM Tris-HCl, pH8, 187.5 mM NaCl, 1.25 mM EDTA, pH8, 1.125 % Triton X-100, 1x proteinase inhibitor). For EBF1 ChIP, 5 µg antibody (ABE1294 or rabbit IgG isotype control) were used. For EBNA2 ChIP, 100 µl hybridoma supernatant (mix of α-EBNA2 (R3), α-EBNA2 (1E6), α-HA (3F10) or isotype control α-GST (6G9) + α-CD23 Dog-CD3) were used. The mix was incubated rolling overnight at 4 °C. Protein A or G beads were pre-blocked with 500 µg/ml salmon testes DNA overnight at 4 °C to reduce unspecific chromatin binding. The next day, the beads were equilibrated with ChIP dilution buffer to a 50 % slurry. 100 µl Protein A beads were added to the EBF1 ChIP and 100 µl Protein

G beads to EBNA2 ChIP and incubated for 3 hours at 4 °C. Washing of beads and reverse cross-link was performed according to Glaser et al, 2017. Reverse cross-linked chromatin was purified using the NucleoSpin™ Gel and PCR Clean-up kit (Macherey-Nagel) according to the manufacturer's instructions. Purified ChIP and input samples were analyzed by qPCR using the primers listed in Table 10.

4.12.7 SDS-PAGE

Proteins were separated using a reducing SDS polyacrylamide gel electrophoresis. The separation gel contained 8-10% and the stacking gel contained 15 % polyacrylamide using 30 % (w/v) acrylamide (Roth, Germany). The respective sample volume was mixed with 2x or 5x Lämmli buffer, boiled at 95 °C for 5 min and loaded onto the gel. Each gel was run at 25 mA in running buffer (250 mM Tris base, 2 M glycine, 1 % SDS) until Bromophenol blue migrated out of the gel.

4.12.8 Western Blot

SDS-PAGE protein gels were transferred to PVDF membranes. For that, the membranes were soaked in 100 % Methanol for 5 min. The gels, activated membranes, Whatman filter paper and sponges were soaked in transfer buffer (250 mM Tris base, 1.92 M glycine, 1 % SDS) until the sandwich assembly. The sandwich was assembled starting from the cathode side as follows: (i) one sponge (ii) 3 layers filter paper (iii) gel (iv) PVDF membrane (v) 3 layers filter paper (vi) one sponge. Air bubbles were removed and the proteins were transferred at 4 °C, 400 mA, for 1 h. The membranes were blocked for 30 min in blocking buffer. Protein detection was done with specific antibodies. Primary antibodies in blocking buffer were incubated with the membrane at 4 °C for 1 h or overnight. After washing 3x with PBS/T for 10 min, the secondary HRP-coupled antibody in blocking buffer was incubated with the membrane at 4 °C for 1 h followed by 4x washing with PBS/T for 15 min and a final wash with PBS. Bound antibodies were detected with the Enhanced Chemiluminescence (ECL) system according to the

manufacturer's instructions. Emitted light was detected with the Fusion FX (Vilber Lourmat) device.

4.13 Nucleic acid related techniques

4.13.1 Isolation of genomic DNA

QIAmp DNA Mini Kit (Qiagen) was used to isolate genomic DNA according to the manufacturer's instructions. Purified DNA was eluted in 50 µl H_2O and the concentration was measured with a nanodrop device.

4.13.2 Preparation of single stranded homology directed repair (HDR) template

500 µl PCR reaction was prepared and split into 50 µl aliquots. The PCR was performed either with SB224 fwd/SB225 rev or SB225 fwd/SB224 rev primers causing the phosphorylation at the 5' end of either the + or – DNA strand. Q5 polymerase was used for the PCR and the mix prepared according to the manufacturer's instructions. pSB212 was used as a DNA template. After the PCR run was complete, an aliquot was loaded onto an agarose gel to check for correct amplification. All 50 µl PCR aliquots of the same PCR were pooled, purified with the NucleoSpin Gel and PCR Clean-Up kit (Macherey & Nagel) and eluted in water. The λ exonuclease (NEB) reaction was prepared according to the manufacturer's instruction to digest the phosphorylated DNA strand. The reaction was incubated for 30 min at 37 °C and the single stranded DNA purified with the NucleoSpin Gel and PCR Clean-Up kit (Macherey & Nagel) and eluted in water. The volume of the eluate was reduced in a speedvac to around 10 µl. The concentrated single stranded DNA could be used for nucleofection.

4.13.3 T7 endonuclease 1 assay

The protocol was adapted from New England Biolab. Genomic DNA was isolated from cells that underwent CRISPR/Cas9 mediated genome editing. In brief, the edited region was PCR amplified using the Q5® Hot Start High-Fidelity DNA polymerase (NEB) and the product was purified with the Marchery-Nagel NucleoSpin® Gel and PCR-Clean up kit according to the

manufacturer's instructions. The annealing reaction was prepared with 200-300 ng DNA and subjected to a thermocycler with the following hybridization conditions: (i) initial denaturation at 95 °C for 5 min (ii) annealing at 95-85 °C, ramp rate -2 °C/second; 85-25 °C, ramp rate -0.1 °C/second (iii) hold at 4 °C. Afterwards, T7 endonuclease 1 was added to the reaction and incubated at 37 °C for 15 min. The digestion of mismatched DNA duplexes was analyzed on a 1.7 % agarose gel.

4.13.4 DNA quantitation

DNA concentration was either measured using the Biophotometer (Eppendorf) device or by Qubit fluorometric quantitation (Thermo Scientific). Both quantitation methods were performed according to the manufacturer's instructions.

4.13.5 Quantitative real time PCR (qPCR)

The qPCR-protocol is a standard protocol in our laboratory. The quantification of cDNA obtained from reverse transcribed RNA and DNA recovered from ChIP experiments was done with the Roche LightCycler480 II device and LightCycler 480 SYBR Green I Master (Roche) reagent according to the manufacturer's instructions. 10 µl per reaction were pipetted into a 96-well plate. 2 µl of sample were added to each well. The master mix contained 5 µl LightCycler 480 SYBR Green I Master, 1 µl 5 µM forward primer, 1 µl 5 µM reverse primer, 1 µl PCR grade H_2O. The cycle conditions are listed in Table 16 LightCycler 480 .

Table 16 LightCycler 480 cycling conditions

Analysis mode	Cycles	Segment	Temperature (°C)	Ramp Rate (°C/s)	Time hold	Acquisition Mode
None	1	Pre-incubation	95	4.4	10 min	none
Quantification	45	Denaturation	95	4.4	3 s	none
		Annealing	63	2.2	10 s	none
		Extension	72	4.4	20 s	single
Melting Curves	1	Denaturation	95	4.4	5 s	none
		Annealing	65	2.2	1 min	none
		Melting	97	0.1	-	continuous
None	1	Cooling	40	2.2	30 s	none

Standard curves with dilutions of defined amounts of the respective PCR product were established and applied to account for varying primer efficiencies. Two replicates for each cDNA, ChIP target and standard dilution were cycled in parallel. The relative quantification of cDNA was normalized to RNA polymerase II signals applying the Roche LightCycler 480 E-Method, which is correcting for primer efficiencies. The percentage of input for ChIP-qPCR signals was calculated for each target as (qPCR signal from specific ChIP/qPCR signal from input) x 100. Mann-Whitney test or Welch's t-test were applied to test for differences. RNA-related techniques

4.13.6 RNA isolation

RNA was isolated from $5x10^6$ to $1x10^7$ cells with the RNeasy® Mini Kit (Qiagen) according to the manufacturer's instructions. β-mercapotethanol in a final concentration of 134 mM was added to the lysis buffer. Cells were lysed in 600 µl lysis buffer. QIAshredder columns (Qiagen) were used to completely disrupt the cells and according to the manufacturer's instructions. Genomic DNA was removed with the RNase-free DNase set (Qiagen) according to the manufacturer's instructions. Purified RNA was eluted in 30-50 µl RNase-free water and the concentration was measured using the Qubit RNA HS kit according to the instructions.

4.13.7 Reverse transcription of RNA

SuperScript™IV First-Strand cDNA Synthesis kit (Thermofisher) was used to transcribe 2-4 µg isolated RNA into cDNA according to the manufacturer's instructions. Random hexamers were used as primers. A reaction without reverse transcriptase was used as a negative control. 1/40 of cDNA (50 ng of input RNA) was used for RT-qPCR.

4.13.8 sgRNA design for CRISPR/Cas9 mediated genome editing

gRNAs for CRISPR/Cas9 experiments were designed using CRISPOR (http://crispor.tefor.net/; Haeussler et al., 2016) and the UCSC genome browser (https://genome.ucsc.edu). The choice for a gRNA was based on the on- and off-target score and position of the gRNA. RNAs were ordered from Integrated DNA Technologies, Inc.

4.13.9 RNA sequencing sample preparation

Sorted B cells in FCS (see section 4.10.8.2) were centrifuged for at 500 g for 5 min, washed with PBS and resuspended in 15-20 µl PBS. Cells were counted and the equivalent of 10,000 cells were pipetted into a 96-well plate containing 50 µl RLT+ Buffer (Qiagen) containing 1 % β-mercaptoethanol. The 96-well plate was immediately stored at -80 °C until all samples from 6 biological replicates were collected. Samples were then transferred to Lucas Wange (AG Enard) who prepared the cDNA library and processed the raw sequencing data. RNA sequencing was performed using the prime-seq method which is based on the scRNA-seq method SCRB-seq (Bagnoli et.al. 2018) and can be used on cell lysates without the intermediate step of RNA isolation. The flow cytometric analysis showed big differences in cell size between the different time points and conditions, which likely corresponds to differences in RNA amounts per cell. As prime-seq uses sample barcoding in the reverse transcription and subsequent pooling of the first strand cDNA, big differences in the RNA input per sample will lead to over-sequencing of some and under-sequencing of other libraries. To avoid this undesirable outcome, a test experiment was performed on a pool of all replicates for each condition. To assess how much RNA is extracted, each pool of lysates in a realistic experiment, prime-seq was performed on each pool individually and the cDNA amplification was performed as a qPCR using SYBR green. The qPCR confirmed the expected differences in RNA content and enabled us to normalize the input accordingly prior to processing with prime-seq by using differing amounts of lysate.

Table 17 Volumes used for cDNA library preparation

Sample (6 replicates each)	Infection	Volume cell lysate [µl]
no_0	uninfected	50
mt_1	EBVΔα1	10
mt_2	EBVΔα1	50
mt_3	EBVΔα1	50
mt_4	EBVΔα1	30
wt_1	EBVwt	30
wt_2	EBVwt	10
wt_3	EBVwt	10
wt_4	EBVwt	10

Having adjusted the input, prime-seq was performed according to protocol (https://www.protocols.io/view/prime-seq-s9veh66). In brief lysates were first treated with Proteinase K (Ambion) followed by a bead clean-up (2:1 beads to lysate ratio) using solid phase reversible immobilization (SPRI) beads (GE Biotech). As SPRI beads unspecifically bind all nucleic acids an on-bead DNase1 digest was performed to remove genomic DNA, followed by another SPRI bead clean up and reverse transcription. Reverse transcription (RT) was performed using Maxima H-RT enzyme (Thermo Fisher Scientific) with well specific barcoded oligo dT primers and a template switch oligo (TSO). After RT, the barcoded first strand cDNA was pooled and cleaned up using magnetic beads as before and remaining oligo dT primers were removed using exonuclease 1 digest. Subsequently cDNA was PCR amplified with Kapa Hifi DNA Polymerase (Roche), quantified using picogreen dye (Thermo Fisher Scientific) and the quality controlled using capillary gel electrophoresis (Agilent Bioanalyzer 2100). Sequencing libraries were generated with the NEB NEXT Ultra II FS kit by using a custom adapter oligo and a primer, specific for the cDNA 3'-end. Fragments between 300 bp and 500 bp were selected using double size selection with SPRIselect beads (Beckman Coulter). Final libraries were quantified and quality controlled again with the Bioanalyzer 2100. Paired End sequencing was performed as 28 bp read 1 and 50 bp read 2 with an 8 bp index read on an Illumina HiSeq 1500 instrument. The first read contains the sample barcode and UMI whereas the second read contains the cDNA sequence. In total 1.5 lanes of a High Out flow cell were sequenced, amounting to 465.9 Mio reads. Raw sequencing data was processed using the zUMIs pipeline (Parekh et. al. 2018) with mapping to the reference Genome using STAR 2.7 (Dobin et al. 2013). As a reference genome hg38 was used with Gencode annotation (v 35) concatenated with the EBV genomes Akata inverted and p6008 (manually assembled, provided by AG Hammerschmidt) including annotation retrieved from (https://github.com/flemingtonlab/public/blob/master/annotation/chrEBV_Akata_inverted_refined_genes_annotation_cleaned.gtf). The processed sequencing data by Lucas Wange was used for bioinformatics analysis of the data.

4.14 Bioinformatic analysis

4.14.1 Quality control and data normalization

The quality control and data normalization were performed with the help of Antonio Scialdone. To initially asses the processed RNA sequencing data, the library size, number of detected cellular genes and the fraction of mitochondrial genes was computes. For subsequent analyses, samples with less than 10,000 detected genes were excluded. Data was normalized using the estimation size factor (Anders and Huber, 2010), which normalizes for sequencing depth and RNA composition.

4.14.2 Principal component analysis (PCA)

PCA was performed with the help of Antonio Scialdone. Normalized and log-transformed expression data of cellular genes was used to perform the PCA.

4.14.3 Differential gene expression (DGE) analysis

Genes with a low expression (cpm <10) were excluded. The DGE analysis was performed using the DESeq2 R package. Differentially expressed genes with an FDR < 0.1 were selected for the gene set enrichment analysis (GSEA).

4.14.4 Gene set enrichment analysis (GSEA)

For the GSEA, the program was downloaded from https://www.gsea-msigdb.org as a desktop application. The hallmark gene sets were selected for the analysis. Lowly expressed genes (cpm <10) were excluded from the analysis and normalized expression data was used for the analysis. Enriched gene sets with a false discovery rate <0.1 were considered significant. The leading-edge subset of significantly enriched sets was used to create Venn diagrams.

4.14.5 Venn diagram

Venn diagrams were generated in R using the VennDiagram package for R (v1.6.20). Genes from the leading-edge subsets were used as input data.

5 References

Adachi, N., So, S., Iiizumi, S., Nomura, Y., Murai, K., Yamakawa, C., Miyagawa, K., and Koyama, H. (2006). The human pre-B cell line Nalm-6 is highly proficient in gene targeting by homologous recombination. DNA Cell Biol. *25*, 19–24.

Åkerblad, P., Rosberg, M., Leanderson, T., and Sigvardsson, M. (1999). The B29 (Immunoglobulin β-Chain) Gene Is a Genetic Target for Early B-Cell Factor. Mol. Cell. Biol. *19*, 392–401.

AlJanahi, A.A., Danielsen, M., and Dunbar, C.E. (2018). An Introduction to the Analysis of Single-Cell RNA-Sequencing Data. Mol. Ther. - Methods Clin. Dev. *10*, 189–196.

Allday, M.J., Bazot, Q., and White, R.E. (2015). The EBNA3 family: Two oncoproteins and a tumour suppressor that are central to the biology of EBV in B cells. In Current Topics in Microbiology and Immunology, (Springer Verlag), pp. 61–117.

Alsaab, H.O., Sau, S., Alzhrani, R., Tatiparti, K., Bhise, K., Kashaw, S.K., and Iyer, A.K. (2017). PD-1 and PD-L1 Checkpoint Signaling Inhibition for Cancer Immunotherapy: Mechanism, Combinations, and Clinical Outcome. Front. Pharmacol. *8*, 561.

Anastasiadou, E., Stroopinsky, D., Alimperti, S., Jiao, A.L., Pyzer, A.R., Cippitelli, C., Pepe, G., Severa, M., Rosenblatt, J., Etna, M.P., et al. (2019). Epstein–Barr virus-encoded EBNA2 alters immune checkpoint PD-L1 expression by downregulating miR-34a in B-cell lymphomas. Leukemia *33*, 132–147.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Bagnoli, J.W., Ziegenhain, C., Janjic, A., Wange, L.E., Vieth, B., Parekh, S., Geuder, J., Hellmann, I., and Enard, W. (2018). Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. Nat. Commun. *9*, 1–8.

Baker, S.J., and Reddy, E.P. (2012). CDK4: A key player in the cell cycle, development, and cancer. Genes and Cancer *3*, 658–669.

Banchereau, J., De Paoli, P., Vallé, A., Garcia, E., and Rousset, F. (1990). Long-term human B cell lines dependent on interleukin-4 and antibody to CD40. Science (80-.). *251*, 70–72.

Banerjee, A., Northrup, D., Boukarabila, H., Jacobsen, S.E.W., and Allman, D. (2013). Transcriptional Repression of Gata3 Is Essential for Early B Cell Commitment. Immunity *38*, 930–942.

Ben-bassats, H., Goldblum, N., Mitrani, S., Goldblum, T., Yoffey, J.M., Cohen, M.M., Bentwich, Z., Ramot, B., Klein, E., and Klein, G. (1977). Establishment in continuous culture of a new type of lymphocyte from a "burkitt-like" malignant lymphoma (line d.g.-75). Int. J. Cancer *19*, 27–33.

Bhende, P.M., Seaman, W.T., Delecluse, H.-J., and Kenney, S.C. (2005). BZLF1 Activation of the Methylated Form of the BRLF1 Immediate-Early Promoter Is Regulated by BZLF1 Residue 186. J. Virol. *79*, 7338–7348.

Bohle, V., Döring, C., Hansmann, M.L., and Küppers, R. (2013). Role of early B-cell factor 1 (EBF1) in Hodgkin lymphoma. Leukemia *27*, 671–679.

Boller, S., and Grosschedl, R. (2014). The regulatory network of B-cell differentiation: a focused view of early B-cell factor 1 function. Immunol. Rev. *261*, 102–115.

Boller, S., Ramamoorthy, S., Akbas, D., Nechanitzky, R., Burger, L., Murr, R., Schübeler, D., and Grosschedl, R. (2016a). Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. Immunity *44*, 527–541.

Boller, S., Ramamoorthy, S., Akbas, D., Nechanitzky, R., Burger, L., Murr, R., Schübeler, D., and Grosschedl, R. (2016b). Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming; Supplements. Immunity *44*, 527–541.

Boller, S., Li, R., and Grosschedl, R. (2018). Defining B Cell Chromatin: Lessons from EBF1. Trends Genet. *xx*, 1–13.

Bornkamm, G.W., Berens, C., Kuklik-Roos, C., Bechet, J.M., Laux, G., Bachl, J., Korndoerfer, M., Schlee, M., Holzel, M., Malamoussi, A., et al. (2005). Stringent doxycycline-dependent control of gene activities using an episomal one-vector system. Nucleic Acids Res. *33*, e137.

Borst, P. (2020). The malate–aspartate shuttle (Borst cycle): How it started and developed into a major metabolic pathway. IUBMB Life *72*, 2241–2259.

Bouamar, H., Abbas, S., Lin, A.-P., Wang, L., Jiang, D., Holder, K.N., Kinney, M.C., Hunicke-Smith, S., and Aguiar, R.C.T. (2013). A capture-sequencing strategy identifies IRF8, EBF1, and APRIL as novel IGH fusion partners in B-cell lymphoma. Blood *122*, 726–733.

Bracken, A.P., Ciro, M., Cocito, A., and Helin, K. (2004). E2F target genes: Unraveling the biology. Trends Biochem. Sci. *29*, 409–417.

Bretones, G., Delgado, M.D., and León, J. (2015). Myc and cell cycle control. Biochim. Biophys. Acta - Gene Regul. Mech. *1849*, 506–516.

Brinkman, E.K., Chen, T., Amendola, M., and Van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. *42*, e168–e168.

Bullerwell, C.E., Robichaud, P.P., Deprez, P.M.L., Joy, A.P., Wajnberg, G., D'Souza, D., Chacko, S., Fournier, S., Crapoulet, N., Barnett, D.A., et al. (2021). EBF1 drives hallmark B cell gene expression by enabling the interaction of PAX5 with the MLL H3K4 methyltransferase complex. Sci. Rep. *11*.

Cen, O., and Longnecker, R. (2015). Latent membrane protein 2 (LMP2). In Current Topics in Microbiology and Immunology, (Springer Verlag), pp. 151–180.

Chabot, P.R., Raiola, L., Lussier-Price, M., Morse, T., Arseneault, G., Archambault, J., and Omichinski, J.G. (2014). Structural and Functional Characterization of a Complex between the Acidic Transactivation Domain of EBNA2 and the Tfb1/p62 Subunit of TFIIH. PLoS Pathog. *10*, e1004042.

Chen, J. (2016). The cell-cycle arrest and apoptotic functions of p53 in tumor initiation and progression. Cold Spring Harb. Perspect. Med. *6*.

Chen, A., DiVisconte, M., Jiang, X., Quink, C., and Wang, F. (2005). Epstein-Barr Virus with the Latent Infection Nuclear Antigen 3B Completely Deleted Is Still Competent for B-Cell Growth Transformation In Vitro. J. Virol. *79*, 4506–4509.

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kühn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat. Biotechnol. *33*, 543–548.

Cohen, J.I., Wang, F., Mannick, J., and Kieff, E. (1989). Epstein-Barr virus nuclear protein 2 is a key determinant of lymphocyte transformation. Proc. Natl. Acad. Sci. U. S. A. *86*, 9558–9562.

Crozatier, M., Valle, D., Dubois, L., Ibnsouda, S., and Vincent, A. (1996). collier, a novel regulator of Drosophila head development, is expressed in a single mitotic domain. Curr. Biol. *6*, 707–718.

Delecluse, H.J., Hilsendegen, T., Pich, D., Zeidler, R., and Hammerschmidt, W. (1998). Propagation and recovery of intact, infectious Epstein-Barr virus from prokaryotic to human cells. Proc. Natl. Acad. Sci. U. S. A. *95*, 8245–8250.

Dias, S., Silva, H., Cumano, A., and Vieira, P. (2005). Interleukin-7 is necessary to maintain the B cell potential in common lymphoid progenitors. J. Exp. Med. *201*, 971–979.

Ding, L., Cao, J., Lin, W., Chen, H., Xiong, X., Ao, H., Yu, M., Lin, J., and Cui, Q. (2020). The roles of cyclin-dependent kinases in cell-cycle progression and therapeutic strategies in human breast cancer. Int. J. Mol. Sci. *21*, 1960.

Dirmeier, U., Neuhierl, B., Kilger, E., Reisbach, G., Sandberg, M.L., and Hammerschmidt, W. (2003). Latent membrane protein 1 is critical for efficient growth transformation of human B cells by epstein-barr virus. Cancer Res *63*, 2982-9.

Dirmeier, U., Hoffmann, R., Kilger, E., Schultheiss, U., Briseño, C., Gires, O., Kieser, A., Eick, D., Sugden, B., and Hammerschmidt, W. (2005). Latent membrane protein 1 of Epstein-Barr virus coordinately regulates proliferation with control of apoptosis. Oncogene *24*, 1711–1717.

Ducker, G.S., and Rabinowitz, J.D. (2017). One-Carbon Metabolism in Health and Disease. Cell Metab. *25*, 27–42.

Dykstra, M.L., Longnecker, R., and Pierce, S.K. (2001). Epstein-Barr virus coopts lipid rafts to block the signaling and antigen transport functions of the BCR. Immunity *14*, 57–67.

Engeland, K. (2018). Cell cycle arrest through indirect transcriptional repression by p53: I have a DREAM. Cell Death Differ. *25*, 114–132.

Epstein, M.A., Achong, B.G., Barr, Y.M., Nat ; Morgan, C., Rose, H.M., Holden, M., Jones, E.P., Hummeler, K., Berkaloff, A., and Holt, S. (1964). VIRUS PARTICLES IN CULTURED LYMPHOBLASTS FROM BURKITT'S LYMPHOMA.

Fahraeus, R., Janssont, A., Rickstent, A., Sjoblomt, A., and Rymott, L. (1990). Epstein-Barr virus-encoded nuclear antigen 2 activates the viral latent membrane protein promoter by modulating the activity of a negative regulatory element .

Feldhaus, A.L., Mbangkollo, D., Arvin, K.L., Klug, C.A., and Singh, H. (1992). Novel Cell-Type- and Stage-Specific Regulator of the B-Lymphocyte Gene mb-i. *12*, 1126–1133.

Frappier, L. (2015). EBNA1. In Current Topics in Microbiology and Immunology, (Springer Verlag), pp. 3–34.

Friberg, A., Thumann, S., Hennig, J., Zou, P., Nössner, E., Ling, P.D., Sattler, M., and Kempkes, B. (2015). The EBNA-2 N-Terminal Transactivation Domain Folds into a Dimeric Structure Required for Target Gene Activation. PLoS Pathog. *11*, 1–24.

Fruehling, S., and Longnecker, R. (1997). The immunoreceptor tyrosine-based activation motif of Epstein-Barr virus LMP2A is essential for blocking BCR-mediated signal transduction. Virology *235*, 241–251.

Fruehling, S., Swart, R., Dolwick, K.M., Kremmer, E., and Longnecker, R. (1998). Tyrosine 112 of Latent Membrane Protein 2A Is Essential for Protein Tyrosine Kinase Loading and Regulation of Epstein-Barr Virus Latency. J. Virol. *72*, 7796–7806.

Fuchs, K.P., Bommer, G., Dumont, E., Christoph, B., Vidal, M., Kremmer, E., and Kempkes, B. (2001). Mutational analysis of the J recombination signal sequence binding protein (RBP-J)/Epstein-Barr virus nuclear antigen 2 (EBNA2) and RBP-J/Notch interaction. Eur. J. Biochem. *268*, 4639–4646.

Gao, H., Lukin, K., Ramírez, J., Fields, S., Lopez, D., and Hagman, J. (2009). Opposing effects of SWI/SNF and Mi-2/NuRD chromatin remodeling complexes on epigenetic reprogramming by EBF and Pax5. Proc. Natl. Acad. Sci. *106*, 11258–11263.

Ge, T., Yang, J., Zhou, S., Wang, Y., Li, Y., and Tong, X. (2020). The Role of the Pentose Phosphate Pathway in Diabetes and Cancer. Front. Endocrinol. (Lausanne). *11*.

Giaimo, B.D., Gagliani, E.K., Kovall, R.A., and Borggrefe, T. (2021). Transcription factor RBPJ as a molecular switch in regulating the Notch response. In Advances in Experimental Medicine and Biology, (Springer), pp. 9–30.

Gisler, R., and Sigvardsson, M. (2002). The Human V-PreB Promoter Is a Target for Coordinated Activation by Early B Cell Factor and E47. J. Immunol. *168*, 5130–5138.

Glaser, L. V., Rieger, S., Thumann, S., Beer, S., Kuklik-Roos, C., Martin, D.E., Maier, K.C., Harth-Hertle, M.L., Grüning, B., Backofen, R., et al. (2017). EBF1 binds to EBNA2 and promotes the assembly of EBNA2 chromatin complexes in B cells. PLoS Pathog. *13*, 1–30.

González-Magaña, A., and Blanco, F.J. (2020). Human PCNA structure, function and interactions. Biomolecules *10*, 570.

Gossen, M., Freundlieb, S., Bender, G., Müller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. Science (80-.). *268*, 1766–1769.

Graham, F.L., Smileyt, J., Russell, W.C., and Nairn, R. (1977). Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5.

Griffin, M.J., Zhou, Y., Kang, S., Zhang, X., Mikkelsen, T.S., and Rosen, E.D. (2013). Early B-cell factor-1 (EBF1) is a key regulator of metabolic and inflammatory signaling pathways in mature adipocytes. J. Biol. Chem. *288*, 35925–35939.

Grogan, E., Jenson, H., Countrymant, J., Heston, L., Gradoville, L., and Miller, G. (1987). Transfection of a rearranged viral DNA fragment, WZhet, stably converts latent Epstein-Barr viral infection to productive infection in lymphoid cells.

Grossman, S.R., Johannsen, E., Tong, X., Yalamanchili, R., and Kieff, E. (1994). The Epstein-Barr virus nuclear antigen 2 transactivator is directed to response elements by the Jκ recombination signal binding protein. Proc. Natl. Acad. Sci. U. S. A. *91*, 7568–7572.

Guilhamon, P., Eskandarpour, M., Halai, D., Wilson, G.A., Feber, A., Teschendorff, A.E., Gomez, V., Hergovich, A., Tirabosco, R., Fernanda Amary, M., et al. (2013). Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2. Nat Commun *4*, 2166.

Györy, I., Boller, S., Nechanitzky, R., Mandel, E., Pott, S., Liu, E., and Grosschedl, R. (2012). Transcription factor EBf1 regulates differentiation stage-specific signaling, proliferation, and survival of B cells. Genes Dev. *26*, 668–682.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. *17*, 148.

Hagman, J., and Lukin, K. (2005). Early B-cell factor "pioneers" the way for B-cell development. Trends Immunol. *26*, 455–461.

Hagman, J., Travis, A., and Grosschedl, R. (1991). A novel lineage-specific nuclear factor regulates mb-1 gene transcription at the early stages of B cell differentiation. EMBO J. *10*, 3409–3417.

Hagman, J., Gutch, M.J., Lin, H., and Grosschedl, R. (1995). EBF contains a novel zinc coordination motif and multiple dimerization and transcriptional activation domains.

Hammerschmidt, W., and Sugden, B. (1989). Genetic analysis of immortalizing functions of Epstein-Barr virus in human B lymphocytes. Nature *340*, 393–397.

Harada, S., and Kieff, E. (1997). Epstein-Barr virus nuclear protein LP stimulates EBNA-2 acidic domain-mediated transcriptional activation. J. Virol. *71*, 6611–6618.

Harder, L., Eschenburg, G., Zech, A., Kriebitzsch, N., Otto, B., Streichert, T., Behlich, A.S., Dierck, K., Klingler, B., Hansen, A., et al. (2013). Aberrant ZNF423 impedes B cell differentiation and is linked to adverse outcome of ETV6-RUNX1 negative B precursor acute lymphoblastic leukemia. J. Exp. Med. *210*, 2289–2304.

Hardy, R.R., Kincade, P.W., and Dorshkind, K. (2007). The Protean Nature of Cells in the B Lymphocyte Lineage. Immunity *26*, 703–714.

Harth-Hertle, M.L., Scholz, B.A., Erhard, F., Glaser, L. V., Dölken, L., Zimmer, R., and Kempkes, B. (2013). Inactivation of Intergenic Enhancers by EBNA3A Initiates and Maintains Polycomb Signatures across a Chromatin Domain Encoding CXCL10 and CXCL9. PLoS Pathog. *9*.

Henle, W., Diehl, V., Kohn, G., Hausen, H. Zur, and Henle, G. (1967). Herpes-type virus and chromosome marker in normal leukocytes after growth with irradiated burkitt cells. Science (80-.). *157*, 1064–1065.

Hermeking, H., Rago, C., Schuhmacher, M., Li, Q., Barrett, J.F., Obaya, A.J., O'Connell, B.C., Mateyak, M.K., Tam, W., Kohlhuber, F., et al. (2000). Identification of CDK4 as a target of c-MYC. Proc. Natl. Acad. Sci. U. S. A. *97*, 2229–2234.

Hertle, M.L., Popp, C., Petermann, S., Maier, S., Kremmer, E., Lang, R., Mages, J., and Kempkes, B. (2009). Differential gene expression patterns of EBV infected EBNA-3A positive and negative human B lymphocytes. PLoS Pathog. *5*.

Hombach, J., Lottspeich, F., and Reth, M. (1990). Identification of the genes encoding the IgM-α and Ig-β components of the IgM antigen receptor complex by amino-terminal sequencing. Eur. J. Immunol. *20*, 2795–2799.

Hsieh, J.J., and Hayward, S.D. (1995). Masking of the CBF1/RBPJ Transcriptional Repression Domain by Epstein-Barr Virus EBNA2. Science (80-.). *268*, 560–563.

Hummeler, K., Henle, G., and Henle, W. (1966). Fine structure of a virus in cultured lymphoblasts from Burkitt lymphoma. J. Bacteriol. *91*.

Hurwitz, R., Hozier, J., Lebien, T., Minowada, J., Gajl-Peczalska, K., Kubonishi, I., and Kersey, J. (1979). Characterization of a leukemic cell line of the pre-B phenotype. Int. J. Cancer *23*, 174–180.

Iso, T., Kedes, L., and Hamamori, Y. (2003). HES and HERP families: Multiple effectors of the Notch signaling pathway. J. Cell. Physiol. *194*, 237–255.

Jackstadt, R., Röh, S., Neumann, J., Jung, P., Hoffmann, R., Horst, D., Berens, C., Bornkamm, G.W., Kirchner, T., Menssen, A., et al. (2013). AP4 is a mediator of epithelial-mesenchymal transition and metastasis in colorectal cancer. J. Exp. Med. *210*, 1331–1350.

Jiang, S., Wang, L.W., Walsh, M.J., Trudeau, S.J., Gerdt, C., Zhao, B., and Gewurz, B.E. (2018). CRISPR/Cas9-Mediated Genome Editing in Epstein-Barr Virus-Transformed Lymphoblastoid B-Cell Lines. Curr. Protoc. Mol. Biol. 31.12.1-31.12.23.

Jin, S., Kim, J., Willert, T., Klein-Rodewald, T., Garcia-Dominguez, M., Mosqueira, M., Fink, R., Esposito, I., Hofbauer, L.C., Charnay, P., et al. (2014). Ebf factors and MyoD cooperate to regulate muscle relaxation via Atp2a1. Nat. Commun. *5*, 1–16.

Johannsen, E., Koh, E., Mosialos, G., Tong, X., Kieff, E., and Grossman, S.R. (1995). Epstein-Barr virus nuclear protein 2 transactivation of the latent membrane protein 1 promoter is mediated by J kappa and PU.1. J. Virol. *69*, 253–262.

Johnston, B.P., and McCormick, C. (2019). Herpesviruses and the unfolded protein response. Viruses *12*, 17.

Jolliffe, I.T., and Cadima, J. (2016). Principal component analysis: a review and recent developments. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. *374*, 20150202.

Kaiser, C., Laux, G., Eick, D., Jochner, N., Bornkamm, G.W., and Kempkes, B. (1999). The Proto-Oncogene c-myc Is a Direct Target Gene of Epstein-Barr Virus Nuclear Antigen 2. J. Virol. *73*, 4481–4484.

Kang, S., Akerblad, P., Kiviranta, R., Gupta, R.K., Kajimura, S., Griffin, M.J., Min, J., Baron, R., and Rosen, E.D. (2012). Regulation of Early Adipose Commitment by Zfp521. PLoS Biol. *10*, e1001433.

Kempkes, B., and Ling, P.D. (2015). EBNA2 and its coactivator EBNA-LP. In Current Topics in Microbiology and Immunology, (Springer Verlag), pp. 35–59.

Kempkes, B., Spitkovsky, D., Jansen-Dürr, P., Ellwart, J.W., Kremmer, E., Delecluse, H.J., Rottenberger, C., Bornkamm, G.W., and Hammerschmidt, W. (1995). B-cell proliferation and induction of early G1-regulating proteins by Epstein-Barr virus mutants conditional for EBNA2. EMBO J. *14*, 88.

Kieser, A., and Sterz, K.R. (2015). The latent membrane protein 1 (LMP1). In Current Topics in Microbiology and Immunology, (Springer Verlag), pp. 119–149.

Kohlhof, H., Hampel, F., Hoffmann, R., Burtscher, H., Weidle, U.H., Hölzel, M., Eick, D., Zimber-Strobl, U., and Strobl, L.J. (2009). Notch1, Notch2, and Epstein-Barr virus-encoded nuclear antigen 2 signaling differentially affects proliferation and survival of Epstein-Barr virus-infected B cells. Blood *113*, 5506–5515.

Kong, N.R., Davis, M., Chai, L., Winoto, A., and Tjian, R. (2016). MEF2C and EBF1 Co-regulate B Cell-Specific Transcription. PLoS Genet. *12*, 1–21.

Kudrycki, K., Stein-Izsak, C., Behn, C., Grillo, M., Akeson, R., and Margolis, F.L. (1993). Olf-l-Binding Site: Characterization of an Olfactory Neuron-Specific Promoter Motif.

Kulwichit, W., Edwards, R.H., Davenport, E.M., Baskar, J.F., Godfrey, V., and Raab-Traub, N. (1998). Expression of the Epstein-Barr virus latent membrane protein 1 induces B cell lymphoma in transgenic mice. Proc. Natl. Acad. Sci. *95*, 11963–11968.

Lam, N., and Sugden, B. (2003). CD40 and its viral mimic, LMP1: Similar means to different ends. Cell. Signal. *15*, 9–16.

Laux, G., Adam, B., Strobl, L.J., and Moreau-Gachelin, F. (1994). The Spi-1/PU.1 and Spi-B ets family transcription factors and the recombination signal binding protein RBP-J kappa interact with an Epstein-Barr virus nuclear antigen 2 responsive cis-element. EMBO J. *13*, 5624–5632.

Li, R., Cauchy, P., Ramamoorthy, S., Boller, S., Chavez, L., and Grosschedl, R. (2018). Dynamic EBF1 occupancy directs sequential epigenetic and transcriptional events in B-cell programming. Genes Dev. *32*, 96–111.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst. *1*, 417–425.

Lin, H., and Grosschedl, R. (1995). Failure of B-cell differentiation in mice lacking the transcription factor EBF. Nature *376*, 263–267.

Lin, J., Hou, Y., Huang, S., Wang, Z., Sun, C., Wang, Z., He, X., Tam, N.L., Wu, C., and Wu, L. (2019). Exportin-T promotes tumor proliferation and invasion in hepatocellular carcinoma. Mol. Carcinog. *58*, 293–304.

Lin, Y.C., Jhunjhunwala, S., Benner, C., Heinz, S., Welinder, E., Mansson, R., Sigvardsson, M., Hagman, J., Espinoza, C.A., Dutkowski, J., et al. (2010). A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. Nat. Immunol. *11*, 635–643.

Ling, P.D., Rawlins, D.R., and Hayward, S.D. (1993). The Epstein-Barr virus immortalizing protein EBNA-2 is targeted to DNA by a cellular enhancer-binding protein. Proc. Natl. Acad. Sci. U. S. A. *90*, 9237–9241.

Ling, X., Xie, B., Gao, X., Chang, L., Zheng, W., Chen, H., Huang, Y., Tan, L., Li, M., and Liu, T. (2020). Improving the efficiency of precise genome editing with site-specific Cas9-oligonucleotide conjugates. Sci. Adv. *6*, eaaz0051.

Liu, F., and Zhou, Z.H. (2007). Comparative virion structures of human herpesviruses. In Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis, (Cambridge University Press), pp. 27–43.

Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D., and Ma, W. (2019). Methodologies for improving HDR efficiency. Front. Genet. *10*, 691.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lu, F., Chen, H.-S., Kossenkov, A. V., DeWispeleare, K., Won, K.-J., and Lieberman, P.M. (2016). EBNA2 Drives Formation of New Chromosome Binding Sites and Target Genes for B-Cell Master Regulatory Transcription Factors RBP-jκ and EBF1. PLOS Pathog. *12*, e1005339.

Lukin, K., Fields, S., Hartley, J., and Hagman, J. (2008). Early B cell factor: Regulator of B lineage specification and commitment. Semin. Immunol. *20*, 221–227.

Maier, H., Ostraat, R., Gao, H., Fields, S., Shinton, S.A., Medina, K.L., Ikawa, T., Murre, C., Singh, H., Hardy, R.R., et al. (2004). Early B cell factor cooperates with Runx1 and mediates epigenetic changes associated with mb-1 transcription. Nat. Immunol. *5*, 1069–1077.

Maier, S., Staffler, G., Hartmann, A., Höck, J., Henning, K., Grabusic, K., Mailhammer, R., Hoffmann, R., Wilmanns, M., Lang, R., et al. (2006). Cellular target genes of Epstein-Barr virus nuclear antigen 2. J. Virol. *80*, 9761–9771.

Mansilla, S.F., De La Vega, M.B., Calzetta, N.L., Siri, S.O., and Gottifredi, V. (2020). CDK-Independent and PCNA-Dependent Functions of p21 in DNA Replication. Genes (Basel). *11*, 593.

Mansson, R., Welinder, E., Åhsberg, J., Lin, Y.C., Benner, C., Glass, C.K., Lucas, J.S., Sigvardsson, M., and Murre, C. (2012). Positive intergenic feedback circuitry, involving EBF1 and FOXO1, orchestrates B-cell fate. Proc. Natl. Acad. Sci. U. S. A. *109*, 21028–21033.

de Martel, C., Georges, D., Bray, F., Ferlay, J., and Clifford, G.M. (2020). Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. Lancet Glob. Heal. *8*, e180–e190.

McFadden, K., Hafez, A.Y., Kishton, R., Messinger, J.E., Nikitin, P.A., Rathmell, J.C., and Luftig, M.A. (2016). Metabolic stress is a barrier to Epstein-Barr virus-mediated B-cell immortalization. Proc. Natl. Acad. Sci. U. S. A. *113*, E782–E790.

Mckenzie, J., and El-Guindy, A. (2015). Epstein-Barr virus lytic cycle reactivation. In Current Topics in Microbiology and Immunology, (Springer Verlag), pp. 237–261.

McManus, S., Ebert, A., Salvagiotto, G., Medvedovic, J., Sun, Q., Tamir, I., Jaritz, M., Tagoh, H., and Busslinger, M. (2011). The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. EMBO J. *30*, 2388–2404.

Medina, K.L., Pongubala, J.M.R., Reddy, K.L., Lancki, D.W., DeKoter, R., Kieslinger, M., Grosschedl, R., and Singh, H. (2004). Assembling a gene regulatory network for specification of the B cell fate. Dev. Cell *7*, 607–617.

Melchers, F., Karasuyama, H., Haasner, D., Bauer, S., Kudo, A., Sakaguchi, N., Jameson, B., and Rolink, A. (1993). The surrogate light chain in B-cell development. Immunol. Today *14*, 60–68.

Mercer, E.M., Lin, Y.C., Benner, C., Jhunjhunwala, S., Dutkowski, J., Flores, M., Sigvardsson, M., Ideker, T., Glass, C.K., and Murre, C. (2011). Multilineage Priming of Enhancer Repertoires Precedes Commitment to the B and Myeloid Cell Lineages in Hematopoietic Progenitors. Immunity *35*, 413–425.

Miller, J.E., and Reese, J.C. (2012). Ccr4-Not complex: The control freak of eukaryotic cells. Crit. Rev. Biochem. Mol. Biol. *47*, 315–333.

Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273.

Mosmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. J. Immunol. Methods *65*, 55–63.

Mrozek-Gorska, P., Buschle, A., Pich, D., Schwarzmayr, T., Fechtner, R., Scialdone, A., and Hammerschmidt, W. (2019). Epstein-Barr virus reprograms human B lymphocytes immediately in the prelatent phase of infection. Proc. Natl. Acad. Sci. U. S. A.

Mullighan, C.G., Goorha, S., Radtke, I., Miller, C.B., Coustan-Smith, E., Dalton, J.D., Girtman, K., Mathew, S., Ma, J., Pounds, S.B., et al. (2007). Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature *446*, 758–764.

Murata, T., Noda, C., Narita, Y., Watanabe, T., Yoshida, M., Ashio, K., Sato, Y., Goshima, F., Kanda, T., Yoshiyama, H., et al. (2016). Induction of Epstein-Barr Virus Oncoprotein LMP1 by Transcription Factors AP-2 and Early B Cell Factor. J Virol *90*, 3873–3889.

Nemerow, G.R., Mold, C., Schwend, V.K., Tollefson, V., and Cooper, N.R. (1987). Identification of gp350 as the viral glycoprotein mediating attachment of Epstein-Barr virus (EBV) to the EBV/C3d receptor of B cells: sequence homology of gp350 and C3 complement fragment C3d. J. Virol. *61*, 1416–1420.

Neuhierl, B., Feederle, R., Hammerschmidt, W., and Delecluse, H.J. (2002). Glycoprotein gp110 of Epstein-Barr virus determines viral tropism and efficiency of infection. Proc. Natl. Acad. Sci. U. S. A. *99*, 15036–15041.

Nikitin, P.A., Yan, C.M., Forte, E., Bocedi, A., Tourigny, J.P., White, R.E., Allday, M.J., Patel, A., Dave, S.S., Kim, W., et al. (2010). 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.

Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K.M., Gregg, A., Noggle, S., and Tessier-Lavigne, M. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature *533*, 125–129.

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2018). zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. Gigascience *7*, 1–9.

Peng, C.W., Xue, Y., Zhao, B., Johannsen, E., Kieff, E., and Harada, S. (2004). Direct interactions between Epstein-Barr virus leader protein LP and the EBNA2 acidic domain underlie coordinate transcriptional regulation. Proc. Natl. Acad. Sci. U. S. A. *101*, 1033–1038.

Pich, D., Mrozek-Gorska, P., Bouvet, M., Sugimoto, A., Akidil, E., Grundhoff, A., Hamperl, S., Ling, P.D., and Hammerschmidt, W. (2019). First Days in the Life of Naive Human B Lymphocytes Infected with Epstein-Barr Virus. MBio *10*, 17.

Pluk, H., Soffner, J., Lührmann, R., and van Venrooij, W.J. (1998). cDNA Cloning and Characterization of the Human U3 Small Nucleolar Ribonucleoprotein Complex-Associated 55-Kilodalton Protein. Mol. Cell. Biol. *18*, 488–498.

Polack, A., Hortnagel, K., Pajic, A., Christoph, B., Baier, B., Falk, M., Mautner, J., Geltinger, C., Bornkamm, G.W., and Kempkes, B. (1996). c-myc activation renders proliferation of Epstein-Barr virus (EBV)-transformed cells independent of EBV nuclear antigen 2 and latent membrane protein 1. Proc Natl Acad Sci U S A *93*, 10411–10416.

Pongubala, J.M.R., Northrup, D.L., Lancki, D.W., Medina, K.L., Treiber, T., Bertolino, E., Thomas, M., Grosschedl, R., Allman, D., and Singh, H. (2008). Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5. Nat. Immunol. *9*, 203–215.

Pope, J.H., Horne, M.K., and Scott, W. (1968). Transformation of foetal human leukocytesin vitro by filtrates of a human leukaemic cell line containing herpes-like virus. Int. J. Cancer *3*, 857–866.

Quentmeier, H., Pommerenke, C., Dirks, W.G., Eberth, S., Koeppel, M., MacLeod, R.A.F., Nagel, S., Steube, K., Uphoff, C.C., and Drexler, H.G. (2019). The LL-100 panel: 100 cell lines for blood cancer studies. Sci. Rep. *9*, 1–14.

Raemaekers, T., Ribbeck, K., Beaudouin, J., Annaert, W., Van Camp, M., Stockmans, I., Smets, N., Bouillon, R., Ellenberg, J., and Carmeliet, G. (2003). NuSAP, a novel microtubule-associated protein involved in mitotic spindle organization. J. Cell Biol. *162*, 1017–1029.

Ramamoorthy, S., Kometani, K., Herman, J.S., Bayer, M., Boller, S., Edwards-Hicks, J., Ramachandran, H., Li, R., Klein-Geltink, R., Pearce, E.L., et al. (2020). EBF1 and Pax5 safeguard leukemic transformation by limiting IL-7 signaling, Myc expression, and folate metabolism. Genes Dev. *34*, 1503–1519.

Reynaud, D., Demarco, I.A., Reddy, K.L., Schjerven, H., Bertolino, E., Chen, Z., Smale, S.T., Winandy, S., and Singh, H. (2008). Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. Nat. Immunol. *9*, 927–936.

Roberts, K.G., Morin, R.D., Zhang, J., Hirst, M., Zhao, Y., Su, X., Chen, S.C., Payne-Turner, D., Churchman, M.L., Harvey, R.C., et al. (2012). Genetic Alterations Activating Kinase and Cytokine Receptor Signaling in High-Risk Acute Lymphoblastic Leukemia. Cancer Cell *22*, 153–166.

Roessler, S., and Grosschedl, R. (2006). Role of transcription factors in commitment and differentiation of early B lymphoid cells. Semin. Immunol. *18*, 12–19.

Roessler, S., György, I., Imhof, S., Spivakov, M., Williams, R.R., Busslinger, M., Fisher, A.G., and Grosschedl, R. (2007). Distinct Promoters Mediate the Regulation of Ebf1 Gene Expression by Interleukin-7 and Pax5. Mol. Cell. Biol. *27*, 579–594.

Ron, D., and Walter, P. (2007). Signal integration in the endoplasmic reticulum unfolded protein response. Nat. Rev. Mol. Cell Biol. *8*, 519–529.

Roth, T.L., Puig-Saus, C., Yu, R., Shifrut, E., Carnevale, J., Li, P.J., Hiatt, J., Saco, J., Krystofinski, P., Li, H., et al. (2018). Reprogramming human T cell function and specificity with non-viral genome targeting. Nature *559*, 405–409.

Schuhmacher, M., Staege, M.S., Pajic, A., Polack, A., Weidle, U.H., Bornkamm, G.W., Eick, D., and Kohlhuber, F. (1999). Control of cell growth by c-Myc in the absence of cell division. Curr. Biol. *9*, 1255–1258.

Seet, C.S., Brumbaugh, R.L., and Kee, B.L. (2004). Early B cell factor promotes B lymphopoiesis with reduced interleukin 7 responsiveness in the absence of E2A. J. Exp. Med. *199*, 1689–1700.

Sigvardsson, M., O'Riordan, M., and Grosschedl, R. (1997). EBF and E47 collaborate to induce expression of the endogenous immunoglobulin surrogate light chain genes. Immunity *7*, 25–36.

Sigvardsson, M., Clark, D.R., Fitzsimmons, D., Doyle, M., Åkerblad, P., Breslin, T., Bilke, S., Li, R., Yeamans, C., Zhang, G., et al. (2002). Early B-Cell Factor, E2A, and Pax-5 Cooperate To Activate the Early B Cell-Specific mb-1 Promoter. Mol. Cell. Biol. *22*, 8539–8551.

Sinclair, A.J., Palmero, I., Peters, G., and Farrell, P.J. (1994). EBNA-2 and EBNA-LP cooperate to cause G0 to G1 transition during immortalization of resting human B lymphocytes by Epstein-Barr virus. EMBO J. *13*, 3321–3328.

Siponen, M.I., Wisniewska, M., Lehtiö, L., Johansson, I., Svensson, L., Raszewski, G., Nilsson, L., Sigvardsson, M., and Berglund, H. (2010). Structural Determination of Functional Domains in Early B-cell Factor (EBF) Family of Transcription Factors Reveals Similarities to Rel DNA-binding Proteins and a Novel Dimerization Motif. J. Biol. Chem. *285*, 25875–25879.

Sjoblom, A., Jansson, A., Yang, W., Lain, S., Nilsson, T., and Rymo, L. (1995). PU box-binding transcription factors and a POU domain protein cooperate in the Epstein-Barr virus (EBV) nuclear antigen 2-induced transactivation of the EBV latent membrane protein 1 promoter. J. Gen. Virol. *76*, 2679–2692.

Smatti, M.K., Al-Sadeq, D.W., Ali, N.H., Pintus, G., Abou-Saleh, H., and Nasrallah, G.K. (2018). Epstein-barr virus epidemiology, serology, and genetic variability of LMP-1 oncogene among healthy population: An update. Front. Oncol. *8*.

Somasundaram, R., Prasad, M.A.J., Ungerbäck, J., and Sigvardsson, M. (2015). Transcription factor networks in B-cell differentiation link development to acute lymphoid leukemia. Blood *126*, 144–152.

Sommermann, T.G., O'Neill, K., Plas, D.R., and Cahir-McFarland, E. (2011). IKKβ and NF-κB transcription govern lymphoma cell survival through AKT-induced plasma membrane trafficking of GLUT1. Cancer Res. *71*, 7291–7300.

Spender, L.C., Cannell, E.J., Hollyoake, M., Wensing, B., Gawn, J.M., Brimmell, M., Packham, G., and Farrell, P.J. (1999). Control of Cell Cycle Entry and Apoptosis in B Lymphocytes Infected by Epstein-Barr Virus. J. Virol. *73*, 4678–4688.

Strobl, L.J., Höfelmayr, H., Stein, C., Marschall, G., Brielmeier, M., Laux, G., Bornkamm, G.W., and Zimber-Strobl, U. (1997). Both Epstein-Barr Viral Nuclear Antigen 2 (EBNA2) and Activated Notch1 Transactivate Genes by Interacting with the Cellular Protein RBP-Jκ. Immunobiology *198*, 299–306.

Styles, C.T., Bazot, Q., Parker, G.A., White, R.E., Paschos, K., and Allday, M.J. (2017). EBV epigenetically suppresses the B cell-to-plasma cell differentiation pathway while establishing long-term latency. PLoS Biol. *15*, 1–30.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.

Szymula, A., Palermo, R.D., Bayoumy, A., Groves, I.J., Ba abdullah, M., Holder, B., and White, R.E. (2018). Epstein-Barr virus nuclear antigen EBNA-LP is essential for transforming naïve B cells, and facilitates recruitment of transcription factors to the viral genome. PLoS Pathog. *14*.

Tanner, J., Weis, J., Fearon, D., Whang, Y., and Kieff, E. (1987). Epstein-barr virus gp350/220 binding to the B lymphocyte C3d receptor mediates adsorption, capping, and endocytosis. Cell *50*, 203–213.

Tarakanova, V.L., Leung-Pineda, V., Hwang, S., Yang, C.W., Matatall, K., Basson, M., Sun, R., Piwnica-Worms, H., Sleckman, B.P., and Virgin IV, H.W. (2007). γ-Herpesvirus Kinase Actively Initiates a DNA Damage Response by Inducing Phosphorylation of H2AX to Foster Viral Replication. Cell Host Microbe *1*, 275–286.

Taylor, G.S., Long, H.M., Brooks, J.M., Rickinson, A.B., and Hislop, A.D. (2015). The Immunology of Epstein-Barr Virus-Induced Disease. Annu Rev Immunol *33*, 787–821.

Tempera, I., and Lieberman, P.M. (2014). Epigenetic regulation of EBV persistence and oncogenesis. Semin. Cancer Biol. *26*, 22–29.

Thal, M.A., Carvalho, T.L., He, T., Kim, H.G., Gao, H., Hagman, J., and Klug, C.A. (2009). Ebf1-mediated down-regulation of Id2 and Id3 is essential for specification of the B cell lineage. Proc. Natl. Acad. Sci. U. S. A. *106*, 552–557.

Thorley-Lawson, D.A. (2015). EBV Persistence-Introducing the Virus. Curr Top Microbiol Immunol *390*, 151–209.

Tomkinson, B., Robertson, E., and Kieff, E. (1993). Epstein-Barr virus nuclear proteins EBNA-3A and EBNA-3C are essential for B-lymphocyte growth transformation. J. Virol. *67*, 2014–2025.

Tong, X., Drapkin, R., Yalamanchili, R., Mosialos, G., and Kieff, E. (1995a). The Epstein-Barr virus nuclear protein 2 acidic domain forms a complex with a novel cellular coactivator that can interact with TFIIE. Mol. Cell. Biol. *15*, 4735–4744.

Tong, X., Wang, F., Thut, C.J., and Kieff, E. (1995b). The Epstein-Barr virus nuclear protein 2 acidic domain can interact with TFIIB, TAF40, and RPA70 but not with TATA-binding protein. J. Virol. *69*, 585–588.

Travis, A., Hagman, J., Hwang, L., and Grosschedl, R. (1993). Purification of early-B-cell factor and characterization of its DNA-binding specificity. Mol. Cell. Biol. *13*, 3392–3400.

Treiber, N., Treiber, T., Zocher, G., and Grosschedl, R. (2010a). Structure of an Ebf1 : DNA complex reveals unusual DNA recognition and structural homology with Rel proteins. Genes Dev. *24*, 2270–2275.

Treiber, T., Mandel, E.M., Pott, S., Györy, I., Firner, S., Liu, E.T., and Grosschedl, R. (2010b). Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. Immunity *32*, 714–725.

Tsai, R.Y.L., and Reed, R.R. (1997). Cloning and functional characterization of Roaz, a zinc finger protein that interacts with O/E-1 to regulate gene expression: Implications for olfactory neuronal development. J. Neurosci. *17*, 4159–4169.

Vilagos, B., Hoffmann, M., Souabni, a., Sun, Q., Werner, B., Medvedovic, J., Bilic, I., Minnich, M., Axelsson, E., Jaritz, M., et al. (2012). Essential role of EBF1 in the generation and function of distinct mature B cell types. J. Exp. Med. *209*, 775–792.

Vousden, K.H., and Lane, D.P. (2007). p53 in health and disease. Nat. Rev. Mol. Cell Biol. *8*, 275–283.

Wadsworth, W.G., and Hedgecock, E.M. (1996). Hierarchical guidance cues in the developing nervous system of C. elegans. BioEssays *18*, 355–362.

Wang, M.M., and Reed, R.R. (1993). Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. Nature *364*, 121–126.

Wang, L., Grossman, S.R., and Kieff, E. (2000). Epstein-Barr virus nuclear protein 2 interacts with p300, CBP, and PCAF histone acetyltransferases in activation of the LMP1 promoter. Proc. Natl. Acad. Sci. U. S. A. *97*, 430–435.

Wang, L.W., Shen, H., Nobre, L., Ersing, I., Paulo, J.A., Trudeau, S., Wang, Z., Smith, N.A., Ma, Y., Reinstadler, B., et al. (2019a). Epstein-Barr-Virus-Induced One-Carbon Metabolism Drives B Cell Transformation. Cell Metab. *30*, 539-555.e11.

Wang, L.W., Wang, Z., Ersing, I., Nobre, L., Guo, R., Jiang, S., Trudeau, S., Zhao, B., Weekes, M.P., and Gewurz, B.E. (2019b). Epstein-Barr virus subverts mevalonate and fatty acid pathways to promote infected B-cell proliferation and survival. PLoS Pathog. *15*, e1008030.

Wang, S., Zhao, Y., Leiby, M., and Zhu, J. (2009). A new positive/negative selection scheme for precise BAC recombineering. Mol. Biotechnol. *42*, 110–116.

Wang, Y., Zolotarev, N., Yang, C.Y., Rambold, A., Mittler, G., and Grosschedl, R. (2020). A Prion-like Domain in Transcription Factor EBF1 Promotes Phase Separation and Enables B Cell Programming of Progenitor Chromatin. Immunity *53*, 1151-1167.e6.

Warming, S., Costantino, N., Court, D.L., Jenkins, N.A., and Copeland, N.G. (2005). Simple and highly efficient BAC recombineering using galK selection. Nucleic Acids Res *33*.

Whitaker, A.M., and Freudenthal, B.D. (2018). APE1: A skilled nucleic acid surgeon. DNA Repair (Amst). *71*, 93–100.

Wiesner, M., Zentz, C., Mayr, C., Wimmer, R., and Hammerschmidt, W. (2008a). Conditional Immortalization of Human B Cells by CD40 Ligation. PLoS One *3*, 1464.

Wiesner, M., Zentz, C., Mayr, C., Wimmer, R., Hammerschmidt, W., Zeidler, R., and Moosmann, A. (2008b). Conditional Immortalization of Human B Cells by CD40 Ligation. PLoS One *3*, e1464.

Woellmer, A., and Hammerschmidt, W. (2013). Epstein-Barr virus and host cell methylation: Regulation of latency, replication and virus reactivation. Curr. Opin. Virol. *3*, 260–265.

Woisetschlaeger, M., Yandava, C.N., Furmanski, L.A., Strominger, J.L., and Speck, S.H. (1990). Promoter switching in Epstein-Barr virus during the initial stages of infection of B lymphocytes. Proc. Natl. Acad. Sci. U. S. A. *87*, 1725–1729.

Wu, D.Y., Krumm, A., and Schubach, W.H. (2000). Promoter-Specific Targeting of Human SWI-SNF Complex by Epstein-Barr Virus Nuclear Protein 2. J. Virol. *74*, 8893–8903.

Xing, M., Ooi, W.F., Tan, J., Qamra, A., Lee, P.H., Li, Z., Xu, C., Padmanabhan, N., Lim, J.Q., Guo, Y.A., et al. (2020). Genomic and epigenomic EBF1 alterations modulate TERT expression in gastric cancer. J. Clin. Invest. *130*, 3005–3020.

Yang, C.Y., Ramamoorthy, S., Boller, S., Rosenbaum, M., Gil, A.R., Mittler, G., Imai, Y., Kuba, K., and Grosschedl, R. (2016). Interaction of CCR4–NOT with EBF1 regulates gene-specific transcription and mRNA stability in B lymphopoiesis. Genes Dev. *30*, 2310–2324.

Yen, T.J., Compton, D.A., Wise, D., Zinkowski, R.P., Brinkley, B.R., Earnshaw, W.C., and Cleveland, D.W. (1991). CENP-E, a novel human centromere-associated protein required for progression from metaphase to anaphase. EMBO J. *10*, 1245–1254.

Yin, H., Qu, J., Peng, Q., and Gan, R. (2019). Molecular mechanisms of EBV-driven cell cycle progression and oncogenesis. Med. Microbiol. Immunol. *208*, 573–583.

Yue, W., Davenport, M.G., Shackelford, J., and Pagano, J.S. (2004). Mitosis-Specific Hyperphosphorylation of Epstein-Barr Virus Nuclear Antigen 2 Suppresses Its Function. J. Virol. *78*, 3542–3552.

Zhang, Z., Cotta, C. V., Stephan, R.P., De Guzman, C.G., and Klug, C.A. (2003). Enforced expression of EBF in hematopoietic stem cells restricts lymphopoiesis to the B cell lineage. EMBO J. *22*, 4759–4769.

Zhao, B., Zou, J., Wang, H., Johannsen, E., Peng, C., Quackenbush, J., Mar, J.C., Morton, C.C., Freedman, M.L., Blacklow, S.C., et al. (2011). Epstein-Barr virus exploits intrinsic B-lymphocyte transcription programs to achieve immortal cell growth. Proc. Natl. Acad. Sci. U. S. A. *108*, 14902–14907.

Zhao, F., McCarrick-Walmsley, R., Akerblad, P., Sigvardsson, M., and Kadesch, T. (2003). Inhibition of p300/CBP by early B-cell factor. Mol. Cell. Biol. *23*, 3837–3846.

Zhou, H., Schmidt, S.C.S., Jiang, S., Willox, B., Bernhardt, K., Liang, J., Johannsen, E.C., Kharchenko, P., Gewurz, B.E., Kieff, E., et al. (2015). Epstein-barr virus oncoprotein super-enhancers control B cell growth. Cell Host Microbe *17*, 205–216.

6 Appendix

6.1 Experimental data

Table S 1 Corresponding to Figure 9D: Cell cycle distribution of EBVwt infected B cells. Primary B cells were isolated from 3 adenoid samples, infected with EBVwt and the cell cycle distribution analyzed by BrdU incorporation and 7-AAD staining on day 0 (uninfected B cells), 2, 4, 6, 8 post infection. The mean was used to generate the plot in Figure 9D. d.p.i. – days post infection

EBVwt					
Donor 6					
d.p.i.	apoptotic	G0/1	early S	S	G2/M
0	0.10	90.74	1.87	5.02	2.26
2	0.30	70.50	19.06	7.91	2.23
4	2.96	52.10	17.78	18.92	8.25
6	2.82	66.96	9.69	13.82	6.70
8	5.19	75.16	6.99	7.29	5.38
Donor 10					
d.p.i.	apoptotic	G0/1	early S	S	G2/M
0	0.15	91.10	2.52	2.61	3.62
2	2.56	73.49	14.74	6.75	2.47
4	4.89	67.91	4.43	11.44	11.33
6	4.37	78.04	3.06	6.81	7.72
8	4.85	82.82	2.54	3.49	6.30
Donor 21					
d.p.i.	apoptotic	G0/1	early S	S	G2/M
0	1.09	90.58	1.89	3.54	2.89
2	3.32	66.66	23.09	5.28	1.64
4	5.63	76.39	6.66	6.35	4.98
6	8.62	75.31	7.75	4.08	4.23
8	11.86	75.37	4.59	1.93	6.25
mean					
d.p.i.	apoptotic	G0/1	early S	S	G2/M
0	0.45	90.81	2.09	3.73	2.92
2	2.06	70.20	18.98	6.64	2.11
4	4.50	65.56	9.58	12.18	8.18
6	5.28	73.44	6.84	8.23	6.21
8	7.31	77.82	4.69	4.21	5.98

Table S 2 Corresponding to Figure 9D: Cell cycle distribution of EBVΔα1 infected B cells. Primary B cells were isolated from 3 adenoid samples, infected with EBVΔα1 and the cell cycle distribution analyzed by BrdU incorporation and 7-AAD staining on day 0 (uninfected B cells), 2, 4, 6, 8 post infection. The mean was used to generate the plot in Figure 9D. d.p.i. – days post infection

EBVΔα1					
Donor 6					
d.p.i.	apoptotic	G0/1	early S	S	G2/M
0	0.10	90.74	1.87	5.02	2.26
2	0.25	75.00	20.26	3.32	1.18
4	4.38	62.04	25.00	6.39	2.18
6	7.46	62.11	19.72	6.18	4.53
8	17.40	54.47	20.64	5.43	2.06
Donor 10					
d.p.i.	apoptotic	G0/1	early S	S	G2/M
0	0.15	91.10	2.52	2.61	3.62
2	2.93	76.15	15.44	3.64	1.84
4	6.61	77.40	7.55	2.90	5.54
6	9.11	70.80	11.85	3.35	4.89
8	10.84	74.53	8.86	1.57	4.21
Donor 21					
d.p.i.	apoptotic	G0/1	early S	S	G2/M
0	1.09	90.58	1.89	3.54	2.89
2	3.62	68.66	20.83	2.96	3.92
4	6.64	66.51	20.25	3.41	3.19
6	9.43	59.43	25.12	3.78	2.23
8	11.59	53.32	28.12	5.80	1.17
mean					
d.p.i.	apoptotic	G0/1	early S	S	G2/M
0	0.45	90.81	2.09	3.73	2.92
2	2.27	73.25	18.86	3.30	2.32
4	5.88	68.64	17.62	4.23	3.63
6	8.67	64.07	18.95	4.44	3.87
8	13.20	60.81	19.26	4.26	2.48

Table S 3 Corresponding to Figure 10C: Cell cycle distribution of uninfected B cells. Primary B cells were isolated from 3 adenoid samples and cultured with (+) or without (-) CD40L expressing feeder cells. The cell cycle distribution of uninfected B cells was analyzed by PI staining of the nuclei and FACS analysis on day 0 (uninfected B cells), 2, 4, 6, 8 post infection. The mean was used to generate the plot in Figure 10C. d.p.i. – days post infection

- CD40L feeder cells					+ CD40L feeder cells				
Donor 8					**Donor 8**				
d.p.i.	apoptotic	G0/1	S	G2/M	d.p.i.	apoptotic	G0/1	S	G2/M
0	0.14	97.93	1.44	0.49	0	0.14	97.93	1.44	0.49
2	10.33	85.16	3.63	0.87	2	13.60	72.25	9.51	4.64
4	16.33	79.14	3.59	0.94	4	9.05	64.10	14.80	12.04
6	22.70	67.70	6.81	2.79	6	11.06	78.92	6.49	3.53
8	26.69	67.23	4.30	1.78	8	19.50	70.78	4.91	4.81
Donor 9					**Donor 9**				
d.p.i.	apoptotic	G0/1	S	G2/M	d.p.i.	apoptotic	G0/1	S	G2/M
0	0.18	97.90	1.43	0.49	0	0.18	97.90	1.43	0.49
2	14.44	80.53	3.67	1.35	2	12.66	73.05	9.22	5.07
4	19.33	76.01	3.54	1.12	4	6.98	65.16	15.78	12.09
6	24.07	67.47	5.52	2.93	6	12.30	77.30	6.04	4.36
8	26.65	66.67	4.02	2.65	8	19.94	69.52	4.98	5.57
Donor 10					**Donor 10**				
d.p.i.	apoptotic	G0/1	S	G2/M	d.p.i.	apoptotic	G0/1	S	G2/M
0	0.21	97.69	1.63	0.47	0	0.21	97.69	1.63	0.47
2	9.16	86.24	3.89	0.71	2	12.99	73.38	8.84	4.79
4	15.36	79.83	3.62	1.18	4	9.26	65.20	14.51	11.04
6	24.50	66.33	6.28	2.89	6	10.70	78.50	6.60	4.21
8	28.07	66.17	4.08	1.67	8	18.67	71.41	5.33	4.59
mean					**mean**				
d.p.i.	apoptotic	G0/1	S	G2/M	d.p.i.	apoptotic	G0/1	S	G2/M
0	0.18	97.84	1.50	0.48	0	0.18	97.84	1.50	0.48
2	12.40	79.70	5.67	2.23	2	13.08	72.89	9.19	4.83
4	17.01	78.32	3.59	1.08	4	8.43	64.82	15.03	11.72
6	23.76	67.16	6.20	2.87	6	11.35	78.24	6.38	4.03
8	27.14	66.69	4.14	2.03	8	19.37	70.57	5.07	4.99

Table S 4 Corresponding to Figure 10C: Cell cycle distribution of EBVwt infected B cells. Primary B cells were isolated from 3 adenoid samples, infected with EBVwt and cultured with (+) or without (-) CD40L expressing feeder cells. The cell cycle distribution of EBVwt infected B cells was analyzed by PI staining of the nuclei and FACS analysis on day 0 (uninfected B cells), 2, 4, 6, 8 post infection. The mean was used to generate the plot in Figure 10C. d.p.i. – days post infection

- CD40L feeder cells					+ CD40L feeder cells				
Donor 8					**Donor 8**				
d.p.i.	apoptotic	G0/1	S	G2/M	d.p.i.	apoptotic	G0/1	S	G2/M
0	0.14	97.93	1.44	0.49	0	0.14	97.93	1.44	0.49
2	5.85	85.99	5.34	2.82	2	7.55	73.50	12.03	6.91
4	9.66	70.71	10.82	8.80	4	6.91	67.54	16.81	8.74
6	12.57	68.51	10.64	8.28	6	18.03	63.22	11.55	7.20
8	17.71	66.19	9.05	7.05	8	29.79	55.41	7.67	7.13
Donor 9					**Donor 9**				
d.p.i.	apoptotic	G0/1	S	G2/M	d.p.i.	apoptotic	G0/1	S	G2/M
0	0.18	97.90	1.43	0.49	0	0.18	97.90	1.43	0.49
2	9.49	82.18	5.55	2.78	2	6.91	72.73	13.06	7.30
4	17.93	70.52	6.89	4.65	4	11.40	60.67	15.59	12.34
6	12.65	68.01	10.53	8.81	6	15.55	64.93	11.81	7.70
8	19.34	64.30	8.59	7.78	8	30.47	55.65	7.11	6.77
Donor 10					**Donor 10**				
d.p.i.	apoptotic	G0/1	S	G2/M	d.p.i.	apoptotic	G0/1	S	G2/M
0	0.21	97.69	1.63	0.47	0	0.21	97.69	1.63	0.47
2	5.65	86.13	5.39	2.83	2	11.80	76.43	8.47	3.30
4	9.68	70.95	10.93	8.44	4	7.05	67.22	18.02	7.71
6	12.84	68.23	10.92	8.02	6	17.73	63.01	11.14	8.12
8	17.61	65.99	9.08	7.32	8	30.24	55.80	7.66	6.30
mean					**mean**				
d.p.i.	apoptotic	G0/1	S	G2/M	d.p.i.	apoptotic	G0/1	S	G2/M
0	0.18	97.84	1.50	0.48	0	0.18	97.84	1.50	0.48
2	7.00	84.77	5.43	2.81	2	8.76	74.23	11.18	5.83
4	12.43	70.73	9.54	7.30	4	8.42	65.19	16.82	9.57
6	12.69	68.25	10.69	8.37	6	17.10	63.72	11.50	7.68
8	18.22	65.49	8.90	7.38	8	30.17	55.62	7.48	6.73

Table S 5 Corresponding to Figure 10C: Cell cycle distribution of EBVΔα1 infected B cells. Primary B cells were isolated from 3 adenoid samples, infected with EBVΔα1 and cultured with (+) or without (-) CD40L expressing feeder cells. The cell cycle distribution of EBVwt infected B cells was analyzed by PI staining of the nuclei and FACS analysis on day 0 (uninfected B cells), 2, 4, 6, 8 post infection. The mean was used to generate the plot in Figure 10C. d.p.i. – days post infection

- CD40L feeder cells					+ CD40L feeder cells				
Donor 8					**Donor 8**				
d.p.i.	apoptotic	G0/1	S	G2/M	d.p.i.	apoptotic	G0/1	S	G2/M
0	0.14	97.93	1.44	0.49	0	0.14	97.93	1.44	0.49
2	9.52	83.53	4.63	2.32	2	10.44	58.55	19.96	11.05
4	14.69	75.36	5.12	4.83	4	6.34	71.03	16.12	6.52
6	18.99	71.12	5.21	4.68	6	13.78	73.30	10.06	2.87
8	20.44	71.90	4.08	3.58	8	16.80	70.51	8.69	4.00
Donor 9					**Donor 9**				
d.p.i.	apoptotic	G0/1	S	G2/M	d.p.i.	apoptotic	G0/1	S	G2/M
0	0.18	97.90	1.43	0.49	0	0.18	97.90	1.43	0.49
2	10.45	82.20	4.80	2.54	2	18.47	53.08	17.09	11.36
4	14.49	74.06	6.38	5.08	4	7.52	68.80	14.21	9.47
6	15.66	73.16	5.58	5.60	6	11.96	75.11	10.04	2.89
8	18.82	72.97	4.19	4.02	8	17.88	69.79	8.66	3.68
Donor 10					**Donor 10**				
d.p.i.	apoptotic	G0/1	S	G2/M	d.p.i.	apoptotic	G0/1	S	G2/M
0	0.21	97.69	1.63	0.47	0	0.21	97.69	1.63	0.47
2	9.24	83.75	4.60	2.40	2	9.57	59.33	19.54	11.56
4	15.14	74.50	5.73	4.62	4	6.24	71.14	16.45	6.18
6	20.62	70.11	5.31	3.96	6	13.78	73.41	10.04	2.78
8	21.32	71.00	3.97	3.71	8	16.92	70.98	8.72	3.38
mean					**mean**				
d.p.i.	apoptotic	G0/1	S	G2/M	d.p.i.	apoptotic	G0/1	S	G2/M
0	0.18	97.84	1.50	0.48	0	0.18	97.84	1.50	0.48
2	9.74	83.16	4.68	2.42	2	12.76	57.03	18.88	11.32
4	14.77	74.64	5.74	4.84	4	6.70	70.33	15.59	7.38
6	18.43	71.46	5.37	4.74	6	13.17	73.94	10.05	2.84
8	20.19	71.96	4.08	3.77	8	17.20	70.43	8.69	3.69

Table S 6 List of significantly enriched gene sets for the GSEA of EBVwt vs uninfected analysis on the indicated days post infection; UPR – unfolded protein response, EMT – epithelial to mesenchymal transition, DN – down

enriched gene sets (FDR < 0.1) for upregulated expression in EBVwt infected B cells compared to uninfected				enriched gene sets (FDR < 0.1) for downregulated expression in EBVwt infected B cells compared to uninfected			
day 1	day 2	day 3	day 4	day 1	day 2	day 3	day 4
MYC target V1	MYC target V1	MYC target V1	MYC target V1	interferon gamma response	interferon gamma response	TNFα signaling via NFκB	TNFα signaling via NFκB
MYC target V2	MYC target V2	oxidative phosphorylation	E2F targets	KRAS signaling up	KRAS signaling up	KRAS signaling up	UV response DN
UPR	oxidative phosphorylation	E2F targets	oxidative phosphorylation	interferon alpha response	allograft rejection	UV response DN	allograft rejection
MTORC1 signaling	MTORC1 signaling	MYC target V2	MYC target V2	allograft rejection	TNFα signaling via NFκB	allograft rejection	KRAS signaling up
	UPR	MTORC1 signaling	MTORC1 signaling	complement	UV response DN	IL6 JAK STAT3 signaling	inflammatory response
	glycolysis	glycolysis	G2M checkpoint	TNFα signaling via NFκB	mitotic spindle	inflammatory response	TGFβ signaling
	cholesterol homeostasis	G2M checkpoint	glycolysis	EMT	apoptosis	interferon gamma response	IL6 JAK STAT3 signaling
	fatty acid metabolism	adipogenesis	fatty acid metabolism	mitotic spindle	EMT	myogenesis	IL2 STAT5 signaling
	adipogenesis	reactive oxygen species pathway	DNA repair	apical junction	TGFβ signaling	apoptosis	estrogen response early
	reactive oxygen species pathway	UPR	reactive oxygen species pathway	apoptosis	interferon alpha response	p53 pathway	myogenesis
	DNA repair	DNA repair	xenobiotic metabolism	coagulation	inflammatory response	heme metabolism	
	xenobiotic metabolism	fatty acid metabolism	adipogenesis	peroxisome	IL6 JAK STAT3 signaling	mitotic spindle	
		xenobiotic metabolism	spermatogenesis	bile acid metabolism	heme metabolism	EMT	
		estrogen response late	UPR	IL6 JAK STAT3 signaling	myogenesis	IL2 STAT5 signaling	
		cholesterol homeostasis	interferon alpha response	inflammatory response	KRAS signaling DN	complement	
		spermatogenesis	cholesterol homeostasis	hypoxia	apical junction		
			estrogen response late	myogenesis	complement		

Table S 7 List of significantly enriched gene sets for the GSEA of EBVα1 vs uninfected analysis on the indicated days post infection; UPR – unfolded protein response, EMT – epithelial to mesenchymal transition, DN – down

enriched gene sets (FDR < 0.1) for upregulated gene expression in EBVΔα1 infected B cells compared to uninfected				enriched gene sets (FDR < 0.1) for downregulated gene expression in EBVΔα1 infected B cells compared to uninfected			
day 1	day 2	day 3	day 4	day 1	day 2	day 3	day 4
MYC target V1	MYC target V1	MYC target V1	oxidative phosphorylation	KRAS signaling up	allograft rejection	interferon gamma response	TNFα signaling via NFκB
MYC target V2	oxidative phosphorylation	oxidative phosphorylation	E2F targets	allograft rejection	interferon gamma response	allograft rejection	KRAS signaling up
UPR	MYC target V2	E2F targets	MYC target V1	interferon gamma response	KRAS signaling up	interferon alpha response	interferon gamma response
MTORC1 signaling	reactive oxygen species pathway	MYC target V2	UPR	complement	interferon alpha response	TNFα signaling via NFκB	inflammatory response
cholesterol homeostasis	UPR	MTORC1 signaling	G2M checkpoint	TNFα signaling via NFκB	TNFα signaling via NFκB	KRAS signaling up	allograft rejection
DNA repair	cholesterol homeostasis	cholesterol homeostasis	DNA repair	apoptosis	inflammatory response	inflammatory response	interferon alpha response
	MTORC1 signaling	adipogenesis	cholesterol homeostasis	mitotic spindle	myogenesis	IL6 JAK STAT3 signaling	UV response DN
	DNA repair	reactive oxygen species pathway	adipogenesis	IL2 STAT5 signaling	apoptosis	hypoxia	IL6 JAK STAT3 signaling
		xenobiotic metabolism	mitotic spindle	G2M checkpoint	IL6 JAK STAT3 signaling		hypoxia
		UPR	glycolysis	EMT	UV response DN		
		coagulation	xenobiotic metabolism	apical junction	IL2 STAT5 signaling		
			estrogen response late	hypoxia			
			spermatogenesis	UV response DN			
			MTORC1 signaling	bile acid metabolism			
			coagulation	peroxisome			
			fatty acid metabolism				
			EMT				

Table S 8 List of significantly enriched gene sets for the GSEA of EBVα1 vs uninfected analysis on the indicated days post infection; UPR – unfolded protein response, EMT – epithelial to mesenchymal transition, DN – down

enriched gene sets (FDR < 0.1) for genes stronger expressed in EBVΔα1 infected cells compared to EBVwt infected cells				enriched gene sets (FDR < 0.1) for genes stronger expressed in EBVwt infected cells compared to EBVΔα1 infected cells			
day 1	day 2	day 3	day 4	day 1	day 2	day 3	day 4
interferon alpha response	mitotic spindle	UV response DN	UV response DN	estrogen response early	MYC target V1	MYC target V1	MYC target V1
	ETM	TGFβ signaling	TNFα signaling via NFκB	MYC target V2	MYC target V2	E2F targets	E2F targets
	TGF ß signaling	TNFα singling	TGF ß signaling	estrogen response late	MTORC1 signaling	MYC target V2	MYC target V2
	WNTβ catenin signaling	ETM	heme metabolism		oxidative phosphorylation	oxidative phosphorylation	interferon alpha response
	UV response DN	apical junction	mitotic spindle		glycolysis	MTORC1 signaling	oxidative phosphorylation
	p53 pathway	apoptosis	myogenesis		UPR	interferon alpha response	MTORC1 signaling
	IL2_STAT5 signaling	coagulation			fatty acid metabolism	DNA repair	interferon gamma response
		estrogen response early			adipogenesis	G2M checkpoint	fatty acid metabolism
		p53 pathway			E2F targets	glycolysis	glycolysis
		mitotic spindle				reactive oxygen species pathway	G2M checkpoint
		heme metabolism				UPR	DNA repair
		myogenesis				fatty acid metabolism	UPR
		KRAS signaling up				interferon gamma response	adipogenesis
		IL2 STAT5 signaling				UV response up	reactive oxygen species pathway
						estrogen response late	xenobiotic metabolism
						adipogenesis	UV response up
						xenobiotic metabolism	estrogen response late
							KRAS signaling DN

Table S 9 Leading edge subset of the 'Unfolded protein response' set identified by GSEA on day 1 post infection

EBVwt vs uninf	EBVΔα1 vs uninf	EBVwt vs uninf	EBVΔα1 vs uninf
XPOT	DDX10	RRP9	LSM1
YIF1A	ASNS	DKC1	RRP9
DDX10	XPOT	SDAD1	EIF4E
SLC7A5	NHP2	ATF4	EIF4G1
TARS1	SLC7A5	NOLC1	SLC1A4
PSAT1	TARS1	SRPRB	NOP56
EIF2S1	DCTN1	DCTN1	GOSR2
NHP2	YIF1A	SLC1A4	XBP1
EXOSC4	EIF2S1	LSM1	MTHFD2
EIF4EBP1	CHAC1	MTHFD2	NOLC1
IARS1	PSAT1	SLC30A5	BANF1
ASNS	IARS1	HERPUD1	NOP14
DDIT4	CEBPG	XBP1	SRPRB
CHAC1	EIF4EBP1	HSPA9	YWHAZ
CEBPG	NPM1	NOP14	HSPA9
NOP56	SDAD1	KHSRP	
EIF4E	SEC31A	YWHAZ	
NPM1	DDIT4	PDIA5	
EIF4G1	EXOSC4	EIF2AK3	
GOSR2	DKC1	HYOU1	

Table S 10 Leading edge subset of the 'MYC V1 targets' set identified by GSEA on day 2 post infection

EBVwt vs uninf	EBVΔα1 vs uninf	EBVΔα1 vs EBVwt	EBVwt vs uninf	EBVΔα1 vs uninf	EBVΔα1 vs EBVwt	EBVwt vs uninf	EBVΔα1 vs uninf	EBVΔα1 vs EBVwt
DDX21	MRPL23	IMPDH2	GNL3	SNRPA1	CANX	CLNS1A	ERH	CYC1
PA2G4	PA2G4	SNRPB2	G3BP1	HDAC2	CCT2	NDUFAB1	PSMA7	RANBP1
MRPL23	HSPD1	VDAC1	NOP16	EIF2S2	PSMB3	KARS1	PSMA2	CCT4
CCT2	CCT2	PWP1	SSB	PSMD7	DHX15	NCBP1	SLC25A3	CCT3
YWHAE	RRP9	NPM1	CDK4	DDX18	SNRPA1	EIF4H	UBE2L3	G3BP1
SNRPG	C1QBP	HSPD1	SRSF1	PSMD3	CCT5	PSMD14	ETF1	SRSF1
XPOT	MYC	AIMP2	PSMA4	SERBP1	POLE3	CDK2	VDAC3	HPRT1
PSMA2	IMPDH2	NOP16	PHB2	CCT5	SYNCRIP	IARS1	CLNS1A	CBX3
SRPK1	NOP16	MYC	PPIA	NOLC1	PRPF31	SSBP1	PHB2	NCBP2
C1QBP	VDAC1	RAN	GSPT1	CYC1	MRPL9	NOLC1	PSMB2	BUB3
HDGF	NPM1	HSP90AB1	NOP56	EIF4E	SRSF3	AP3S1	HPRT1	H2AZ1
PHB	PHB	PSMD1	CCT4	SSB	U2AF1	COX5A	DHX15	HNRNPA3
EIF2S1	XPOT	HSPE1	TARDBP	GOT2	LDHA	PGK1	CBX3	MCM7
EIF3J	NHP2	SSBP1	NPM1	PSMD8	COX5A	SNRPB2	KARS1	EPRS1
SYNCRIP	SRPK1	MRPL23	PRDX3	AIMP2	XRCC6	RNPS1	ACP1	KARS1
EIF2S2	HSP90AB1	PSMC6	SNRPA1	SNRPG	DDX18	RUVBL2	PABPC1	PSMA4
CCT7	NME1	PSMA1	VDAC1	EXOSC7	SRPK1	EIF3D	PSMB3	RACK1
HSPD1	PSMC6	NME1	PSMD7	LDHA	EXOSC7	EIF3B	ILF2	GOT2
TFDP1	SYNCRIP	C1QBP	SLC25A3	FBL	EIF2S1	RAD23B	HDGF	
ABCE1	EIF2S1	NOLC1	HSP90AB1	EPRS1	SRSF7	CBX3	TOMM70	
CCT3	SNRPB2	SERBP1	PSMA7	SRSF1	CTPS1	UBA2	PRDX3	
EEF1B2	RUVBL2	HNRNPA1	PTGES3	CCT4	USP1	VDAC3	YWHAQ	
ODC1	PWP1	NHP2	PSMD8	ABCE1	EIF4E	FAM120A	SRM	
MYC	GNL3	PA2G4	CCT5	PRDX4	VDAC3	PRPS2	RANBP1	
ILF2	EIF3J	PSMD14	TUFM	PSMA1	CCT7	GLO1	TFDP1	
NHP2	PPIA	PHB	IMPDH2	TARDBP	EIF3J	PPM1G	CAD	
POLE3	RAN	TCP1	TOMM70	COX5A	PSMD7	SF3A1	EEF1B2	
YWHAQ	CDK4	PPIA	SET	PSMA4	ACP1	XRCC6	PSMC4	
PRDX4	CCT7	GNL3	CAD	SNRPD1	IARS1	PSMC6	CSTF2	
PCBP1	POLE3	COPS5	SNRPD2	PSMD1	HDAC2	RAN	CANX	
EIF4E	POLD2	SNRPD1	RRP9	EIF3B	NOP56	RANBP1	RACK1	
USP1	TUFM	TUFM	GOT2	IARS1	KPNA2	VBP1	RNPS1	
EPRS1	NOP56	RUVBL2	EXOSC7	YWHAE	SSB	SRSF2	ODC1	
LSM7	PSMD14	FBL	PABPC1	GSPT1	PSMC4	PWP1	DDX21	
NME1	HSPE1	ETF1	SRM	SNRPA	CLNS1A	PSMB2	PCBP1	
CYC1	G3BP1	CDK4	TRIM28	HNRNPA1	EIF4H		SNRPD3	
KPNB1	USP1	POLD2	SNRPA	EIF4H	APEX1		NDUFAB1	
DDX18	CCT3	ERH	XPO1	TCP1	PGK1			
HDAC2	SSBP1	PSMD3	PSMD3	PGK1	PSMD8			
UBE2L3	U2AF1	RRP9	HPRT1	XRCC6	PSMB2			

Table S 11 Leading edge subset of the 'E2F targets' set identified by GSEA on day 3 post infection

EBVwt vs uninf	EBVΔα1 vs uninf	EBVΔα1 vs EBVwt	EBVwt vs uninf	EBVΔα1 vs uninf	EBVΔα1 vs EBVwt	EBVwt vs uninf	EBVΔα1 vs uninf	EBVΔα1 vs EBVwt
RRM2	TOP2A	H2AZ1	RFC1	LBR	NME1	ZW10	TP53	XRCC6
HMGB2	TUBB	HMGB2	PTTG1	CENPM	SLBP	DEK	EXOSC8	SYNCRIP
TOP2A	HELLS	BIRC5	RAN	SMC3	SRSF2	TRIP13	BARD1	SMC4
HELLS	USP1	CDC20	MCM4	PCNA	NOP56	MCM2	DUT	NAA38
TK1	HMMR	AURKB	PRDX4	PLK1	H2AX	PAICS	RFC3	ATAD2
HMMR	RRM2	DUT	GINS4	ZW10	HMMR	SLBP	NUP205	POP7
CDC20	TK1	ASF1B	CKS2	PA2G4	TUBG1	NOP56	TFRC	CDCA3
HMGA1	PSMC3IP	RPA3	LMNB1	POLD2	KPNA2	SHMT1	RPA3	MCM4
BIRC5	POLD3	PAICS	RFC3	EZH2	SNRPB	RANBP1	RAD51C	STMN1
H2AZ1	HMGB2	CKS2	SYNCRIP	GINS4	UNG	POP7	MLH1	SRSF1
PA2G4	CDCA8	HMGA1	JPT1	MCM4	NASP	TIMELESS	RAN	PRDX4
KPNA2	RAD51AP1	RFC1	BRCA1	DEK	RAD51C	MTHFD2	RANBP1	BRCA1
USP1	CDKN3	TIPIN	CDCA3	MCM3	MCM7	SMC3	MCM7	DCTPP1
HMGB3	CBX5	LYAR	NME1	CDK4	RAD51AP1	UBE2T	POLD1	RRM2
CDK1	CENPE	JPT1	MCM7	ATAD2	PLK1	XRCC6	AURKB	TRA2B
TUBB	STMN1	CDK1	TACC3	MAD2L1	HMGB3	DNMT1	DDX39A	SPC25
CDCA8	KPNA2	PTTG1	DDX39A	SYNCRIP	PRKDC	MCM6		CDKN3
PCNA	HMGB3	CTPS1	EZH2	CDKN2A	CSE1L	KIF22		POLD2
RAD51AP1	RAD50	RFC3	POLD2	RFC2	ASF1A	CBX5		CDCA8
RPA3	BIRC5	MTHFD2	EIF2S1	CKS1B	TACC3	LBR		TOP2A
DUT	TIMELESS	RAN	PSMC3IP	CDCA3	CKS1B			UBR7
CENPE	CDK1	HELLS	CKS1B	SRSF1	MCM5			LMNB1
STMN1	CDC20	MAD2L1	SRSF1	ASF1B	RNASEH2A			CDKN1B
PLK1	EIF2S1	TRIP13	PRKDC	SMC4	NOLC1			
NUDT21	LMNB1	CDKN1A	RAD51C	POLE4	EZH2			
CDKN3	SPC25	NUDT21	SNRPB	TUBG1	RANBP1			
CENPM	HMGA1	ORC6	NAA38	SHMT1	MCM2			
ASF1B	RACGAP1	UBE2T	RACGAP1	NAA38	PCNA			
MAD2L1	BRCA1	PA2G4	TIPIN	PRKDC	MYC			
POLD3	KIF22	TK1	ATAD2	MCM6	LIG1			
SPC25	PRDX4	DNMT1	SMC4	TRIP13	GINS4			
AURKB	EED	TBRG4	CDK4	MCM2	CENPM			
TUBG1	NUDT21	DDX39A	CSE1L	RBBP7	SSRP1			

Table S 12 Leading edge subset of the 'G2M checkpoint' set identified by GSEA on day 4 post infection

EBVwt vs uninf	EBVΔα1 vs uninf	EBVΔα1 vs EBVwt	EBVwt vs uninf	EBVΔα1 vs uninf	EBVΔα1 vs EBVwt
TOP2A	TOP2A	UBE2C	CDC6	DBF4	PML
HMMR	SMC2	H2AZ1	CCNA2	UBE2C	DTYMK
UBE2C	NUSAP1	BIRC5	LMNB1	PRMT5	KIF20B
CDC20	HMMR	CDC20	SYNCRIP	EZH2	TACC3
HMGA1	MT2A	AURKB	JPT1	CCNA2	CKS1B
BIRC5	TFDP1	CKS2	MNAT1	DTYMK	MCM5
MT2A	H2AZ2	HMGA1	HMGN2	TENT4A	CDC27
H2AZ1	MARCKS	JPT1	PRMT5	MCM3	NOLC1
KIF23	BUB1	CDK1	TACC3	CDK4	ORC5
BUB1	CDKN3	PTTG1	DDX39A	MAD2L1	EZH2
KPNA2	MNAT1	KIF23	EZH2	SYNCRIP	MCM2
HMGB3	CENPE	MAD2L1	CKS1B	CKS1B	MYC
CDK1	STMN1	UCK2	SRSF1	KIF15	FBXO5
SMC2	TPX2	ORC6	HSPA8	SRSF1	SYNCRIP
TPX2	KPNA2	SNRPD1	RACGAP1	UPF1	SMC4
NUSAP1	HMGB3	DDX39A	SMC4	SMC4	STMN1
TFDP1	HMGN2	SRSF2	SNRPD1	SQLE	SRSF1
CENPE	KIF23	H2AX	CDK4	G3BP1	PRMT5
STMN1	BIRC5	DBF4	NCL	MCM6	TPX2
GINS2	GINS2	CCNA2	MCM2	BUB3	DKC1
PLK1	CDK1	HMMR	H2AZ2	MCM2	TRA2B
DBF4	CDC20	GINS2	KIF20B	TOP1	CDC6
CDKN3	LMNB1	KPNA2	CCNF		SLC7A5
FBXO5	HMGA1	NASP	MCM6		CDKN3
MAD2L1	RACGAP1	BUB1	KIF22		CCNF
AURKB	KIF22	NCL	BUB3		TOP2A
PTTG1	LBR	HSPA8	LBR		LMNB1
DTYMK	PLK1	PLK1			CDKN1B
CKS2	FBXO5	HMGB3			

Table S 13 Leading edge subset of the 'Glycolysis' set identified by GSEA on day 4 post infection

EBVwt vs uninf	EBVΔα1 vs uninf	EBVΔα1 vs EBVwt	EBVwt vs uninf	EBVΔα1 vs uninf	EBVΔα1 vs EBVwt
MIF	HMMR	LDHA	SRD5A3	GALK1	ALDOA
HMMR	G6PD	PPIA	POLR3K	SRD5A3	GMPPB
PPIA	STMN1	GLRX	HAX1	PDK3	MPI
TPI1	MIF	TPI1	COPB2	MDH1	KIF2A
LDHA	GFPT1	ISG20	MED24	ALDOA	STMN1
CDK1	CDK1	FAM162A	ISG20	AKR1A1	TALDO1
TXN	PKM	CDK1	CLN6	B4GALT2	BPNT1
STMN1	TXN	SOD1	PGK1	TPI1	RBCK1
GLRX	TALDO1	PGAM1	PGLS	PPP2CB	PGK1
PGAM1	GALE	GOT2	NASP		RPE
ENO1	DLD	MDH2	GNPDA1		SDHC
GALE	ENO1	HAX1	GMPPB		RARS1
GOT2	GYS1	MIF	CHST12		NDUFV3
TALDO1	HDLBP	GALE	COG2		ALG1
PKM	CLN6	HMMR	PMM2		ME2
FAM162A	PGAM1	NASP	GMPPA		AK3
DLD	RARS1	CHST12	TSTA3		B4GALT7
RPE	PPIA	ENO1	RBCK1		PKM
MDH2	GMPPA	TXN	MDH1		TSTA3
GFPT1	MED24	FKBP4			COPB2
ALDOA	GOT2	GALK1			DLD
RARS1	KIF2A	PGM2			GNPDA1
GALK1	COPB2	PSMC4			SRD5A3
KIF2A	RPE	POLR3K			GUSB
SOD1	ARPP19	DDIT4			B3GAT3
G6PD	PGLS	PMM2			

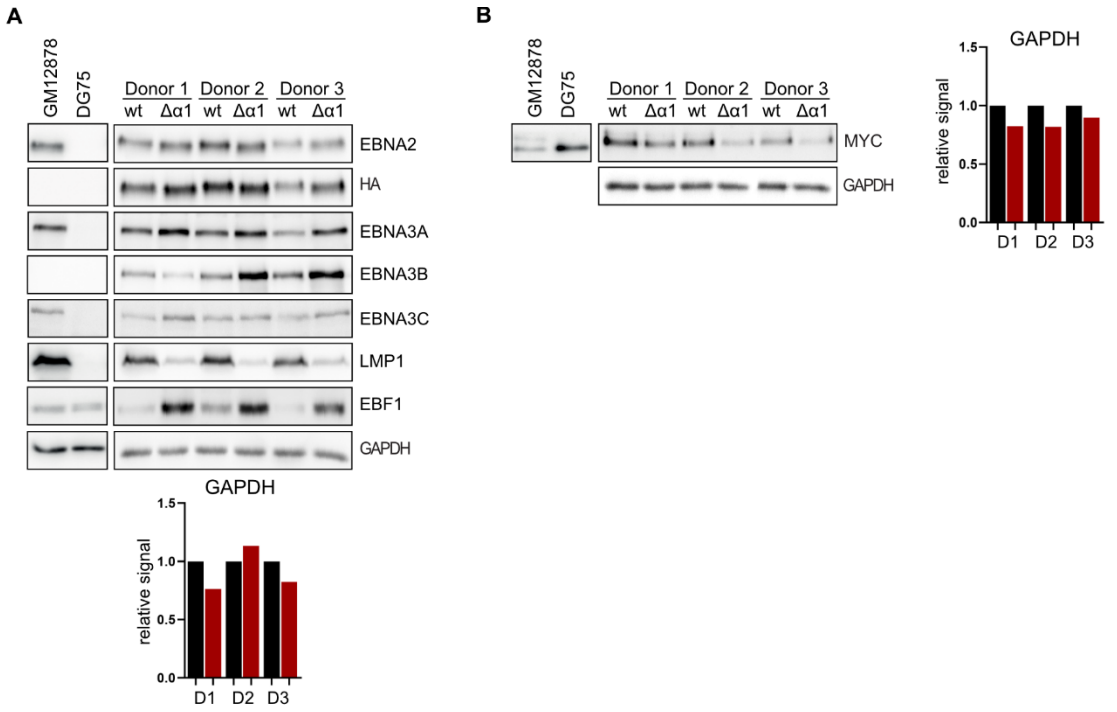


Figure S 1 Corresponding GAPDH Western blot signals for the quantification of relative Western blot signals in Figure 11C. The Western blot signals were quantified with Bio1D and the GAPDH signals for each donor LCLwt was set to 1 as a reference. A) GAPDH corresponding to EBNA2, HA, EBNA3A/B/C, LMP1, EBF1 B) GAPDH corresponding to MYC

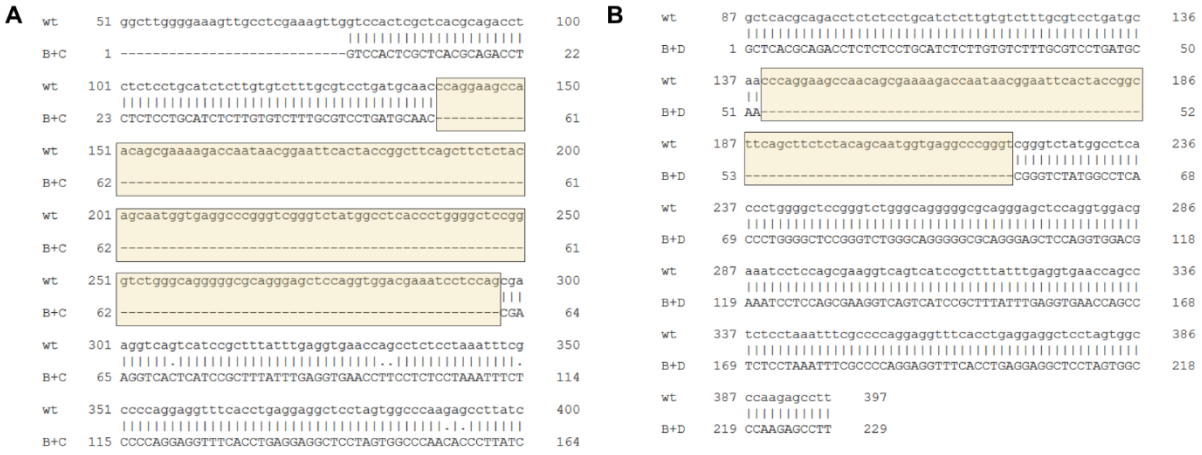


Figure S 2 Sequence alignment of the knock-out of EBF1 exon 3 with RNP combinations A) B+C and B) B+D, The yellow area indicates the deleted region, which includes exon 3.

6.2 Affirmation

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass die vorliegende Arbeit mit dem Titel

„The role of the EBNA2-EBF1 complex in EBV driven B cell transformation"

von mir selbständig und ohne unerlaubte Hilfsmittel angefertigt wurde, und ich mich dabei nur der ausdrücklich bezeichneten Quellen und Hilfsmittel bedient habe. Die Arbeit wurde weder in der jetzigen noch in einer abgewandelten Form einer anderen Prüfungskommission vorgelegt.

München, den 01.03.2021 _____

Sophie Beer

Erklärung

Hiermit erkläre ich, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich anderweitig einer Doktorprüfung ohne Erfolg nicht unterzogen habe.

München, den 01.03.2021 _____

Sophie Beer

6.3 Acknowledgement

First of all, I would like express my gratitude to Prof. Dr. Bettina Kempkes who supervised me and my project; who was always approachable with questions and concerns I had; who pointed out the good parts and small successes of my research project that I couldn't see; who trusted me and my abilities. Thank you!

I would like to thank Prof. Dr. Wolfgang Enard for being the second referee for my thesis.

I would also like to thank Dr. Sebastian Bultmann and Prof. Dr. Robert Schneider for participating as advisors in my thesis advisory committee and for the fruitful discussions and feedback during our meetings.

A big thank you to Lucas Wange and Dr. Antonio Scialdone for the successful collaboration in the RNA sequencing project. Lucas, thank you for your time and advice regarding the sample preparation and processing of the data. Antonio, I would have been help- and clueless without your advice regarding the data analysis. I am grateful for your time and patience to explain and clarify the many questions I had and for your support during the various analysis steps.

Dear Xiang and Nini, having you as my lab colleagues was a fantastic experience. Thank you for sharing techniques and reagents but more importantly thank you for the interesting and funny conversations we had and for broadening my horizon by sharing stories. Dear Conny, I am sincerely grateful for your friendly, joyful and extremely helpful manner, which made working in the lab a lot easier and enjoyable in many different ways. I'm glad I had the chance to work with you!

Laura, Ezgi and the Irmela-Girls, lab life would not have been the same without you. You brightened up my coffee breaks, made me laugh, were always up for a chat and shared your knowledge with me. Your presence and shared time carried me through the good and not so good times! Thank you!

Mein größter Dank geht an meine Familie, allen voran Mama, die nie aufhört, an mich zu glauben und mich gehen lässt, wohin es mich auch immer zieht. Danke für die fortwährende Unterstützung und Liebe!

Last but not least, mein Nikolaus, mit dir an meiner Seit sind schöne Momente noch schöner und schwierige Zeiten leichter zu ertragen! Danke für deine Liebe, Ermutigungen und Geborgenheit!