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

TARGET GENE REGULATION BY EBV LATENT TRANSCRIPTION FACTORS – EXPLOITING CELLULAR ENHANCER ELEMENTS

Laura V. Glaser

München, September 2016

Erstgutachter:	Prof. Dr. Bettina Kempkes
Zweitgutachter:	Prof. Dr. Wolfgang Enard
Tag der Abgabe:	13.09.2016
Tag der mündlichen Prüfung:	21.03.2017

Zusammenfassung

Die latente Epstein-Barr-Virus (EBV) Infektion und die damit einhergehende Expression von EBV Nukleären Antigen (EBNA) Proteinen sind mit verschiedenen B-Zell Tumoren assoziiert. Die Infektion von ruhenden primären B-Zellen *in vivo* leitet Differenzierungsprogramme ein, die typischerweise mit B-Zellaktivierung und -proliferation korrelieren. Letztendlich entstehen persistierende, nicht proliferierende B-Gedächtniszellen, welche durch die minimale Expression viraler Gene gekennzeichnet sind. *In vitro* wird dieser Prozess in einem proliferativen Stadium blockiert, phänotypisch ähnlich einem B-Zellblasten, und lymphoblastoide Zelllinien (LCLs) entstehen. EBNA2, EBNA3A und EBNA3C (E2, E3A und E3C) fungieren als Transkriptionsfaktoren (TF), die in der Lage sind die Transkription der Wirtszelle zu manipulieren. Des Weiteren sind sie essentiell für die B-Zelltransformation und EBV-getriebene Proliferation. Differenzielle Genexpressionsstudien, unter der Verwendung von Deletionsmutanten oder konditionellen Expressionssystemen, enthüllten teilweise überlappende sowie einzeln regulierte EBNA Zielgene. Um gemeinsame oder unabhängige EBNA Funktionsweisen zu untersuchen, wurden rekombinante EBV Genome und anschließend LCLs hergestellt. Diese wurden zur Identifizierung von EBNA Bindestellen im humanen Genom durch ChIP-seq Methoden verwendet. Die dabei erhobenen Daten wurden mit Informationen zu Chromatinstruktur und TF-Bindestellen aus LCLs, die im Rahmen des ENCODE Projekts publiziert wurden, kombiniert und verglichen. Ein bioinformatischer Ablauf wurde spezifisch für diesen Zweck, unter Verwendung der Galaxy Plattform, entworfen. Dabei konnte gezeigt werden, dass E2 bereits bestehende B-Zellenhancer bindet. Ein neuartiger Ansatz, der eine quantitative Bewertung und den Vergleich von Bindestellen einschließt, konnte die reziproke Besetzung von B-Zellenhancern durch E2 und E3 Proteine zeigen. Diese werden begleitet durch unterschiedliche TF Kombinationen, welche die spezifischen Bindemuster für das jeweilige EBNA Protein definieren. Des Weiteren indiziert die auffallende Korrelation der Besetzung und Anreicherungsmuster von E3A und E3C Bindestellen eine mögliche Kooperation in der gezielten Chromatinbindung. Alle drei EBNA Proteine sind nicht in der Lage DNA direkt zu binden, sondern bedienen sich zellulärer Adapter, wobei CSL/CBF1, das zentrale Effektormolekül des Notch-Signalwegs, den am umfassendsten beschriebenen darstellt. Die Untersuchung von EBNA2 Bindestellen in CSL/CBF1 negativen Burkitt-Lymphom Zelllinien konnte einen massiven Besetzungsverlust aufzeigen und damit CSL/CBF1 als die Haupt- jedoch nicht einzige Determinante für den Zugang von E2 an Chromatin bestätigen. Analysen zur Motifanreicherung und der Vergleich mit ChIP-seq Daten aus LCLs identifizierten den für B-Zellen charakteristischen TF EBF1 als einen wichtigen Faktor für die E2 Bindung. Letztendlich konnte die Interaktion von E2 und EBF1 in B-Zellen gezeigt werden. Zusammenfassend tragen die Ergebnisse dieser Arbeit zu einem besseren Verständnis der zellulären Mechanismen bei, die von den EBNA Proteinen instrumentalisiert werden, um selektiv Zielgene der Wirtszelle zu regulieren. Außerdem stellen sie einen Ausgangspunkt für die Identifizierung von deterministischen Voraussetzungen für das selektive Binden von E2, E3A oder E3C an Chromatin dar.

Abstract

Latent Epstein-Barr virus (EBV) infection and the accompanied expression of EBV Nuclear Antigen (EBNA) proteins are associated with multiple B cell malignancies. Infection of resting primary B cells *in vivo* triggers differentiation programs typically linked with B cell activation and proliferation. Eventually, persistent non-proliferating memory B cells arise, characterized by minimal expression of viral genes. This process is blocked *in vitro* at a proliferating state, phenotypically similar to a B cell blast, resulting in lymphoblastoid cell lines (LCLs). EBNA2, EBNA3A, and EBNA3C (E2, E3A, and E3C) operate as transcription factors (TFs) which are able to manipulate host cell transcription and are vital to B cell transformation and EBV driven proliferation. Differential gene expression studies, employing knock-out mutants or conditional expression systems, revealed partially overlapping as well as uniquely regulated EBNA target genes. In order to assess concerted or independent EBNA functions, recombinant EBV genomes were constructed and subsequently LCLs were generated which were employed for identification of EBNA binding sites within the human genome applying ChIP-seq methods. These data were combined with and compared to information on host cell chromatin organization and cellular TF binding sites in LCLs, published by the ENCODE project. A bioinformatics workflow specific for this purpose was established, using the Galaxy platform. This could show the predominant targeting of preexisting B cell enhancer elements by E2 and identify co-occurring TFs. A novel approach, which includes a quantitative binding site evaluation and comparison, revealed the reciprocal occupation of B cell enhancers by E2 versus E3 proteins accompanied by a distinct set of co-occurring cellular factors, defining binding patterns specific for each EBNA protein. Furthermore, the striking correlation of E3A and E3C binding site occupancy and enrichment distribution in particular indicated a possible cooperation in chromatin targeting. All three investigated EBNA proteins are not able to directly target DNA, but employ adaptor proteins instead, the cellular TF CSL/CBF1, the major down-stream effector of the Notch signaling pathway, being the most extensively described among them. Investigation of E2 binding site occupancy in CSL/CBF1 knock-out Burkitt's lymphoma cell lines disclosed an extensive loss of these sites, confirming CSL/CBF1 as a major, but not exclusive determinant for E2 chromatin accession. Motif enrichment analyses and comparison with published ChIP-seq data in LCLs revealed the early B cell TF EBF1 as an important factor for mediating E2 binding. Finally, the interaction of E2 and EBF1 in B cells were demonstrated. In conclusion, the findings of this thesis contribute to a better understanding of the cellular mechanisms exploited by EBNA proteins to selectively regulate host cell target genes. They also provide an initial starting point for the identification of deterministic prerequisites for selective E2 or E3A and E3C binding to chromatin.

Table of Contents

Registers

List of Figures

List of Tables

List of Abbreviations

2-DOG	2-deoxy-galactose
°C	Degree Celsius
α	Alpha (anti)
β-ME	β-Mercaptoethanol
Δ	Delta (deletion)
µCi	Microcurie
µg	Microgram
µl	Microliter
µM	Micromolar
A	Alanine/2'-deoxyadenosine 5'-phosphate
aa	Amino acid
AIDS	Acquired immunodeficiency syndrome
Amp	Ampicillin
APS	Ammonium persulfate
BAC	Bacterial artificial chromosome
BL	Burkitt's lymphoma
bp	Base pair(s)
BSA	Bovine Serum Albumin
C	Cysteine/2'-deoxycytidine 5'-phosphate
cat	Chloramphenicol acetyltransferase
CBF1	C promoter binding factor 1
CBP	CREB-binding protein
CD	Cluster of differentiation
CDK	Cyclin dependent kinase
cDNA	complementary DNA
ChIP	Chromatin Immunoprecipitation
ChIP-seq	ChIP and next generation sequencing
chr	Chromosome
cm	Centimeter
CMV	Cytomegalo virus
Co-IP	Co-Immunoprecipitation
Cp	C promoter
CsCl	Caesium chloride
CSL	CBF1/Su(H)/Lag-1 family
CtBP	C-terminal binding protein
D	Aspartic acid
d	day(s)
DBD	DNA binding domain
DIM	Dimerization domain
DMSO	Dimethyl sulfoxide
DNA	2'-deoxyribonucleic acid
DNaseI	Deoxyribonuclease
dNTP	3'-deoxyribonucleotide-5'-phosphat
Dox	Doxycycline
DTT	Dithiothreitol
e	Exon
E1	EBNA1
E2	EBNA2
E3A	EBNA3A
E3B	EBNA3B
E3C	EBNA3C
EBER	EBV encoded small RNAs
EBNA	EBV Nuclear Antigen
EBV	Epstein-Barr virus
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
E-LP	EBNA-LP
ENCODE	Encyclopedia of DNA elements
EtBr	Ethidium bromide
F	Phenylalanine
FACS	Fluorescence activated cell sorting
FC	Fold change
FCS	Fetal Calf Serum
Fig.	Figure
fwd	forward
G	Glycine/2'-deoxyguanosine 5'-phosphate
g	gravitational constant
galK	Galactokinase
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GC	Germinal center
GFP	Green florescent protein
gp	Glycoprotein
GRU	Green Raji Unit
GST	Glutathione S-transferase
H	Histidine
h	hour(s)
HA	Hemagglutinin
HCl	Hydrogen chloride
HDAC	Histone deacetylase
HHV-4	Human herpesvirus 4
HIV	Human immunodeficiency virus
HL	Hodgkin lymphoma
hpt	Hygromycin phosphotransferase
HRP	Horseradish peroxidase
HS	Hypersensitive sites
Hyg	Hygromycine
I	Isoleucine
IP	Immunoprecipitation
kb	Kilobasepairs
kDa	Kilodalton
ko	Knock out
L	Leucine
LB	Luria-Bertani
LCL	Lymphoblastoid cell line
LMP1	Latent Membrane Protein 1
LMP2A	Latent Membrane Protein 2A
LMP2B	Latent Membrane Protein 2B
M	Molar
mA	Milliampere
me	Methylated
min	Minute
m.c.	monoclonal
MHC	Major histocompatibility complex
ml	Milliliter
mM	Millimolar
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger RNA
NaCl	Sodium chloride
NaOH	Sodium hydroxide
ng	Nanogram
NK cell	Natural killer cell
nm	Nanometer
NMR	Nuclear Magnetic Resonance
NPC	Nasopharyngeal carcinoma
ORF	Open reading frame
ori	Origin of replication
P	Proline
PBS	Phosphate buffered saline
p.c.	polyclonal
PCNSL	Primary central nervous system lymphoma
PcG	Polycomb group
PCR	Polymerase chain reaction
PIC	Proteinase inhibitor cocktail
Pol	Polymerase
PTLD	Post-transplant lymphoproliferative disease
PVDF	Polyvinylidene fluoride
qPCR	Quantitative PCR
rev	Reverse
RMA	Robust multichip average
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Rounds per minute
rRNA	ribosomal RNA
RT	Reverse transcription/Room temperature
S	Serine
s	Second(s)
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS Polyacrylamide gel electrophoresis
SEM	Standard error of mean
T	Threonine/2'-deoxythymidine 5'-phosphate
TAD	Transactivation domain
TAE	Tris acetate EDTA
TE	Tris EDTA buffer
TEMED	Tetramethylethylenediamine
Temp.	Temperature
TF	Transcription factor
Tris	Tris(hydroxymethyl)aminomethane
TSS	Transcriptional start site
UV	Ultraviolet
V	Volt
vs.	versus
W	Tryptophan
WB	Western blot
Wp	W promoter
wt	wild type
Y2H	Yeast two-hybrid

Acknowledgements

First, I would like to express my sincere gratitude to my doctoral adviser Prof. Dr. Bettina Kempkes for providing me this project, giving me the opportunity to further arrange it freely and to grow with the requirements, for her excellent supervision and frequent discussions, the possibility to visit another laboratory abroad, the continuous support and the trust in my abilities and me as a person.

I thank Prof. Dr. Wolfgang Enard for being my second referee for this thesis.

I would also like to thank the members of my thesis committee, Prof. Dr. Vigo Heissmeyer and PD Dr. Philipp Korber, for their insightful comments and encouragement but also for their critical assessment and constructive suggestions.

My sincere thanks also goes to Prof. Dr. Rolf Backofen and Dr. Björn Grüning, who provided an opportunity to visit their team at the University of Freiburg and gave me access to their Galaxy bioinformatics server. In particular I'm grateful to Björn, who taught me a lot and was always available for questions and fixing bugs.

I thank my present fellow labmates Simone Rieger and Conny Kuklik-Roos and former colleagues Marie Harth-Hertle, Sybille Thumann, and Barbara Scholz for the stimulating discussions, sharing of techniques and reagents and most of all, for the great atmosphere and fun we had in the lab. Particular thanks to Simone, who supported me very much during the DG75 project, was always in for discussions, and is a curious and motivated scientist; it was a great time I won't forget. Thanks also to Conny, the best technician you can wish for and whose good mood is infectious to those around her. My gratitude goes also to Marie, who was always there, even after she left the lab, who is a role model to me. Thanks for the constant support, the encouraging words and postcards, and singing in the cell culture; it was the best time.

My most sincere thanks goes to parents and my sister, for the many many ways they supported me during my whole studies, who believe in me and what I'm doing without questioning or any doubts. It would not have been possible without them.

Last but not least, I want to thank Daniel, for his love, his constant support and motivation, for listening and counseling, for pushing me when I needed it; we are the best team.

Dedication

To Elmar, who seeded curiosity and taught to question.

To Rita, who taught empathy and to see the bright side.

To Jana, who brought joy and purpose.

To Daniel, who loves and understands.

To the giants, whose shoulders I stand upon.

"The scientist does not study nature because it is useful; he studies it because he delights in it,
and he delights in it because it is beautiful.
If nature were not beautiful, it would not be worth knowing,
and if nature were not worth knowing, life would not be worth living."

Jules Henri Poincaré (1854-1912)

1 Introduction

In 1958 the British medical officer and surgeon Denis Burkitt described the occurrence of characteristic tumors in children in Uganda and Equatorial Africa, which are known today as Burkitt's lymphoma (Burkitt, 1958). Since the prevalence of this specific lymphoma associated strikingly with distinct climatic and geographical factors, the involvement of a biological infectious agent was proposed (Burkitt, 1962a, Burkitt, 1962b). Soon after, in 1964 the team of M. A. Epstein, Y.M. Barr, and B.G. Achong was able to maintain suspension cultures from patient samples and could describe the presence of viral particles in these cells for the first time (Epstein et al., 1964). Subsequently the fine structure of the virus was described by Hummeler and the virologist couple Henle, and they recognized and assigned it to be a member of the herpes viruses (Hummeler et al., 1966) which was later named Epstein-Barr Virus (EBV) (Henle et al., 1968). Shortly after it was discovered that EBV is able to transform B cells *in vitro* (Henle et al., 1967, Pope et al., 1968) and it was classified as the first human tumor inducing virus.

In the last 50 years since the discovery of EBV and its transforming properties, several other malignancies apart from Burkitt's lymphoma, e.g. diverse lymphomas and carcinomas, were described to be associated with this virus (reviewed in Thorley-Lawson et al., 2015, Rickinson and Kieff, 2007). Since the vast majority of the human population worldwide is latently infected with EBV, intensive research is conducted to elucidate the tumor inducing or passenger properties of EBV.

1.1 Epstein-Barr Virus

EBV, also termed human herpesvirus 4 (HHV-4), is a member of the γ-herpesvirinae, which is a ubiquitously distributed human pathogen, with over 95% of adults being infected. Common for all herpesviruses is its structure composed of a lipid containing envelope, a tegument, and an icosaedric nucleocapsid containing the double stranded DNA. The EBV genome consists of approx. 172 kb and encodes 80-100 different proteins. Characteristic for EBV is the ability to infect resting B cells *in vitro* and *in vivo* and to induce proliferation. *In vivo* this process is controlled by the immune system and the infection, which mostly happens during early childhood, remains asymptomatic. However, an infection later in life can result in 35-50% of the cases in infectious mononucleosis due to an excessive cytotoxic T cell response, which is a benign and self-limiting disease (reviewed in Kieff and Rickinson, 2007). Since some of the infected cells can escape from the immune response, EBV achieves a lifelong persistence in the host. Here, no virus particles are produced, a state termed latency. Characteristic for γ-herpesviruses is the ability to regulate

proliferation of the host cell also in this state. Sporadically, the virus can enter the lytic phase, where virus particles are being produced and new B cells can be infected. In healthy adults the viral induced proliferation as well as the lytic reactivation are strictly monitored and controlled by the immune system. However, in immunocompromised individuals a spontaneous reactivation or primary infection with EBV can lead to a pathological proliferation of transformed cells. For instance, in immunosuppressed transplant recipients a highly malignant *post-transplant lymphoproliferative disease* (PTLD) can occur (reviewed in Gottschalk et al., 2005). Also HIV positive late stage AIDS patients are at risk to develop EBV induced lymphoproliferative diseases, e.g. AIDS-associated *primary central nervous system lymphoma* (PCNSL) (Carbone et al., 2009). While for PTLD and PCNSL EBV could be identified as the direct causing agent, it might display a passenger in other malignancies (see chapter 1.1.3). In general, environmental triggers including additional pathogens and specific genetic predisposition are considered to contribute to EBV pathogenesis by complex mechanisms.

1.1.1 The life cycle of EBV

Characteristic for EBV and other herpesviruses is the separation of its life cycle into a latent and a lytic phase. After transmission through saliva EBV enters the epithelium of the Waldeyer's ring and infects resting naïve B cells there (Fig. 1). EBV glycoproteins gp350 and gp220 play an important role in recognizing and binding the CD21 receptor at the B cell surface (Nemerow et al., 1987, Tanner et al., 1987). Subsequently, endocytosis of the virus is mediated by the interaction of gp42 with MHC class II molecules, which leads to fusion of the viral envelope with the cell membrane (Silva et al., 2004). Then the viral capsid travels to the nucleus, the viral DNA is released into the nucleus and the viral genome circularizes and is established as an episome. In the following EBV triggers B cell differentiation and proliferation programs, which are regularly activated only after encountering an antigen but do not need further external signals (reviewed in Thorley-Lawson, 2001, Thorley-Lawson and Allday, 2008, Thorley-Lawson et al., 2015) (summarized in Fig. 1).

After the infection of a naive B cells these get activated and differentiate into proliferating B cell blasts. In the absence of EBV this process is activated upon antigen binding to the B cell and further needs T cell signals to proceed. However, in the case of EBV infection the expression of only 11 viral latent genes is sufficient. Among these are six *Epstein-Barr Nuclear Antigens* (EBNA1, -2, -LP, -3A, -3B, and -3C), three *Latent Membrane Proteins* (LMP1, -2A- and -2B) and two genes encoding small non-polyadenylated RNAs (EBER1 and -2). This viral expression program is termed latency III or *growth program*. The latency III associated viral expression profile can be found in PTLD or PCNSL and is also expressed in *in vitro* established

LCLs. Latency III triggers excessive proliferation, however all of the expressed viral proteins but EBNA1 exhibit epitopes, which can be presented by MHC molecules of the infected cell. This in turn activates an immune response, where cytotoxic T cells eliminate the uncontrollably proliferating B cells.

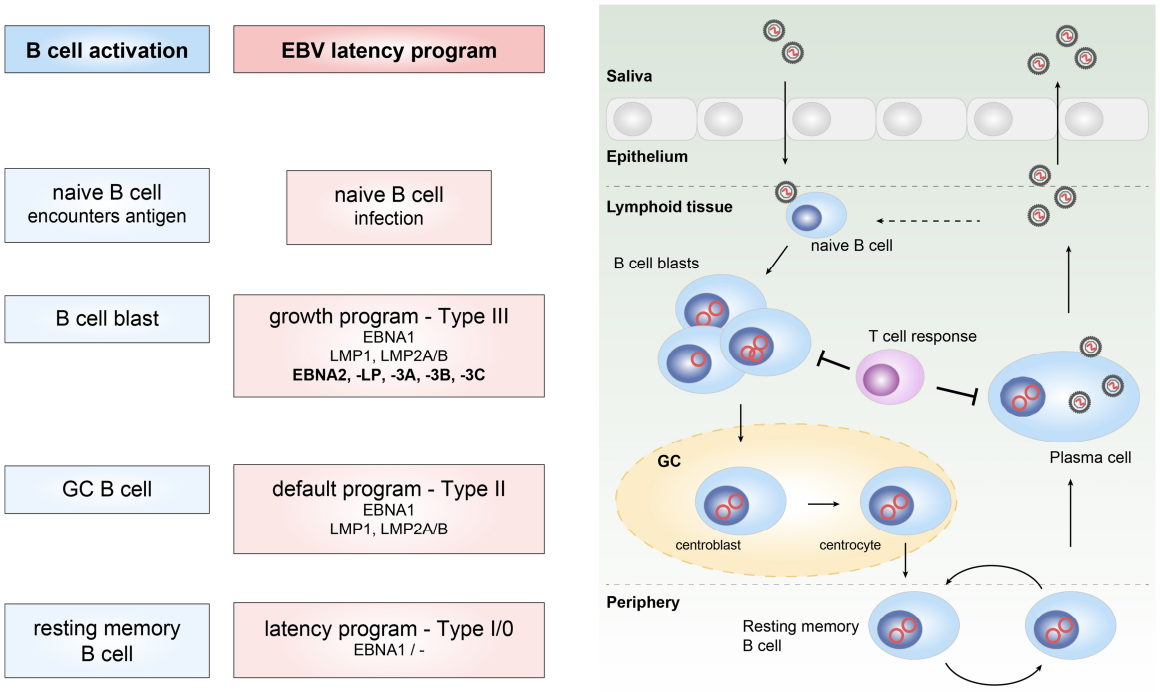


Figure 1. Schematic representation of the EBV life cycle. EBV enters the lymphatic tissue of the oropharynx by passing through epithelial cells, where it supposedly initiates lytic infection which results in amplification of the virus. Then these virions infect resting naïve B cells and drive them to become proliferating lymphoblasts. These lymphoblasts are vastly eliminated by the T cell response, due to EBV antigen exposure. However, some cells escape and travel to the germinal center (GC) were they undergo a GC reaction and differentiate into memory B cells. If these memory B cells further differentiate into plasma cells, the lytic cycle is induced and EBV particles are produced and released. Each step of the EBV latency program and the associated viral expression patterns are indicated (red boxes) as well as the single steps of the regular B cell activation, which are imitated by EBV infection (blue boxes). Figure adapted from Thorley-Lawson and Allday, 2008.

However, a certain percentage of EBV infected cells escape the elimination by the immune system and travel in the germinal centers (GC) of lymph nodes, where they differentiate into centroblasts, centrocytes, and finally memory B cells. This process is accompanied by the gradual deactivation of the expression of latent genes. In the latency II stage, also called *default program,* expression of EBNA2 and the EBNA3 proteins is shut down first and only EBNA1, the LMPs, and EBER RNAs are expressed. The LMP proteins provide pivotal survival signals during the GC reaction, which are essential for the differentiation into memory B cells. When this step is accomplished no more viral proteins are being expressed, a state which is designated latency 0 or *latency program*. These resulting latently infected memory B cells circulate within the periphery and cannot be recognized by the immune system, since no EBV antigens are being expressed. In this state the virus can persist lifelong in the host. When the memory B cell is dividing, only EBNA1

is expressed, which ensures tethering of the viral genome to the host DNA during cell division but is not presented on MHC molecules. This state is called latency I or *EBNA1-only program*.

The process of viral reactivation is not completely solved to date. Supposedly, this is achieved by regular physiological signals, which cause the memory B cell to differentiate into a plasma cell. Here, the lytic replication program takes place, and all viral genes are being expressed (Laichalk and Thorley-Lawson, 2005). New virus particles are being produced and released, which are in turn able to infect further naïve B cells or are transmitted by saliva to a new host. Since the viral titers in the salvia of infected individuals are rather high, and EBV is also able to infect epithelial cells (reviewed in Hutt-Fletcher, 2007) it was hypothesized that lytic replication can also take place in epithelial cells of the oropharynx and supports virus amplification (reviewed in Thorley-Lawson and Allday, 2008).

1.1.2 EBV latent genes

The infection of B cells *in vitro*, without the appropriate immune response, leads to the outgrowth of continually proliferating *lymphoblastoid cell lines* (LCLs), which exhibit a latency III type expression pattern of viral genes. Since this can also be observed early after infection *in vivo*, these cells display a cell culture model for the process of transformation. Furthermore, they are also considered a model system for diseases as PTLD and PCNSL, which are also associated with latency III expression pattern. The function of all 11 latent genes has been subject to extensive research in the past and is still a matter of interest. The current knowledge on these proteins is shortly summarized below (reviewed in Young and Murray, 2003).

EBNA1 is a DNA binding protein, which tethers the viral episomal genome, by binding to its own *origin of plasmid replication* (oriP), to the human genome during cell division (reviewed in Frappier, 2015).

EBNA2, in the following abbreviated to E2, is the central transactivator of EBV driven gene regulation. Together with EBNA-LP it is the first viral gene to be expressed after infection. The expression of both proteins is initially under control of the W promoter (Wp), but switches to C promoter (Cp) mediated by E2 action. Cp also controls the expression of EBNA1 and the EBNA3 proteins. Furthermore, E2 induces expression of the LMPs and regulates several cellular genes. E2 is not able to bind directly to DNA but utilizes cellular CBF1, the key downstream effector of Notch signaling, to access DNA (reviewed in Kempkes and Ling, 2015). Since E2 plays an essential role in this thesis, its function is further elaborated in chapter 1.2.1.

EBNA-LP cooperates with E2 in transcriptional regulation and increases the activation of viral target genes. However, known cellular target genes of E2 are not affected by EBNA-LP action, which indicated that EBNA-LP only displays a coactivator for a subset of E2 target genes.

The interaction between E2 and EBNA-LP has not been solved to date. It has been proposed that EBNA-LP acts as a coactivator by displacing repressive NCoR complexes from enhancers (Portal et al., 2011). The analysis of EBNA-LP binding sites revealed only a moderate overlap with E2 binding sites and a preference for promoter sites over enhancer regions, implying a different mode of action than only being a coactivator of E2 function (reviewed in Kempkes and Ling, 2015).

The members of the EBNA3 protein family, hereafter abbreviated E3, E3A, -3B, and -3C were initially described as repressors of transcription. All three proteins are able to bind CBF1 as well and therefore an antagonism of E2 function was proposed several times in the past. Since E3A and E3C display the leading actors of this thesis, together with E2, their known functions will be described in more detail in chapter 1.2.2.

LMP1, -2A, and -2B are trans-membrane proteins which mediate signaling in a ligand independent fashion. LMP1 mimics a constitutively active CD40 receptor, a key protein is the activation a differentiation of B cells, and delivers proliferation and survival signals independent of T cell interaction (Kieser and Sterz, 2015). LMP2A imitates the B cell receptor and provides survival signal for the cell in the absence of antigen (Cen and Longnecker, 2015).

The EBERs, EBER1 and -2, are small non-polyadenylated RNAs whose function is not fully understood yet. An immunomodulatory and anti-apoptotic role was supposed for these highly abundant latent transcripts (Skalsky and Cullen, 2015).

None of the latent genes described above is able to induce B cell immortalization independently. Thus a coordinated cooperation between the EBNAs and LMPs is needed, where the individual contribution of the single factors is very different. Infection studies employing recombinant viruses with knock-outs for the single latent genes gave insight on the dependence of EBV induced immortalization on individual genes. These were subsequently classified as essential, critical, or non-essential for immortalization and outgrowth of LCLs. E3B, E-LP, LMP2A, -2B, and the EBERs are non-essential, while EBNA1 is critical for immortalization since EBNA1 deficient LCLs can only be established with low frequencies and need special culture conditions (Humme et al., 2003). However, E2, E3A, E3C, and LMP1 were described to be absolutely essential for B cell immortalization and therefore display a most interesting subject to study EBV induced transformation. However, research conducted in our laboratory could show that E3A is in fact dispensable for immortalization and knock-out LCLs can be established on a regular basis, although they exhibit disabled proliferation and elevated apoptosis rates (Hertle et al., 2009).

1.1.3 EBV associated tumors

EBV displays a rather harmless pathogen for the healthy individual, while it can induce highly malignant immunoblastic B cell lymphomas in immunocompromised patients, as mentioned above, since an appropriate T cell response is not provided here. Among these are PTLD and PCNSL as well as immunoblastic lymphomas in patients with hereditary immunodeficiencies (reviewed in Carbone et al., 2008). However, EBV is also associated with tumors of immuno-competent patients, including e.g. Burkitt's or Hodgkin lymphoma, NK- and T cell lymphomas, and also epithelial tumors (reviewed in Rochford and Moormann, 2015, Murray and Bell, 2015, Jha et al., 2016, Raab-Traub, 2015). However, these tumors are not exhibiting the growth program of latency III, which is the latency state investigated in this thesis, but it is speculated that these tumor cells, which show latency II or I expression pattern, underwent a latency III phase at some point. The described tumors can also show no viral expression except for EBNA1 and EBV could be supportive rather than driving in the multiple step tumor progression. The most common EBV associated tumor malignancies are shortly described below.

Burkitt's lymphoma (BL) is characterized by the reciprocal translocation of chromosome 8 and chromosome 14, 2, or 22. The translocations place the proto-oncogene *c-myc* under the control of the immunoglobulin enhancers which results in a constitutive high level activation of *c-myc*. Endemic BL, found in equatorial Africa, which led to the identification of EBV, is associated with EBV infection in over 95% of the cases. The disease characteristically manifests as a fast growing tumor involving the jaw or other facial bones and the abdomen. Before the AIDS pandemic age, BL displayed the most frequently occurring childhood tumor in Africa. The geographical pattern of BL led to the suggested association with Malaria infection, which might cause a reactivation of the latent viral infection. The sporadic type of BL, occurring in adults in moderate climate zones, is very rare and only associated with EBV in approx. 20% of the cases, while the prevalence of EBV in AIDS-associated type was reported to reach up to 50%. Interestingly, in EBV positive BLs only EBNA1 and the EBERs are expressed, exhibiting latency I, which is not fully understood to date (reviewed in Rochford and Moormann, 2015).

Hodgkin lymphoma (HL) involves the secondary lymphatic organs, as lymph nodes and the spleen, and is histologically characterized by multinuclear Reed-Sternberg cells. These malignant cells represent only a minority of the tumor mass, a cellular infiltrate comprised of non-neoplastic cells including T and B cells. HL represents approx. 20% of all lymphomas in the western world, where 40-50% are associated with EBV while in developing countries the prevalence reaches 90-100%. HIV associated HL are virtually always associated with EBV as well. EBV positive HL express EBNA1, the LMPs, and the EBERs, a pattern characteristic for

latency II. Also in this case the contribution of EBV to the development of HL is not fully understood (reviewed in Murray and Bell, 2015).

Also the nasopharyngeal carcinoma (NPC) could be associated with EBV and involves cells of the nasopharynx region and displays one of the most common tumors in southern China, and Mediterranean Africa. Undifferentiated NPC is constantly associated with EBV and displays latency II expression pattern. Due to the geographic distribution, genetic predispositions (e.g. HLA type) and certain environmental factors, as nitrosamine containing food, were discussed to play important roles in the pathogenesis of NPC (Raab-Traub, 2015).

Even 10% of gastric carcinomas are associated with EBV, which display one of the most common human cancers. Supposedly, EBV only plays a role in a late stage of pathogenesis where it infects neoplastic epithelial cells of the stomach. However, EBV positive gastric carcinomas do not display a consistent latent expression pattern and range between latency I and II (Zur Hausen et al., 2004).

1.2 Epstein-Barr Virus Nuclear Antigen 2 and 3 family

E2 and two members of the E3 family, namely E3A and E3C, were described to be essential for B cell transformation by EBV, which was largely disproved for E3A, and function as transcription factors (TFs) by modulating target gene transcription. The main focus of this thesis is on these three TFs, E2, E3A, and E3C, which are described in more detail in this section.

1.2.1 EBNA2

1.2.1.1 The EBNA2 protein

EBNA2 (E2) displays the key transactivator of EBV in immortalization and it is absolutely essential for B cell infection and proliferation. E2, together with EBNA-LP, is the first latent protein to be expressed after infection and subsequently activates further viral and also cellular genes. First E2 is expressed from the viral W promoter (Wp), which is then switched to the C promoter (Cp) due to E2 binding and activation. Also EBNA1 and the E3 transcripts are initiated from this Cp. Furthermore E2 induces the expression of the LMPs from different viral promoters (reviewed in Kempkes and Ling, 2015). The E2 protein which was studied most extensively derives from EBV strain B95.8 and consists of 487 aas (Skare et al., 1982, Baer et al., 1984) and is schematically depicted in Figure 2.

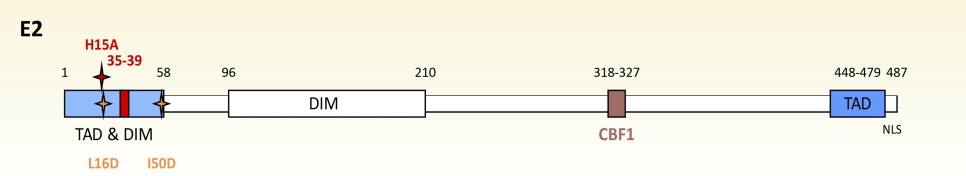


Figure 2. Schematic representation of E2 protein and its functional domains. E2 (B95.8 strain) exhibits two domains which mediate dimerization (DIM) as well as two transactivation domains (TAD, blue); one is located at the N-terminal the other at the C-terminal region of the protein. The N-terminal region (aas1-58) was studied in further detail i.a. by structural analysis and the essential role of aas L16 and I50, which are located at the interface of an E2 homo dimer could be demonstrated (orange stars). The substitution of each aa, from a hydrophobic aa to negatively charged aspartic acid (D), led to a complete loss of dimerization. Furthermore H15 and aas 35-39 (red), which form an α-helix, are exposed on the protein surface and were found to be very important in E2 mediated transactivation (Friberg et al., 2015). CBF1 binding was mapped to aas 318-327 with the E2 protein (Ling and Hayward, 1995).

E2 features two dimerization domains (DIM), which mediate homo dimerization, where the N-terminal one (aas 1-58) was studied in more detail by heteronuclear NMR-spectroscopy. Subsequent structure-guided mutational analysis revealed two aas within the hydrophobic homodimer interface (L16 and I50) to be essential for dimerization. Furthermore, the surface exposed aa H15 and an α-helix consisting of aas 35-39 were found to be important for E2 mediated transactivation (Friberg et al., 2015). This N-terminal DIM was also found to exhibit transactivation function (TAD) and was also described to interact with EBNA-LP (Gordadze et al., 2004, Harada et al., 2001, Peng et al., 2004). The second TAD at the C-terminal region (aas 448-479) of E2 was described to bind to TFIIB, TAF40, and TFIIH, factors of the transcription initiation complex, as well as RPA70, the replication protein A (Tong et al., 1995b, Tong et al., 1995a). Both TADs are able to recruit histone acetyltransferases CBP, p300, and PCAF as well (Wang et al., 2000), and the structure of this E2 TAD with CBP/300 and TFIIH was recently solved by NMR (Chabot et al., 2014). The structure of full-length E2 could not be solved to date, due to high proline content and RG repeats, which most likely prevent structured folding in the absence of specific binding partners.

1.2.1.2 DNA accession of EBNA2

Interestingly, E2 is not able to directly access DNA but utilized the ubiquitously expressed cellular transcription factor, *C promoter binding factor 1* (CBF1), which is also termed *Suppressor of Hairless* (Su(H)) in *D. melanogaster*, Lag-1 in *C. elegans* and therefore summarized as CSL, and is sometimes also referred to as RBPJ (Grossman et al., 1994, Henkel et al., 1994). CBF1, a sequence specific DNA binding protein, is the downstream effector of Notch signaling pathway and is described to recruit co-repressor complexes to DNA in the absence of Notch to repress specific target genes. These co-repressor complexes include combinations of proteins including SMRT, NCoR, HDAC1/2, Sin3A, SAP30, CIR, SKIP, and CtBP (reviwed in Lai, 2002). E2 is

able to displace this co-repressor complex, supposed to bind to CBF1 in complex with DNA, and recruits co-activators of transcription in a second step (Hsieh and Hayward, 1995). Since CBF1 plays a pivotal role in Notch signaling and also displays the DNA adaptor for Notch, a potential mimicry of Notch function by E2 was proposed and investigated (reviewed in Hayward et al., 2006). Indeed it could be shown that E2 and Notch both bind to a hydrophobic pocket within the repression domain of CBF1, yet to distinct aas, (Fuchs et al., 2001, Kovall and Hendrickson, 2004) and therefore binding of these two TFs is mutually exclusive. However, target gene comparison of E2 and Notch revealed a negligible overlap and favored a scenario where E2 rather hijacks CBF1 as a DNA adaptor than fully mimicking Notch signaling. A genome wide search for E2 and CBF1 binding sites by Chromatin Immunoprecipitation experiments followed by deep sequencing of the associated DNA fragments (ChIP-seq) revealed an overlap of approx. 70% of E2 and CBF1 sites and indicated CBF1 as the major DNA adaptor for E2. Furthermore, the cellular B cell lineage defining TF PU.1 was described to mediate E2 binding to DNA, in concert with CBF1, in activation of the viral *LMP1* promoter (Johannsen et al., 1995, Laux et al., 1994a, 1994b). However, this was not reported for cellular regulatory elements and complex formation of both proteins could only be demonstrated once (Yue et al., 2004).

1.2.2 EBNA3A and EBNA3C

1.2.2.1 E3A and E3C proteins – Regulators of transcription

The E3 gene family consists of three members, E3A, E3B, and E3C, which are only expressed in latency III and are located as a tandem array in the EBV genome. All E3 transcripts are initiated from the viral promoter Cp, which can be activated by E2 action, gives rise to all EBNA transcripts by different splicing events, and is only active during latency III. These three genes are thought to be derived from gene duplication events due to their similarity in genomic and protein structure, which display a conserved region of approx. 30% aa identity, specific for this family, in the N-terminal region (Allday et al., 2015) (Fig. 3). All three members were described to be regulators of transcription which are able to bind to the DNA binding protein CBF1 as well. However, a functional redundancy within the E3 family could not be confirmed (O'Nions and Allday, 2004). Furthermore, E3B is not essential for B cell immortalization while E3A and E3C were described to be indispensable. However, this statement was disproven for E3A, since LCLs deficient for E3A could be established in our laboratory on a regular basis. These LCLs are impaired in proliferation and showed elevated apoptosis rates during the first three months post

infection (Hertle et al., 2009). Due to their very critical functions in B cell immortalization E3A and E3C were further investigated in this thesis.

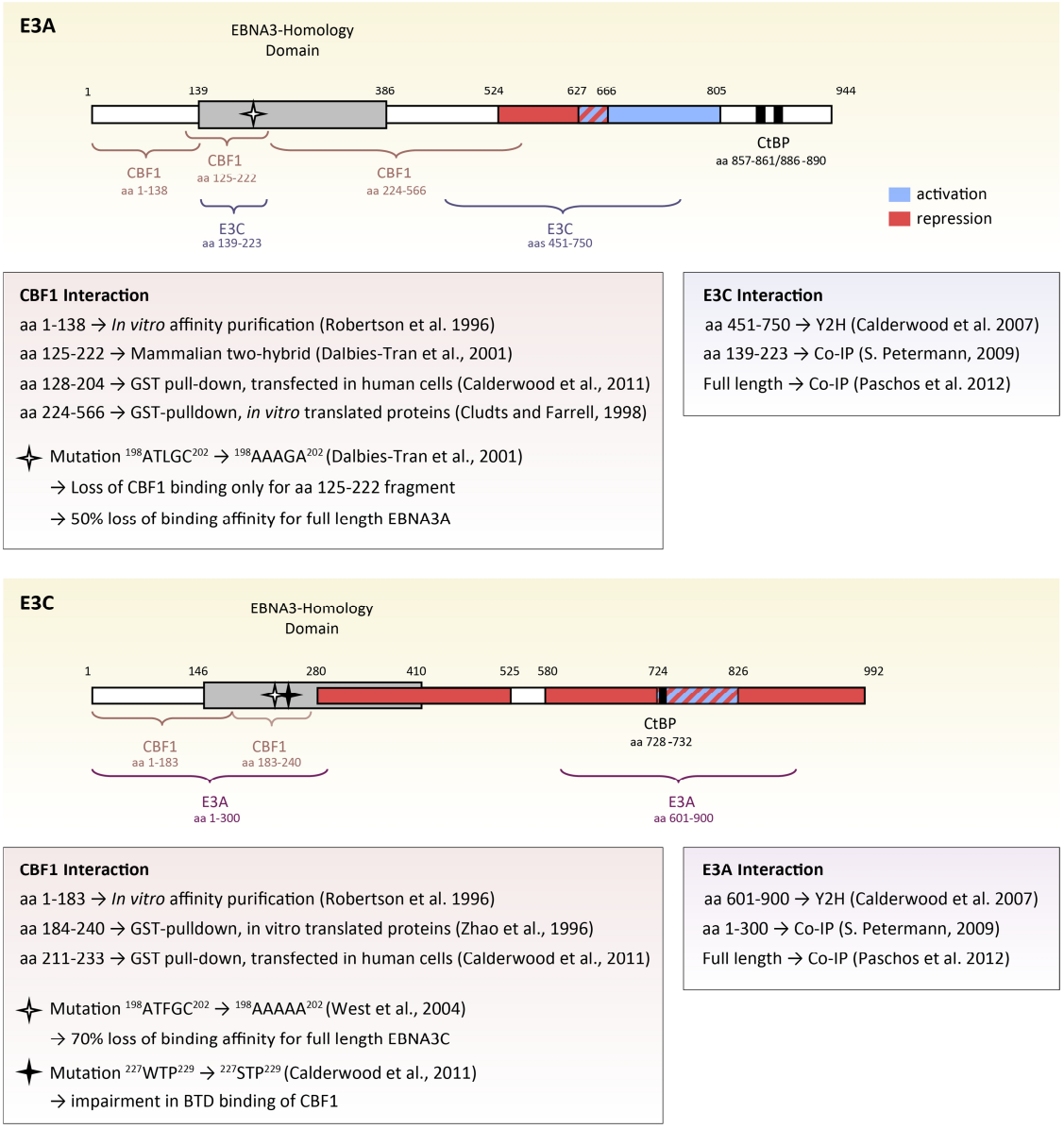


Figure 3. Schematic representation of E3A and E3C proteins and summarized information on CBF1 interaction and hetero dimerization. B95.8 strain E3A and E3C proteins and the aas involved in divers protein-protein interaction are displayed linearly. The E3 homology domain is indicated (grey) as well as potential activation (blue) or repression domains (red). Aas described to be important for CBF1 interaction are indicated and the experimental evidence is specified in the boxes below. Also the aas mediating hetero dimerization of E3A and E3C are highlighted and the underlying data described in the boxes below.

E3A and E3C were initially described as regulators of transcription in GAL4 reporter assays, when tethered to DNA by the fusion to GAL4-DBD (Bain et al., 1996, Bourillot et al., 1998, Cludts and Farrell, 1998, Marshall and Sample, 1995). Hence, a functional repression domain was mapped to aas 524-666 of E3A (Bourillot et al., 1998) but also activating properties could be assigned to a fragment of aas 627-805 (Dalbies-Tran et al., 2001). Similarly, a repression domain comprising aas 280-525 of E3C was described as well as a second, less potent one, including

aas 580-992. Also in the case of E3C an "activation domain" (aas 724-826) could be mapped, residing within the C-terminal repression domain (Bain et al., 1996). Since both proteins, E3A and E3C, mediate repression as full length proteins in reporter assays, it seems very likely that the identified "activation domains" are actually masked within the secondary protein structure and therefore are not important in gene regulation.

1.2.2.2 Protein-protein interactions of E3A and E3C

E3A and E3C are able to form hetero-dimers by interaction of the C-terminal regions, respectively, as initially described by yeast two-hybrid (Y2H) experiments (Calderwood et al., 2007), which was confirmed by Co-IP experiments in LCLs for full length proteins (Paschos et al., 2012). Mutational analyses using recombinant proteins expressed in human cells mapped the interaction domains within the N-terminal part of the E3 proteins (dissertation S. Petermann, 2009) (Fig. 3). Therefore, a functional cross-talk between E3A and E3C seems possible.

Both viral TFs were shown to bind to the cellular co-repressor of transcription CtBP (*C-terminal binding protein*) and binding could be mapped to aas 857-861 and 886-89 within E3A (Hickabottom et al., 2002) and aas 728-732 within E3C (Touitou et al., 2001). In the case of E3A this interaction was described to be critical for B cell transformation (Maruo et al., 2005). Furthermore, E3A and E3C interact with various other co-factors repressing transcription, including histone deacetylases HDAC1 and HDAC2, Sin3A and NCoR (Radkov et al., 1999, Knight et al., 2003). Especially E3C was extensively analyzed in search for interacting proteins, and indeed several cellular factors could be identified, including cyclin A, cyclin D1, SUMO1/3, SCF, RB, MYC, p300, MDM2, CHK2, and H2AX (reviewed in Allday et al., 2015). However, most approaches consisted of pull-down assays after co-transfection of recombinant proteins which were not confirmed in endogenous settings and therefore were not further elaborated in this thesis.

1.2.2.3 DNA accession of E3A and E3C

As mentioned above, E3A and E3C are also able to bind to the cellular TF CBF1 and involved aas within the viral proteins could be mapped to the E3 homology domain, respectively, and showed a somewhat contradictory picture of CBF1 binding with E3 proteins (Fig. 3) (Robertson et al., 1996, Zhao et al., 1996, Cludts and Farrell, 1998, Dalbies-Tran et al., 2001). However, mutation analyses confirmed the importance of aas 198-202 of full-length E3A or E3C proteins for CBF1 binding (Dalbies-Tran et al., 2001, West et al., 2004). Additionally, a WTP motif discovered in E3C which resembles WΦP motif of Notch (WFP) or E2 (WWP), known to mediate CBF1 interaction, was shown to be responsible for the interaction with the beta trefoil

domain (BTD) of CBF1. E3A and E3C mutants deleted for fragments in the E3 homology domain, including these aas, did not interact with CBF1 and failed to repress E2 mediated activation of reporter genes and more importantly could not maintain lymphoblastoid cell growth (Maruo et al., 2005, 2009, Lee et al., 2009). Interestingly, E3A and E3C bind to the same site as E2 within the CBF1 protein (Robertson et al., 1995, 1996) and therefore E2 and E3 interaction with CBF1 is most likely mutually exclusive. However, the structures of the different protein complexes have not been solved to date and even if E2 and E3 proteins do not bind to the very same aas within CBF1, displacement would be possible.

The regulation of the bidirectional viral *LMP1/LMP2B* promoter by E3C in concert with E2 was shown to be dependent on a PU.1 binding site and a direct interaction between E3C and PU.1 was demonstrated *in vitro* (Zhao and Sample, 2000). Subsequently, recruitment of E3C to this promoter could be shown in E3C inducible EBV positive B cells, a mechanism which did not apply to E2 regulated Cp which was responsive to E3C in reporter assays (Jimenez-Ramirez et al., 2006).

1.2.3 Partly antagonistic gene regulation by E2 and E3 proteins

In the past, the regulation of individual genes by the different EBNAs was reported applying divers assays and cell lines, yet did not reveal a general strategy in gene regulation. Not until the application of micro-arrays, genome wide differential gene expression patterns of knock-out (ko), mutant, or conditional EBV positive cell lines could be revealed. These included E2 target genes identified using conditionally active E2 in the EBV positive LCL background (Spender et al., 2006, Zhao et al., 2006) as well as conditional E2 expression in EBV negative BL cell lines (Maier et al., 2006, Lucchesi et al., 2008). Summarized, in these studies *FCER23* (CD23), *CR2* (CD21), *CCR7*, *HES1*, *BATF*, *BCL2A1*, *FCRL5*, *ABHD6*, *CCL3*, *CCL4*, *CDK5R1*, *DNASE1L3*, *MFN1*, *RAPGEF2*, *RHOH*, *SAMSN1*, *SLAMF1*, and *CXCR7* could be identified as E2 target genes in EBV negative B cells, independent of the expression of other viral factors. In the EBV positive cells the proto-oncogene *MYC*, the p55α subunit of *PIK3R1*, *FCER23*, *CR2*, *RUNX3*, and *FCRL5* were shown to be direct targets of E2, since their induction was independent of de novo protein synthesis, while the induction of *CCND1* (cyclin D), *CDK4*, *CSF3*, and *LT* (lymphotoxin alpha) genes required additional cellular or viral factors.

Also for the E3 proteins the application of micro arrays revealed cellular target genes on a genome wide scale. To this end EBV negative BL cells, conditionally expressing E3A or E3C (White et al., 2010, McClellan et al., 2012), E3A ko LCLs (Hertle et al., 2009), or LCLs expressing conditionally active E3C (Zhao et al., 2011a, Skalska et al., 2013) were applied and showed many target genes involved in apoptosis, cell cycle progression and lymphocyte differentiation

(reviewed in Allday et al., 2015). Interestingly, the overlap of target genes was very small comparing different cellular backgrounds and already pointed towards the importance of certain cellular factors, as TFs and chromatin landscape, in EBNA specific target gene regulation.

Most interestingly, the comparison of E2, E3A, and E3C target genes, identified in the Kempkes laboratory, revealed a significant overlap for these gene sets (Fig. 4). It could be shown that 16.2 and 13.1% of E3A and E3C target genes, respectively, are counter-regulated by E2 action. But also co-regulation for 9.1 and 6.3% of E3A and E3C targets, respectively, by E2 was demonstrated (Fig. 4A and B). Furthermore, the comparison of E3A and E3C targeted revealed significant cooperation in target gene regulation with 12.2 or 16.2% of co-regulated genes, depending on the reference target gene set due to differences in absolute numbers of E3A and E3C regulated genes. But also some E3A and E3C counter-regulated genes could be identified (Fig. 4C and D).

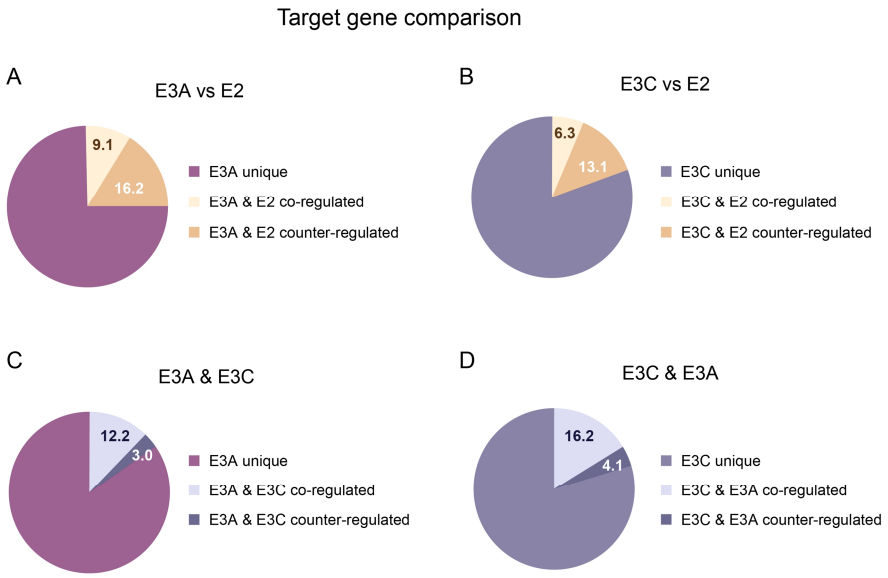


Figure 4. Comparison of E2, E3A, and E3C target genes identified in the Kempkes laboratory. E2 target genes derived from BJAB and BL41 EBV negative cell lines expressing inducibly active E2 (Maier et al., 2006), while E3A and E3C target genes were obtained by comparing wt with ko LCLs (Hertle et al., 2009, diploma thesis A. Nowak, 2008). Percentages of identified (A) E3A or (B) E3C target genes which are also co- or counter-regulated by E2. The overlap between E3A and E3C regulated genes using (C) E3A or (D) E3C target gene set as basis for analysis (percentages differ due to different absolute numbers of regulated genes). Percentages of co- and counter-regulated genes are indicated (Harth-Hertle, unpublished).

In summary, these data indicate the possibility of a functional antagonism of E2 and the E3 proteins but the mechanisms underlying these observations are still not clear to date. Two popular mechanisms have been proposed, one which favors a model where the E3 proteins destabilize the interaction of the E2-CBF1 complex with DNA (Robertson et al., 1995, Robertson et al., 1996, Waltzer et al., 1996), and a second, which pictures the E3 proteins to replace E2 on DNA-bound CBF1 and subsequently recruit co-repressors of transcription

(Radkov et al., 1999, Touitou et al., 2001, White et al., 2010). However, both might not be mutually exclusive and it might be possible that both apply for distinct sets of target genes.

1.3 Objectives

In 2011, the year this thesis was initiated, very little was known about the mechanisms by which E2 and E3 proteins access chromatin and subsequently regulate target gene transcription. CBF1 had been identified as a potential DNA anchor shared by the viral proteins. Most studies had focused on the viral genome and only selected cellular genomic loci had been studied with respect to E2 and E3 binding, co-occurring cellular transcription factors, chromatin signatures, and transcriptional activity. One example, which was extensively investigated in the Kempkes laboratory, is the CXCL9 and -10 gene locus. Here, E3A was shown to compete with E2 for chromatin binding to an intergenic enhancer region and to repress transcription by a CBF1 dependent process (Harth-Hertle et al., 2013). In the Kempkes laboratory extensive information on EBNA target genes was collected and, supported by published data from other laboratories, formed a picture, where E2 and E3 proteins seem to share a significant subset of counter- and even co-regulated genes. This suggested that CBF1 might be the common determinant of E2 and E3 functions. On the other hand, the majority of E2, E3A and E3C target genes are regulated uniquely by a single viral transcription factor, indicating that additional cellular or viral factors are critical determinants for E2 and E3 function. These determinants could either influence the chromatin state of respective target sites or serve as anchors for E2 or E3 proteins to promote chromatin binding.

In order to define the specific and unique cellular determinants for E2 and E3 binding to cellular chromatin, a genome wide screen for E2, E3A, and E3C binding sites was performed by Chromatin Immunoprecipitation (ChIP) experiments followed by deep sequencing of the associated DNA fragments (ChIP-seq). The resulting data sets were analyzed in context of publicly available information on LCLs provided by the Encyclopedia of DNA Elements (ENCODE) consortium (ENCODE_Consortium, 2012) including ChIP-seq experiments for 84 TFs (by June 2015, start of bioinformatics analysis) and extensive information on the chromatin state. For this thesis, a novel bioinformatics strategy had to be developed. Based on genome wide quantitative correlation analyses an unbiased complex picture of the specific composition of the different subsets of EBNA binding sites needed to be generated.

2 Material

2.1 Cell lines

Table 1. General and commercially available cell lines

Cell Line	Description	Reference
721	Human lymphoblastoid cell line, immortalized with EBV type I strain B95.8	Kavathas et al. (1980)
DG75	Human EBV negative Burkitt's lymphoma (BL) cell line	Ben-Bassat et al. (1977)
HEK293	Human embryonic kidney epithelial cell line, transformed by DNA fragments of adenovirus type 5	Graham et al. (1977)
Raji	Human EBV positive BL cell line	Pulvertaft (1964)

Table 2. Lymphoblastoid cell lines

Cell Line			
Donor	Recombinant EBV	Clone	Internal Designation
D1	Flag-E3A	1	LG309.1.1
		2	LG309.1.2
	Flag-E3C	1	LG309.2.1
		2	LG309.2.2
	wt	1	LG309.3.1
		2	LG309.3.2
D2	Flag-E3A	1	LG395.1.1
		2	LG395.1.2
	Flag-E3C	1	LG395.2.1
		2	LG395.2.2
	wt	1	LG395.3.1
		2	LG395.3.2
D1	Flag-E3A	1	LG396.1.2
		2	LG396.1.5
	Flag-E3C	1	LG396.2.3
		2	LG396.2.6
	wt	1	LG396.3.1
		2	LG396.3.2

All LCLs listed above were generated in this thesis.

Table 3. HEK293 based EBV producer cell lines

Cell Line	Recombinant EBV	(BACmid Designation)	Internal Designation
HEK293/Flag-E3A*	Flag-E3A	(pFlag-E3A)	LG267 A 5.2
HEK293/Flag-E3C*	Flag-E3C	(pFlag-E3C)	LG267 A 3.1
HEK293/2089	wt	(p2089)	2089
HEK293/ΔEBNA2	EBNA2 deletion	(pΔEBNA2)	KG481-2

(* generated in this thesis)

2.2 BAC constructs and plasmids

Table 4. Recombinant EBV BACmids

BACmid	Description	Internal Designation
pFlag-E3A*	Triple Flag-tag integrated in frame and N-terminal of EBNA3A gene	pLG174.1a.1
pFlag-E3C*	Triple Flag-tag integrated in frame and N-terminal of EBNA3C gene	pLG191.2.1
pgalk-E3A*	galK integrated in frame and N-terminal of EBNA3A gene, target site for Flag-tag, intermediate product for positive/negative selection during Recombineering	pLG154.1a
pgalK-E3C*	galK integrated in frame and N-terminal of EBNA3C gene, target site for Flag-tag, intermediate product for positive/negative selection during Recombineering	pLG154.2a
p2089	wt EBV, strain B95.8 (Delecluse et al., 1998)	p2089
pΔEBNA2	EBNA2 deletion	p2491

(* generated in this thesis)

Table 5. Plasmids

Plasmid	Description	Reference	Internal Designation
pgalK	Galactokinase (*galK*) expression plasmid for amplification of *galK* for *Recombineering* cloning technique	Warming et al. (2005)	-
pBZLF1	Mammalian vector for the constitutive expression of Zta/BZLF1 (pCMV backbone)	AG Hammerschmidt	p509
pBALF4	Mammalian vector for the constitutive expression of gp110/BALF4 (pCMV backbone)	AG Hammerschmidt	p2670

2.3 DNA constructs

Table 6. DNA constructs for Recombineering

Construct	Description	Sequence	Internal Designation
E3A/H1-Flag-H2	Triple Flag-tag flanked by 150 bp homologous to region upstream of E3A including ATG (H1) and 150 bp of E3A gene (H2)	CCGTGAGATGGATCAGGCTCTGGATGGTGTACTGACACACAAGCAAGG CTGCCTCCATTGTCTCGGCACCGATTTCTAGGCAGCATCCTCTTTAAT AGGTACAAGGGGGGTGCGGTGTTGGTGAGTCACACTTTTGTTGCAGAC AAA*ATG***GACTACAAAGACCATGACGGTGATTATAAAGATCATGACATC GACTACAAGGATGACGATGACAAG**GACAAGGACAGGCCGGGTCCCCCG GCCCTGGATGACAACATGGAAGAAGAAGTCCCATCTACCTCGGTTGTG CAGGAACAGGTATCGGCGGGAGATTGGGAAAATGTCCTCATAGAGTTA TCAGATAGCAGCTCAGAAAAGGAAGCAGAA	LG177
E3C/H1-Flag-H2	Triple Flag-tag flanked by 150 bp homologous to region upstream of E3C including ATG (H1) and 150 bp of E3C gene (H2)	TCTGAAACATCGAACGATGAGTGATTTCGCCCATGTAACAAGAACTGG GATGAACCCTGGGGCAACAGACTGCGGGGAGGAGGGGGGCAGTGATAA GTCATGACAATTTTAGATGAGGTAGAAATTTTGCATATTTTCAGACCC ACC*ATG***GACTACAAAGACCATGACGGTGATTATAAAGATCATGACATC GACTACAAGGATGACGATGACAAG**GAATCATTTGAAGGACAGGGGGAC TCTAGACAGTCACCCGACAATGAGCGGGGAGATAATGTACAGACTACC GGCGAGCATGATCAGGACCCTGGGCCGGGGCCTCCATCCAGTGGGGCT TCTGAGAGATTGGTACCAGAAGAGTCATAC	LG178

DNA constructs were ordered at MWG Operon (Ebersberg, Germany), already cloned into pCR2.1, and only used for amplification via PCR. The Flag-tag encoding sequence is highlighted in bold, primer sites for amplification (150 bp or 100 bp homology arms possible) are underlined, and the respective start codon is shown in italic, underlined letters.

2.4 Bacteria

Table 7. Bacterial strains

E. coli	Description	Application	Reference
DH5α	F⁻ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*Δ M15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^- m_K^+$), λ⁻	Default plasmid amplification and cloning	Hanahan (1985)
SW105	DH10B [λ*c*1857 (*cro-bioA*)<>*Tet*] *gal490* (*cro-bioA*)<>*araC*-P_{BAD}*Flpe gal*⁺ Δ*galK*	Recombineering	Warming et al. (2005)

2.5 Primers

All primers used in this thesis were ordered at Metabion AG (Martinsried, Germany) and designed via Primer3 web-based software (http://frodo.wi.mit.edu) and tested negative for off-target hits by BLAT search (http://genome.ucsc.edu/cgi-bin/hgBlat). Primers used for transcript quantification were designed to amplify across exon-exon junctions if possible. If multiple Refseq transcripts (hg19) were assigned for one gene, primers were chosen to cover as many transcript variants as possible by targeting shared exons. In silico PCR for these primer pairs were performed to exclude off-target effects in genomic DNA.

Table 8. Primers for amplification of galK flanked by sequence specific homology arms (50 bp) for Recombineering as overhangs

Primer	Internal Designation	Sequence	Annealing (°C)	Product (bp)
E3A/H1(50bp 5' E3A +ATG)-galK(first 20bp)	Be859 (for)	ACAAGGGGGGTGCGGTGTTGGTGAG TCACACTTTTGTTGCAGACAAA*ATG* **CCTGTTGACAATTAATCATCGGCA**	60	1,331
galK(last 20bp)-E3A/H2(first 50bp -ATG)	LG150 rev	TCTTCTTCCATGTTGTCATCCAGGG CCGGGGGACCCGGCCTGTCCTTGTC **TCAGCACTGTCCTGCTCCTT**	60	
E3C/H1(50bp 5' E3C +ATG)-galK(first 20bp)	Be861 (for)	TGACAATTTTAGATGAGGTAGAAAT TTTGCATATTTTCAGACCCACC*ATG* **CCTGTTGACAATTAATCATCGGCA**	60	1,331
galK(last 20bp)-E3C/H2(first 50bp -ATG)	Be862 (rev)	CGCTCATTGTCGGGTGACTGTCTAG AGTCCCCCTGTCCTTCAAATGATTC **TCAGCACTGTCCTGCTCCTT**	60	

Primers were designed to amplify *galK* from pgalK and generate homology arms (H1 and H2) specific for E3A or E3C N-terminal integration site by including 50 bp overlaps. GalK specific sequences are highlighted in bold and start codons are indicated in italic letters.

Table 9. Primers for amplification of Flag constructs from plasmids

Target	Internal Designation	Sequence	Annealing (°C)	Product (bp)
E3A/H1-Flag-H2/150bp	LG179a for	CCGTGAGATGGATCAGGCT	55	366
	LG179a rev	TTCTGCTTCCTTTTCTGAGCT	55	
E3A/H1-Flag-H2/100bp	LG179b for	GCCTCCATTGTCTCGGCA	55	266
	LG179b rev	CCCAATCTCCCGCCGATA	55	
E3C/H1-Flag-H2/150bp	LG180a for	TCTGAAACATCGAACGATGAG	55	366
	LG180a rev	GTATGACTCTTCTGGTACCAAT	55	
E3C/H1-Flag-H2/100bp	LG180b for	TGAACCCTGGGGCAACAGA	55	266
	LG180b rev	CCGGCCCAGGGTCCTGAT	55	

These primers were used to amplify the Flag construct with the specific homology arms from ordered plasmids listed in Table 6.

Table 10. Primers for diagnostic PCR of Flag constructs in EBV background

Pair	Target	Internal Designation	Sequence	Annealing (°C)	Product (bp)
A	Flanking Flag 5' E3A	LG179a for	CCGTGAGATGGATCAGGCT	55	300(-Flag) /366(+ Flag)
		LG179a rev	TTCTGCTTCCTTTTCTGAGCT	55	
B	Only Flag 5' E3A	LG179a for	CCGTGAGATGGATCAGGCT	55	212
		LG158 rev	TCATCGTCATCCTTGTAGTCG	55	
C	Flanking Flag 5' E3C	LG180a for	TCTGAAACATCGAACGATGAG	55	300(-Flag) /366(+ Flag)
		LG180a rev	GTATGACTCTTCTGGTACCAAT	55	
D	Only Flag 5' E3C	LG158.2 for	CGGAGGAACTGCTAAACAGG	55	282
		LG158 rev	TCATCGTCATCCTTGTAGTCG	55	

Diagnostic PCR primers were used for verification of EBV genomes in recombinant EBV BACmids (Fig. 7), HEK293 producer cell lines (Fig. 8), and LCLs (Fig. 9).

Table 11. Primers for transcript quantification

Gene	Internal Designation	Sequence	Annealing (°C)	Product (bp)
E3A	MH277 for	GAAACCAAGACCAGAGGTCC	63	276
	MH277 rev	CCCAGGGCCGGACAATAGG		
E3C	LG321 for	GACAGTCACCCGACAATGAG	63	344
	LG321 rev	TTGCAGGTGCGATTGCTTG		
BCL2L11	BimEL for	GCTGTCTCGATCCTCCAGTG	60	128
	BimEL rev	GTTAAACTCGTCTCCAATACG		
CXCL9	MH2146 for	GCATCATCTTGCTGGTTCTG	63	255
	MH2146 rev	TTTGGCTGACCTGTTTCTCC		
CXCL10	MH2145 for	TGACTCTAAGTGGCATTCAAGG	63	239
	MH2145 rev	CCTTTCCTTGCTAACTGCTTTC		
GAPDH	BS688 for	GAAGGTGAAGGTCGGAGTC	63	152
	LG90 rev	TGGGTGGAATCATATTGGAAC		

Table 12. Primers for quantification of DNA recovered by ChIP experiments

Target	Internal Designation	Sequence	Annealing (°C)	Product (bp)
SDAD1 TSS	MH1729 fw	CTCGTGTTTCCGGGTATGAC	63	95
	MH1729 rv	TGAGGCTTCCGTAGCATAGC		
CXCL9 TSS	MH1726 fw	AGCTGAGCTAACTAAATTGACCAC	63	81
	MH1726 rv	ACATGCAGAAATTCCCTTGG		
CXCL E1	MH2348.B fw	CAGGGACGGTAAGAGCCTTC	63	82
	MH2348 rv	AAATTCAAACAGGCCTGGAG		
CXCL E3	MH2350 fw	GTGTTTGCTCAAGGCCCTAC	63	77
	MH2350 rv	TGCTTGCAGGGAAGGATATAAG		
CXCL10 TSS	MH824 fw	TCCCTCCCTAATTCTGATTGG	63	138
	MH1008 rv	AGCAGAGGGAAATTCCGTAAC		
CXCL11 TSS	MH1721 fw	TGAGTCATGCACCTTTCCTG	63	162
	MH1721 rv	AAGAAGGCTGGTTACCATCTG		
CXCL E4	MH2352 fw	AGTTGGTGGCTGGGTATGTG	63	128
	MH2352 rv	GCCACATGGGAGACATTAAAC		
CXCL E5	LG465 fw	ACACACAAACACAACAAACCTG	63	117
	LG465 rv	GCCACAATTCCTGCTGTTTAC		
ADAM28 TSS	LG567 fw	ATTGTTGCAGGACCACAGC	63	112
	LG567 rv	TGCCTCCTCTCCAGTGAGAC		
ADAM28 +20kb	LG460 fw	ACACCTCATCTGTCCCGAAC	63	107
	LG460 rv	TGGATCAGCACATTTCTTGC		
ADAM E1	MH2675 fw	CTTCATGGCTACAGACTCTTGG	63	93
	MH2675 rv	CCTATGTCTCGCTTCCTGCT		
ADAMDEC1 TSS	MH2752 fw	CCCCAATCTCACACGAAAAG	63	99
	MH2752 rv	AAGTTGTGGTCTCCCCAGTG		
ADAM E2	MH2676 fw	GTTTGGCAAGCCTTCTTCTG	63	89
	MH2676 rv	GAGCCTGTGTCTCAGAGGTG		
MED13L TSS	LG587 fw	GAAGTGCGACCCAGAATCC	63	129
	LG587 rv	ATCGTCTCTCTCTCGCCTTG		
MED13L -75kb	LG649 fw	CCATTCATGCAACAGTGAGG	63	114
	LG649 rv	GCAACCTCCAACTTCTGGTC		
MED13L E1	LG613 fw	GGCTTCTTGACGGTTACTGC	63	108
	LG613 rv	CATGATGCTCAGCTCTGTGG		
MED13L E2	LG614 fw	CACTGGCACCTTCCTTTCTC	63	130
	LG614 rv	CTGGGCTGAGCTAGAAGTGG		
cntrl (RPL30)	ST122 fw	CTGGTCTGACGCTCCTGACT	63	120
	ST122 rv	CAGTGCCCAGAATTCCAGAT		
CD23 p	ST156 fw	TGTGATCGGCCATAGTGGTA	63	101
	ST156 rv	TTAAGCAGCAAGTTCCCACA		

2.6 Antibodies

Table 13. Primary Antibodies

Specificity	Official Designation	Species	Miscellaneous	Application	Reference
α-BATF	B-ATF H-19	Goat	Pc IgG, product no.: sc-15280	WB	Santa Cruz Biotechnology
α-CBF1	RBP-J 7A11	Rat	Mc, IgG2b	WB	E. Kremmer, IMI
α-EBF1	EBF C-8	Mouse	Mc, IgG2a, product no.: sc-137065	WB, IP	Santa Cruz Biotechnology
α-E1	E1B5 1H4-1-4	Rat	Mc, IgG2a	WB	E. Kremmer, IMI
α-E2	1E6	Rat	Mc, IgG2a	ChIP	E. Kremmer, IMI
α-E2	R3	Rat	Mc, IgG2a	WB, ChIP	E. Kremmer, IMI
α-E3A	E3AN 4A5-1111	Rat	Mc, IgG2a, epitope within aas 1-50	WB	E. Kremmer, IMI
α-E3B	E3B2 6C9-1-1	Rat	Mc, IgG2a	WB	E. Kremmer, IMI
α-E3C	A10 P2-583	Mouse	Mc, epitope aas 682-686 (WAPSV)	WB	E. Kremmer, IMI
α-Flag	M2 F3165	Mouse	Mc, IgG1, epitope DYKDDDDK	WB	Sigma-Aldrich (F3165)
α-GAPDH	MAB374 6C9	Mouse	Mc, IgG1	WB	Millipore, USA (MAB374)
α-GST	6G9	Rat	Mc, IgG2a, Isotype control	ChIP	E. Kremmer, IMI
α-HA	3F10	Rat	Mc, IgG1	ChIP	E. Kremmer, IMI
α-IRF4	IRF4 H-140	Rabbit	Pc IgG, product no.: sc-28696	WB	Santa Cruz Biotechnology
α-LMP1	S12	Mouse	Mc, IgG2a	WB	E. Kremmer, IMI
Mouse IgG1 Isotype Control		Mouse	Mc, IgG1	ChIP	Invitrogen (MA5-14453)

Mc: Monoclonal antibody, Pc: polyclonal antibody, WB: Western Blot, IP: Immunoprecipitation, ChIP: Chromatin-Immunoprecipitation, IMI: Institute for molecular Immunology, Helmholtz Zentrum München

Table 14. Secondary Antibodies

Specificity	Species	Miscellaneous	Application	Reference
α-mouse IgG	goat	HRP coupled	WB	Santa Cruz Biotechnology, USA (sc-2005)
α-rat IgG	goat	HRP coupled	WB	Santa Cruz Biotechnology, USA (sc-2006)

HRP: horseradish peroxidase

2.7 Cell culture material

Reagent	Distributor / Reference
Dimethyl sulfoxide (DMSO)	Merck, Germany
Doxycycline (Dox)	Sigma-Aldrich, USA
Fetal Calf Serum (FCS)	PAA Laboratories, Austria
Hygromycin B	Invitrogen, UK
L-Glutamine	GIBCO, UK
OptiMEM Medium	GIBCO, UK
Penicillin/Streptomycin	GIBCO, UK
Puromycin	Merck (Calbiochem), Germany
RPMI 1640-Medium	GIBCO, UK
Trypsin	GIBCO, UK

2.8 Bacterial culture material

Reagent	Distributor / Reference
Agar	Bacto™, BD, USA
Ampicillin	Sigma-Aldrich, USA
Chloramphenicol	Sigma-Aldrich, USA
D-Biotin	Sigma-Aldrich, USA
Galactose	Sigma-Aldrich, USA
Glycerol	Merck, Germany
L-Leucine	Sigma-Aldrich, USA
MacConkey Agar Base	Difco, BD, USA
M9 Minimal Salts	Sigma-Aldrich, USA
M63 Minimal Salts	Sigma-Aldrich, USA
Tryptone	Bacto™, BD, USA
Yeast Extract	Bacto™, BD, USA
2-Deoxy-galactose (2-DOG)	Sigma-Aldrich, USA

2.9 Enzymes and reaction kits

Reagent	Distributor / Reference
ECL	GE Healthcare (Amersham), UK
High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems, Thermo Fisher Scientific, USA
LightCycler 480 SYBR Green I Master	Roche Diagnostics, Germany
NucleoSpin Plasmid	Macherey-Nagel, Germany
peqGold Taq Polymerase, all inclusive	PEQLAB, Germany
Phusion® High-Fidelity DNA Polymerase	New England Biolabs, USA
Proteinase K (PCR grade)	Roche Diagnostics, Germany
QIAamp DNA Mini Kit	QIAGEN, Germany
QIAquick PCR Purification Kit	QIAGEN, Germany
Qubit® dsDNA HS Assay Kit	Invitrogen, UK
Restriction Enzymes & Buffers	New England Biolabs, USA
RNase A	Sigma-Aldrich, USA
RNase-free DNase Set	QIAGEN, Germany
RNeasy Mini Kit	QIAGEN, Germany

2.10 Chemicals and reagents

Reagent	Distributor / Reference
Acrylamid 30%	Roth, Germany
Agarose	Invitrogen, UK
APS	MP Biomedicals, Germany
BSA	MP Biomedicals, Germany
Complete Protease Inhibitor	Roche Diagnostics, Germany
Ethidium bromide	Merck, Germany
Ficoll-Paque Plus	GE Healthcare, UK
Formaldehyde, 37%	Merck, Germany
Glycogen	Sigma-Aldrich, USA
Isopropanol	Roth, Germany
Milk powder	AppliChem, Germany
MS2 RNA	Roche Diagnostics, Germany
Polyethylenimine (PEI)	Sigma-Aldrich, USA
Protein G-Sepharose	GE Healthcare, UK
TEMED	GE Healthcare, UK
Triton X-100	Sigma-Aldrich, USA
Trypan blue	GIBCO, UK

All chemicals which are not listed above were purchased at Merck, MP Biomedicals, Roth, and Sigma-Aldrich.

2.11 Software and databases

Name	Reference
CellQuestPro	BD Biosciences, USA
Clone Manager 9 Professional	Scientific & Educational Software, USA
Ensembl Genome Browser	http://www.ensembl.org
LightCycler 480 Software, Version 1.5	Roche Diagnostics GmbH, Germany
Primer3	http://frodo.wi.mit.edu
UCSC Genome Browser	http://genome.ucsc.edu

2.12 Bioinformatic Tools

Name	Reference
Bowtie2	Langmead and Salzberg (2012)
deepTools package	Ramirez et al. (2014a)
FastQC	Andrews (2010)
Illumina Demultiplex	AG Blum, Gene Center Munich, Germany
MACS2	Zhang et al. (2008)
Trim Galore!	Krueger (2012)

3 Methods

3.1 Mammalian cell culture methods

3.1.1 Cell culture

All cell lines were cultivated at 37°C and 6% CO_2 in RPMI 1640 Medium supplemented with 100 U/ml Penicillin, 100 µg/ml Streptomycin, 4 mM L-Glutamine, and 10 or 20% FCS respectively. Cell density was determined by using a Neubauer counting chamber. To this end, cells were diluted 1:2 with trypan blue, added to the chamber and living unstained cells were counted under the microscope. The cell density was calculated as follows: cells/ml = mean no. cells of all four big squares x 2 (dilution factor) x 10^4.

Suspension cell lines

Cell lines 721, DG75 and DG75 descendant cell lines were cultured with 10% FCS. Additionally, 1 µg/ml Puromycin was added to the medium of DG75 cell lines harboring pRTR vectors for selection. Cells were maintained at 2-4 x 10^5 cells/ml and reseeded and supplied with fresh medium every 3-4 days. Primary B cell preparations infected with recombinant EBV and established LCLs were supplemented with 20% FCS. Within the first 3-5 weeks after infection in 96 well plates, containing lethally irradiated MRC5 fibroblasts as "feeder cells", 50% of the medium was exchanged once a week. Established LCLs were cultured without feeder cells and medium was exchanged every 3-4 days adjusting the cell density to 2 x 10^5 cells/ml to ensure standardized culture conditions. Suspension cells were always centrifuged at 300 g for 10 min at RT for reseeding purposes or at 500 g for 5 min prior to harvest.

Adherent cell lines

HEK293 cells, stably transfected with recombinant EBV BACmids were cultured in medium supplemented with 10% FCS and 100 µg/ml Hygromycin B for BACmid selection. MRC5 fibroblasts received 20% FCS supplemented in the medium. To detach adherent cells from the culture dishes, cells were washed briefly with PBS, subsequently moistened with trypsin and incubated at 37°C for approx. 3 min, and cell detachment was controlled under the microscope. Cells were diluted 1:3 – 1:10 with fresh medium and reseeded every 3-4 days.

PBS	137 M NaCl, 2.7 M KCl, 7.3 M Na_2HPO_4, 1.5 M KH_2PO_4, pH 7.4

3.1.2 Long term cell storage

To preserve cells for a longer period of time, cells were frozen in liquid nitrogen. To this end, 1 x 10^7 cells were collected (suspension cells by centrifugation and adherent cells with preceding trypsin treatment), resuspended in 1.5 ml freezing medium and transferred to 1.8 ml Cryotubes (NUNC). Using a propanol freezing container cells were slowly cooled to -80°C and stored there for approx. one day. Subsequently tubes were transferred to liquid nitrogen. To re-cultivate frozen cells, these were thawed rapidly in a 37°C waterbath, washed with 30 ml medium to remove DMSO, and resuspended in fresh medium. Required selection additions were added the day after to the medium.

Freezing medium	40% culture medium, 50% FCS, 10% DMSO

3.1.3 Generation of HEK293 cells stably transfected with recombinant EBV

The day before transfection 6 x10^5 HEK293 cells were plated per well and transfection reaction of a 6-well plate. For the transfection two reaction batches were pre-mixed: A) 300 µl OptiMEM with 1 µg of the desired BACmid DNA and B) 300 µl OptiMEM with 4 µg PEI (1 mg/ml). Subsequently the two reactions were mixed and incubated for 20 min at room temperature (RT). For transfection the cell culture medium was replaced by 1 ml OptiMEM and the reaction mix was slowly added to the cells by dropping. Cells were incubated for 4 h at 37 °C and then the reaction solution was replaced by 3 ml of regular culture medium containing 10% FCS without selection. The day after the cells were trypsinized and reseeded in a 14 cm diameter culture dish and supplied with medium containing 100 µg/ml hygromycin B for selection of transfected cells. Approx. four weeks after GFP expressing clonal colonies are growing out derived from single transfected cells. These colonies were identified by fluorescence microscope analysis. To pick colonies the culture medium was removed, the cells were washed were slowly and carefully with PBS, and small pieces of filter paper, sterilized by autoclaving and pre-incubated in trypsin, were carefully put on the desired colonies with forceps. After 1 min incubation the filter piece was removed and immediately placed in a well of a 6-well plate pre-filled with 3 ml of warm medium supplemented with hygromycin. Clonal HEK cell lines were checked daily by microscope for cell density and diluted accordingly. After 2-3 weeks the clonal cell lines could be diluted and reseeded on a regular basis of 2-3 days and could be used for the production of viral particles. For each recombinant EBV genome several clones were picked and cultivated to check for differences in virus titers production. Potent clones were maintained and used for further particle production.

3.1.4 Transfection of HEK293 cells for the production of infectious viral particles

Clonal HEK293 producer cell lines, stably transfected with recombinant EBV genomes and generated as described above, were transiently transfected with BZLF1 (p509) and BALF4 (p2670) expression plasmids, to induce the lytic cycle of EBV. To this end HEK293 cells were plated on 10 cm diameter culture dishes with approx. 50-60% confluency the day prior to transfection. For transfection the culture medium was preplaced by 3 ml of OptiMEM and two reaction batches prepared as follows: A) (3 µg p509 + 3 µg p2670 + 600 µl OptiMEM)/dish and B) (24 µl PEI (1 mg/ml) + 600 µl OptiMEM)/dish. Subsequently the two reaction batches were mixed and incubated for 20 min at RT. Then 1.2 ml reaction was added per dish very carefully by dropping and the cells subsequently incubated for 4 h at 37 °C, then the supernatant was removed and replaced by 10 ml of regular RPMI medium supplemented with 10% FCS without hygromycin. After 3 days of incubation the virus particle containing supernatants were harvested and filtrated (pore size 0.8 µm) to remove potential cell contaminations and stored at 4 °C.

3.1.5 Quantification of viral titers in cell supernatants

For quantification of viral titers 3 x10^5 Raji cells in 1 ml reactions were infected with 25, 50, 100, 250, and 500 µl virus containing supernatant, respectively and further cultivated. The day after infection the culture medium was replaced by fresh one. Four days after infection the cells were harvested by centrifugation, twice washed in PBS/5% FCS and GFP expressing cells were quantified by FACS analysis. Viral titers were quantified as *green Raji units* (GRUs) per ml supernatant and the mean of all five dilution steps was used as further reference.

3.1.6 Preparation of primary B cells from cord blood

Cord blood from anonymized donors was retrieved from the Klinikum der Universität München (LMU) and lymphocytes were isolated by Ficoll density centrifugation preparation as follows: 20 ml Ficoll-Paque Plus was prepared in 50 ml reaction tubes and carefully overlayed with of 20 ml 1:3 with PBS diluted cord blood (the Ficoll volume was adjusted to match diluted blood volume but did not exceed 20m; for larger volumes, samples were split in two). The blood was added very slowly to not disturb the Ficoll layer and two phases were formed. The tubes were centrifuged at 300 g for 40 min and no brake was applied to stop the centrifuge. After this step the lymphocytes remain at the interface between plasma (top layer) and Ficoll phase (bottom) while the red blood cells are pelleted on the tube bottom. Lymphocytes are aspirated carefully and pooled if blood was split prior to Ficoll centrifugation. The cells were washed twice with

PBS/Versen (1:5000) and finally resuspended in a small volume of RPMI culture medium containing 20% FCS, counted, and immediately used for infection experiments or frozen for long time storage. Prior to infection the rate of B cells in the lymphocyte preparation was quantified by FACS analysis: To this end 4 reactions à 10^6 cells were prepared, washed with PBS/5% FCS and resuspended in 100 µl PBS/5% FCS each and 2 µl of the following FACS specific antibodies were added: 1) APC Mouse anti-human CD19 (BD Pharmingen, 555415), for the detection of B cells, 2) Mouse IgG1 negative control:APC (AbD Serotec, MCA928APC), as background control for 1), 3) PE mouse anti-human CD3 (BD Pharmingen, 555333) specific for T cells, and 4) mouse IgG1 negative control:RPE (AbD Serotec, MCA928PE) as background control for 3). The reactions were incubated for 40 min at 4 °C in the dark, subsequently washed twice with PBS/5% FCS and finally resuspended in 500 µl PBS/5% FCS and APC and PE positive cell percentages quantified via FACS analysis. All samples used for generation of LCLs in this thesis showed approx. 10% of B cells and 30% T cells within the lymphocyte preparation. Samples with lower percentages were discarded.

3.1.7 Infection of primary B cells with recombinant EBV for the generation of LCLs

One day prior to the actual infection experiment, MRC5 fibroblast cells were trypsinized, resuspended in medium, transferred to a 50 ml tube and lethally irradiated using a γ-radiation source applying 5,000 cGy. Subsequently the cells were washed with medium and reseeded to approx. 80% confluency in 96-well plates in 200 µl medium. For infection 150 µl of the culture supernatant per well were removed, and 3 x 10^5 lymphocytes in 50 µl were added per well. Virus containing supernatants were added to 5,000 GRUs/well and did not exceed 100 µl/well. For each recombinant EBV at least half a 96-well plate was infected. After infection the culture medium containing 20% FCS was exchanged once per week. After 2-3 weeks the cultures were transferred to new successively larger plates, cultivated without feeder cells, and diluted and reseeded on a regular basis.

3.2 Bacterial culture methods

3.2.1 Propagation and storage of bacteria

Bacteria were cultivated as suspension cultures in LB medium or for separation of colonies cultured on LB agar plates at 37 °C. All mediums and reagents were autoclaved prior to usage if not indicated otherwise. Transformed bacteria were selected by addition of antibiotics

appropriate for the respective resistance gene. Short term storage of bacteria was conducted at 4 °C, while 100 µl DMSO were added to 900 µl freshly overnight grown suspension culture derived from a single colony and frozen at -80 °C for long term storage.

LB medium	1% Tryptone, 0.5% Yeast Extract, 1% NaCl, pH 7.4
LB agar	LB medium supplemented with 1.5% Agar
Antibiotics	100 µg/ml Ampicillin or 20 µg/ml Chloramphenicol, respectively (sterile filtrated, dissolved in ethanol)

3.2.2 Generation of chemically transformation competent bacteria

500 ml LB medium were inoculated with 5 ml overnight culture of *E. coli* DH5α and incubated under vigorous shaking at 37 °C until an OD_{595} of 0.3-0.4 was reached. Subsequently the culture was divided in precooled 50 ml tubes, incubated on ice for 10 min and pelleted by centrifugation (1,600 g, 7 min, 4 °C). Each pellet was resuspended in 10 ml ice cold $CaCl_2$ solution, incubated for 30 min on ice and again pelleted by centrifugation (1,100 g, 5 min, 4 °C). Finally, the cell pellets were each resuspended in 2 ml ice cold $CaCl_2$ solution, aliquoted in 200 µl per pre-cooled 1.5 ml reaction tube, shock frozen in liquid nitrogen and stored at -80 °C.

$CaCl_2$ solution	60 mM $CaCl_2$, 10 mM PIPES, 15% Glycerol, sterile filtrated (0.22 µm)

3.2.3 Heat shock transformation of *E. coli*

100 µl of chemically competent *E. coli* were thaw on ice and 50-100 ng of the DNA of interest were added, mixed, and incubated for 30 min on ice. Then the bacteria were heat shocked for 50 s at 42 °C, shortly incubated on ice, and 900 µl LB medium were added. To enable expression of the resistance gene, the reaction was incubated 1 h at 37 °C under vigorous shaking prior to plating various dilutions on LB agar plates containing the appropriate antibiotic.

3.2.4 Recombineering

In order generate recombinant EBV genomes in the BACmid background the *recombineering* technique (Warming et al., 2005) was applied for cloning strategies. The distinct purpose of this cloning and the single steps are further explained in chapter 4.1.1, while the technical details are described in the following section.

Step I – Targeted integration and galK positive selection

E. coli strain SW105 already transformed with wt EBV BACmid p2089 was a gift from W. Hammerschmidt and cultivated in chloramphenicol containing LB medium or agar. For the first step of the recombineering protocol 5 ml LB medium containing chloramphenicol were

inoculated with a single colony and incubated over night at 32 °C under vigorous shaking. The day after 25 ml LB medium containing chloramphenicol were inoculated with 500 µl of the overnight culture and further incubated at 32 °C under vigorous shaking until an OD_{595} of 0.6 was reached. Subsequently 10 ml of the culture were transferred to a new Erlenmeyer flask and incubated for 15 min in a 42 °C water bath under shaking in order to express the heat sensitive integrated phage genes *exo*, *bet* and, *gam*, which are needed for mediating homologous recombination. The remaining culture was kept at 32 °C as not induced negative control. Then both flasks, induced and not induced cultures, were shortly incubated in iced water and subsequently the cultures were pelleted by centrifugation using 15 ml round bottom tubes (4,000 rpm, 5 min, 4 °C). The supernatants were aspirated and cell pellets were gently resuspended in 1 ml ice cold H_2O by swirling, then 9 ml H_2O were added and cells pelleted by centrifugation. This washing procedure was repeated once and then the supernatant was carefully and completely aspirated and the bacterial pellets were placed on ice. For the transformation 60 ng of the desired PCR product, consisting of the *galK* gene flanked by each 50 bp homologous to the targeted DNA sequence, were submitted to a BioRad cuvette (0.1 cm). Also 100 ng of pgalK plasmid DNA were used as positive control for transformation efficiency. Subsequently, 25 µl of the respective bacteria were added to the DNA and the mixture was electroporated applying 25 µF, 1.75 kV, and 200 Ω using a BioRad Gene Pulser® II device. After electroporation immediately 1 ml of LB medium was added and the reaction was transferred to a new 15 ml round bottom tube and incubated for 1 h at 32 °C under shaking. Then the cells were transferred to a 1.5 ml reaction tube, pelleted by centrifugation (13,200 rpm, 15 s, RT) and resuspended in 1 ml M9 minimal salts (M9) solution. This washing procedure was repeated twice and finally the pellet was resuspended in 1 ml M9 solution and 1:10 as well as 1:100 dilutions in M9 solution were made and 100 µl of each dilution was spread on M63 minimal salts (M63) agar plates containing galactose as the only carbon source. Only bacteria which successfully integrated *galK* are now able to process galactose and to grow on these minimal plates (Gal+). The plates were incubated at 32 °C for approx. 5 days until single colonies were growing out. Colonies were further checked for successful *galK* integration by replica plating on fresh M63 plates and MacConkey base agar plates serving as Gal+ indicator plates (bright red-pink colonies). For each transformation reaction at least 10 colonies were picked and overnight cultures in 5 ml LB medium containing chloramphenicol were made. DNA was recovered for integrity check by restriction enzyme digestion diagnostic PCR as described in 3.2.6.1.

Step II – Exchange of *galK* and *galK* negative selection

Bacterial clones which were identified for the correct integration of *galK* at the destination of interest were now used for the second step, where *galK* is exchanged by the sequence of interest,

here coding for a triple Flag-tag. To this end selected clones were induced to express lambda encoded, recombination mediating genes and made competent for transformation as described above. Also the transformation was performed as described above, this time employing 200 ng of the desired constructs. Here the coding sequence for the Flag-tag including 150 bp homology arms for the integration sites of interest was amplified from plasmids containing this artificially synthesized DNA fragment (Table 6) using primers described in Table 9. Both constructs, using 100 or 150 bp homology arms resulted in successful recombination, while usage of a construct with 50 bp homology arms, as described in the original publication, was not successful. This time after transfection the bacteria were incubated for 4.5 h at 32 °C under vigorous shaking and subsequently resuspended and wash with M9 solution as described above. Finally, bacteria were spread on M63 agar plates containing 2-deoxy-galactose (2-DOG) and glycerol as carbon sources. Bacteria still harboring *galK* metabolize 2-DOG to a toxic product thereby *galK* can now be used as negative selection marker. Again, at least 10 single colonies were picked and checked correct insertion of the Flag construct by diagnostic PCR, and also subjected to restriction enzyme digestion.

M9 minimal salts solution	1x solution (11.3 g 5x salts diluted in 1 l H_2O)
M63 minimal salts agar + galactose	1x M63 salts, 1.5% agar, added after cool down: 1mM $MgSo_4 \times 7H_2O$, 2.5 mg Biotin (sterile filtrated), 45 mg Leucine (sterile filtrated), 20 µg/ml Chloramphenicol, 0.2% Galactose
MacConkey base agar	40 g/l MacConkey agar base, 1.5% agar, 20 µg/ml Chloramphenicol, 0.2% Galactose
M63 minimal salts agar + 2-DOG	as M63 recipe, but 0.2% 2-DOG and 0.2% Glycerol instead of Galactose

3.2.5 Plasmid recovery from bacterial cultures

To recover plasmid DNA transformed and amplified in *E. coli* DH5α the NucleoSpin Plasmid kit was applied according to the manufacturer's instructions.

3.2.6 BACmid recovery from bacterial cultures

3.2.6.1 Small scale preparation for integrity check

For a fast recovery of BACmids from *E. coli* SW105 in order to check for integrity and success of recombination, simultaneously checking several clones, 5 ml overnight cultures (which were also frozen at -80 °C for long time storage until integrity check) were streaked on half a LB agar plate containing chloramphenicol and incubated at 32 °C overnight. The next day, a small area of confluent bacteria of approx. 1.5 cm^2 was scratched of the plate with the tip of a microliter pipet and resuspended in 200 µl Binding Buffer. Then 200 µl Lysis Buffer were added and the reaction

was mixed and bacteria were lysed by carefully inverting the tube 6-8 times and incubation for 5 min on ice. The reaction was neutralized by adding 200 µl Neutralization Buffer to the reaction and carefully mixed by inverting the tube 6-8 times. The lysate was cleared by centrifugation (16,000 g, 10 min, 4 °C) and the supernatant was transferred to a new reaction tube. This step was repeated once to fully clear the lysate. Then 400 µl isopropanol were added and the reaction was mixed by inverting. The BACmid DNA was precipitated by centrifugation (16,000 g, 10 min, RT) and the DNA pellet was washed once with 80% ethanol (16,000 g, 10 min, RT). Finally the supernatant was aspirated completely and the pellet was shortly air dried before it was dissolved in 20 µl TE Buffer. The complete preparation was used for one diagnostic restriction digest reaction.

Binding Buffer	50 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 100 µg/ml RNase A
Lysis Buffer	200 mM NaOH, 1% SDS
Neutralization Buffer	3.1 M Potassium acetate (CH3CO2K), pH5.5
TE Buffer	10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0

3.2.6.2 High purity large scale BACmid preparation for transfection

In order to obtain high quantities of transfection grade supercoiled BACmid, large volumes of bacterial culture were harvested and recovered BACmid DNA was subsequently subjected to CsCl density gradient centrifugation. First 50 ml LB medium containing chloramphenicol were inoculated with one colony harboring the desired BACmid and incubated at 32 °C under vigorous shaking overnight. The next day, 6x 400 ml LB medium plus chloramphenicol and 24 ml of 5M NaCl were inoculated with each 1 ml of the overnight culture and again incubated at 32 °C overnight under vigorous shaking. Cells were harvested by centrifugation (4,600 g, 15 min, 4 °C), the supernatant was discarded and the bacterial pellets were (at least) shortly frozen at -80 °C. Pellets were resuspended in each 10 ml Solution I by pipetting, transferred to 200 ml conical tubes, and filled up to 45 ml final volume. From now on tubes were constantly kept on ice. For each tube 10 mg Lysozyme were added, mixed by gently inverting the tube and incubated on ice for 10 min. Cells were lysed by addition of 58 ml freshly prepared Solution II and subsequent mixing by gentle 5-6 times inversion of the tube. Then the tubes were incubated for 5 min on ice. The reaction was neutralized by addition of 70 ml Solution III and mixed by gentle inversions of the tube until the lysate was cleared and kept of at least 30 min on ice. Then the non-soluble fraction was pelleted by centrifugation (4,600 rpm, 45 min, 4 °C). The supernatants were cleared by filtration trough filter paper and transferred and distributed into new conical 200 ml tubes with a maximal volume of 130 ml. BACmid DNA was precipitated by addition of 0.75x volume of isopropanol, mixed by inversions, and incubated for 30 min at RT. The DNA

was pelleted by centrifugation (4,600 rpm, 60 min, RT), the supernatant was discarded, and each pellet washed with 50 ml 80% ethanol. After centrifugation for 20 min, the supernatant discarded, the pellets were air dried and resuspended in a total of 40 ml TE Buffer. DNA pellets were not pipetted but dissolved by gentle rocking overnight. After complete resuspension of the DNA, all tubes were combined and 400 µg RNase A was added and incubated at 37 °C for 15 min. Then 6 mg Proteinase K were added and the reaction incubated at 50 °C for 45 min. The reaction volume was split into two 50 ml tubes, the net weight was determined and the same quantity of CsCl salt in g (+ 1 g to compensate for the 1 ml ethidium bromide (EtBr) to be added) was slowly added to each tube in several portions and warmed to 50 °C in between for better dissolving. When the CsCl was completely dissolved, 1 ml EtBr was added to each tube. Each reaction was transferred in one 35 ml Sorvall ultracentrifugation tube (#03989) and the tubes were completely filled with 1.55 g/ml CsCl solution, carefully balanced against each other, and sealed. The tubes were subjected to ultracentrifugation at 38,000 rpm, for three days at RT without applying a brake to stop the rotor. DNA was shortly visualized under UV light (312 nm) and the lower DNA band (supercoiled DNA) was slowly and carefully aspirated using a 14 gauge needle and syringe. Prior to aspiration of the DNA a second needle was carefully put through the upper part of the tube to ensure pressure compensation upon volume reduction in the tube. The DNA containing solutions were combined, filled up to 11.5 ml with 1.55 g/ml CsCl solution, and again subjected to density gradient centrifugation at 38,000 rpm, for three days at RT, without brake applied. Supercoiled DNA was aspirated as described above. Subsequently EtBr was completely removed by Isobutanol solvent extraction and CsCl was removed by dialysis with 2x 2 l TE Buffer. Recovered BACmid DNA was checked for integrity by restriction enzyme digest, diagnostic PCR, and was subjected to sequencing for critical regions e.g. the inserted regions.

Solution I	50 mM Glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA, pH 8.0, 100 µg/ml RNase A
Solution II	200 mM NaOH, 0.4% SDS
Solution III	3 M Potassium acetate (CH3CO2K), pH5.5

3.3 RNA related techniques

All RNA involving assays and experiments required the usage of filter tips, RNase free reaction tubes and reagents, and were conducted on ice.

3.3.1 Isolation of RNA from mammalian cells

RNA was isolated from 5 x 10^6 to 10^7 cells using RNeasy mini extraction kit (Qiagen) according to the manufacturers' instructions. For lysis β-Mercaptoethanol (β-ME) was added to buffer RLT (134 mM final) and cells were lysed applying 700 µl RLT+ β-ME. For complete disruption of cells QIAshredder columns were used according to the manufacturers' instructions. Additionally, RNase-free DNase set (Qiagen) was applied according to the manufacturers' instructions to eliminate residual genomic DNA in the preparation. Finally, RNA was eluted in 50 µl H_2O and RNA concentration and purity was determined using a nanodrop device.

3.3.2 RNA agarose gel electrophoresis

RNA preparations were subjected to denaturing RNA agarose gel electrophoresis to assess RNA quality. High RNA quality is associated with sharp 28 and 18S rRNA bands visible under UV light, while degraded RNA produces rather smeared bands. To this end 1.2% agarose was melted in autoclaved water and after cooling to approx. 60 °C, formaldehyde and MOPS were added to a final concentration of 2.2 M and 1 x, respectively. For each sample 5 µg RNA were subjected to electrophoresis. The reaction was prepared on ice, where 2 µl 5 x MOPs, 3.5 µl 37% formaldehyde, 10 µl deionizing 100% formamide, and 0.08 µl EtBr were added to each sample, mixed by pipetting and denatured at 56 °C for 15 min. Subsequently samples were shortly incubated on ice, 2 µl of RNA sample Buffer were added and samples were subjected to electrophoresis. 1 x MOPS was applied as running buffer and electrophoresis was conducted at 1-2 V/cm.

10 x MOPS	0.4 M MOPS, pH 7.0, 0.1 M Na-Acetate, 0.01 M EDTA, pH 8.0
RNA sample Buffer	50% Glycerol, 1 mM EDTA, pH 8.0, 0.4% Bromophenol blue

3.3.3 Reverse transcription of RNA

RNA was reverse transcribed (RT) to obtain cDNA as target for PCR analyses (RT-PCR and RT-qPCR). To this end the High-Capacity cDNA Reverse Transcription Kit was applied according to the manufacturers' instructions using 2 µg RNA as input. As a standard procedure for each analyzed sample, two reactions were set up, one containing reverse transcriptase and the second without enzyme, serving as negative control for genomic DNA contamination. For qPCR analysis, 1/80 or 1/40 of the cDNA reaction (corresponding to 25 or 50 ng input RNA, respectively) was used as template for quantification of GAPDH and the other analyzed transcripts, respectively.

3.4 DNA related techniques

3.4.1 Preparation of genomic DNA from mammalian cells

For the preparation of complete genomic DNA QIAamp DNA Mini Kit was applied according to the manufacturers' instructions using 5 x 10^6 cells as input material. Finally, DNA was eluted in 100 µl H_2O and concentration and purity was determined using a nanodrop device.

3.4.2 Restriction enzyme digestion of DNA

Plasmid or BACmid DNA was controlled for integrity and cloning success by restriction enzyme digestion according to the manufacturers' instructions of the respective enzyme. To this end 0.5-1 µg purified DNA or the complete BACmid mini preparation was used as template.

3.4.3 DNA Gel electrophoresis

DNA fragments were separated on agarose gels with the appropriate agarose concentrations (0.8-1.5%) which contained 0.01% (v/v) EtBr and 1 x TAE Buffer was used as gel and running buffer. The DNA samples were mixed with 1/6 final volume of DNA Loading Buffer and electrophoresis was conducted applying 5-8 V/cm. Finally, gels were analyzed under UV light and captured for documentation.

TAE	40 mM Tris-Acetate, 1 mM EDTA, pH 8.0
6x DNA Loading Buffer	15% Glycerol, 0.25% Bromophenol blue, 0.25% Xylene cyanol

3.4.4 Sequencing of DNA

Sequencing of DNA was conducted at MWG Operon, Ebersberg, Germany and analyzed using the Chromas Lite software.

3.4.5 Conventional PCR

Diagnostic PCR analysis for confirming Flag insertions was conducted applying peqGold Taq Polymerase all-inclusive kit according to the manufacturers' instructions using 100 ng DNA as template.

3.4.6 Quantification of cDNA and DNA by quantitative PCR (qPCR)

cDNA obtained from reverse transcribed RNA and DNA recovered from chromatin-immunoprecipitation (ChIP) experiments was quantified using a Roche LightCycler 480 II

instrument and LightCycler 480 SYBR Green I Master (Roche) reagent according to the manufacturers' instructions. In particular qPCR was performed using 96 well plates in 10 µl reaction volume. A mastermix consisting of 5 µl LightCycler 480 SYBR I Green, 1 µl 5 µM Forward Primer, 1 µl 5 µM Reverse Primer, and 1 µl H_2O per well was prepared and pipetted in the wells and 2 µl sample was added last in the appropriate wells and mixed by pipetting. Sample volume did not exceed 2 µl and was adjusted to 2 µl in cases of lower volume. Cycle conditions are listed in Table 15.

Table 15. Cycle conditions for qPCR at the LightCycler 480 II device

Analysis Mode	Cycles	Segment	Temperature (°C)	Ramp Rate (°C/s)	Time*	Acquisition Mode
None	1	Pre-Incubation	95	4.4	10 min	None
Quantification	45	Denaturation	95	4.4	3 s	None
		Annealing	60-63	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	1.5	10 s	None

* time hold after reaching the indicated temperature

Standard curves using dilutions defined amounts of the respective PCR product as templates were made to account for differences in primer efficiencies. To this end at least two "standard" samples of defined PCR product particles per primer pair were applied for each run and used for absolute particle quantification and normalization between runs. For analysis of ChIP samples this absolute quantification analysis was performed and % input was calculated as described in chapter 3.5.4. Relative expression levels were calculated by normalization of transcripts of interest to GAPDH applying the relative quantification mode of the LightCycler 480 software (based on the ΔC_t method but using the measured primer efficiency instead of 2 as default).

3.4.7 Library preparation for deep-sequencing of ChIP associated DNA fragments

DNA fragments recovered from chromatin immunoprecipitation (ChIP) experiments were subjected to next generation sequencing to gain genome wide information on TF binding sites. To this end recovered DNA, and also an input sample as negative control, were quantified with a Qubit® dsDNA HS (high sensitivity) Assay Kit using a Qubit® Fluorometer (Invitrogen). Samples were further processed in Dr. Blums laboratory at the Gene Center of the LMU Munich. A maximum of 100 ng ChIP as well as the same amount of input DNA were subjected to library preparation using NEBNext Ultra DNA Library Prep Kit (New England Biolabs) according to

the manufacturers' instructions. Up to eight samples per lane were sequenced separated by different barcodes. Sequencing was conducted using an Illumina HiSeq 1500 device producing 50 bp single-end reads.

3.5 Protein biochemistry related techniques

3.5.1 Generation of whole mammalian cell lysates

For the generation of whole cell lysates 10^7 cells were harvested by centrifugation, washed once with PBS, and resuspended and lysed in 100-200 µl NP-40 Lysis Buffer. The reaction was incubated for 1 h one ice and subsequently sonicated 3 x for 10 s applying 10% amplitude (3 mm conical microtip, Branson Sonifier). Cell debris was pelleted by centrifugation (20,000 g, 15 min, 4 °C),the supernatant was transferred to a new 1.5 ml reaction tube and stored at -80 °C. The protein content of lysates was quantified by Bradford method using a defined serial dilution (1-10 µg) of BSA as reference. To this end 5x Bradford Solution was diluted 1:5 with H_2O just prior to usage and 1-2 µl of the lysates were added and mixed by inversion of the cuvette. Adsorption of the mixtures was measured at 595 nm using a spectral photometer and applying 1x Bradford Solution without protein as blank value.

NP-40 Lysis Buffer	50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% NP-40, 1x Proteinase Inhibitor Cocktail (Roche)
5x Bradford Solution	100 mg Coomassie Brilliant Blue G-250, 47% Methanol, 42.5% Phosphoric acid

3.5.2 SDS Polyacrylamide gel electrophoresis

To separate proteins by electrophoresis reducing SDS containing polyacrylamide gels were applied. Separation gels contained 8, 10, or 15% and stacking gels 5% polyacrylamide, respectively, using a 30% (w/v) acrylamide (Sambrook and Gething, 1989). The appropriate amount of whole cell protein lysate (1.5-30 µg, depending on protein of interest) was mixed with 2x or 5x Lämmli Buffer, boiled at 95 °C for 5 min, and loaded on a gel next to a protein molecular weight standard (Prestained Protein Ladder, MBI Fermentas, Germany). Separation was conducted applying 25 mA per gel for approx. 1 h.

2x Lämmli Buffer	4% SDS, 20% Glycerin, 5% β-Mercaptoethanol, 120 mM Tris-HCl, pH6.8, 1 spatula tip Bromophenol blue
5x Bradford Solution	10% SDS, 50% Glycerin, 12.5% β-Mercaptoethanol, 300 mM Tris-HCl, pH6.8, 1 spatula tip Bromophenol blue
Running Buffer	25 mM Tris Base, 0.2 M Glycine, 0.1% SDS

3.5.3 Western Blot

SDS-PAGE separated proteins were transferred to PVDF membranes for specific protein detection by antibodies. First, membranes were activated by incubation in 100% methanol for 5 min and subsequently equilibrated, together with Whatman blotting (WB) paper, and sponges, in Transfer Buffer. The blotting sandwich was set up, starting on the cathode side, as follows: One sponge, two layers WB paper, the running gel, PVDF membrane, two layers WB paper, and another sponge. The transfer was conducted at 400 mA for 1 h. Membranes were rinsed with PBS and incubated for 30-60 min at 4 °C under rolling in Blocking Buffer for protein saturation. Then membranes were incubated with primary antibodies diluted in Blocking Buffer for 1 h at RT or overnight at 4 °C. After several washing steps with PBS/Tween and a final wash step with PBS, membranes were incubated with the appropriate horseradish peroxidase (HRP) coupled secondary antibodies, specific for the used primary antibody, diluted in Blocking Buffer. Again, membranes were washed several times with PBS/Tween and once with PBS before bound antibodies were detected using an Enhanced Chemiluminescence (ECL) system (GE Healthcare) according to the manufacturers' instructions. The emitted light, resulting from the HRP mediated oxidation of luminol, was detected by applying Hyperfilm ECL films (Amersham, GE Healthcare).

Transfer Buffer	25 mM Tris Base, 192 mM Glycine, 0.1% SDS, 20% Methanol
Blocking Buffer	50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5% non-fat milkpowder
PBS/Tween	PBS plus 0.05% Tween

3.5.4 Chromatin Immunoprecipitation (ChIP)

3.5.4.1 ChIP in LCLs

The basis of this approach forms a protocol commonly used in our laboratory (Ciccone et al., 2004) with minor modifications as indicated below. In brief, 2 x 10^7 cells were harvested and washed twice in ice cold PBS, resuspended in 20 ml RPMI 1640 and cross-linked first with disuccinimidyl glutarate (DSG, 2 mM final) for 23 min at RT and then formaldehyde (1% final) was added for additional 7 min. The reaction was stopped by addition of glycine (125 mM final) and gentle shaking for 5 min at RT. Cells were pelleted and washed twice in ice cold PBS. Nuclei were isolated by washing the cells 3x with 10 ml of ice cold Lysis Buffer and subsequent centrifugation (300 g for 10 min at 4 °C). Nuclei were resuspended in 1 ml Sonication Buffer I and incubated on ice for 10 min. Chromatin was sheared to an average size of 300 bp by four (qPCR analysis as final readout) or five (samples designated for deep sequencing, resulting in fragments with average 200 bp in size) rounds of sonication for 10 min (30 sec pulse, 30 sec

pause) using a Bioruptor® device (Biogenode). Cell debris was separated by centrifugation at maximum speed for 10 min at 4 °C and chromatin containing supernatants were stored at -80 °C or directly used for IP. For preparation of input DNA 25 µl aliquots (1/10 of the amount used per IP) were saved at -80 °C. For IPs 250 µl chromatin (equals 5 x 10^6 cells) were diluted 1:4 with IP Dilution Buffer I and incubated with 5 µg of antibody or 100 µl of hybridoma supernatant on a rotating platform at 4 °C overnight. Antibodies used for ChIP are listed in Table 13. Protein G sepharose (GE Healthcare) was equilibrated with IP Dilution Buffer I, added to the lysate and incubated at 4 °C for 4 h with constant rotation. Beads were extensively washed with 2x Wash Buffer I, 1x Wash Buffer II, and 1x Wash Buffer III for 5 min under rotation. Then washed 2x with TE for 1 min. Protein-DNA complexes were eluted with 2x 150 µl Elution Buffer at 65 °C for 15 min. Input samples were adjusted to 300 µl with Elution Buffer. Eluates and input samples were incubated with Proteinase K (1.5 µg/µl final, Roche) for 1 h at 42 °C. Cross-linking was reversed by incubation at 65 °C overnight. DNA was recovered using QIAquick PCR purification kit (Qiagen) according to the manufacturers' instructions. For sequencing purposes four ChIP samples for the same protein of interest were pooled using one QIAquick column.

The DNA amount in input samples and after IP with specific antibody or an unspecific isotype-matched IgG control was quantified by qPCR using primers listed in Table 12. To account for differences in amplification efficiencies a standard curve was generated for each primer pair using serial dilutions of fragmented DNA (input) as template. DNA quantities detected in input samples were adjusted to the amount of chromatin used per IP by multiplication with 10. Enrichment was indicated as percentage of input and calculated as (DNA from specific IP corrected for IgG control background/ DNA input) x 100.

DSG	Pierce #20593, using freshly prepared 0.5 M stock solution in DMSO
Lysis Buffer	10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 0.5% NP-40, 1x proteinase inhibitor cocktail (PIC, Roche)
Sonication Buffer I[1]	50 mM Tris-HCl, pH 8.0, 5 mM EDTA, pH 8.0, 0.5% SDS, 0.5% Triton X-100, 0.05% sodium deoxycholate, 1x PIC
Dilution Buffer I	12.5 mM Tri-HCl, pH 8.0, 187.5 mM NaCl, 1.25 mM EDTA, pH 8.0, 1.125% Triton X-100, 1 x PIC
Wash Buffer I	20 mM Tris-HCl, pH 8.0, 2 mM EDTA, pH 8.0, 1% Triton X-100, 150 mM NaCl, 0.1% SDS, 1 x PIC
Wash Buffer II	20 mM Tris-HCl, pH 8.0, 2 mM EDTA, pH 8.0, 1% Triton X-100, 500 mM NaCl, 0.1% SDS, 1 x PIC
Wash Buffer III	10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1 x PIC
Elution Buffer	25 mM Tris-HCl, pH 7.5, 10 mM EDTA, pH 8.0, 0.5% SDS

[1] Adopted composition from an commercially available buffer (SDS lysis buffer, upstate, EZ ChIP protocol, catalog #17-371) and the sonication buffer used by Kouskouti et al. (2004).

3.5.4.2 ChIP in DG75 cell line

The ChIP protocol for DG75 cells is based on the ChIP protocol for LCLs described above with minor modifications. Cross-linking of cells was achieved by only using formaldehyde (1% final, 7 min incubation). Nuclei were resuspended in 1 ml Sonication Buffer II and incubated on ice for 10 min. Chromatin was sheared to an average size of 200-300 bp by four rounds of sonication for 10 min (30 sec pulse, 30 sec pause) using a Bioruptor® device (Biogenode) for all down-stream applications. Here, Dilution Buffer II was used instead of Dilution Buffer I.

Sonication Buffer II	50 mM Tris-HCl, pH 8.0, 10 mM EDTA, pH 8.0, 0.5% SDS, 1x PIC
Dilution Buffer II	12.5 mM Tris-HCl, pH 8.0, 1.25% Triton X-100, 212.5 mM NaCl, 1x PIC

3.6 Bioinformatic methods

All bioinformatic analyses steps were conducted independently, using the Galaxy platform hosted at the Bioinformatics Department of the University of Freiburg, if not indicated otherwise. Generated workflows were downloaded from the Galaxy server for documentation and are accessible at the HMGU server accessible via the following link:
https://hmgubox.helmholtz-muenchen.de:8001/d/2dcf3ec670/
The directories of the single files are described in the following sections.

3.6.1 Peak calling and generation of normalized ChIP-seq signals

TF ChIP-seq data

The main procedure of processing ChIP-seq data and purpose of each step is explained in chapter 4.2.1.2 (summarized in Fig. 13). The initial step of demultiplexing the obtained data according to applied barcodes was conducted at the Galaxy of the sequencing facility of the Blum laboratory using an in-house script. The subsequent quality control by FastQC (Andrews, 2010), read trimming applying TrimGalore (Krueger, 2012), mapping to hg19 applying Bowtie2 (Langmead and Salzberg, 2012), peak calling applying MAC2 (Zhang et al., 2008) and subsequent filter steps, and the generation of input normalized signal tracks applying bamCompare of the deepTools package (Ramirez et al., 2014a) are documented in the following workflow:

Dissertation_LG/Galaxy_Workflows/Galaxy-Workflow-peak_calling_and_signal_track_LG

For E2, E3A, and E3C ChIP-seq in LCL the analyzed data for called peaks and signals are also available via the HMGU server under the following directories:

Dissertation_LG/EBNA_ChIP-seq_LCLs/hg19_peaks
Dissertation_LG/EBNA_ChIP-seq_LCLs/hg19_input_norm_signals

For analysis of binding to the EBV genome, reads which did not map to hg19 were extracted and mapped to the EBV genome (HHV-4 type I, NC_007605.1). The workflow for peak calling and normalized signal generation can be found in the following workflow:

Dissertation_LG/Galaxy_Workflows/Galaxy-Workflow-peak_calling_and_signal_track_EBV_LG

Analyzed LCL derived data for EBV can be found at:

Dissertation_LG/EBNA_ChIP-seq_LCLs/HHV4_peaks
Dissertation_LG/EBNA_ChIP-seq_LCLs/HHV4_input_norm_signals

The E2 ChIP performed in DG75 was analyzed applying the same workflows and analyzed data can be found at:

Dissertation_LG/E2_ChIP-seq_DG75/E2_DG75_peaks
Dissertation_LG/E2_ChIP-seq_DG75/E2_DG75_input_norm_signals

Histone modification ChIP-seq data

Histone modifications are known to produce rather broad peaks than sequence specific TF, therefore the workflow for peak calling had to be adapted and is available at:

Dissertation_LG/Galaxy_Workflows/Galaxy-Workflow-peak_calling_and_signal_histone_mods_LG

To this end, mapped reads (bam) as published by the ENCODE project (for discrete files see Table S1) or other studies (Table S2) were downloaded and the appropriate replicate ChIP-seq or input files were merged, independent of the absolute number of bam files (between 1-5), prior to peak calling.

DNase-seq data

DNaseI hypersensitive sites (HS) in LCL and DG75 were analyzed as well (file list, see Table S2) applying a separate workflow due to the absence of input sample, which can be found at:

Dissertation_LG/Galaxy_Workflows/Galaxy-Workflow-peak_calling_and_signal_DNaseI_HS_LG

3.6.2 Generation of anchor plots for comparison of signals at different peak sets

To compare a distinct normalized signal between two sets of peaks, anchor plots were generated for visualization and the underlying data was used for statistical analyses. To this end computeMatrix and profiler tools of the deepTools package (Ramirez et al., 2014a) were applied. Depending on the question 2 (TF peaks), 5 (DNaseI HS), 10 or 20 kb (most histone modification analyses) in each direction of the peak center were analyzed. An example workflow for three peak sets and 2 kb extension from the peak center in both directions is available for download:

Dissertation_LG/Galaxy_Workflows/Galaxy-Workflow-Achor_plots_LG

3.6.3 Generation of heatmaps for comparison of different signals at the same peak set

In order to visualize different signals at the same peak set heatmaps were generated applying computeMatrix and heatmapper tools of the deepTools package (Ramirez et al., 2014a). To this end, as for the anchor plot generation, distinct regions of 2 kb in each direction from the peak centers were analyzed. In some cases (Fig. 20) the analyzed peak set was sorted by the mean signal of each peak in a descending manner, and subsequently this sorted peak list was used as reference for other ChIP-seq signals without changing the order of the peaks this time. An example workflow was deposited at:

Dissertation_LG/Galaxy_Workflows/Galaxy-Workflow-Heatmaps_LG

3.6.4 Correlation analyses

For correlation analyses and matrix generation bamCorrelate of the deepTools package was applied (Ramirez et al., 2014a) using Spearman correlation method. An example workflow with only three input data sets to compare was stored at the HMGU data deposit for demonstration of parameters:

Dissertation_LG/Galaxy_Workflows/Galaxy-Workflow-bamCorrelate_LG

In order to investigate and correlate two signals at one distinct peak set (as in Fig. 23), a different approach was chosen where for each peak the mean signal was calculated by computeMatrix and the resulting data was plotted applying an R script. The workflow and R script can be found at:

Dissertation_LG/Galaxy_Workflows/Galaxy-Workflow-Correlation_peak_set_LG
Dissertation_LG/Galaxy_Workflows/R_script_correlation_peak_set

3.6.5 Peak cluster analyses

To investigate the occurrence of preselected TFs at distinct peak sets cluster analyses were performed applying either Jaccard index or k means based clustering. Jaccard clustering was conducted by Björn Grüning (Universität Freiburg) while k means cluster was performed using Galaxy tool Numeric Clustering (Pedregosa et al., 2011).

Dissertation_LG/Galaxy_Workflows/Galaxy-Workflow-k_means_clustering_LG

Analyzed data of the E2 and E3 cluster can be found at:

Dissertation_LG/Cluster_Analyses/E2_cluster
Dissertation_LG/Cluster_Analyses/E3_cluster

4 Results

The results of this thesis are structured into three main parts:
In the first part (chapter 4.1), the generation of LCLs expressing epitope tagged E3 proteins is described and the cell lines are confirmed and characterized.

The establishment of the ChIP-assay for E3 proteins, the reliability of this assay, the bioinformatic analyses, the actual identification of EBNA binding sites in the viral and human genome, the further characterization of those sites in the human genome and pattern formation as well as the identification of associated TFs are depicted and explained in the second part (chapter 4.2).

In the last results part (chapter 4.3) the dependency of E2 chromatin accession on CBF1 is further examined in studies using knock-out cell lines and a potential novel adaptor is identified and characterized.

4.1 Introducing a new experimental system: LCLs infected with recombinant EBV encoding epitope tagged E3A or E3C protein

To investigate the chromatin binding properties of E3A and E3C proteins, a robust and reliable ChIP assay displays the basis for all further experiments and analyses. One crucial step of this assay, as for all immunological experiments, is the choice and specificity of the respective antibody. For the E3 proteins several commercially available antibodies, as well as such from Dr. Elisabeth Kremmer's laboratory at the HMGU, have been tested in the Kempkes' laboratory and failed to reach ChIP assay standards in enrichment and specificity (data not shown). Therefore an epitope tag was fused to either E3A or -3C ORF in the viral genome to gain LCLs which express the recombinant proteins at an endogenous level. Furthermore this epitope tag ensures a highly efficient immunoprecipitation.

The process of infection and immortalization of primary B cells by EBV and the subsequent outgrowth of LCLs is well described in literature (Delecluse et al., 1998, Delecluse et al., 2008, Feederle et al., 2010) and also established in our laboratory. The different steps of this procedure are depicted in Figure 5.

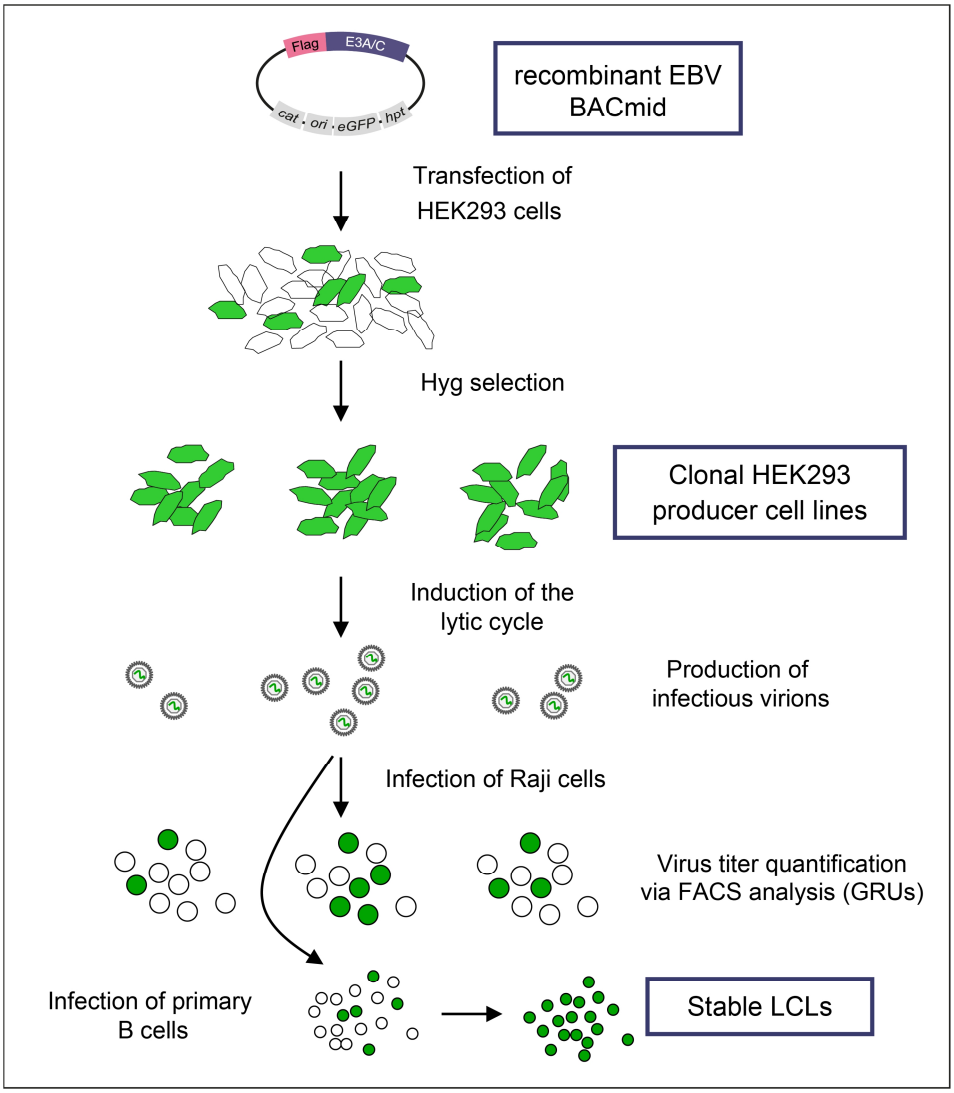


Figure 5. Generating LCLs by infection of primary B cells with recombinant EBV. Flow scheme depicting the multistep process to generate LCLs starting from recombinant EBV genomes.

Starting with transfection of the desired EBV genome (as described in 4.1.1) in HEK293 cells and subsequent hygromycin selection, EBV positive clones can be selected and further cultivated to become stable so called "EBV producer cell lines". The lytic cycle of EBV and consequently the production of infectious particles can be induced by the lytic switch protein BZLF1. Infection efficiency of virus containing supernatants can be measured by infecting Raji B cells and monitoring eGFP expression via FACS (Delecluse et al., 1998) (see 4.1.2). Viral titers can be calculated and potent supernatants can be used for infection of primary B cells derived from lymphoid tissue or as in this case cord blood which is free of endogenous virus.

The single steps and results of this work are described in the following chapters with the details for the experiments outlined in the methods section (chapter 3.2.4).

4.1.1 Generation and characterization of recombinant EBV genomes via "Recombineering"

The overall idea was to insert an epitope tag N-terminally and in frame fused to E3A or E3C respectively in the EBV genome using the EBV BACmid (p2089) (Delecluse et al., 1998) as wildtype reference. Originally in this publication, the EBV genome was extracted from B95.8 cell line and hygromycin phosphotransferase (*hpt*) resistance gene for selection in eukaryotic cells, chloramphenicol acetyltransferase (*cat*) resistance gene for selection in bacteria, enhanced green fluorescent protein (*eGFP*) as a reporter and an origin of replication (*ori*, derived from F-plasmid) for propagation in bacteria were inserted.

For the epitope a Flag-tag consisting of a triple repetition of the Flag peptide was chosen, since many ChIP assay compatible antibodies are commercially available specific for this construct.

Since the assay for cloning in the viral genome, which was already established in the laboratory, includes the permanent introduction of a reporter/selection gene (Cherepanov and Wackernagel, 1995), a different protocol was chosen: "Recombineering" makes it possible to insert or to delete sequences and also to introduce mutations in bacterial artificial chromosomes (BACs) without leaving any other sequences in the vector of interest (Warming et al., 2005). In brief, a special bacterial strain (here *E coli* SW105) deficient for the *galK* gene, which encodes galactokinase one of the enzymes crucial for metabolizing galactose, and the usage of different minimal media makes it possible to use *galK* in a two-step procedure as both, positive and negative selection marker. Additionally *E. coli* SW105 harbors an integrated defective lambda prophage whose expression is temperature sensitive and which is coding for three proteins (*exo*, *bet,* and *gam*) mediating homologous recombination. First a PCR construct consisting of *galK* flanked by two homologous arms of the region of interest is transformed into the bacteria and subsequently expression of the lambda genes and thereby recombination is induced (Fig. 6). Here two different constructs were used, ensuring the integration 3' of the start codon of E3A or E3C respectively. Using minimal media with galactose as the only carbon source, selection of clones with an integrated *galK* gene was ensured. This intermediate product was also checked via restriction digest and gel electrophoresis for genomic integrity (Fig. 6, plasmid no. 2). In a second step, a PCR product of the described Flag-tag flanked by the same homology arms as used before was transformed, lambda expression induced and the bacteria plated on minimal media containing 2-deoxy-galactose (2-DOG), allowing *galK* to be used for negative selection this time. Bacteria still harboring *galK* metabolize 2-DOG to a toxic product while *galK* deficient ones can use it as carbon source. The resulting products were two EBV BACs with a Flag-E3A or -E3C fusion gene.

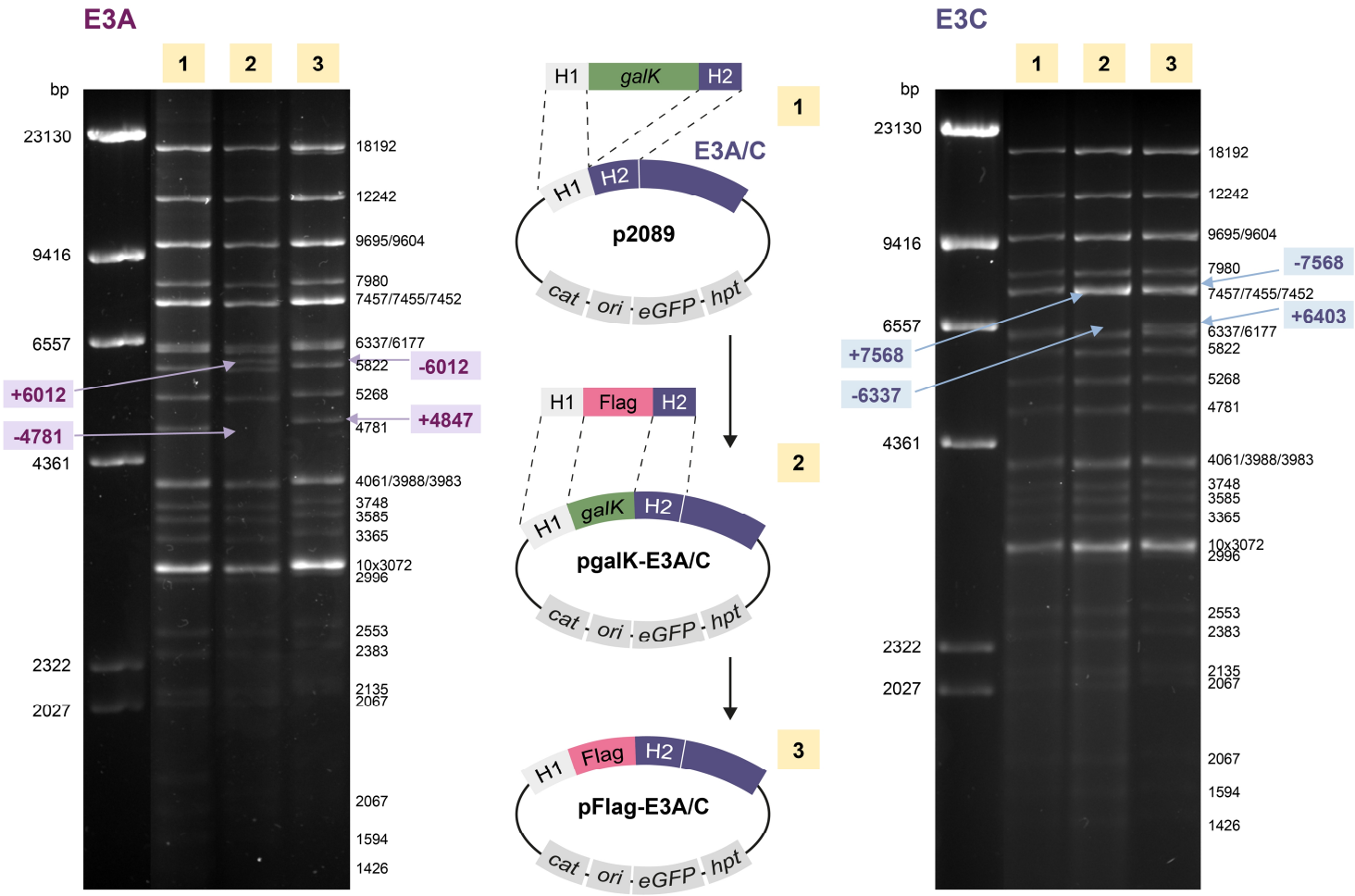


Figure 6. Generation of two recombinant EBV genome BACmids harboring Flag-E3A and -E3C fusion genes. Schematic overview of the two-step cloning procedure to generate N-terminal Flag-tag insertions for E3A and E3C in the EBV genome applying the recombineering protocol (middle panel). Wildtype EBV BACmid (1, p2089) which encodes chloramphenicol acetyltransferase (*cat*), the origin of replication from an F-plasmid (*ori*), *eGFP* as a reporter and hygromycin phosphotransferase (*hpt*) was used as starting point. After transformation of *galK* flanked by homologous regions H1 and H2 for E3A or E3C (E3A/C), induction of recombination and positive selection for *galK* the intermediate product (2, pgalK-E3A or -E3C) was obtained. In the second step a construct of the Flag-tag flanked by homologous regions was transformed in the bacteria, recombination was induced and after *galK* negative selection the final EBV BACmid with the respective desired fusion gene were produced (3, pFlag-E3A or -E3C). Gel electrophoretic separation of BglII restriction digest for all three plasmid steps as performed for E3A (left panel) and E3C targeting (right panel) is shown and expected fragment sizes are indicated. The arrows highlight restrictions fragments of particular interest, since they shift in size upon galK or Flag insertion.

The integrity of the recombinant EBV genomes was monitored by restriction digest and gel electrophoresis (Fig. 6, plasmid no. 3) and a diagnostic PCR was established to quickly check for correct Flag insertion (Fig. 7). This PCR assay was also used at various steps of this work to verify the identity of recombinant EBV genomes in different cell lines.

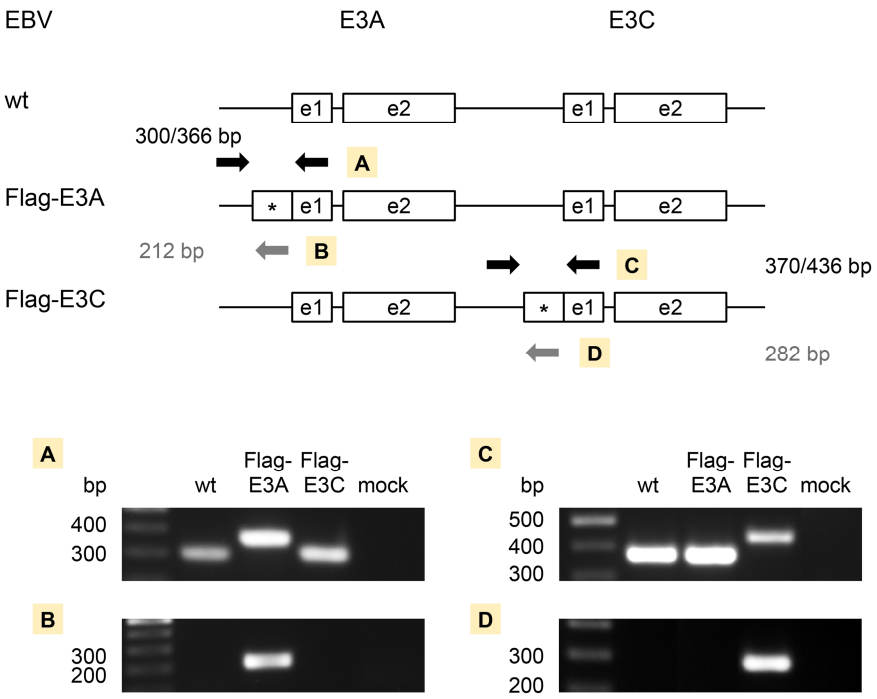


Figure 7. Diagnostic PCR confirming the correct insertion of Flag-tag 5' of E3A or E3C in the EBV BACmid genome. Schematic representation of the genomic background of the constructed recombinant EBV BACmid genomes (upper panel) and the positions and product sizes for four PCR primer combinations (A-D). The depiction of exons and primers is not in scale. In the lower panel gel electrophoretic separation and UV light detection of the obtained PCR products are shown.

Furthermore the critical regions where the insertion took place were sequenced (chapter 3.4.4) to further ensure genome accuracy. Finally correct recombinant EBV BACmids were used to generate cell lines.

4.1.2 Generation and characterization of HEK293 EBV producer cell lines

In a second step the novel recombinant EBV BACmids were transfected into HEK293 cells and outgrowth of EBV positive clones was ensured using hygromycin selection. Single clone colonies were selected and separately cultivated until stable cell lines were growing out. Those cell lines could now be used as a tool to generate infectious virions. Upon transient transfection of expression plasmids for EBV proteins BZLF1, the lytic switch protein which induces the lytic cycle (Countryman et al., 1987), and BALF4, the viral glycoprotein gp110 which enhances packaging and thereby infection efficiency (Neuhierl et al., 2002), viral particles were produced and released into the cell culture medium. To determine viral titers and infection efficiency of different producer cell line clones and supernatant batches, the EBV positive B cell line Raji is incubated with those supernatants and eGFP expression from recombinant EBV is monitored by FACS analysis.

For the purpose of this work several HEK293 EBV producer cell lines with the two different recombinant EBV genomes encoding Flag-E3A and -E3C respectively, could be generated and characterized to be efficient producers of infectious EBV particles. Also at this point the diagnostic PCR, as described in chapter 4.1.1, was used to determine the integrity of the EBV genome of the different clones (Fig. 8A).

Virus titer determination via FACS analysis allowed the identification of efficient EBV producer cell lines (Fig. 8B). Therefore different dilutions of the virus particle containing supernatants were incubated with Raji cells, the mean percentage of eGFP positive cells was calculated and the virus titer defined as *green raji units* (GRUs) was determined.

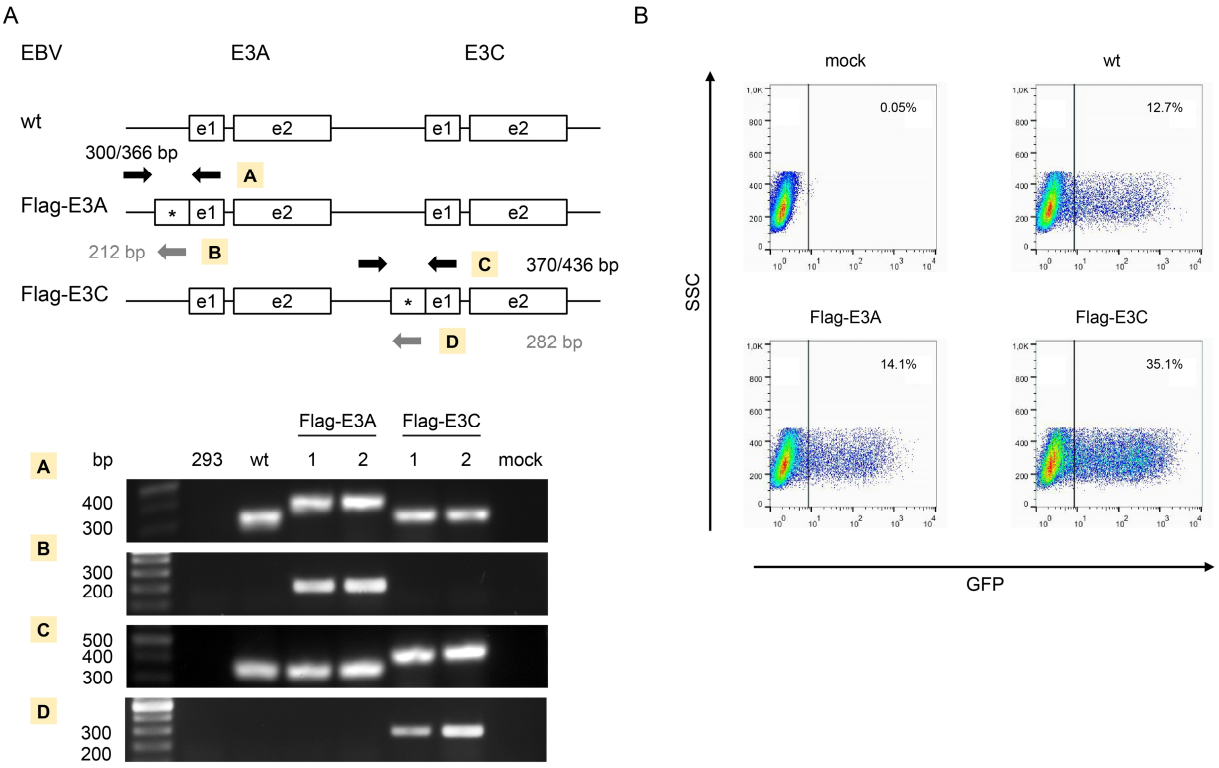


Figure 8. Stable HEK293 producer cell lines efficiently generate infectious recombinant EBV particles with Flag-E3A and E3C fusion genes. (A) Diagnostic PCR verifying the recombinant EBV genome of different established producer cell lines. Only after PCR confirmation cell lines were used for virus production. The parental HEK293 cell line is used as negative control (293) and p29089 is used for wt EBV reference. For each recombinant EBV genome two different HEK293 producer clones were analyzed. (B) FACS analysis showing the percentage of eGFP positive Raji cells after 24 h incubation with EBV producer cell line supernatants containing infectious particles. Here the readouts of one representative supernatant sample per recombinant EBV using 1:2 dilutions are shown. Virus titers (GRUs) were determined using the mean of 5 dilution steps and the most potent supernatants were used in further experiments.

4.1.3 Generation and characterization of LCLs expressing Flag-E3A or -E3C fusion proteins

The recombinant EBV particles could now be used to infect primary B cells derived from cord blood and stable LCLs expressing Flag-E3A and -E3C respectively were generated for three different donors. For each donor an LCL infected with wt EBV (p2089) was generated in parallel as phenotype reference, since the introduction of the Flag-tag should not impair the EBV variants in infection efficiency, the LCL in viability and proliferation or the respective E3 in target gene regulation and interaction behavior. When using B cells collected from tonsils which might be derived from an EBV positive donor, it is possible that some cells, infected with endogenous virus, spontaneously grow out and give rise to LCLs. Thus, as a negative control, primary B cells were infected with a recombinant EBV harboring an EBNA2 deletion, since EBNA2 is an important EBV encoded transactivator and essential for immortalization. This control is actually not necessary when using cord blood as B cells source, since EBV cannot pass through the placenta and infection shortly after birth is very unlikely. However, using EBNA2 deletion mutant as negative control for infection displays standard procedure in our laboratory and was applied as standard procedure in infection assays.

During infection experiments, at least 48 wells with 3 x 10^5 B cells were infected with wt, Flag-E3A, Flag-E3C, and E2 deletion EBV and no impairment of the Flag-E3 EBVs in comparison to wt could be observed (data not shown). Infection with EBNA2 deletion virus did not result in immortalization in all cases; the cells did not enter cell cycle and underwent apoptosis after approximately 2 weeks.

As expected, all of the established LCLs passed the diagnostic PCR analyses verifying the particular EBV genome (Fig. 9A/B). RT-qPCR analyses for E3 expression levels showed no significant differences between LCLs with different recombinant EBV genomes (Fig. 9C). Western Blot analyses (Fig. 9D) confirmed the expression of Flag-E3A and -E3C fusion genes in the respective cell lines and also showed no aberrant expression level of other EBV latent proteins due to Flag insertion. Thus the expression levels of the Flag fusion genes are comparable to the respective wildtype E3 protein. Also full-length proteins are expressed since the apparent protein sizes in SDS-PAGE are comparable to the respective wildtype E3, but show a small shift to a higher molecular weight due to the 22 aa Flag-tag. Shorter variants could not be observed by Western Blot analyses.

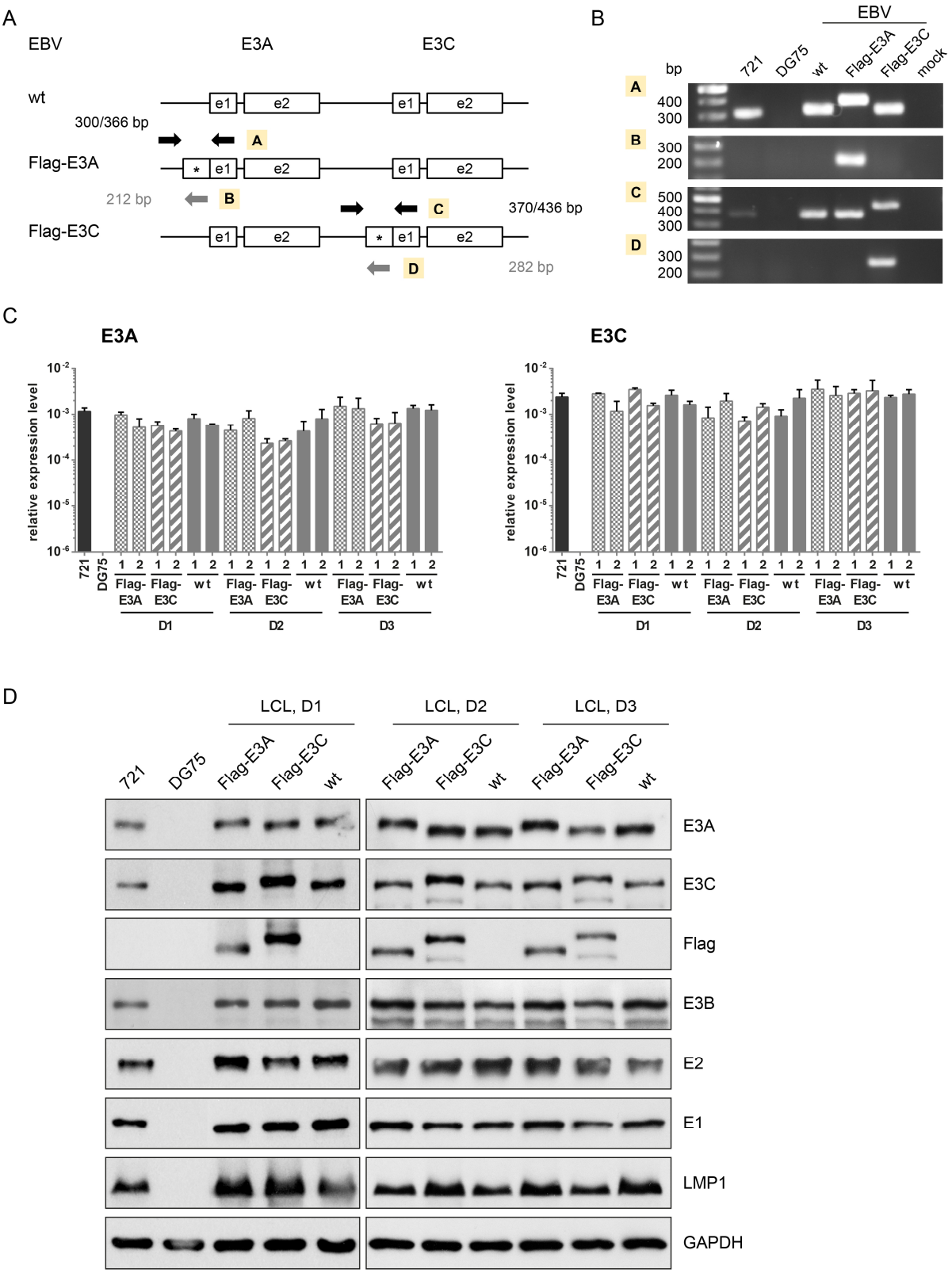


Figure 9. Established LCLs expressing Flag-tagged E3A or E3C show wildtype levels of EBV latent protein expression. Characterization of LCLs derived from three different donors. Schematic representation (A) and results (B) of diagnostic PCRs confirming the genomic background of recombinant EBVs used for infection. RT-qPCR (C) demonstrating E3 expression levels for LCLs with different EBV genome background. Cell lines derived from three different B cell donors (D1-D3) and for each donor two clones (1 and 2) were examined. Means and standard deviations (SD) from two independent experiments consisting of technical duplicates are shown. cDNA levels were normalized to GAPDH expression levels. Western Blot analysis (D) showing expression levels of six EBV latent proteins. An anti-Flag antibody was used to confirm the expression of recombinant E3 proteins. Equal amounts of

total protein lysates were loaded and GAPDH served as internal loading control. For both analyses the established wt LCL (721) was used as positive control while the EBV negative Burkitt's lymphoma cell line DG75 served as negative control.

Proliferation rates of the Flag-E3A and -E3C LCLs were comparable to wt LCLs. Also, viability was determined using MTT-assays and did not show disadvantage of Flag-E3 LCLs compared to wt LCL derived from the same donor as control (data not shown).

To assess the protein-protein interaction capabilities of the Flag-E3 LCLs the well described interaction with CBF1, a cellular DNA binding protein, was examined. CBF1 is known to physically interact with E2 (Henkel et al., 1994) as well as with E3A and E3C (Robertson et al., 1995, Robertson et al., 1996) and to mediate their recruitment to DNA (reviewed in (Kempkes and Ling, 2015, Allday et al., 2015). In immunoprecipitation (IP) experiments (Fig. 10) the interaction of both Flag-tagged E3 proteins with CBF1 could be verified and also a heterodimer of E3A and E3C could be detected. Interestingly this experiment revealed that E3B can form a heterodimer with E3A as well, but failed to interact with E3C. This finding was not reported before and might be an interesting interaction to be further studied since it could also have an impact on E3 functions. However, this interaction was not further analyzed in this work. The transcriptional repressor function of the Flag-E3 proteins was also reviewed with special attention to the introduced Flag-tag. To this end expression levels of three well described target genes were monitored. *BCL2L11* (or Bim), a proapoptotic tumor suppressor, is repressed by E3A and E3C in a cooperative fashion (Anderton et al., 2008) and *CXCL9* and *-10*, encoding chemokines, which are both repressed by E3A (Hertle et al., 2009) and E3C (our data, not published and for *CXCL10* (McClellan et al., 2012)). Expression levels for all three target genes did moderately fluctuate between samples, but no impact of the Flag-tag on gene expression could be observed (Fig. 11).

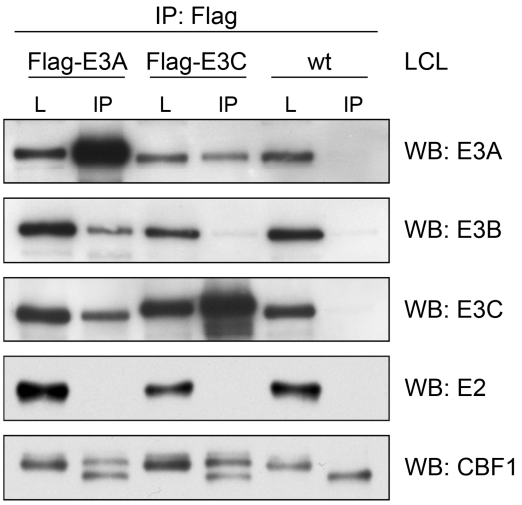


Figure 10. Flag-E3 proteins do interact with the DNA adaptor CBF1 in recombinant LCLs. IP analyses using a Flag specific antibody in Flag-E3A, -E3C and wt LCLs. Total cell lysates (L) display 5% of the cells used for IP samples. One representative experiment is shown (n=3).

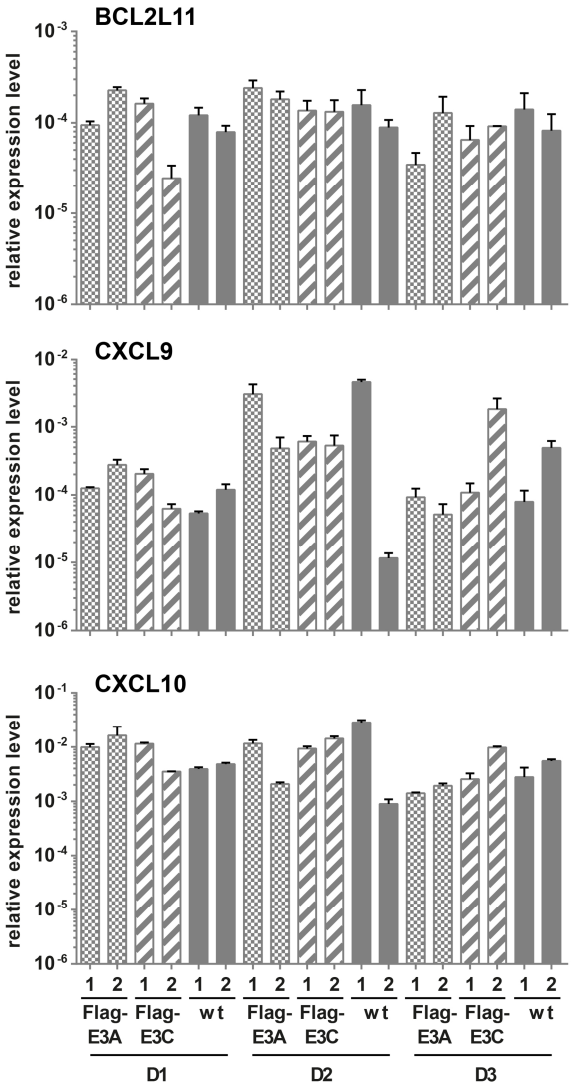


Figure 11. Flag-tag does not impair E3 target gene regulation. Expression levels of E3 target genes were quantified using RT-qPCR. Cell lines derived from three different B cell donors (D1-D3) and for each donor two clones (1 and 2) were examined. Shown are means and SD from two independent experiments consisting of technical duplicates. cDNA levels were normalized to GAPDH expression levels.

Therefore the Flag-E3 expressing LCLs developed and described in this work are an excellent tool to investigate chromatin binding mechanisms of the E3 proteins, which is the main focus of part II of this thesis. In addition these cell lines represent a versatile tool for different applications involving IP steps for the E3 proteins in a wt EBV and latency III expression background.

4.2 EBNA transcription factors – Exploiting enhancer elements

In the following chapter the process of how to get from ChIP-seq data to a better understanding of chromatin accession of EBNA proteins and eventually target gene regulation is pictured. It starts with the establishment of the necessary ChIP protocol, then the bioinformatic analysis pipeline is explained. Finally, upon comparison with data on TF binding, information on histone modifications and functional elements derived from the ENCODE project, conclusions on chromatin accession of EBNA proteins and contributing factors will be drawn.

4.2.1 Identification of E2, E3A, and E3C binding sites by ChIP-seq

This section of the thesis is divided into two different parts: One is the biochemical experimental part, which required several steps of adaption to the specific question as well as to the deep sequencing readout. The second part, which is very important as well, demands many control and optimization steps and careful evaluation of the results: the bioinformatic analyses of the obtained data. Since this part was performed independently, using the Galaxy Platform (Giardine et al., 2005) hosted and maintained by the Bioinformatics Department of the University of Freiburg, and very specific thresholds and parameters were used, a very detailed description is provided here in the results section. The details for each tool that was used and specific parameters are listed in the methods section (chapter 3.6).

4.2.1.1 Biochemistry

Optimization of the ChIP-assay – Cross-link

To elucidate subsets of EBNA binding sites and their characteristic TF occupancies a robust and reliable ChIP-assay had to be established. Here different protocols were combined to achieve the best results for the specific case of TFs which access DNA in an indirect manner. The basis of this approach forms a protocol commonly used in our laboratory (Ciccone et al., 2004) with some adaptions as described in the methods sections (chapter 3.5.4).

The most crucial adaption displays the optimization of the crosslinking process since formaldehyde only bridges a distance of 2 Å between two amino (or imido) groups, such as side chains of lysine and arginine by a covalent but reversible bond (Jackson, 1999). The E3 proteins but probably also E2 are expected to act in bigger protein complexes and to be recruited to DNA by cellular proteins. Therefore it is important for this specific assay to cover protein-protein as well as DNA-protein interactions.

It has been described that the usage of different cross-linking agents such as bifunctional imidoesters (Fujita and Wade, 2004) or N-hydroxysuccinimide (NHS)-esters (Nowak et al., 2005) could significantly improve the DNA recovery in ChIP assays, especially for proteins which form complexes on DNA. Those two different classes of reagents have a longer effective bridging distance between functional groups (approx. 8-16 Å) in common and are thereby thought to cross-link especially protein-protein interactions, which are not covered by formaldehyde only. Imidoesters show a higher reactivity for alkyl amines (as in lysine) than for aromatic amines (as in DNA bases), therefore favoring the cross-link of protein-protein over DNA-protein interactions (Fujita and Wade, 2004). Furthermore, both reagent classes are soluble in DMSO and can freely permeate cell membranes making them convenient for ChIP.

In this study the imidoester dimethyl 3, 30-dithiobispropionimidate (DTBP) and the two NHS-esters disuccinimidyl glutarate (DSG) and ethylene glycol bis(succinimidylsuccinate) (EGS) were tested for their ability to improve cross-linking efficiency. ChIP was performed for E2 and Flag-E3A in Flag-E3A LCLs and known E2 binding sites (Zhao et al., 2011b) in the well-studied E3A and E3C controlled region encompassing *CXCL9* and *-10* (Harth-Hertle et al., 2013) genes were analyzed (Fig. 12). In this publication from our laboratory the competitive binding of E2 and E3A for enhancers in this region could be shown. Therefore this locus displays an optimal example for ChIP assay improvement.

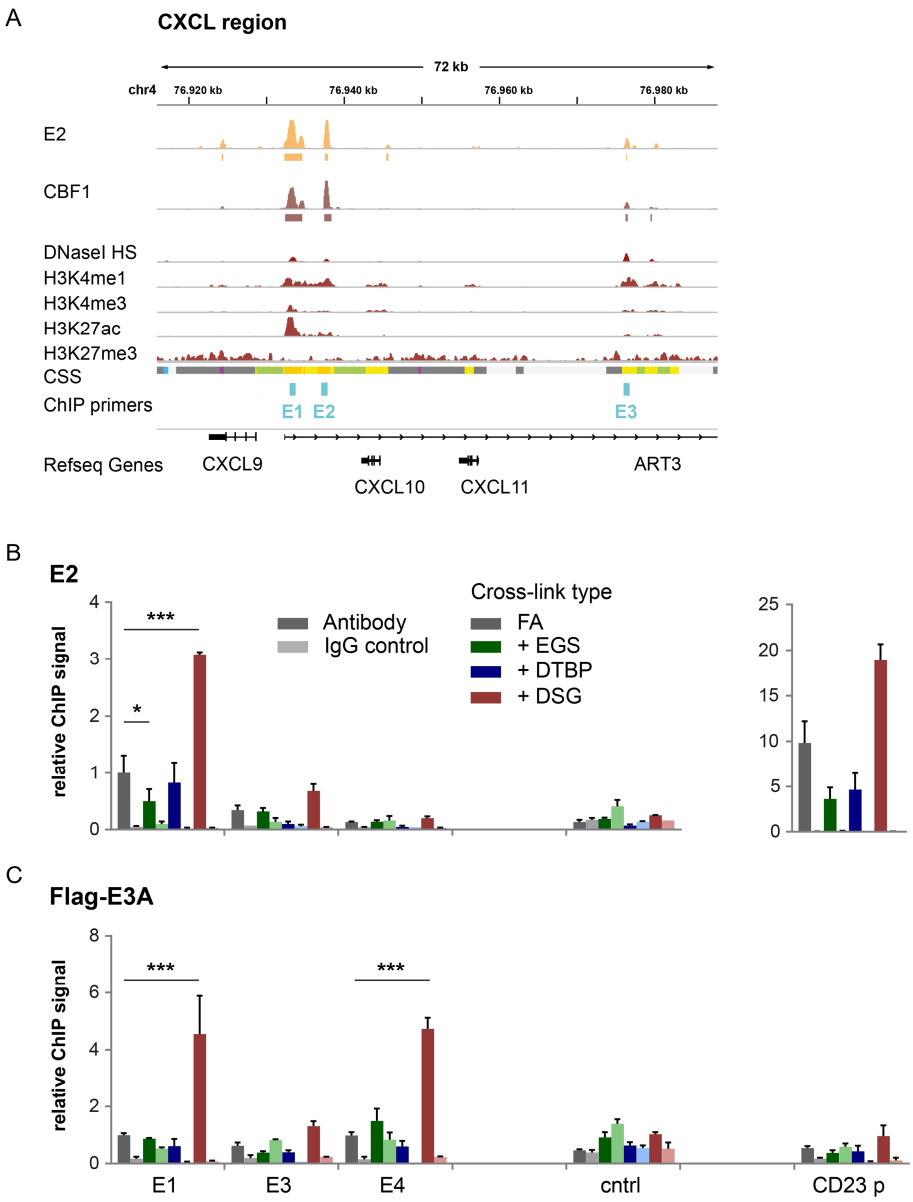


Figure 12. The impact of dual cross-linking on specific E2 and Flag-E3C ChIP enrichment. Three different cross-linking agents in addition to formaldehyde (FA) treatment were analyzed for their effect on specific DNA enrichment in E2 and Flag-E3A ChIP assays in the Flag-E3A LCL: the imidoester dimethyl 3, 30-dithiobispropionimidate (DTBP) and the two NHS-ester disuccinimidyl glutarate (DSG) and ethylene glycol bis(succinimidylsuccinate) (EGS). (A) Schematic view of a genomic region on chromosome 4 (hg19 coordinates)

encompassing *CXCL9* and *-10* genes, which are repressed by E3A and E3C while induced by E2. E2 and CBF1 ChIP signals and peaks are shown. The raw data derived from another study (Zhao et al., 2011b) but was processed with own bioinformatics pipeline. DNaseI HS, histone modification marks and chromatin state segmentation (css) for GM12878 cell line (LCL) from ENCODE are shown. All signals were normalized to input sample and RPKM but DNaseI HS only to coverage. 25x the mean signal at respective peaks was set as maximum for visualization. Positions of primers used for ChIP-qPCR are indicated and are located at strong (orange) or weak (yellow) enhancer regions as determined by ENCODE css. ChIP-qPCR for (B) E2 and (C) Flag-E3A using different cross-linking strategies. Cells were treated with 1.5 mM EGS for 20 min, 2 mM DTBP or 2 mM DSG for 30 min prior to FA cross-link for 7 min or with FA only for 7 min. Isotype matched antibody controls (IgG control) were used in both ChIP assays as negative controls. Three regions with known E2 binding (E1-E3) were chosen for investigation along a negative control locus (cntrl) were no E2 or CBF1 binding could be observed as well as the CD23 promoter (CD23 p), a well described E2 binding site as positive control for effective E2 ChIP intensity. Means of biological and technical duplicates with SD are shown. Significances of differences of means were assessed applying an unpaired two-tailed t-test (* $p<0.05$, *** $p<0.0005$).

The regular ChIP protocol was already successful in detecting E2 and Flag-E3A at the three investigated and described (Harth-Hertle et al., 2013) binding sites. Only the application of DSG prior to FA cross-linking could increase ChIP-qPCR signals for E2 and Flag-E3A at known binding sites E1, E2 and E3 (Harth-Hertle et al., 2013) while unspecific DNA recovery at a negative control region or using an isotype matched antibody as control did not increase (Fig. 12B and C). Thereby a 1.5 to 3.1 fold increase for E2 and 2.2 to 4.8 fold increase for Flag-E3A could be detected. E2 detection could also be enriched at the CD23 promoter region, a well described E2 binding site (Wang et al., 1991, Ling et al., 1994, Zhao et al., 2011b), while E3A was not significantly enriched here. This observation was expected since CD23 is only a described target gene for E2 (reviewed in Kempkes and Robertson, 2015) but not E3A or E3C. Application of EGS or DTBP did rather decrease DNA recovery than enrich the output. Especially EGS seems to work not as reliable as the other reagents since in one experiment very high DNA recoveries for the Flag-E3A ChIP could be detected in a very unspecific way (data not shown). The evaluation of the single experiment using EGS, which did not result in those huge enrichments, showed no significant improvement over sole FA cross-link. Thus a combination of DSG and FA was chosen for cross-link in all further ChIP-assays.

Optimization of the ChIP-assay for subsequent sequencing (ChIP-seq) - general protocol

But not only the cross-link displays a crucial point for possible optimization of the ChIP-assay. Especially ChIP samples which are intended for sequencing purposes need to pass certain criteria which have been carefully assessed and evaluated by the ENCODE consortium (Landt et al., 2012) and the ChIP-seq experiments conducted in this work vastly rely on those criteria.

One benchmark is of course the choice of the specific antibody, which should pass at least two different characterization assays. By using an epitope tag with commercially available well characterized antibodies, as described in this work, this step can be circumvented. As

described in chapter 4.1.3 the successful detection of Flag-E3A and -E3C in immunoblot assays (Fig. 9) as well as in IP experiments (Fig. 10) using a Flag specific antibody was possible.

In contrast to ChIP-assays with qPCR as readout (ChIP-qPCR), for ChIP-seq the size of DNA fragments after chromatin fragmentation is very crucial to the outcome of the experiment. Since only the two ends of a fragment are being sequenced (single-end sequencing) the absolute length of the DNA fragments ultimately limits the possible resolution for ChIP-seq. Hence, average fragment sizes of 100-300 bp are recommended for sequencing purposes. Since the application of DSG prior to FA cross-link results in stronger overall cross-linking in the cell the sonication process was analyzed and optimized carefully. Finally five rounds of sonication were applied to obtain fragment sizes desired for sequencing, while four rounds were enough for subsequent qPCR analyses (data not shown).

Furthermore, it was shown that more than two biological replicates did not significantly improve binding site discovery (Rozowsky et al., 2009) but sequencing depth has a great positive impact on site detection (ENCODE_Consortium, 2011). Especially low density peaks indicating weak or indirect DNA interactions and therefore display interesting data as well are only detected with higher sequencing depth. The ENCODE consortium therefore advises a minimum depth of 20 million reads for point-source TFs, which was exceeded in all experiments of this work (see Table 16).

Reproducibility and binding site detection are also depended on the complexity of the ChIP-seq library as defined by the nonredundant fraction of mapped reads (Landt et al., 2012). Calculation of the nonredunadant fraction as described by ENCODE was not performed here but all sequencing experiments conducted in this work showed percentages uniquely mapping reads of 62-72 % (with alignment rates of 94-99 %), which is described in detail in chapter 4.2.1.2.

4.2.1.2 Bioinformatic experimental design – from reads to peaks

There are many ways to get information on TF binding sites from ChIP-seq data using diverse bioinformatic tools and several guidelines for analyses were published, for instance by the ENCODE consortium (Landt et al., 2012). Having access to the Galaxy Server, hosted and maintained by the Bioinformatics Department of the University of Freiburg under the supervision of and kindly provided by Prof. Dr. Rolf Backofen, an independent analysis of the ChIP-seq data obtained in this thesis was possible. Galaxy per se is a publically accessible platform developed and hosted by researchers at Penn State University and John Hopkins University (galaxyproject.org) (Giardine et al., 2005). The Galaxy Platform does not only give access to diverse bioinformatic tools, they are also presented in a very comprehensible way and most notably workflows can be saved and shared with other users thereby making the whole process reproducible.

In the following the strategy for identifying binding sites for E2, E3A, and E3C is explained briefly with focus on the choice and importance of each step (Fig. 13), while the details are listed in the methods section (chapter 3.6.1).

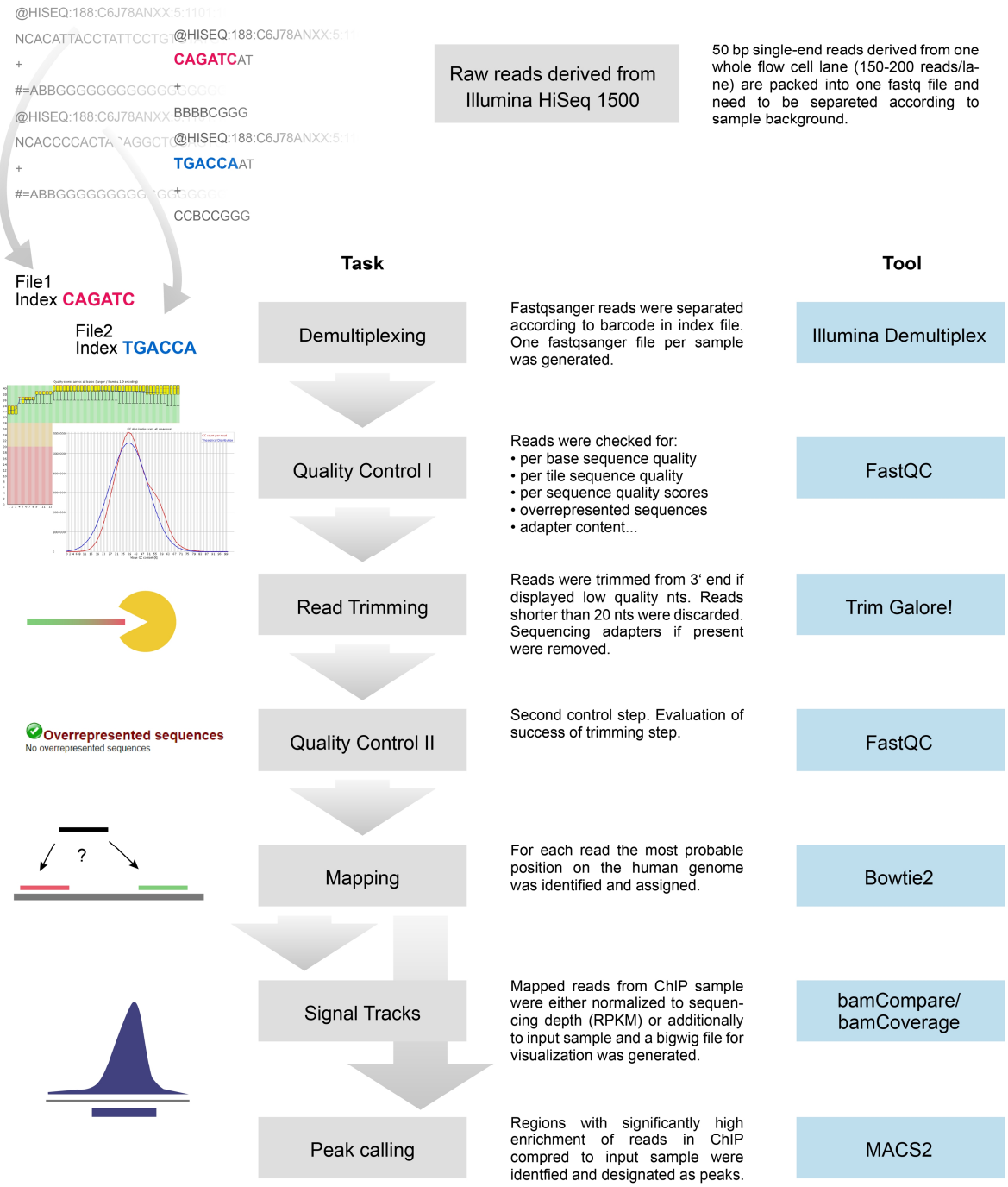


Figure 13. Schematic workflow of peak detection from ChIP-seq raw data. Each step of the analysis is explained briefly. The purpose of each task (grey boxes) is described and the respective used bioinformatic tools (blue boxes) are indicated.

For each protein of interest two independent biological replicates, consisting of the ChIP and a chromatin input control sample, were prepared and single-end sequencing libraries were prepared and subjected to next generation sequencing using an Illumina HiSeq 1500 machine, set to produce 50 bp reads as output, at Dr. Helmut Blums laboratory at the Gene Center of the LMU

Munich. ChIP samples were prepared by myself, while library preparation and sequencing process were conducted by Dr. Blums laboratory.

The obtained raw data was provided at their own Galaxy instance alongside an in-house developed Demultiplexing tool. Since samples were barcoded (for distinct identification) and then pooled, up to 8 samples per flow cell lane, the raw data displays a composite of all reads from one flow cell lane. For each flow cell lane one fastqsanger file, containing the actual reads plus their identifiers and one index file, containing the identifiers plus the detected barcode, are generated. Upon demultiplexing the data is split in separate files according to the designated barcodes which are then ready for further processing.

Table 16. Obtained reads from ChIP-seq after demultiplexing

ChIP	Replicate – Sample Type	Read Count	Overrepresented Sequences (%)	Internal Designation
E2	E2-I-ChIP	21,451,466	0.37	LG562.2_E2
	E2-I-input*	35,575,701	-	LG562.2_input
	E2-II-ChIP	26,478,900	0.79	LG568.1_E2
	E2-II-input	29,618,835	-	LG568.1_input
Flag-E3A	E3A-I-ChIP	30,675,192	-	LG470_Flag-E3A
	E3A-I-input	37,311,996	0.40	LG470_input
	E3A-II-ChIP	26,637,528	1.14	LG562.1_Flag-E3A
	E3A-II-input	28,509,282	-	LG562.1_input
Flag-E3C	E3C-I-ChIP	82,905,413	-	LG478_Flag-E3C
	E3C-I-input	38,575,786	0.89	LG478_input
	E3C-II-ChIP	45,447,840	1.81	LG562.2_Flag-E3C
	E3C-II-input*	35,575,701	-	LG562.2_input

For each protein of interest two independent biological replicates (I and II), each consisting of the actual ChIP sample and an input control, were generated and subjected to deep sequencing. Reads were demultiplexed according to the designated barcode using a tool from the Gene Center specific for this purpose and a single fastqsanger file was written for each sample. The absolute number of reads for each sample is listed here alongside the percentage of overrepresented sequences as calculated by FastQC. * E2-I-input and E3C-II-input are actually the same sample since E2 and Flag-E3C ChIPs were performed using the same chromatin preparation.

In a first Quality Control step the obtained reads for each sample were analyzed for their basic features as per base sequence quality, per tile sequence quality, per sequence quality, per base sequence content, per sequence GC content, per base N content, sequence length distribution, sequence duplication levels, overrepresented sequences, adapter content, and Kmer content. Thereby the FastQC tool (Andrews, 2010) rates each quality control step in three categories (good, intermediate and failed) and all assessed samples rated intermediate at worst for all criteria.

Nevertheless it was noticeable that some samples showed an elevated percentage of overrepresented sequences (Table 16), which appeared to be derived from adaptor contamination and therefore a trimming step was included prior to mapping the reads. For the Read Trimming TrimGalore (Krueger, 2012) was applied and set to remove Illumina adaptor sequences if found and to trim the reads from the 3' end if low quality is detected. If reads became shorter than 20 nt after adapter and quality trimming they were discarded.

In a second Quality Control step applying FastQC the trimmed reads were checked and showed no overrepresented sequences at all, indicating the Illumina adaptor has been responsible for those. Just a very small fraction of reads for each sample was discarded due bad quality or length after quality trimming (Table 17, column 4).

Next the reads were mapped to the human genome (hg19) using the Bowtie2 software (Langmead and Salzberg, 2012). This algorithm identifies the most probable location on the assigned genome for each read, which can be one distinct location (uniquely mappable read) or several locations which display different assigned probabilities. My ChIP-seq data reached 93.8 - 98.8% of reads Mapping to the human genome, including around 70% of uniquely mapping reads, reflecting good sample qualities (Table 17). Reads not mapping to the human genome were written in a separate output file and were subsequently used for mapping against the EBV genome (HHV-4 type I, NC_007605.1) as depicted in Table 18.

Table 17. Reads after different workflow steps and mapping to the human genome

		Read count			
ChIP	Replicate - Sample Type	Demultiplex	Trimming (% of Demultiplexed)	Mappable Reads (% of Trimmed)	Uniquely Mappable Reads (% of Trimmed)
E2	E2-I-ChIP	21,451,466	21,321,357 (99.4)	95.8	71.3
	E2-I-input*	35,575,701	35,502,629 (99.8)	98.8	70.1
	E2-II-ChIP	26,478,900	26,212,962 (99.0)	96.5	70.0
	E2-II-input	29,618,835	29,558,726 (99.8)	98.8	72.0
Flag-E3A	E3A-I-ChIP	30,675,192	30,331,627 (99.9)	97.0	71.2
	E3A-I-input	37,311,996	36,668,892 (98.3)	97.9	71.3
	E3A-II-ChIP	26,637,528	26,234,960 (98.5)	97.5	71.3
	E3A-II-input	28,509,282	28,434,767 (99.7)	98.7	69.8
Flag-E3C	E3C-I-ChIP	82,905,413	82,871,787 (99.96)	97.9	72.5
	E3C-I-input	38,575,786	38,409,666 (99.6)	95.8	61.7
	E3C-II-ChIP	45,447,840	44,451,409 (97.8)	93.8	70.0
	E3C-II-input*	35,575,701	35,502,629 (99.8)	98.8	70.1

Reads obtained after demultiplexing were subjected to trimming (percentages of remaining reads are indicated) and subsequently to Bowtie2 for mapping to the human genome (hg19). * E2-I-input and E3C-II-input are actually the same sample since E2 and Flag-E3C ChIPs were performed using the same chromatin preparation.

Table 18. Reads mapping to the EBV genome

ChIP	Replicate - Sample Type	Read Count		
		Not Mapping to hg19	Mapping to EBV (HHV-4)	Uniquely Mappable (%)
E2	E2-I-ChIP	895,965	7,791	6,809 (87.4)
	E2-I-input*	424,267	13,358	10,649 (79.7)
	E2-II-ChIP	906,486	32,365	29,007 (89.6)
	E2-II-input	351,825	10,233	8,444 (82.5)
Flag-E3A	E3A-I-ChIP	905,260	15,347	12,855 (83.8)
	E3A-I-input	761,709	33,599	25,974 (77.3)
	E3A-II-ChIP	644,486	6,588	5,478 (83.1)
	E3A-II-input	359,120	11,162	9,229 (82.7)
Flag-E3C	E3C-I-ChIP	1,728,003	60,067	47,693 (79.4)
	E3C-I-input	1,609,662	142,042	108,330 (76.3)
	E3C-II-ChIP	2,745,597	8,305	6,864 (82.6)
	E3C-II-input*	424,267	13,358	10,649 (79.7)

Reads from ChIP-seq experiments conducted in LCLs which did not map to the human genome were mapped to the EBV genome (HHV-4 type I, NC_007605.1). * E2-I-input and E3C-II-input are actually the same sample since E2 and Flag-E3C ChIPs were performed using the same chromatin preparation.

E2, E3A, and E3C ChIP-seq experiments were performed as independent duplicates, each time an input sample was sequenced as well. There is no standard procedure of dealing with duplicates in ChIP-seq experiments. Some researchers call peaks for the individual replicates and then make an intersection and only further analyze those peaks. This kind of analysis tends to focus on peaks with high enrichment rates but smaller peaks, which are present in only one experiment get lost. Since not only the peak positions but also enrichment for quantitative analyzes was of interest for this thesis, duplicates were merged prior to peak calling. Peaks with low enrichment and significances deriving from only one replicate, most likely due to noise, are "flattened" out by this kind of analysis. Low enrichment peaks due to indirect DNA interaction or weak TF binding but present in both replicates are expected to be be included in the output.

For Peak Calling MACS2 (Zhang et al., 2008) was applied using merged ChIP and input files. The specific settings had to be adjusted for each experiment (e.g. due to different fragment lengths) (see MM chapter 3.6.1). In Table 19 absolute numbers of called peaks on the human genome are listed while those on the EBV genome are listed in Table 20.

Table 19. Peaks identified in the human genome using MACS2

ChIP	Subjected to MACS2	Read Count		Allowed Duplicate Tags	Redundancy Rate (%)	Peaks
		Merged Mapped Reads	Filtered			
E2	E2-ChIP	45,731,868	44,783,737	2	2.1	23,314
	E2-input	64,285,263	63,034,015	2	1.9	
Flag-E3A	E3A-ChIP	55,016,841	53,573,894	2	2.6	14,858
	E3A-input	63,982,830	62,781,693	2	1.9	
Flag-E3C	E3C-ChIP	122,849,596	120,155,588	3	2.2	12,504
	E3C-input	71,878,366	70,492,793	3	1.9	

Mapped reads of replicates were merged and subjected to MACS2 peak calling algorithm. Here reads were filtered for allowed duplicate tags, which represent maximum permitted reads mapping to the exact same position. This value is calculated by MACS2 in accordance with absolute read count and genome coverage. The redundancy rate is indicating the percentage of duplicate reads not allowed and displays a measurement for library complexity.

Table 20. Peaks identified in the EBV genome using MACS2

ChIP	Subjected to MACS2	Read Count		Allowed Duplicate Tags	Redundancy Rate (%)	Peaks
		Merged Mapped Reads	Filtered			
E2	E2-ChIP	40,156	31,347	4	22.0	7
	E2-input	23,591	23,591	4	0.0	
Flag-E3A	E3A-ChIP	21,935	21,864	4	0.3	10
	E3A-input	44,761	44,756	4	0.0	
Flag-E3C	E3C-ChIP	68,372	68,327	5	0.1	15
	E3C-input	155,400	155,279	5	0.1	

Mapped reads of replicates were merged and subjected to MACS2 peak calling algorithm. Here reads were filtered for allowed duplicate tags, which represent maximum permitted reads mapping to the exact same position. This value is calculated by MACS2 in accordance with absolute read count and genome coverage. The redundancy rate is indicating the percentage of duplicate reads not allowed and displays a measurement for library complexity.

In the peak calling process the files containing the merged duplicates of already mapped reads were filtered for duplicate reads at the same position. MACS2 calculates the number of maximum allowed duplicates at the exact same position including the information on sample and genome size. Using large input datasets as in this example, MACS2 rates two or even three reads at the same position in hg19 to be due to sample size rather than to low library complexity and PCR artefacts. Redundancy rates are calculated based on this assumption and exceeding duplicate reads are excluded from peak calling. Peak calling for human and EBV genome differs noticeably: Less reads could be aligned to the EBV genome but more duplicate reads at the exact same position are allowed due to the significantly smaller EBV genome (180 kb) were the overall read coverage is higher than for the human genome (data not shown).

Besides peak calling also the quantification and visualization of ChIP-seq results in read coverage density displays an important analysis step. Many laboratories, also the ENCODE project, are providing such Signal Tracks but in most cases the ChIP-seq signal here is only normalized to genome coverage and the input sample is treated and shown in parallel for comparison. Here, in this thesis, one further normalization step was included by using bamCompare, a tool from the deepTools package (Ramirez et al., 2014a), where the input was subtracted from the actual ChIP sample in addition to normalization to fragments (reads) per kb per million (RPKM) to account for genome coverage. By normalization to the input file, not only for the peak calling but also the signal track generation, the resulting visualization is more informative and regions with high enrichment for both, specific ChIP and input, e.g. small repeats are not showing signal enrichment anymore. This procedure was used as standard procedure in this work, unless indicated otherwise when e.g. using data derived from other laboratories with no respective control sample available bamCoverage (of deepTools) was used to generate signal tracks normalized to coverage only.

During this bioinformatic analysis and constant monitoring of the obtained results in genome browsers the observation was made that in some cases a peak was called for a certain position but the signal track was not matching, but in fact showed negative amplitude. This event could be observed in particular at pericentromeric or -telomeric regions, which are enriched in repetitive elements and are harder to map. Therefore one further step was established to filter peaks called by MACS2 for "negative peaks" and such falling in "unmappable regions": Peaks were sorted according to their mean signal (using bamCompare signals and normalized to peak length) and discarded if below 1.5 (to get rid of negative peaks and marginal cases). Then peaks which were located on black-listed regions, as assigned by ENCODE DAC and Duke to account for unmappable regions (Derrien et al., 2012), were excluded, resulting in fewer but high confidence peaks (Table 21). Since the data should be comparable with those from ENCODE for GM12878 LCL, peaks located on unknown chromosomal locations (e.g. chrUn) and the Y chromosome, since GM12878 derived from a female and the LCL used here from a male donor, were excluded as well.

Table 21. Signal and mappability corrected peaks in the human genome

ChIP	Identified by MACS2	Signal corrected	Blacklist corrected	GM12878 compatible	% of MACS2 peaks
E2	23,314	22,857	22,715	22,500	96.5
Flag-E3A	14,858	14,553	13,579	13,490	90.8
Flag-E3C	12,504	11,134	8,898	8,733	69.8

Peaks identified by MACS2 were further filtered to exclude peaks which display a negative amplitude, fall on blacklisted regions or a chromosome not compatible with GM12878, the LCL used by ENCODE.

This evaluation could not be applied to the peaks identified by MACS2 in the EBV genome, but due to the small genome size and few peak numbers, the genome wide peak assessment is here more feasible than for the human genome. None of the detected peaks shows a negative amplitude and or is located at the known repetitive EBV regions and therefore could be used for further investigation as provided by MACS2.

Later in this part of the results, the data obtained during this work is compared to other NGS data, mostly ENCODE data. Here the self-designed standard procedure was applied and for signal tracks, replicates (if available) were merged (for ChIP sample and input) and ChIP normalized to input and genome coverage as explained above. Also for peak calling (if applied) the merged replicates were used. For both analyses, peak calling and generation of signal tracks, already mapped reads (conducted by the respective laboratories) were used. If peak files were available e.g. from the ENCODE project, they were subjected for further analyses. In some indicated cases peak calling was conducted by myself e.g. when comparing data derived from different laboratories.

One exception displays the data on E2 and CBF1 DNA binding by Zhao et al., which was only provided as reads mapped to hg18. In this special case fastqsanger files were used and processed applying the self-designed workflow.

4.2.2 Characterization of E2, E3A, and E3C binding sites in the EBV genome

Potential binding of E2, E3A, and E3C to the EBV genome should be investigated in this thesis as well, since regulation of EBV genes by the EBNA proteins has been described and characterized intensively in the past. E2 is known to induce expression of viral genes *LMP1*, *LMP2A/B* as well as such derived from the C promoter (Cp), which gives rise to a polycistronic RNA coding for all six EBNA proteins (reviewed in Kempkes and Ling, 2015).

The E3 proteins were also described to interfere with EBV transcription but the complete picture is still not clear to date. On one hand the E3 proteins, E3A, E3B, and E3C, were described to repress the E2 mediated activation of the *LMP1* promoter (Le Roux et al., 1994), but E3C was also found to CBF1 independently activate the *LMP1* promoter in cooperation with E2 (Lin et al., 2002). Furthermore, E3C could be located at the *LMP1* promoter (Jimenez-Ramirez et al., 2006) as the only E3 protein so far. In reporter assays E3C, as well as E3A, were described to repress Cp derived transcription (Radkov et al., 1997, Waltzer et al., 1996), which could not be verified in an EBV positive B cell line with inducible E3C expression (Jimenez-Ramirez et al., 2006).

Analyzing the ChIP-seq data derived from LCLs in this thesis, 7 E2, 10 E3A, and 15 E3C binding sites could be identified (Table 20) and were further investigated for their locations within the EBV genome. Due to its relatively small size it is possible to depict the whole EBV genome in one map for an overview of EBNA binding behavior (Fig. S1). Here, E2, E3A, and E3C data from this work were compared with published data on E2 and CBF1 (Zhao et al., 2011b), the best described DNA adaptor for the EBNAs. To this end the publicly provided raw data was used to generate signal tracks and peaks applying standards described above leading to the identification of 3 E2 and 4 CBF1 binding sites (Fig. 14). The analysis of those published ChIP-seq data revealed significantly less reads mapping to the EBV genome than was discovered in the experiments conducted in this thesis (approx. 10% of total read count, data not shown). Accordingly, genome read coverage was much lower and peak calling is also influenced by this circumstance.

Two of the identified E2 binding sites showed a very prominent enrichment at the bidirectional *LMP1/LMP2B* promoter site and the *LMP2A* promoter 3 kb further upstream (Fig. 14, red columns). E2 could be identified at the same positions using the data from Zhao et al. supporting the significance of those binding sites. Interestingly, the signal enrichment at the *LMP1/LMP2B* promoter is higher in this study compared to Zhao et al., which could be due to better read coverage. Also CBF1 could be located at both sites, showing the same enrichment pattern as E2 derived from Zhao et al. Both sides are also positive for E3A and E3C, with significant peaks at the *LMP1/LMP2B* promoter but only for E3C at the *LMP2A* promoter.

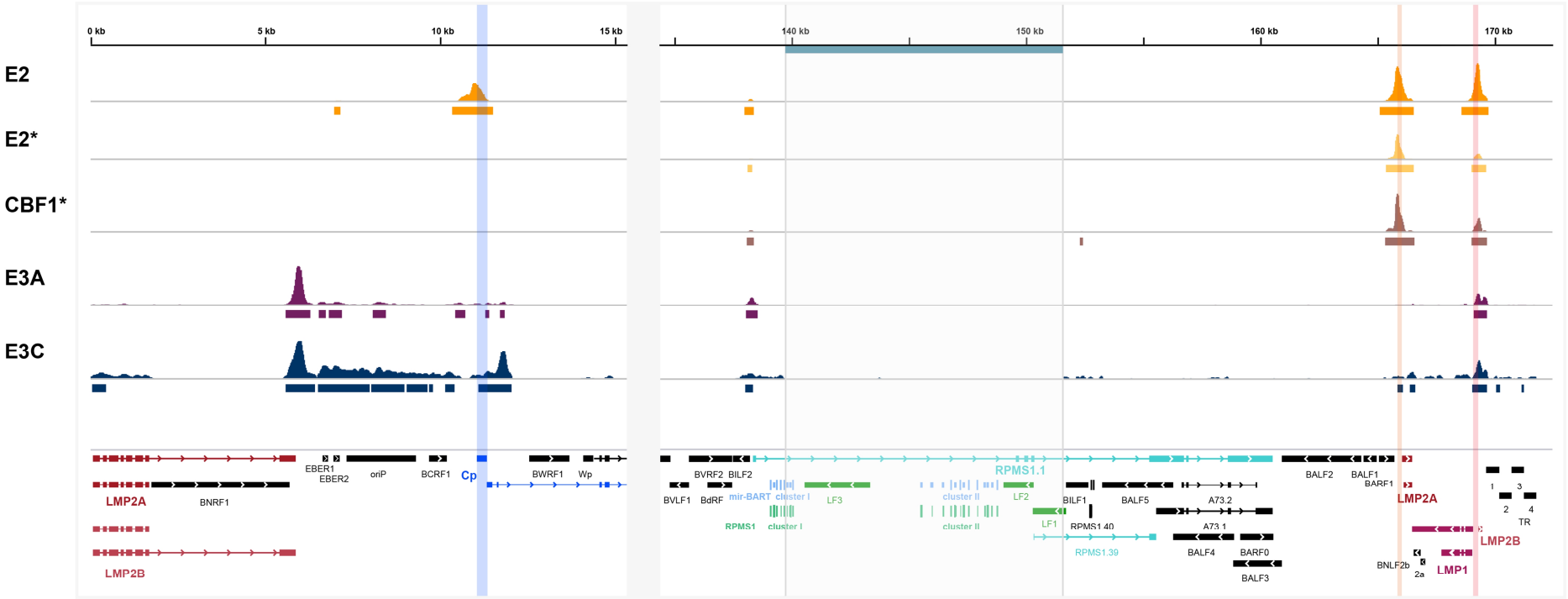


Figure 14. Identification of E2, E3A, and E3C binding sites in the EBV genome. Schematic map depicting two details of the EBV genome (HHV-4 type I, NC_007605.1, map provided by the EBV portal (Arvey et al., 2012)). Genes expressed during the lytic cycle are depicted in black and genes expressed during latency are highlighted in color. Also marked is the EBNA regulated Cp, which gives rise to different (polycistronic) splice variants coding for all EBNAs, including proteins of interest E2, E3A, and E3C. The light grey box to the right encompasses a region, which is deleted in the B95.8 EBV genome used for EBV BACmid generation compared to HHV-4 type I reference genome. Genes affected by the B95.8 deletion are highlighted in blue and green. Thus it is not possible that reads from ChIP-seq analysis derive from this genomic region. EBNA regulated *LMP1*, *LMP2A*, and *LMP2B* genes are shown in red, with the bidirectional promoter controlling *LMP1* and *LMP2B* expression as well as the *LMP2A* promoter highlighted with light red columns. In the upper panels ChIP-seq signal profiles and underneath peaks called by MACS2 for E2, E3A, and E3C are shown. (*) Additionally published data for E2 and CBF1 (Zhao et al., 2011b) is shown for comparison. All signal tracks were set to show maximal intensities of the respective ChIP-seq.

Another region of interest displays the C promoter (Fig. 14, blue column), which harbors an E2 but also E3A and E3C binding sites, which do not overlap precisely. This E2 binding site could not be identified using the data from Zhao et al. and also no CBF1 enrichment was detected here. Furthermore the E3 proteins, especially E3C, show a very broad stretch of ChIP-seq signal enrichment a region spanning oriP, the *EBERs*, and up to the last exon of *LMP2A/B*. This behavior is not typical for TF factors and is more common among histone modification marks or histone variants.

Strikingly, the Cp and *LMP1/LMP2B* promoter show alternate binding behavior for E2 and E3 signal enrichment. Both regions harbor significant binding sites for all proteins but the relative enrichment is inverted, with high E2 signals at the *LMP1/LMP2B* and *LMP2A* promoter and high E3C signal at the C promoter and upstream region. Thereby E2 and CBF1 show a very similar binding pattern with one additional, low enrichment CBF1 peak within the *BART* region. Also the E3 proteins demonstrate a very similar enrichment profile, with a few more called peaks for E3C mainly at the oriP region but also at the terminal repeats (TR) and exons 2 and 3 of *LMP2A/B*.

Also all four investigated TFs could be identified to bind at the promoter of the full length transcript of *RPMS1* (Fig. 14, turquoise), a putative ORF whose translation to a protein could not be confirmed to date but gives rise to *BART* ncRNAs and BART miRNAs.

The contribution of cellular TFs on EBNA binding to the EBV genome was not assessed in this study, since the major focus was on interaction with the cellular genome and TFs in this context. Nevertheless, the findings on EBNA binding to the EBV genome display a new piece of information which could contribute to further characterizations of gene regulation in the viral background.

4.2.3 Preferential targeting of enhancer modules in the human genome by E2, E3A, and E3C

The main focus of this study is on the interaction of EBNA proteins with the cellular genome in order to regulate target gene transcription. As described in the introduction, several laboratories published datasets on target genes for EBNA proteins and for some distinct examples the regulation processes were characterized in detail: E2 induces viral the expression of Cp derived transcripts and *LMP1* by targeting promoters (reviewed in Kempkes and Ling, 2015). A genome wide analysis of E2 and CBF1 binding sites revealed the conjointly targeting of regulatory elements such as enhancers rather than promoters (Zhao et al., 2011b). Within the *CXCL9* and *-10* genomic locus the direct reciprocal targeting of intergenic enhancers by E2 and E3A could be shown for the first time (Harth-Hertle et al., 2013). Also a genome wide search for E3

binding sites revealed that E3 proteins primarily target promoter distal elements (McClellan et al., 2012). Although one has to mention that in this particular study all three E3 proteins were precipitated together in one ChIP-seq experiment and no genome wide conclusion for single E3 binding behavior can be drawn. Furthermore, this experiment was performed in Mutu III cell line, which derived from an EBV positive Burkitt's lymphoma but showing latency III type expression of viral genes, and not in an LCL background.

To investigate which functional elements are targeted by the single EBNA proteins within the human genome, the chromatin state segmentation (css) dataset from ENCODE was used (Ernst et al., 2011). Here, 9 histone modifications associated with distinct functional regions, binding sites for CTCF a sequence specific chromatin insulator protein, PolII and histone variant H2A.Z associated with nucleosome free regions derived from 9 different cell lines commonly used by ENCODE, including the LCL GM12878, were used to generate chromatin state maps of the human genome consisting of 15 distinct states. Those are divided in active, weak and poised promoters, strong and weak enhancers, putative insulators, transcribed regions, polycomb repressed regions, and heterochromatin.

Comparison of detected binding sites with css shows that all three EBNAs and the adaptor protein CBF1 (data from Zhao et al. 2011) primarily target enhancer regions in the human genome (Fig. 15A). Interestingly, the percentage of targeted enhancer regions is higher for both E3 proteins compared to E2 and CBF1. The second most targeted functional elements by all EBNAs are promoter regions. Here, E2 and CBF1 show more binding sites at promoter regions than both E3 proteins.

The segmentation of the human genome into functional elements by css is exclusively based on histone modification marks and PolII occurrence, including the determination of promoter positions without referring to the genomic positions of annotated genes. Besides, it was described that active enhancers are frequently transcribed (reviewed in Plank and Dean, 2014, and Kulic et al., 2015) which makes it possible that those transcribed enhancers get annotated as promoters by css. Such incidents could be observed when visualizing the obtained ChIP-seq data in a genome browser (data not shown).

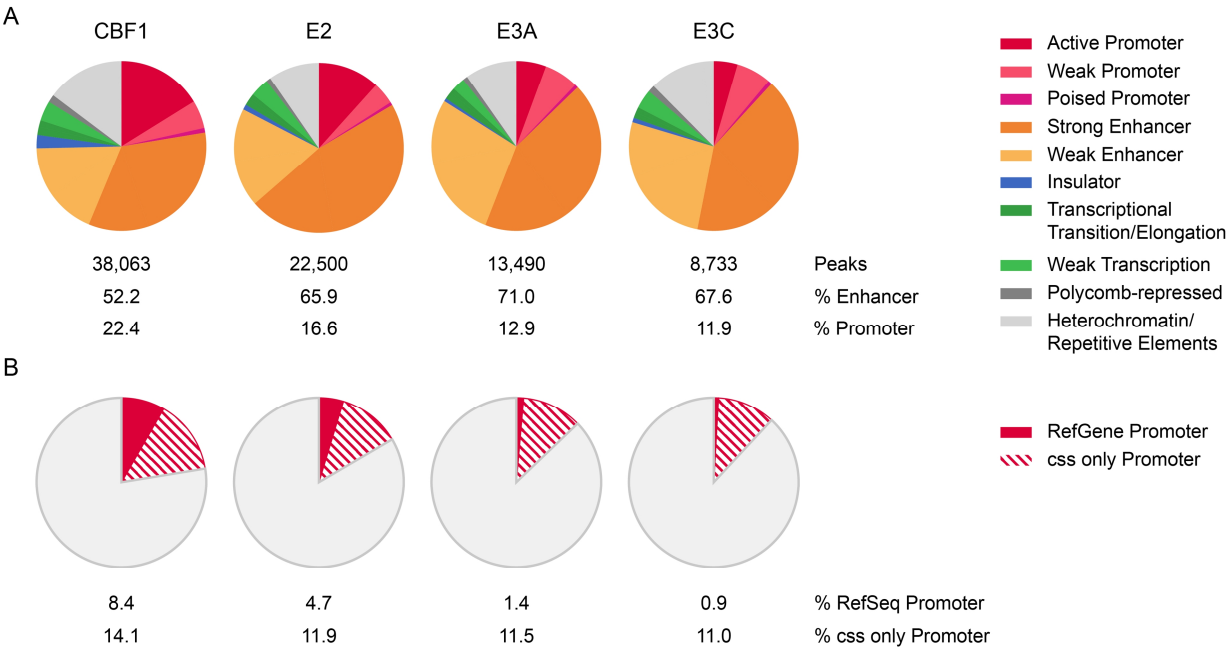


Figure 15. EBNA proteins target enhancer elements rather than promoter regions. Locations of CBF1, E2, E3A, and E3C peaks were analyzed in respect to functional DNA elements using the peak center as decisive position criterion. (A) Chromatin state segmentation (css) performed by the ENCODE consortium (Ernst et al., 2011) in GM12878 cell line was used as information for the location of functional DNA elements. Here histone modifications and polymerase occupation were used to identify the different states. Absolute number of peaks and percentages located on enhancer elements are indicated below. (B) Peaks which are located on one of the three promoter states from css (22.4, 16.6, 12.9, and 11.9% for CBF1, E2, E3A, and E3C respectively) were further analyzed for their location relative to annotated Refseq genes. To this end the regions of 1 kb upstream of each Refseq gene in hg19 were considered as Refseq promoters. Peaks which were previously described to be located on promoter elements by css but do not fall on a Refseq promoter are designated as peaks on css only promoters. Percentages of peaks located on those subsets are indicated below.

In order to investigate this phenomenon, the promoter associated peaks were further analyzed for the presence of annotated genes by Refseq (Pruitt et al., 2005) (Fig. 15B). Interestingly, only 8.4% of CBF1 and 4.7% of E2 binding sites were located at promoters as defined 1 kb upstream of Refseq TSS (*RefSeq promoters*) when the genome wide css prediction identified 22.4 and 16.6% peaks at promoters for CBF1 and E2, respectively. This effect was even more prominent for E3 binding sites. When further dissecting E3A and E3C peaks located at promoter regions (12.9 and 11.9% respectively) only 1.4 and 0.9% respectively were found at Refseq promoters. These findings imply that part of the regions targeted by EBNA proteins and declared as promoters by css are actually promoters of enhancers which give rise to non-coding RNAs which are not yet included in the Refseq genes catalogue. Thus the percentage of targeted enhancers is even higher than anticipated in the initial analysis and reveals a picture where E3 proteins are almost exclusively targeting enhancers but almost no promoter regions. And also the percentage of E2 and CBF1 bound promoters is lower than estimated first but still accounts for a subset not to be ignored.

To further characterize the regions bound by EBNA proteins, histone modifications H3K4me1 and H3K4me3, characteristic for enhancers, as well as H3K27ac, typical for active enhancer elements, derived from ChIP-seq experiments in GM12878 by ENCODE were used to generate anchor plots depicting signal distribution at the different binding sites (Fig. 16).

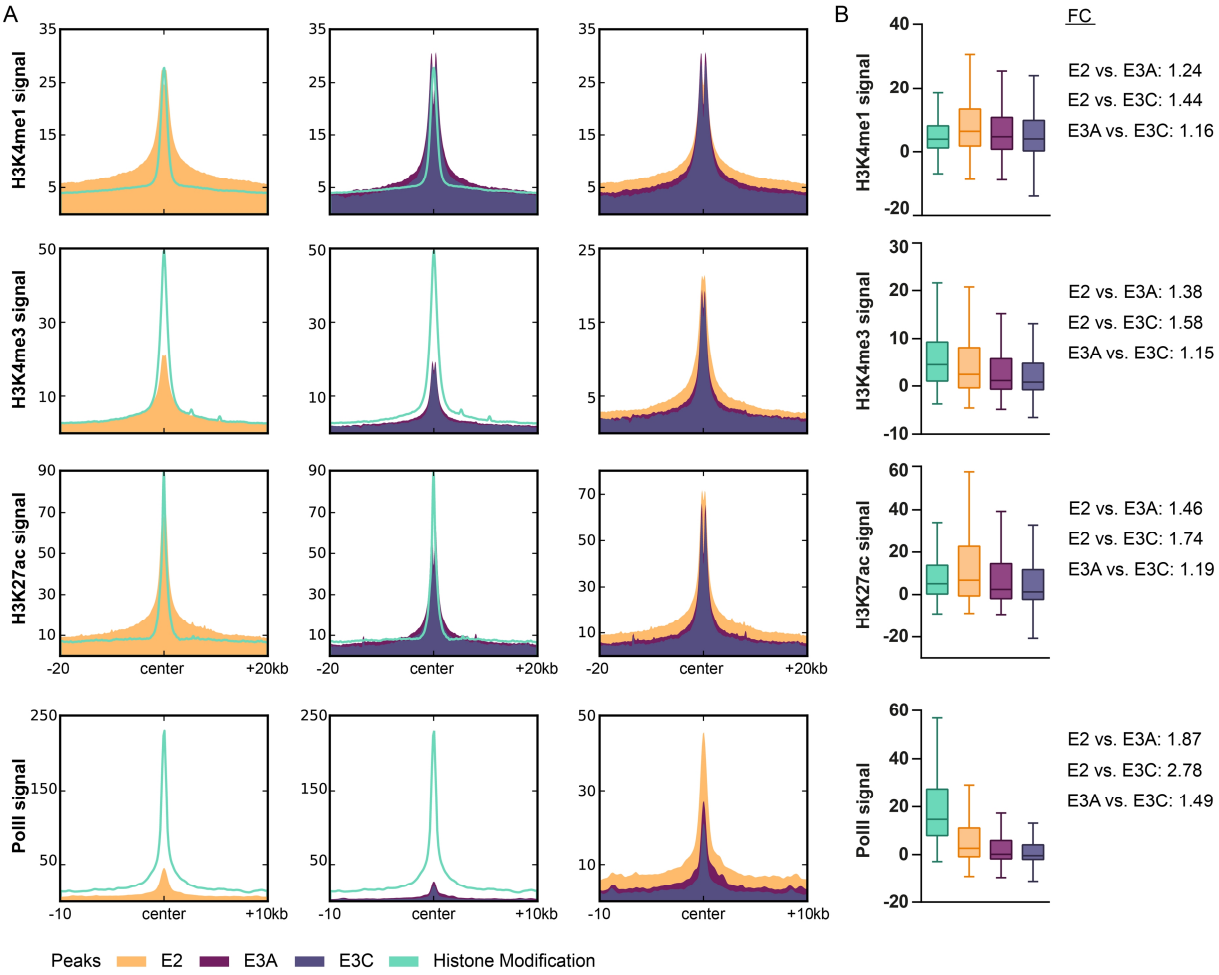


Figure 16. E2 binding sites show stronger enrichment for enhancer marks than E3 binding sites. Signal intensities for histone modification marks H3K4me1, H3K4me3, and H3K27ac as well as PolII derived from ENCODE ChIP-seq and were normalized for input and genome coverage (RPKM). (A) Anchor plots were generated showing signal distributions at regions flanking 20 or 10 kb (PolII) in each direction of EBNA peak centers. (B) The data underlying (A) were used to generate boxplots depicting distributions of signal intensities. For internal comparison of the different signal intensities peaks of each histone modification or PolII were used as positive control. Fold changes of signal intensities between E2 and E3 peaks are indicated and all comparisons are statistically significant with $p<0.0001$ applying two-tailed t-test with Welch's correction. Boxplot whiskers extend to 1.5x interquartile range.

Interestingly, all enhancer marks as well as PolII showed higher signal enrichments at E2 peaks compared to E3 peaks. The phenomenon was most distinctive for PolII followed by H3K27ac, indicating that E2 can be found more frequently at activated enhancers than E3. The E3 proteins are even targeting more enhancers percentage wise than E2 in the css analysis (Fig. 15A), which indicates that E2 is binding to strong enhancers more frequently than E3.

However, subsets of E2, E3A, and E3C peaks are located at RefSeq promoters, as described above (Fig. 15B). In order to investigate the relationship between the positions of EBNA binding sites and their regulated genes, a relative distance analysis was performed (Fig. 17). To this end different data sets from the Kempkes laboratory on EBNA target genes were used: The information on E2 target genes derived from an inducible system in BL41 (Maier et al., 2006), E3A target genes were analyzed in LCL, comparing E3A ko and wt cells (Hertle et al., 2009), and E3C targets were also investigated by comparison of E3C ko and wt cells (diploma thesis Agnes Nowak).

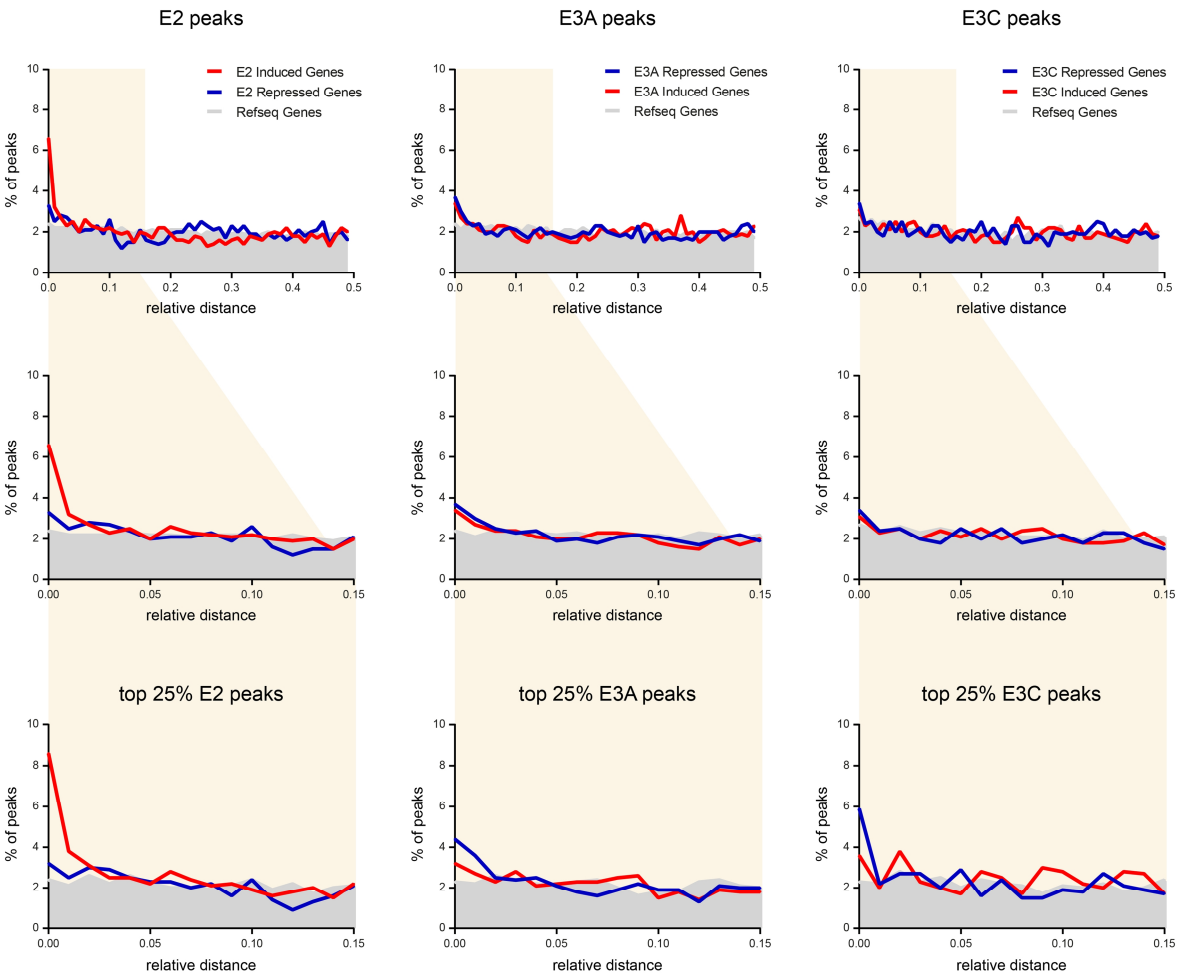


Figure 17. Subsets of E2, E3A, and E3C target genes are directly bound by the regulating EBNA. Relative distance analysis showing the relationship between E2, E3A, and E3C peaks and their regulated genes. The distance of each peak was to the nearest gene within the analyzed data set was assessed and relativized by diving this distance through the distance between the two genes the peak is located between. The relative distance measure ranges between 0, which displays a perfect hit of the peak on the nearest gene, to 0.5, which displays the perfect center between to genes. The percentage of peaks showing distinct relative distance values (from 0 to 0.5) is indicated. Therefore, if no spatial correlation between peaks and the analyzed gene set exists an uniform distribution of relative distance values from 0 to 0.5 is expected but if they are closer than expected by chance, a shift towards small values of relative distance are expected. The upper panel shows the whole range of relative distances for EBNA peaks and their target genes, the middle panel displays a zoom-in to relative distance values 0-0.15, and in the bottom panel the top 25% peaks, sorted by mean signal are assessed for their relative distance to different gene sets.

This relative distance analysis revealed that E2 peaks showed an elevation of peaks showing very low relative distance towards E2 induced genes, including 6.6% of E2 peaks located directly at an induced gene, while this could not be observed towards E2 repressed genes (Fig. 17, top and middle panel). This effect was even more pronounced when focusing only on the top 25% of E2 peaks, defined by E2 signal enrichment, even reaching 8.6% peaks directly located at an induced gene. Interestingly, this phenomenon could not be observed for E3A or E3C peaks towards their regulated genes. Only when the top 25% of E3A or E3C peaks were used for this analysis, an elevation of E3A and E3C peaks with very low relative distances towards E3A or E3C repressed genes, respectively, could be observed (Fig. 17, bottom panels). For all analyses the distribution of EBNA peaks towards the whole RefSeq gene set was used as a negative control which resulted in a random distribution of relative distances each time.

4.2.4 Enhancer signature is a prerequisite for accession of E2 to chromatin and is enriched upon E2 expression

The finding that EBNA proteins primarily target enhancer elements and especially that E2 associates with stronger enhancer signatures than identified for E3 proteins raised the question if this strong enhancer signature is a prerequisite for E2 accession to chromatin or a result of E2 chromatin binding and subsequent recruitment of e.g. histone acetyltransferases which mediate active marks.

Another study already showed H3K4me1, the histone modification most prominent at enhancer elements, to be present not only at E2 binding sites in LCL but also at the same genomic positions in CD19+ primary B cells, which are EBV negative and model the situation before infection (Zhao et al., 2011b). It was concluded that E2 is binding to enhancer elements already existing in primary B cells, with overall lower H3K4me1 intensities. However, this analysis did compare absolute values of normalized H3K4me1 signals at E2 binding sites derived from two different experiments, conducted by two different research groups (CD19+ cells by Roadmap Epigenomics (Bernstein et al., 2010) and LCL by ENCODE (ENCODE_Consortium, 2012) with two different antibodies being used for ChIP-seq.

Due to the inconsistencies listed above, it seems hard to draw profound conclusions from comparison of absolute numbers, which do not include the relative value of a signal within a dataset with an unknown signal distribution and pattern across the genome. To account for the differences between the two experiments, a different approach was chosen here to investigate E2 chromatin accession. To this end not only H3K4me1 but also H3K4me3, enhancer and promoter mark, H3K27ac, active enhancer mark, and DNaseI hypersensitive sites (DNaseI HS), typical for open chromatin, derived from CD19+ primary B cells, conducted by Roadmaps

Epigenomics project (Bernstein et al., 2010) and from LCL GM12878 (ENCODE_Consortium, 2012) were used for comparison applying one further normalization step. For each ChIP-seq or DNase-seq experiment peaks were called applying the self-generated workflow (see chapter 3.6) and signal distributions at histone modification or DNaseI HS peaks were used as reference for a positive signal. This step allowed a relative comparison to signal intensities at positive sites and therefore a comparison between the two experiments is now possible (Fig. 18).

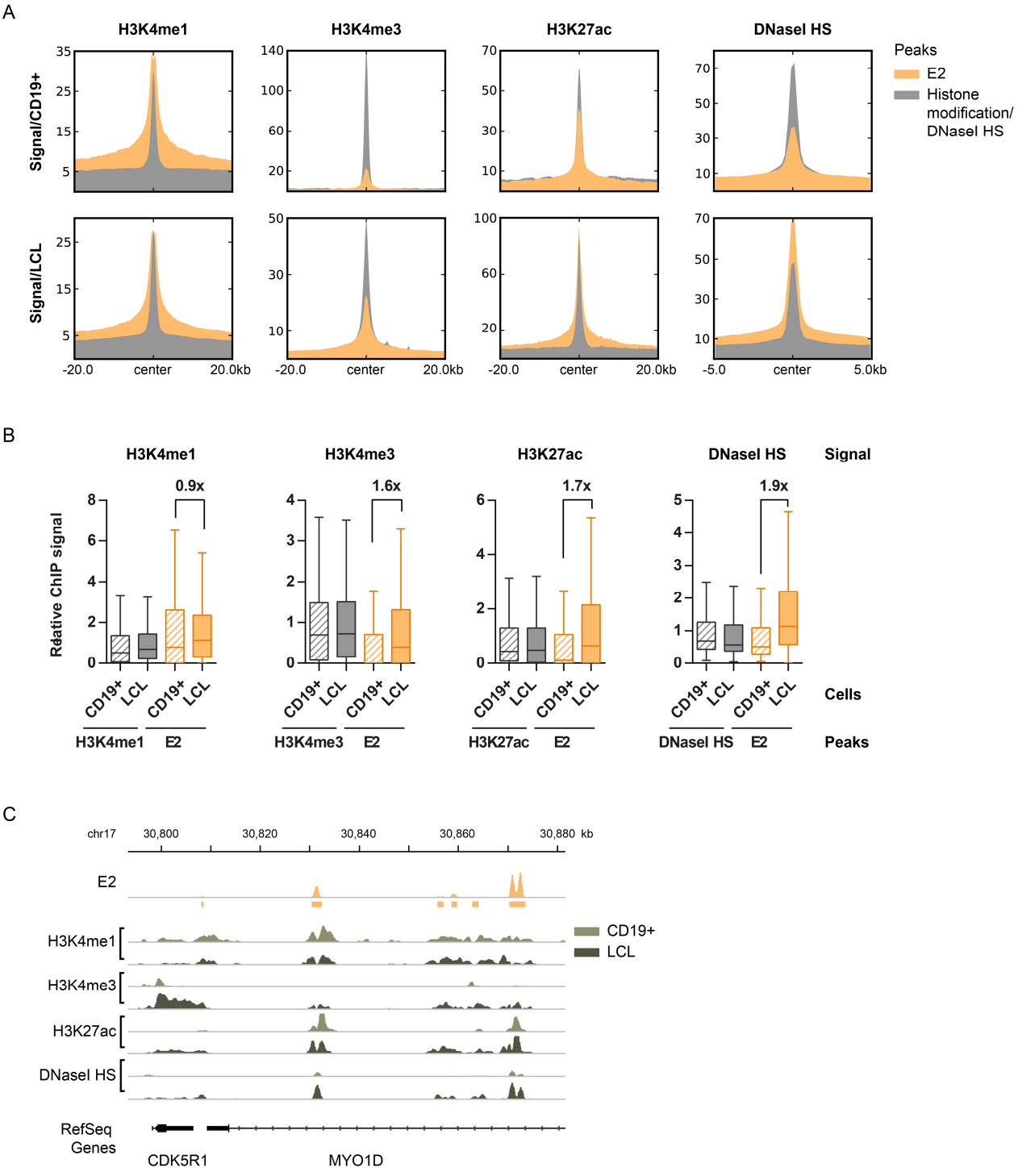


Figure 18. E2 binding sites already exhibit enhancer specific histone modifications in EBV negative B cells (CD19+) which increase in the presence of E2. ChIP-seq experiments for histone modifications and DNaseI HS performed in CD19+ primary B cells (Roadmap Epigenomics,(Bernstein et al., 2010)) and the LCL GM12878 (ENCODE_Consortium, 2012) were used to compare chromatin signature changes upon EBV infection and E2 expression. To this end raw data (already mapped reads) were analyzed applying own standard workflows using Galaxy platform and tools. (A) Anchor plots depicting the respective histone modification or DNaseI HS signal at E2 binding sites in CD19+ cells (upper panel) and LCL (lower panel). For comparison each time the peaks of the respective histone modification or DNaseI HS are shown as well, so average signal strength at peaks can be used as indirect reference. Regions of 20 kb, or 5kb for DNaseI HS, in each direction from peak center were analyzed. ChIP-seq signals of histone modifications were normalized to input and coverage, DNaseI HS signal only to coverage. (B) Analysis of mean ChIP-seq signal strength at E2 binding sites in CD19+ cells compared to LCL. Again the regions of 20 kb, or 5 kb for DNaseI HS sites, in both directions from peak centers were used. The mean signals at E2 binding sites were normalized to the mean signal at the respective histone modification or DNaseI HS peaks so

E2 signal derived from two different cells and experiments can be compared. Box plots whiskers extend to 1.5x the interquartile range and p-values were calculated using unpaired two-tailed t-test (all comparisons between average signal at positive sites and E2 binding sites and all comparisons of E2 peaks in different cell background: $p<0.0001$). The fold changes of the mean relative signals at E2 binding sites in LCL compared to CD19+ cells are shown. (C) Graphic representation of histone modification and DNaseI HS signals at the E2 binding site in proximity to CDK5R1 target gene. The position on hg19 is indicated on top. Called peaks for E2 are shown as yellow bars below the signal track. For all histone modification and DNaseI HS tracks the maximum of the scale is set to 50x the mean signal at called peaks for the respective signal. This allows the comparison of tracks derived from different cell lines.

This analysis showed that E2 binding sites in CD19+ primary B cells do indeed display H3K4me1 enrichment, also in comparison to mean H3K4me1 enrichment at H3K4me1 peaks. Relative comparison to data from LCL revealed even a very small decrease in H3K4me1 enrichment at E2 sites. Analysis of H3K4me3, H3K27ac, and DNaseI HS showed signal enrichment for all three marks at E2 binding sites in CD19+ cells and also a relative enrichment of all three marks compared to LCL. Since H3K4me3 is a histone modification associated with active transcription, H3K27ac is associated with activated enhancers, and DNaseI HS is associated with accessible chromatin, these findings imply that E2 is binding to preexisting enhancers in CD19+, which subsequently get activated. In summary, E2 is not introducing enhancer signatures upon chromatin binding, rather targets existing enhancers for an initial accession to chromatin and in turn activates them.

4.2.5 Distinct combinations of cellular TFs characterize E2 versus E3 predominated chromatin regions

As described in the introduction, a significant overlap of E2 and E3 target genes as well as E3A and E3C target genes could be observed comparing expression array data from our laboratory, which could indicate a competition for CBF1 binding to achieve chromatin accession and thereby operating the same target genes. Nevertheless, the majority of EBNA target genes are uniquely regulated by one EBNA protein and also studies from other laboratories tend to show a separate mode of action for E2 and E3 proteins.

4.2.5.1 Comparing genomic positions of significant EBNA binding sites revealed only moderate overlaps

Comparing the different EBNA binding sites identified in this study, including published CBF1 ChIP-seq data, for their positions in the human genome also shows moderate overlaps (Fig. 19).

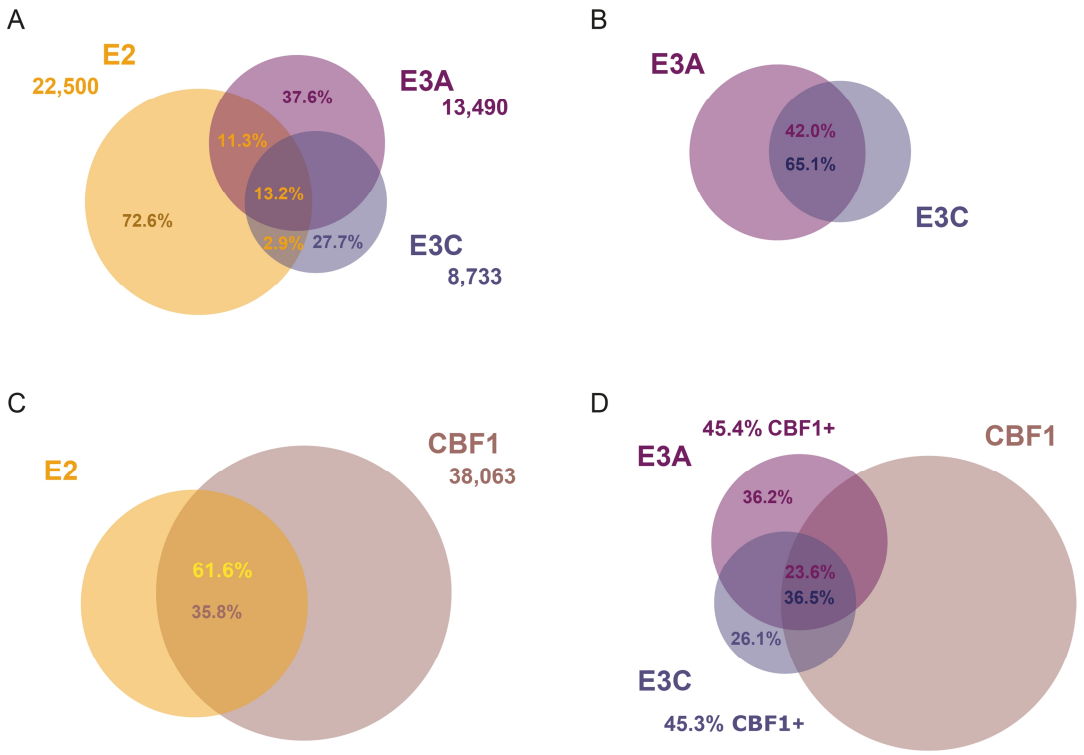


Figure 19. Binding site intersections for EBNA proteins and CBF1. Called peaks were extended by 50 bp in each direction before performing intersection analyses. Absolute numbers of called peaks are indicated as well as percentages for important subsets.

The intersection analyses revealed two important pieces of information. First, on the one hand, significant overlaps can be observed between E2 and both E3 peak sets (Fig. 19A) as well as between E3A and E3C (Fig. 19B), supporting the CBF1 competition model. Nevertheless, the majority of identified peaks are unique for one EBNA protein. Second, when looking at CBF1 co-occupation, E2 displays a far bigger overlap with CBF1 sites than both E3 proteins do (Fig. 19C and D), displaying a first hint for a more important role of CBF1 in recruitment of E2 to DNA than for E3 proteins.

4.2.5.2 Quantitative analysis of signal intensities at EBNA binding sites reveals significant positive correlation patterns

In this very simple comparison of binding site occupation, one important piece of information from ChIP-seq experiments is missing: A quantitative examination of binding sites. To achieve a more complete overview on E2 and E3 binding sites in the human genome, a different approach, including quantitative values was introduced. To this end EBNA binding sites were ordered by their own signal intensities and compared to the ChIP signals from other TFs (Fig. 20A-C). Here, a more complex picture emerged: E2 does not only show a larger overlap with CBF1 binding sites in comparison to E3, but also displays a correlation in CBF1 signal enrichment at E2 binding sites (Fig. 20A, column 2). However, CBF1 signal at E3 binding sites does not display

this kind of positive correlation but seems rather to be distributed randomly (Fig. 20B and C, columns 3). Interestingly, E3A and E3C exhibit a very prominent positive correlation of signal intensities at both, E3A and E3C peaks, indicating a more intense cooperation on DNA binding level than assumed by intersection analyses only. Also E2 and the E3s do not show a significant correlation of signal intensities at E2 (Fig. 20A, columns 3 and 4) or E3 binding sites (data not shown). So despite a moderate overlap in significant binding sites no correlation in signal intensities could be observed.

This phenomenon could be detected at various genomic regions when visualizing ChIP-seq data in a genome browser, e.g. at the well described genomic region encompassing *CXCL9* and -*10* (Fig. 20D), confirmed by ChIP-qPCR analysis (Fig. 20E). E2 and CBF1 show a very similar ChIP signal distribution at this region as well as E3A and E3C do. As already demonstrated on a genome wide scale by histone modification analysis at different EBNA peaks (Fig. 16), also the E2 enriched enhancer "E1" in this example exhibits higher H3K27ac signals as the E3 dominant enhancer regions "E4" and "E5". In fact, E2 and the two E3 proteins show reciprocal signal enrichment at bound enhancer regions within this analyzed region.

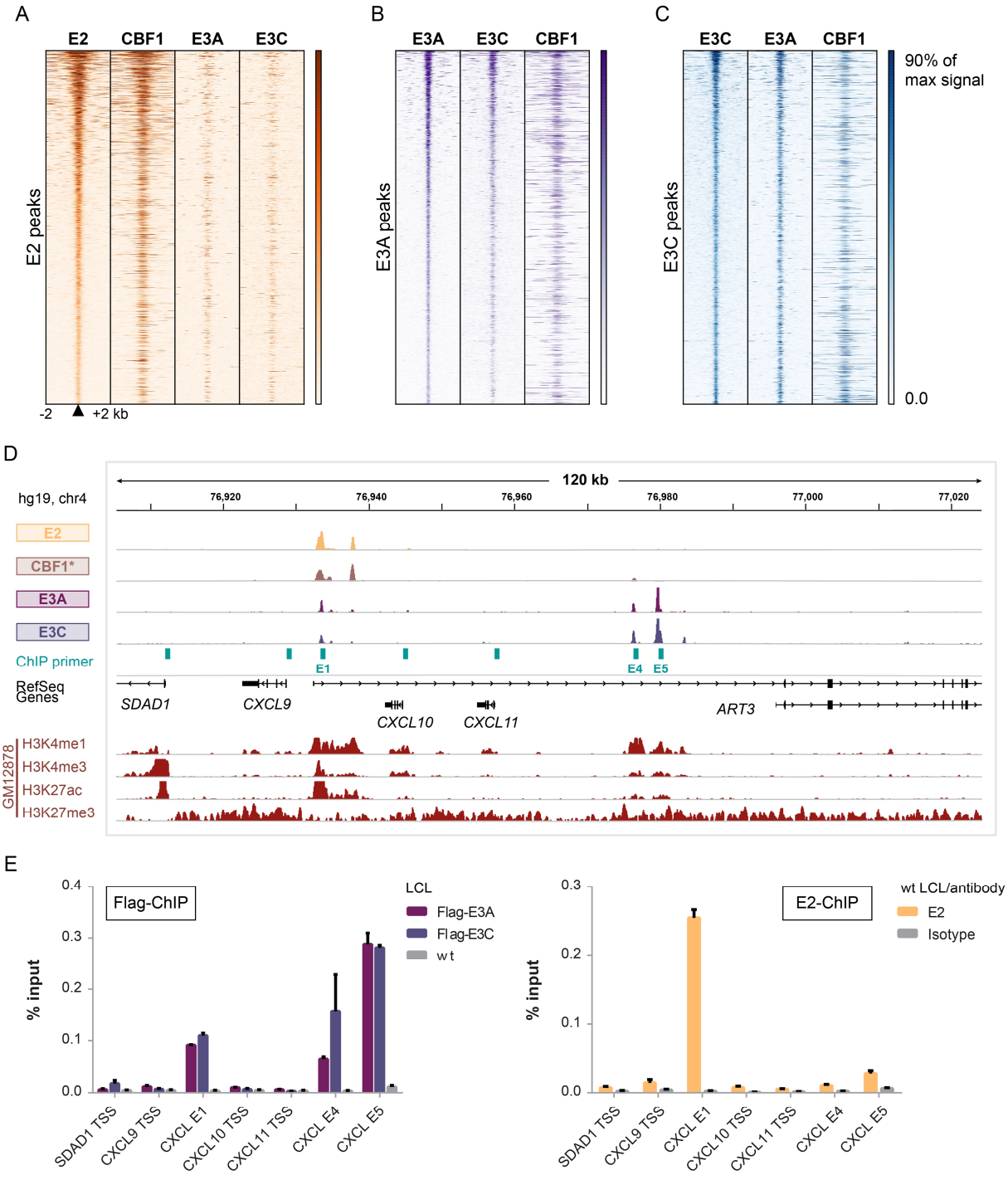


Figure 20. CBF1 signal positively correlates with E2 signal at E2 binding sites but not with E3 signals at E3 sites. (A) E2, (B) E3A, and (C) E3C peaks were sorted in a descending order according to their own mean normalized signal (first lane of each panel) and then compared to signal intensities of other TFs, listed on top, without changing order. (D) Graphic representation of E2, CBF1 (* raw data from Zhao et al., 2011), E3A, and E3C signals at a genomic region encompassing E2, E3A, and E3C target genes *CXCL9* and *-10*. The position on hg19 is indicated on top. Primers used for ChIP-qPCR are shown as turquois bars below the signal tracks and primers at enhancer regions are labeled. Annotated RefSeq genes are shown below. For all histone modification tracks the maximum of the scale is set to 10x the mean signal at called peaks for the respective signal. (E) ChIP-qPCR analysis using primers highlighted in (D). A Flag-ChIP was performed in Flag-E3A and -E3C LCLs as well as in the wt LCL derived from the same donor as a negative control. The E2-ChIP was performed in the wt LCL, derived from the same donor as used for the Flag-ChIP. Here, an isotype matched antibody was used as negative control. Means and SEM of two independent ChIP experiments with technical duplicates are shown.

4.2.5.3 A genome wide correlation analysis of transcription factor binding patterns reveals distinct sets of E2 and E3 associated factors

In this study a hypothesis was formed, which pictures subsets of EBNA regulated genes and a definition of those by the TF occupancies of their regulatory elements. In this picture CBF1 is not sufficient for subset determination rather than co-occurring factors and their combinations, which define such subsets. To include information on further TF binding in the most unbiased fashion, all TF ChIP-seq experiments performed in GM12878 by ENCODE at the time when this analysis was performed (chapter 3.6.4) were included in a genome wide approach to identify potential TFs important for EBNA accession of (specific) DNA elements.

To this end a genome wide correlation analysis was performed applying bamCorrelate, which is part of the deepTools package (Ramirez et al., 2014a), and signal distributions over the whole genome for all three EBNAs, CBF1, and 84 ENCODE TFs were included. BamCorrelate performs a one to one comparison for all possible combinations of submitted samples, where the genome is split in bins of distinct size and reads for each bin and sample are counted. Comparing reads/distinct bin for two samples a regression curve and correlation coefficient (here Spearman correlation, r_s) can be calculated indicating the degree of "similarity". An r_s of 1 displays perfect positive correlation, 0 no correlation at all, and -1 indicates perfect anti-correlation. A detailed information on this tool can be found at the developers github page (Ramirez et al., 2014b).

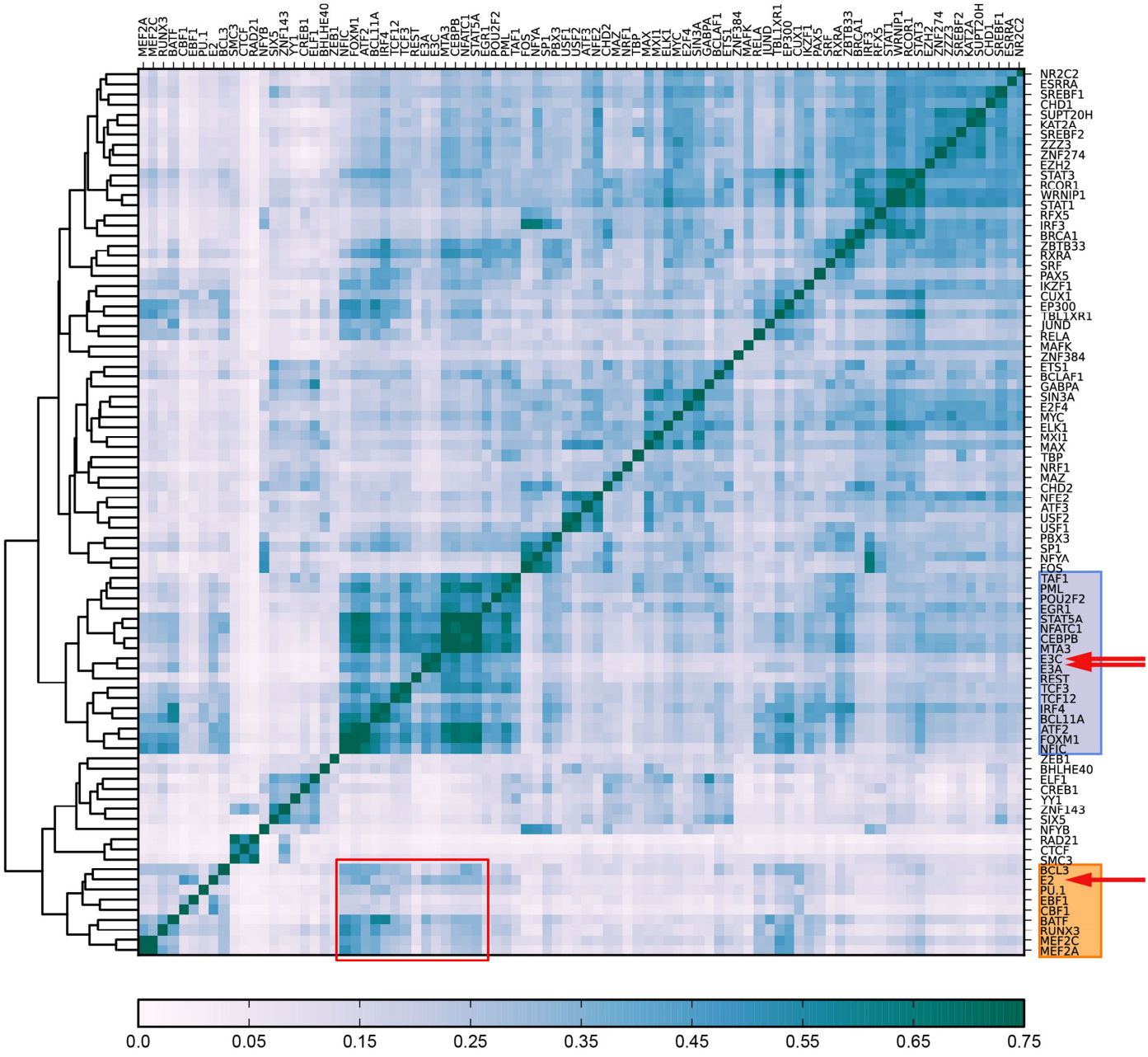


Figure 21. Genome wide correlation analysis reveals distinct clusters for E2 and E3 proteins. All 84 TFs ChIP-seq experiments available in GM12878 to date (ENCODE_Consortium, 2012) and the CBF1 ChIP-seq were compared with E2 and E3 proteins for their genome wide overall "similarity" in signal enrichment applying bamCorrelate. To this end the human genome (hg19) was split in 100 bp bins, reads of each sample for each bin were counted, one to one comparison of all samples looking at reads/distinct bin were performed, and then a correlation coefficient, applying Spearman correlation, for each possible sample pair combination was calculated. The matrix shows relations between single TFs as predicted by hierarchical clustering by bamCorrelate. The color intensities representing correlation coefficients are depicted below the matrix as color bar. E2 and E3 associated TF subsets are highlighted by colored frames; investigated EBNAs are highlighted by red arrows. The red frame highlights the correlation values between E2 and E3 subsets.

The resulting data matrix (Fig. 21) provides a huge amount of information on TFs clusters and specific combinations in LCL, which were not fully analyzed in this thesis. Here, the focus is mainly on TFs factors displaying a high positive correlation with the single EBNAs. Taking a first

look at the results, one feature attracts attention immediately: E2 and the two E3 proteins are located in two separate clusters, each associated with specific TFs.

Here, E2 was found in one cluster together with the described DNA adaptor CBF1, EBF1, PU.1, BCL3, BATF, and in a more distant branch of this cluster RUNX3, MEF2A, and MEF2C with assigned r_s in comparison to E2 of 0.50, 0.42, 0.17, 0.32, 0.20, 0.25, 0.16, and 0.16 respectively (Fig. 21, orange box). Taking a closer look at the single r_s values revealed that not all of the TFs show a high correlation to E2 but rather show a similar pattern in the genome wide comparison to the investigated TFs. The characteristics and quality of E2 and CBF1 binding on DNA was characterized in detail above (Fig. 20), where a strong positive correlation at identified E2 binding sites could be shown, indicating that an r_s of 0.50 in a genome wide comparison displays a solid positive correlation. Only the comparison to EBF1 reached a similar r_s of 0.42 within the E2 cluster.

E3A and E3C were placed in a different subset by bamCorrelate, which includes 16 further TFs besides the E3s (Fig. 21, blue box). Interestingly, the comparison of E3A and E3C shows a very high r_s of 0.70, the highest r_s each E3 reaches in comparison to all investigated TFs. Members of the cohesion complex SMC3 and RAD21, which are known to be recruited to DNA together, reached an r_s of 0.71 and are clustered together with CTCF, which interacts with cohesion on DNA to link regulatory elements with their targets (reviewed in Merkenschlager and Odom, 2013). Also TFs, which are known to act in heterodimers, such as BATF/IRF4 (Ravasi et al., 2010, Glasmacher et al., 2012) and MEF2A/MEF2C (Li et al., 2015) combinations, score r_s values in the range of E3A with E3C (0.60 and 0.78 respectively). This finding implies that E3A and E3C act very closely together on DNA and might even operate as a heterodimer. It has to be mentioned though, that some TF combinations, which are known to act as heterodimers, like AP-1 factors Jun and Fos only show a low r_s (0.06) in this genome wide correlation analysis. Several reasons could contribute to this finding: some heterodimer combinations are cell type specific and often distinct combinations only access a certain subset of binding sites detectable for the single components of such a heterodimer, which would result in lower r_s values on a genome wide level.

Both E3 proteins exhibit r_s values of at least 0.3 for all TFs within the E3 cluster, with TF12 as one exception (r_s of 0.26 and 0.25 for E3A and E3C respectively). In general, the E3 cluster forms the most prominent cluster within the genome wide matrix, where all factors display high rs values to each other. This phenomenon indicates the existence of an LCL specific TF network which is exploited by E3 proteins for chromatin accession.

The E2 cluster in contrast does not exhibit high r_s values for all TFs included but is rather formed by pattern similarities in the overall comparison. Interestingly, looking at the relationship

of the E2 and the E3 clusters to each other, a slight but noticeable positive correlation is detectable (Fig. 21, red frame) pointing out a relationship between the two clusters.

A drawback of the matrix and the hierarchical clustering is the formation of branches and clusters by pattern similarities, which sometimes results in separation of factors which display high r_s values per se but fall into patterns of different hierarchies. Of course this approach is very useful for identifying patterns also in big datasets, but in this special case the TFs with significant positive correlation to the EBNAs were of particular interest, more than the cluster of higher magnitude they are fitting in.

Thus the next analysis step only focused on TFs, which might by relevant for chromatin accession of the EBNA proteins. To this end TFs identified by the genome wide correlation analysis, which scored a high positive correlation (r_s > 0 35) with at least one of the investigated EBNA proteins were included in a second genome wide correlation analysis. Here, the same parameters were applied and already calculated r_s values did not chance, but the pattern discovery differed from the previous one, including all TFs examined by ENCODE. In this new matrix, only including preselected TFs (by r_s threshold) as slightly different pattern emerges (Fig. 22).

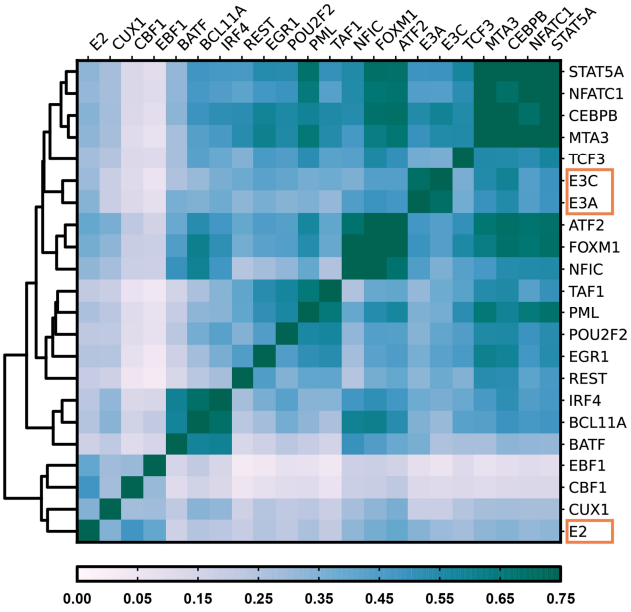


Figure 22. Genome wide correlation analysis of preselected TFs reveal two separate clusters for E2 and E3 proteins associated with distinct TFs. TFs with a genome wide rs > 0.35 in comparison to at least one investigated EBNA protein were used to generate a new matrix. BamCorrelate was used to calculate rs and perform hierarchical clustering applying the same standards as in Fig. 21. The color intensities representing correlation coefficients are depicted below the matrix as color bar. E2 and E3s are highlighted by colored frames.

Again two separate clusters for E2 and E3 proteins could be identified, which are associated with distinct sets TFs of a very similar composition as in the first analysis. E2 still clusters with CBF1 and EBF1, but newly emerged CUX1, which was previously added in a branch together with IKZF1 and p300. The comparison with the E3 cluster shows that CBF1 and EBF1 signals are exclusively correlating with E2 and CUX and exhibit almost no cross interaction with the E3 cluster.

The E3 cluster now consists of 16 TFs forming five subsets consisting of (i) STAT5, NFATC1, CEBPB, MTA3, and TCF3, (ii) E3A and E3C, (iii) ATF2, FOXM1, and NFIC, (iv) TAF1, PML, POU2F2, EGR1, and REST, as well as (v) IRF4, BCL11A, and BATF. Subclusters (i), (ii), (iii), and (v) show

very strong signal correlations, while subcluster (iv) displays weaker correlations.

The TFs now identified in a very unbiased genome wide approach to positively correlate with EBNA ChIP-seq signal intensities displayed a starting point for further characterization of the TF composition of specific subsets of binding sites and associated chromatin accession.

4.2.5.4 Anti-correlation of E2 and E3 signal intensities at combined binding sites

The correlation analysis described above gave a great overview on a genome wide interaction network of distinct TF sets associated with the EBNA proteins. To refine the resolution of this kind of analysis and to further concentrate on TFs associated with EBNA proteins only, a second more specific correlation analysis was performed. To this end, not the whole genome but only regions which are binding sites for at least one of the investigated EBNA proteins (*EBNA peaks*) were considered as reference regions and the TFs identified in the genome wide correlation analysis, which showed a good positive correlation were reanalyzed (Fig. 23).

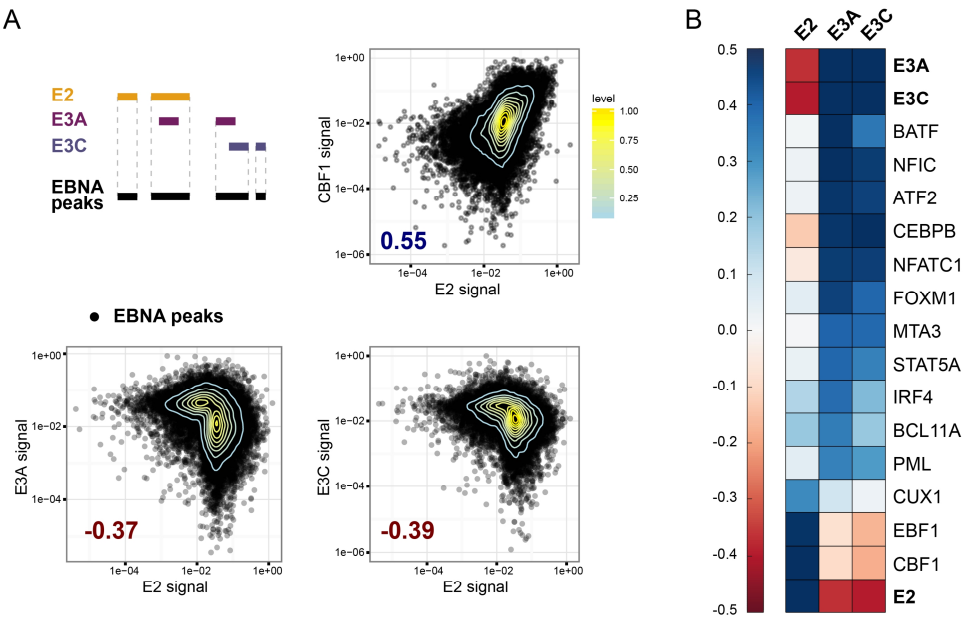


Figure 23. Correlation analysis of TF signal intensities at EBNA binding sites only revealed anti-correlation of E2 and E3 proteins and largely confirmed associated TF clusters. Correlation analysis for EBNA proteins, CBF1, and the 19 TFs identified in 4.2.5.3 was repeated (as performed in 4.2.5.3), but restricted to genomic sites bound by at least one EBNA protein (EBNA peaks). (A) Schematic representation of EBNA peak file generation. Called E2, E3A, and E3C peaks were merged (32,671 resulting regions) and subsequently used as reference regions for correlation analysis applying Spearman correlation. Scatter plots (plotted by Simone Rieger) depicting E2 versus CBF1, E3A, or E3C signals at EBNA peaks. Each dot represents one EBNA peak. Correlation coefficients r_s are indicated. (B) Preselected TFs which displayed an $r_s > 0.3$ in comparison to at least one investigated EBNA protein were used to generate a new matrix. The color intensities representing r_s values are depicted next to the matrix as color bar.

This analysis, restricted to the EBNA peaks confirmed the presence and composition of the two distinct clusters of TFs, one for E2, which is again strongly associated with CBF1 (r_s = 0.55) and

EBF1 (r_s = 0.49), and another cluster for E3A and E3C, which again show a very strong positive correlation (r_s = 0.62) to each other. TFs POU2F2, REST, EGR1, TCF3, and TAF1, which were identified in the genome wide approach, were not included in the generation of the new EBNA peak restricted correlation matrix, since they scored r_s values below the set threshold to all three investigated EBNA proteins. This could be due to a general presence of those TFs at EBNA peaks which is recognized as a significant correlation in the genome wide point of view (Fig. 21 and 22), but the signals do not positively correlate with either E2 or the E3 signals at the peak restricted analysis.

In general, on a genome wide scale, a inter-correlation between E2 and E3 clusters can be observed, since all TFs involved mainly bind to enhancer elements which only represent a very small fraction of the entire genome. In this broad context, the interconnection between the E2 and E3 cluster is quite significant (Fig. 21, red frame). But when the picture was narrowed down on only the fragments of interest, which are actually occupied by the EBNA proteins (EBNA peaks), a more specific picture emerges. Here, E2 and E3 proteins display an anti-correlation relationship which characterized by two distinct sets of TFs. The described E2 and E3 specific associated TF compositions could be observed frequently when visualizing the signal tracks in a genome browser and the E2 and E3 anti-correlation in signal intensities could be verified by ChIP-qPCR for several loci. For illustration two well described gene loci were chosen and the signals for TFs identified above were used for comparison (Fig. 24). *ADAM28* and *ADAMDEC1* are two described target genes known to be repressed by E3A and E3C (Hertle et al., 2009, McClellan et al., 2012, our unpublished data from E3C ko LCL gene expression array) and induced by E2 action (unpublished data by Sybille Thumann from gene expression array and conditional expression of E2 in DG75 B cell line). Here, the two described correlation trends are visible: CBF1, EBF1, and CUX signal distribution is very similar to E2 signal intensities, while the TFs derived from the E3 cluster follow the E3 signal intensity distribution at inter- and intragenic enhancers E1 and E2. E3A and E3C, as well as their correlating TFs, each show the higher signal at E1 and a lower enrichment at E2. For E2 and its associated TFs it is the other round; E2 is the enhancer with the higher signal enrichment (Fig. 24A). This phenomenon could also be observed at enhancers in proximity to E2 and E3 target genes *CXCL9* and *-10* (described above in more detail): E2 and associated TFs show high enrichments at enhancer E1, but only low enrichment at enhancers E4 and E5, which are significant E2 binding sites as well. E3 proteins and the TFs from the E3 cluster showed more prominent enrichment at the distal enhancers E4 and E5 over intergenic E1 (Fig. 24D).

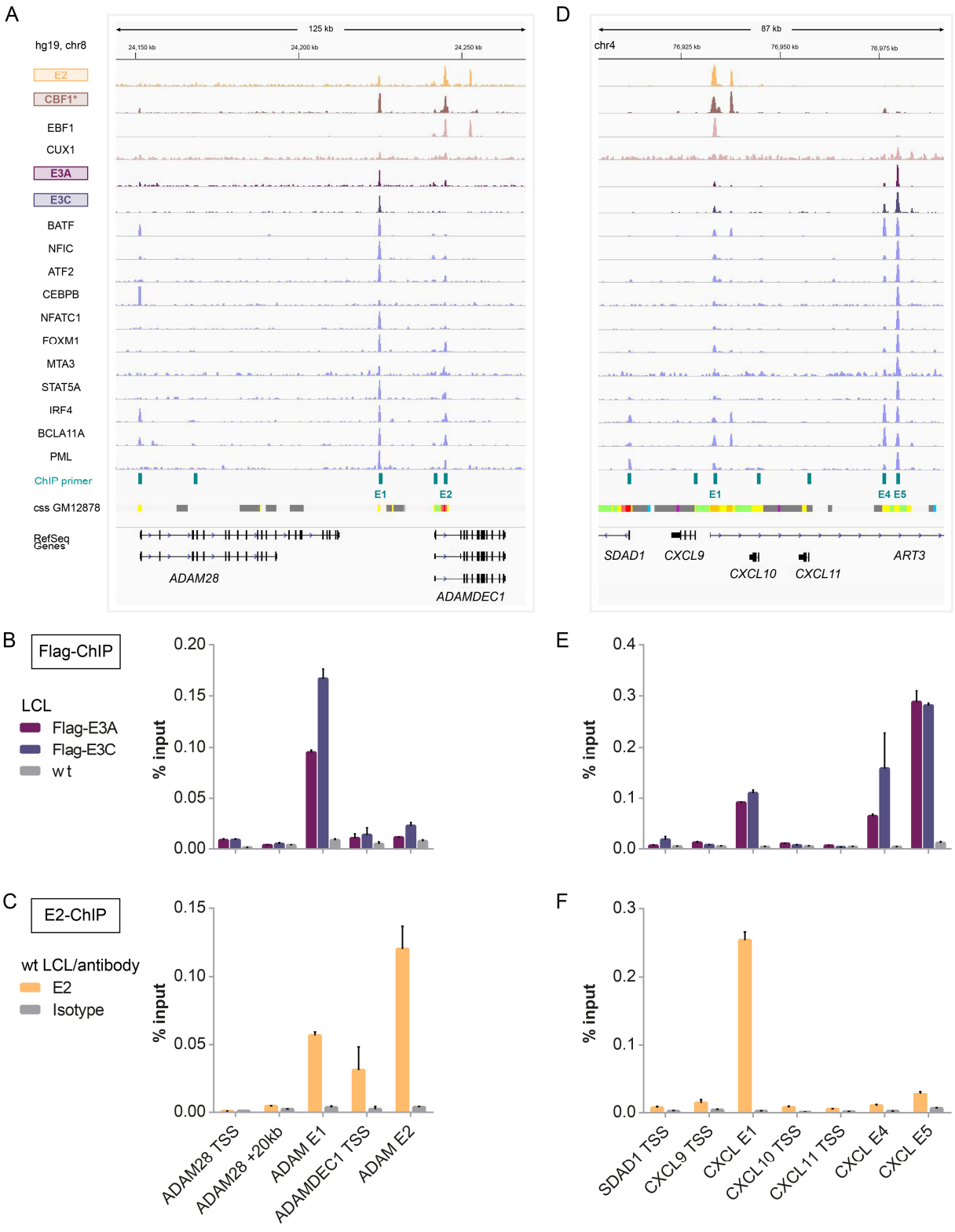


Figure 24. E2 and E3 specific associated TF sets identified in correlation analyses show reciprocal binding patterns at EBNA peaks at two model loci. (A, D) Graphic representation of E2, CBF1 (* raw data from Zhao et al., 2011), E3A, and E3C signals as well as signals of the 13 TFs identified in 4.2.5.4 at two genomic model loci. The position on hg19 is indicated on top. Primers used for ChIP-qPCR are shown as turquois bars below the signal tracks and primers at enhancer regions are labeled. Annotated RefSeq genes are shown below. For all ChIP-seq tracks the scale was set to the local maximum of the depicted regions. (B, E) Flag and (C, F) E2 ChIP-qPCR analysis using primers highlighted in (D). The Flag-ChIP was performed in Flag-E3A and -E3C LCLs as well as in the wt LCL derived from the same donor as a negative control. The E2-ChIP was performed in the wt LCL, derived from the same donor as the LCL applied for the Flag-ChIPs. Here, an isotype matched antibody was used as negative control. Means and SEM of two independent ChIP experiments with technical duplicates are shown.

4.2.5.5 Characterization of EBNA binding sites by cluster analyses including preselected TFs

The performed correlation analyses described above revealed TFs which are positively correlating in signal intensities with the investigated EBNA proteins and therefore might be contributing to recognition and specificity of EBNA binding site subsets. In order to stress the idea that subsets exist, which are defined by their TF composition, clusters of TF combinations were searched at EBNA binding sites. Together with Björn Grüning (University of Freiburg), intersection analyses of called peaks of TFs, identified in the correlation analysis 4.2.5.3, at the EBNA peaks were performed and Jaccard similarity coefficients were calculated to compare similarity and diversity of the included sample sets. Hence, a cluster analysis was performed describing the relationship of the selected TFs at EBNA peaks, but also the EBNA peaks were sorted according to the identified clusters (Fig. 25). This analysis does not include quantitative measurements like the correlation analysis, where signal intensities were used, but is only based on called peaks and therefore is able to sort the EBNA peaks according to their TF occupancies and identified clusters.

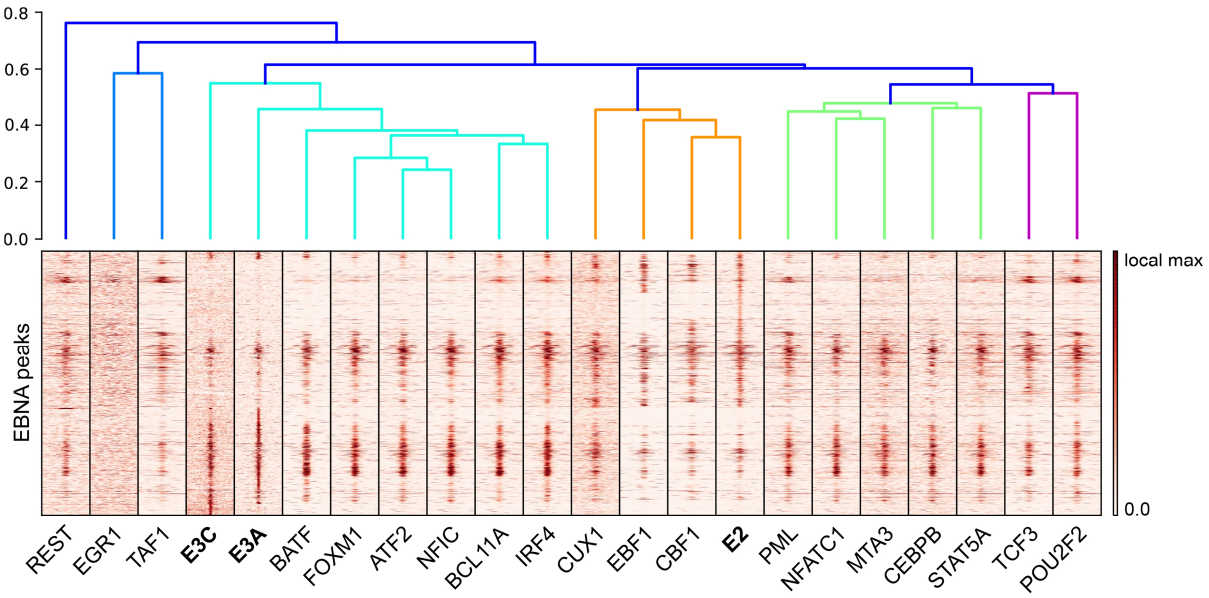


Figure 25. Cluster analysis at EBNA peaks identified hierarchies of associated TFs at EBNA peaks. TFs identified to positively correlate (r_s > 0.35) with at least one EBNA protein in the genome wide correlation analysis comparing TF signal intensities were included in a new cluster analysis in order to sort EBNA peaks according to compositions of associated TFs. To this end, the EBNA peaks were used as reference regions for an intersection analysis creating a matrix which depicts hits for each selected TFs (using peaks identified by ENCODE, see Table S1) at every EBNA peak. The resulting matrix was used as template for cluster search applying Jaccard similarity correlation index (performed by Björn Grüning). The resulting identified relations between the investigated TFs are depicted in the dendrogram in the upper panel. The EBNA peaks were sorted according to the identified TF clusters, and heatmaps for each TF were generated. Sorted EBNA peaks were centered and genomic regions of 2 kb in each direction from peak center are show. The scale of each heatmap was set to depict the whole range of detected signal at the investigated EBNA peaks.

Here, the 19 TFs which were already identified in the genome wide correlation analysis were used to generate this cluster analysis, where not only TFs are clustered according to their relation at EBNA peaks but also the respective EBNA peaks were sorted.

Interestingly, TFs REST, EGR1, and TAF1 were placed at the very outer level of relationship towards the other factors including the EBNA proteins, which could indicate that they play a more general role at EBNA binding sites, rather than defining E2 or E3 binding site subsets. TCF3 and POU2F2 hierarchically fall into the E2 sub-branch, but are located to the outer level there.

The E3 sub-branch includes TFs BATF, FOXM1, ATF2, NFIC, BCL11A, and IRF4, while the other TFs which were previously sorted in the E3 cluster, now are predicted to have a more close relationship to E2.

E2 builds the strongest cluster with CBF1, EBF1, and CUX1. This E2 core-branch is identical with the TF composition of the E2 cluster in the genome wide signal correlation analysis, as well as with the one from the EBNA peak restricted correlation analysis. Since both approaches, signal correlation and plain peak intersection analysis, identify the same TFs to be strongly associated with E2, they were chosen for further investigations.

The heatmaps generated for the sorted EBNA peaks and the investigated TF signals show a prominent pattern, shared by factors within one sub-branch, but no sharp clusters of peaks defined by TF occupation could be identified. This finding is probably due to the relatively large set of investigated TFs, including some, which show rather an overall EBNA peak binding than preference for distinct subsets. To get a better understanding of the TF composition of E2 versus E3 binding sites, both peak sets were investigated separately using the TFs identified in this approach.

4.2.5.5.1 Cluster analyses for E2 binding sites reveal subsets defined by combinatorial TF sets

The TFs identified in the cluster analysis of TFs present at EBNA peaks (Fig. 25) to be closely related to E2, were now used to cluster E2 peaks according to their TF compositions. Again an intersection analysis was performed, this time using E2 peaks as reference regions, where each region was evaluated for intersection with CBF1, EBF1, and CUX1. The resulting matrix was used as template for cluster search and eight (out of 16 possible) distinct clusters of E2 peaks could be identified which are characterized by specific combinations of the investigated TFs (Fig. 26A). The identified clusters of E2 peaks do not only show distinct compositions of the investigated TFs, but also show different signal intensities for those within each cluster (Fig. 26B).

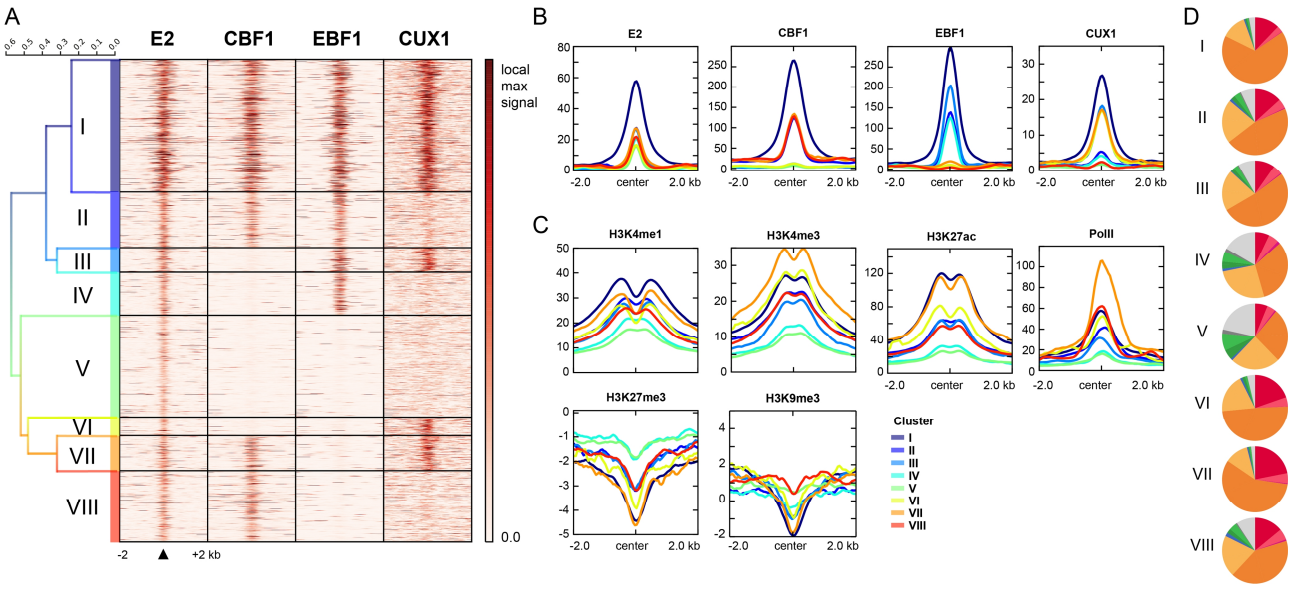


Figure 26. Cluster analysis for E2 peaks identified eight distinct clusters of TF combinations which are associated with different histone modifications. TFs identified to cluster with E2 in the EBNA peak wide TF cluster analysis described in 4.2.5.5 were used to generate a new cluster analysis in order to sort E2 peaks according to compositions of associated TFs. To this end, the E2 peaks were used as reference regions for an intersection analysis creating a matrix which depicts hits for each selected TFs (using peaks identified by ENCODE, see Table S1) at every E2 peak. The resulting matrix was used as template for cluster search applying Jaccard similarity correlation index (performed by Björn Grüning). (A) The E2 peaks were sorted according to the eight identified TF clusters and heatmaps for each TF were generated. Sorted E2 peaks were centered and genomic regions of 2 kb in each direction from peak center are show. The scale of each heatmap was set to the maximum signal detected at an E2 peak. Anchor plots depicting mean signal distributions of (B) E2 and the three cluster determining TFs as well as (C) histone modifications and PolII at the different E2 peak clusters. As in (A) a region of 2 kb in each direction of the peak center was analyzed. ChIP-seq signals from ENCODE were normalized to their respective input samples and RPKM (see chapter 3.6.1). (D) E2 peaks of the eight different clusters were analyzed for their location on functional chromatin elements as determined by ENCODE css. Centers of E2 peaks were used to assign chromatin states.

Cluster I displays by far the highest signal intensities for E2 and all three defining TFs. The remaining clusters are either positive or negative for each defining TF, but the mean signal intensities for those do not differ to the extent observed for cluster I. Interestingly, the clusters do not only differ in their TF composition, but also they show different chromatin signatures as defined by histone modifications (Fig. 26C and Fig. S2). Three clusters emerge to be of special interest for further studies of functionality:

Cluster I, positive for all three investigated TFs, shows the strongest enrichment for H3K4me1 as well as for H3K27ac, indicating an association of these binding sites with active enhancers. Consistent with this finding, the majority of E2 peaks from this cluster are located at strong enhancers (66.7%) according to css by ENCODE in GM12878 (Fig. 26D). Hence, this cluster shows the strongest enrichment of strong enhancer associated peaks of all identified clusters.

Cluster VII, characterized by co-occupation of E2 binding sites with CBF1 and CUX1 but no EBF1 binding, exhibits the highest enrichment for H3K4me3 and PolII signals, which are associated with transcribed promoters. Also the css analysis of E2 peaks within this cluster shows

an overrepresentation of promoter regions of 27.3% (Fig. 26D) of which 10.3% are located within 1 kb upstream of a Refseq gene (data not shown). This subset of E2 peaks could therefore display a TF composition which determines E2 accessible promoter regions. Furthermore, E2 binding sites of this cluster are also enriched for H2AFZ, a histone variant associated with active or poised promoters (Ku et al., 2012) and H3K79me2, a histone modification associated with active transcription as well as DNA repair (reviewed in Nguyen and Zhang, 2011) supporting the connection to active promoter regions (Fig. S2).

Cluster V, which is comprised of E2 binding sites with no significant peaks for CBF1, EBF1, nor CUX1, shows the lowest E2 signal as well as the lowest enrichment of all investigated histone modifications associated with transcriptional or positive regulatory activity out of all clusters. However, cluster V displays the weakest depletion for H3K27me3, a histone modification associated with repression of transcription by polycomb group proteins (PcG) (Fig. 26C) and also H3K9me, associated with repression of transcription, is not as much locally depleted as for the other clusters. All other investigated histone modifications show the lowest enrichment for this very E2 peak cluster (Fig. S2).

The majority of the remaining TFs investigated by ENCODE, not included in this cluster formation, show the strongest signal enrichments at E2 peaks of cluster I or VII compared to the other E2 peak clusters (Fig. S3 and S4). No TF emerged to be as strongly enriched for one of the other clusters. Strikingly, the vast majority of TFs showed the lowest signals for cluster V and IV.

A de novo motif search for E2 binding sites of the eight different clusters was performed as well, to scan for further factors which might contribute to their specificity and to identify probable determining DNA sequences (Fig. 27). Interestingly, EBF1 showed up as the most significantly enriched motif of E2 peaks derived from cluster I to IV, all four clusters out of eight which are actually positive for EBF1 binding in ChIP-seq (Fig. 26A). CBF1 on the other hand, which only displays significant binding at E2 sites of cluster I, II, VII and VIII, can be found as an enriched motif for all eight clusters.

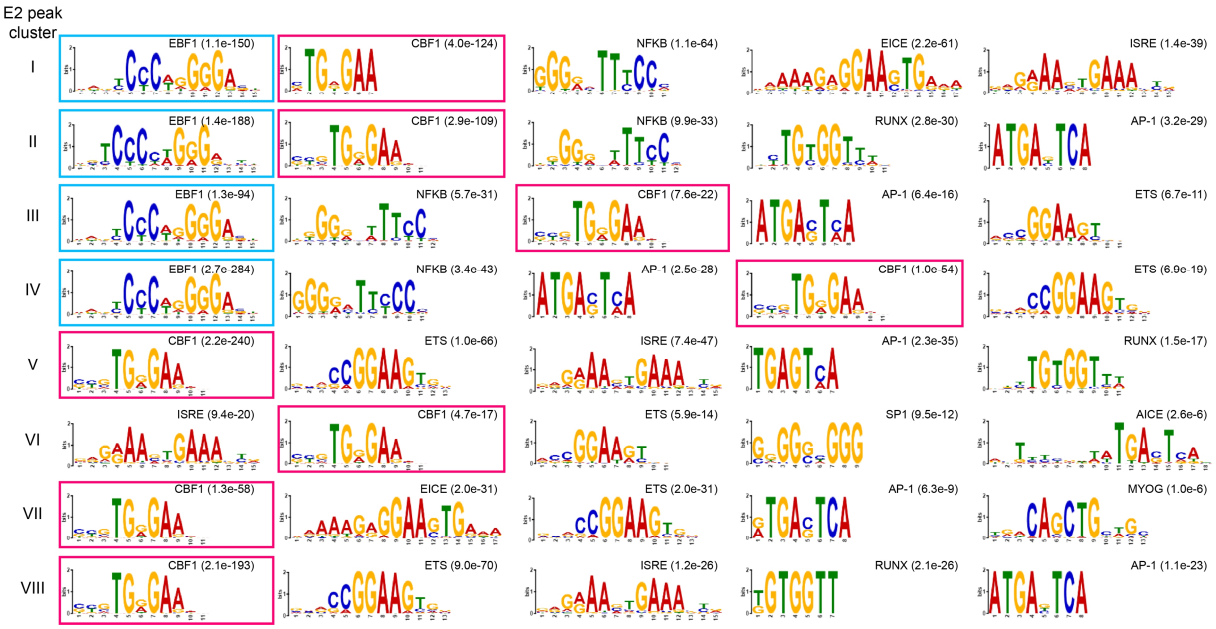


Figure 27. Cluster of E2 binding sites are characterized by specific compositions of enriched DNA motifs. The DNA sequences of E2 binding sites (500 bp in each direction from peak center) for the eight different clusters of TF composition (as identified in 4.2.5.5.1) were subjected to de novo motif search using MEME-ChIP analysis tool (Machanick and Bailey, 2011). Depicted are the top five enriched de novo identified motifs (E-values for significance of enrichment within input dataset are shown in brackets). The TF (or TF family) most likely to recognize the identified motifs are indicated and were predicted by TOMTOM motif comparison tool (Gupta et al., 2007) (scanning hocomoco v9 database). EBF1 and CBF1 motifs are highlighted by blue and pink frames, respectively.

However, the E-value for significance of enrichment and the ranking of CBF1 motifs varies between the different clusters, while EBF1 motif, if enriched, displays the top candidate for clusters I to IV. For clusters V to VIII, lacking EBF1 binding and motif enrichment, CBF1 motif takes up top ranking positions.

For all eight clusters similar motifs show up in this de novo motif enrichment search, highlighting the importance for TFs of e.g. the AP-1, ETS, NF$\varkappa$B, RUNX, and IRF families to chromatin accessibility and determination of E2 binding. Yet, the respective motifs emerge in different combinations and rankings for the different clusters and some patterns can be recognized.

E2 binding sites of clusters I to IV, characterized by EBF binding and motif enrichment, are enriched for the NFkB motif as well, while the EBF1 negative clusters do not show this characteristic, pointing at a potential role for TFs of the NFkB family in E2 binding at those sites.

The sequence motif recognized by members of the interferon regulatory factors (IRFs) family could be discovered within peaks of clusters I, V, VI, and VII and even displays the most significantly enriched motif for cluster VI. Also the recognition motifs for TFs of the ETS family could be identified in all clusters but I and II.

Interestingly, the ETS and ISRE composition motif (EICE) was also enriched for binding sites of clusters I and VII, the strong enhancer and promoter clusters respectively. EICEs are

recognition sites for ETS factors as PU.1 and SpiB, both being expressed in B cells, which recruit IRF4 or IRF8 to DNA and are very well described (Brass et al., 1996, Brass et al., 1999, Eisenbeis et al., 1995).

Also the consensus recognition motif for TFs of the AP-1 family could be observed in all clusters but clusters I and VI in the motif enrichment analysis. This superfamily includes TFs of the c-Fos, c-Jun, ATF, and JDP families of TFs and many of them are included in the E3 cluster of TFs as identified by genome wide signal pattern correlation analyses (Fig. 21). Here, the E2 cluster showed a slight positive correlation towards the E3 cluster (Fig. 21, red frame). But, the AP-1 and ISRE composition motif (AICE), which can be recognized by AP-1/IRF complexes (Glasmacher et al., 2012) is enriched for E2 binding sites of cluster VI, leaving cluster I the only E2 peak cluster without top enriched AP-1 motifs. Since cluster I represents the E2 peaks with the highest signal intensities, it is most likely that AP-1 TFs are not necessary are prerequisite for E2 accession to chromatin, but could facilitate binding in scenarios missing one or more important factors, like CBF1 or EBF1.

Furthermore, consensus motifs for TFs SP1 (cluster VI), RUNX family (cluster II, V, and VIII), and MYOG (cluster VII) could be observed among to the top enriched motifs.

Noticeably, the consensus motif of CUX1 could not be observed to be enriched in any of the analyzed E2 binding site clusters. As a control analysis CUX1 binding sites, as determined by ENCODE (experiment: ENCSR000DYR, peaks: ENCFF001VDY, peak count: 40,246), were subjected to de novo motif search using MEME-ChIP applying the same parameters as for E2 peak analyses (data not shown). Here, the CUX1 consensus motif (vbRvndATYRRTbb, TRANSFAC20112:136) was not among the top 5 enriched motifs as well but only identified as the sixth most enriched motif (E-value: 1.0e-70), while ETS (8.7e-388), RUNX (1.6e-158), ISRE (2.1e-151), and MEF (5.9e-101) motifs were more significantly enriched. In Summary, eight clusters of E2 binding sites characterized by distinct TF binding events and motif occurrences and associated with different chromatin marks could be identified. These clusters can now contribute to further elucidate the determining prerequisites for E2 accession to chromatin.

4.2.5.5.2 Cluster analyses for E3 binding sites reveal subsets defined by combinatorial TF sets

The combinatorial approach as performed for E2 binding sites was applied to E3A and E3C peaks as well. To this end, TFs which could be identified in the cluster analysis of TFs present at EBNA peaks (Fig. 25) and were clustered within the “E3 sub-branch”, were now used to cluster E3A and E3C peaks according to their TF compositions. Also in this setting an intersection analysis was performed, this time using *E3 peaks* (merge of E3A and E3C peaks) as reference

regions, where each region was evaluated for intersection with BATF, ATF2, BCL11A, FOXM1, NFIC and IRF4. The resulting matrix was used as template for cluster search which could identify eight clusters (out of 256 possible) of E3 peaks with different TF combinations (Fig. 28).

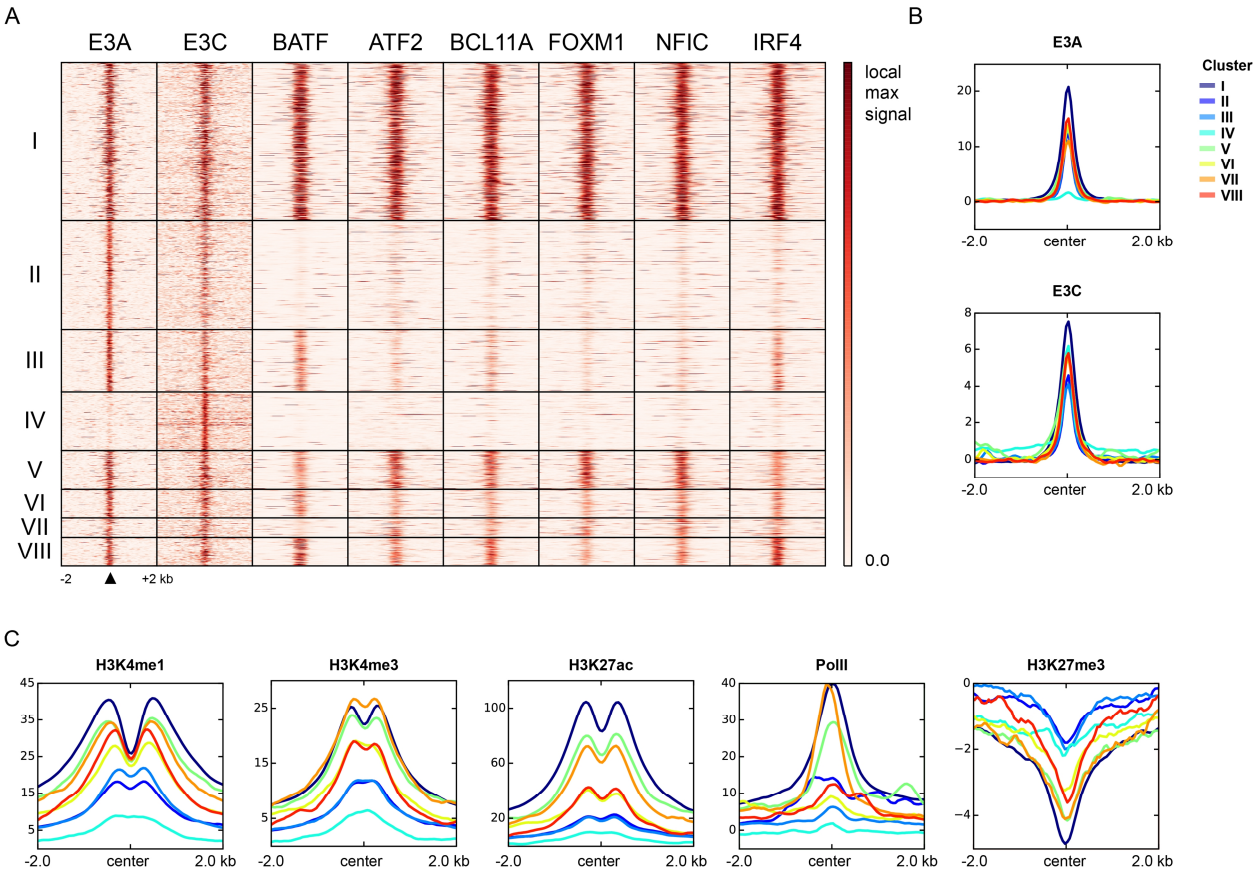


Figure 28. Cluster analysis for E3 peaks identified several sub-clusters of TF combinations which are associated with different histone modifications. TFs identified to cluster with E3A and E3C in the EBNA peak wide TF cluster analysis described in 4.2.5.5 (Fig. 25) were used to generate a new cluster analysis in order to sort *E3 peaks* (merge of E3A and E3C peaks) according to compositions of associated TFs. To this end, the E3 peaks were used as reference regions for an intersection analysis creating a matrix which depicts hits for each selected TFs (using peaks identified by ENCODE, see Table S1) at every E3 peak. The resulting matrix was used as template for k-means clustering. (A) The E3 peaks were sorted according to the eight identified TF clusters, and heatmaps for each TF were generated. Sorted E3 peaks were centered and genomic regions of 2 kb in each direction from peak center are show. The scale of each heatmap was set to the maximum signal detected at an E3 peak. Anchor plots depicting mean signal distributions of (B) E3A, E3C, and the six cluster determining TFs as well as (C) histone modifications and PolII at the different E3 peak clusters. As in (A) a region of 2 kb in each direction of the peak center was analyzed. ChIP-seq signals from ENCODE were normalized to their respective input samples and RPKM.

E3 peak cluster I is positive for both, E3A and E3C, as well as all six investigated TFs, which show the highest enrichment for this cluster. Cluster III depicts E3A and E3C shared binding sites, with higher E3A enrichment, and associated TFs BATF and IRF4. Clusters V, to VIII appear to be very similar but display distinct combinations of E3A and E3C associated factors: Cluster V is depleted for IRF4 signal, cluster VI shows significant enrichment for BATF and IRF4 but shows random occurrence of the other TFs, cluster VII shows at depletion for BATF and IRF4, and cluster VIII is depleted for FOXM1 signal. Also an E3A specific cluster, with only

very low E3C enrichment and depletion for all investigated TFs (cluster II) as well as an E3C specific cluster, with no other associated factors (cluster IV) could be detected.

Cluster I represents the strongest enhancers in this analysis, since it shows the highest enrichment for active enhancer associated chromatin marks H3K4me1, H3K27ac, and even PolII. Cluster V and VII are very similar to cluster I in the presence of enhancer specific histone modifications, but less enriched. However, it has to be pointed out that the absolute signal enrichment for PolII at these three clusters is much less pronounced than for E2 cluster VII and is more comparable to the slight enrichment at the other E2 clusters. Clusters VI and VIII display enhancer signatures as well, but not as highly enriched and without strong H3K27ac or PolII marks characteristic for weak enhancers. Clusters II and III exhibit the lowest enrichment of all investigated histone modifications but repressive H3K27me3, which is representative for PcG mediated repression, while cluster IV is almost devoid of any histone modifications, which is distinctive for heterochromatin.

Therefore, the eight identified clusters of E3 peaks reveal a combinatorial TF co-occupation of E3A and E3C binding sites, characterized by different histone modification patterns, and display a great basis for further functional analyses.

4.3 CBF1 as a determining factor for E2 access to chromatin?

The interaction of E2 and CBF1 has been studied extensively and the interaction of CBF1 and E2 together with DNA could be demonstrated in different approaches (Grossman et al., 1994, Henkel et al., 1994, reviewed in Hayward et al., 2006). A more recent study could show that also on a genome wide level E2 and CBF1 binding sites are significantly overlapping (Zhao et al., 2011b) and also the findings presented in this thesis show a strong positive correlation of E2 and CBF1, not only in binding site occupation but also in signal intensity.

Furthermore, in the context of viral gene regulation, another cellular factor, PU.1 (or Spi-1) was described to be important for E2 driven activation of the LMP1 promoter (Laux et al., 1994b, Laux et al., 1994a, Johannsen et al., 1995). Nevertheless, the interaction of E2 and PU.1 was only reported once (Yue et al., 2004) in a co-immunoprecipitation (Co-IP) experiment applying whole cell lysates which allows no conclusion on a direct interaction. Also the different correlation analyses of TF binding intensities performed in this thesis did not identify PU.1 to correlate with E2 binding patterns.

The dependency of E2 on CBF1 in its function as transcriptional activator has long been subject to studies in our laboratory. To assess the whole extend of CBF1s contribution to E2 mediated gene regulation on a genome wide level, Sybille Thumann of our group, performed gene expression analyses, using GeneChip® Human Gene 2.0 ST (Affymetrix) chip arrays, comparing CBF1 wt and ko cell lines (manuscript in preparation). The utilized DG75 cell lines, with CBF1 wt or ko genetic background, expressing E2 fused to the hormone binding domain of the Estrogen receptor (ER) (in the following ER/E2) have been published (Maier et al., 2005). Here, E2 activity can be induced by addition of Estrogen to the cell culture media. This study revealed a total of 136 at least 4-fold ($p<0.001$) E2 regulated transcripts in the $DG75^{ER/E2}$/CBF1 wt cell line. Interestingly, also in the CBF1 ko situation 21 E2 regulated ($\geq$ 4x, $p<0.001$) transcripts could be identified. The majority of CBF1 independently regulated transcripts are regulated by E2 in the wt background as well.

These findings were leading to the questions of how E2 mediates gene regulation and how it gains access to chromatin in the absence of the cellular anchor protein CBF1. To this end, further ChIP-seq studies for E2 binding in DG75 cell lines, with CBF1 wt and ko background, were performed as part of this work.

4.3.1 DG75 cell lines inducibly expressing HA-tagged E2 as a model system

In pursuance of studying E2 binding to DNA in the absence of CBF1 expression, the DG75 cell line harboring a somatic knock-out for CBF1 constructed in our laboratory (Maier et al., 2005)

was again the system of choice. Since the precipitation of E2 in the DG75 cellular background was rather inefficient using standard antibodies (Master thesis Jasmin Schwarz, 2014 and experiments by Sybille Thumann) an HA-tagged E2 was introduced in the pRTR vector (Fig. 29A) which was subsequently transfected in DG75/CBF1 wt and ko parental cell lines respectively (conducted by Cornelia Kuklik-Roos). E2 inducibility of the obtained cell lines was monitored by FACS analysis and cell line integrity was confirmed by western blot experiments (Fig. 29B and C).

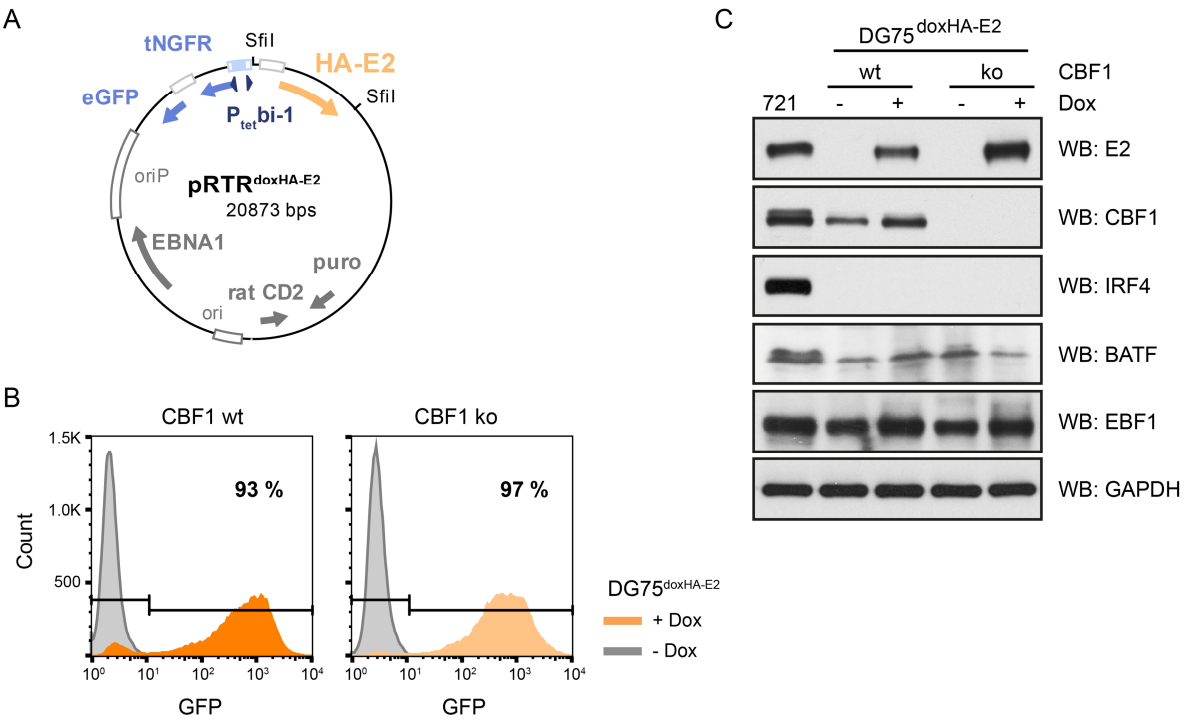


Figure 29. Stable DG75 EBV negative B cell lines proficient or deficient for CBF1 conditionally express HA-E2. (A) Simplified schematic map of the pRTR$^{doxHA-E2}$ vector used to generate stable DG75 cell lines. The coding sequence for E2 fused to a N-terminal HA-tag (HA-E2) with a preceding intron of the beta-globin gene for enhanced expression was cloned into the pRTR vector (Jackstadt et al., 2013, Bornkamm et al., 2005) using SfiI restriction sites. The bidirectional promoter (P_{tet}bi-1) simultaneously drives the expression of HA-E2 in one and the bicistronic reporter construct of a truncated nerve growth factor receptor gene (tNGFR) and enhanced green fluorescent protein (eGFP) gene in the other direction upon doxycycline induction. A truncated CD2 gene from rat which is constitutively expressed from SV40 promoter allows further selection of transfected cells. (B) Expression of HA-E2 was induced with 1 µg/ml doxycycline for 24 h and monitored by quantifying eGFP expression via flow cytometry and scored at least 89% with a maximum of 5% difference between DG75/CBF1 wt and ko. Data from one representative experiment (n=3) and percentages of induced cells are shown. (C) Western Blot analysis confirming the expression of HA-E2 in DG75$^{doxHA-E2}$ cell lines upon 24 h induction with 1 µg/ml doxycycline and the absence of CBF1 expression in DG75$^{doxHA-E2}$/CBF1 ko cell line. GAPDH was used as internal loading control. EBV positive LCL 721 lysate serves as a positive control for protein expression levels. Same amount of total protein lysate was loaded for each sample but for the E2 blot. Here a 1:10 dilution of DG75 lysates compared to 721 was loaded due to high E2 expression levels.

E2 expression could only be detected upon addition of doxycycline, in FACS analyses monitoring the surrogate marker GFP as well as in western blot experiments, confirming a reliable expression system. E2 expression levels did not noticeably differ between DG75$^{doxHA-E2}$/CBF1 wt or ko lines but were approx. ten-fold elevated over LCL (721) levels. To

confirm the correct parental DG75 cell lines, CBF1 expression status was evaluated as well. Furthermore, the expression levels of TFs IRF4, BATF, and EBF1, which were identified in this work to correlate with E2 or E3 binding patterns, were assessed. IRF4 could not be detected at all and BATF was reduced compared to LCL, while EBF1 expression was comparable to LCL levels (Fig 29C).

Next, the ChIP protocol had to be optimized for the application in DG75 cells as well as for the used antibodies (see chapter 3.5.4.2). The combination of an HA-tag specific antibody and two different antibodies directed against E2 were the most efficient approach for ChIP in this case (data not shown). The success of the E2 ChIP in DG75$^{doxHA\text{-}E2}$ in CBF1 wt and ko background was determined by ChIP-qPCR evaluating regions which were identified as E2 binding sites in the E2 ChIP-seq in LCL (Fig. 30).

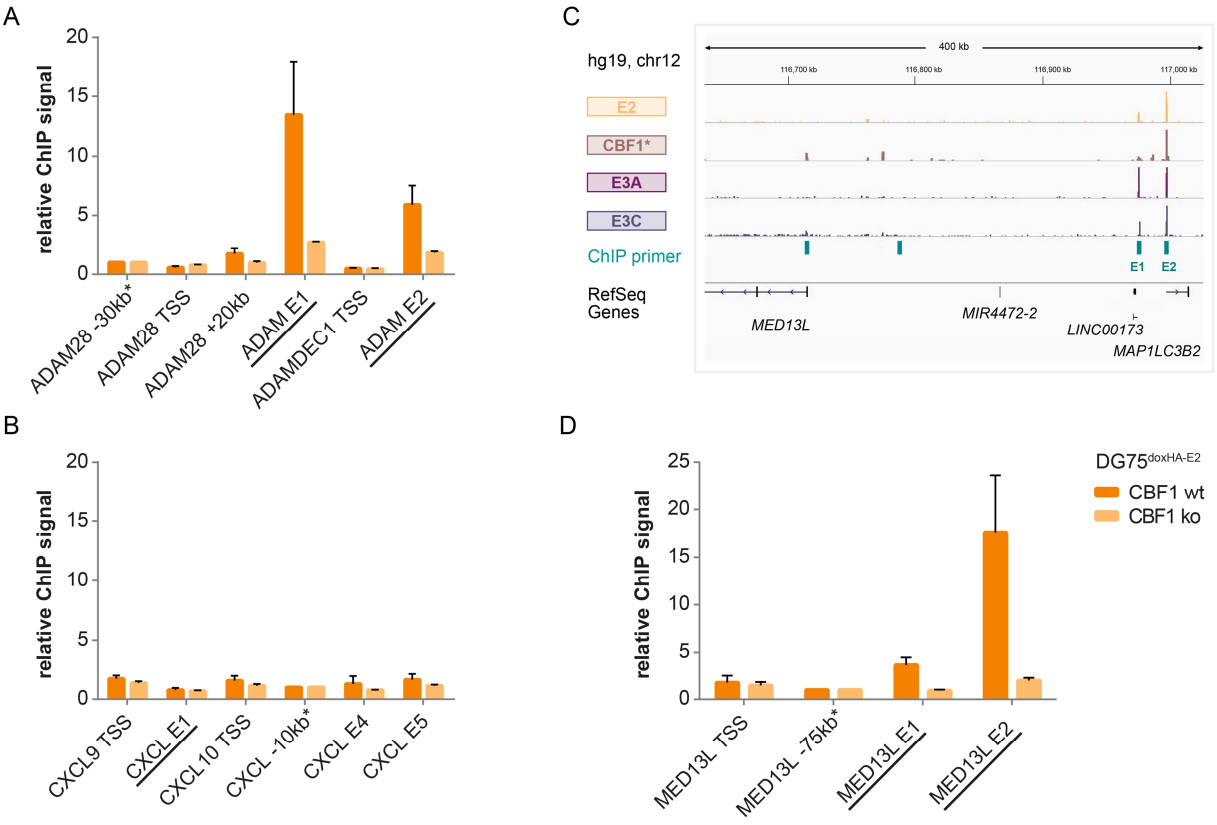


Figure 30. Successful detection of specific HA-E2 chromatin interactions by ChIP-qPCR in the inducible DG75 cell system. E2 ChIP was performed in DG75$^{doxHA\text{-}E2}$/CBF1 wt and ko respectively upon induction of E2 expression as described in chapter 3.5.4.2. Recovered DNA was analyzed for enriched regions by qPCR applying primers specific for known E2 binding sites (underlined, detected in LCL, chapter 4.2). As negative control for unspecific DNA enrichment, the E2 ChIP was performed using uninduced cells for chromatin preparation as well. Percent input values for each genomic region were calculated and these background values were subtracted from values of induced samples respectively. Relative ChIP enrichment over negative control regions (*) with no known TF binding site in LCL were calculated. One representative experiment is shown (n=2). Binding sites in (A) proximity to *ADAM28* and *ADAMDEC1* as well as (B) *CXCL9* and *-10* have been described (chapter 4.2.5.4, Fig. 24). (C) Schematic representation of the genomic region of approx. 300 kb upstream of *MED13L* including two significant E2 binding sites in LCL (E1 and E2). Signal intensities of E2, CBF1 (Zhao et al., 2011b), E3A, and E3C are shown with the scale set to 3x the mean value of each track. Positions of ChIP-qPCR primers are indicated and RefSeq genes are shown. (D) E2 ChIP-qPCR analysis of the genomic region depicted in (C).

Here, E2 could be detected at the intergenic enhancer (E1) between *ADAM28* and *ADAMDEC1* and at the *ADAMDEC1* intragenic enhancer (E2) in DG75$^{doxHA\text{-}E2}$/CBF1 wt cell line and binding was severely impaired in the CBF1 ko but still above unspecific background levels (Fig. 30A). Interestingly, at the intergenic enhancer E1 within the genomic region encompassing *CXCL9* and *CXCL10*, a significant E2 binding site in LCL, no E2 binding in neither DG75$^{doxHA\text{-}E2}$/CBF1 wt nor ko cell line could be detected (Fig. 30B). At a third genomic region, upstream of E3A and E3C target gene *MED13L* (Hertle et al., 2009, and our unpublished results from E3C ko LCLs), two E2 binding sites were identified in LCL and could now also be detected in DG75$^{doxHA\text{-}E2}$/CBF1 wt cell line (Fig. 30C). E2 binding in DG75$^{doxHA\text{-}E2}$/CBF1 ko at E1 and E2 upstream of *MED13L* could not be detected any more.

These selective E2 ChIP-qPCR analyses did show a successful and specific precipitation of E2 in the DG75$^{doxHA\text{-}E2}$/CBF1 wt and even some enrichment in the CBF1 ko background, at least at one genomic locus. Therefore, the same samples analyzed by qPCR were subjected to deep-sequencing (ChIP-seq) for a genome wide analysis of CBF1 dependent E2 chromatin binding.

4.3.2 Identification of E2 binding sites in DG75 cell lines proficient or deficient for CBF1

The data obtained from E2 ChIP-seq experiments in DG75$^{doxHA\text{-}E2}$/CBF1 wt and ko cell lines was subjected to the same bioinformatic pipeline as described in chapter 4.2.1.4 (schematic depiction in Fig. 13) to identify binding sites in the human genome and to generate signal tracks. However, in this case overrepresented sequences were rarely detectable and the overall quality of the obtained reads as assessed by FastQC quality control tool was very good (data not shown), so demultiplexed reads were directly subjected to mapping to the human genome (hg19) without a prior trimming step. An overview of the sequenced samples and data obtained directly from sequencing and mapping is shown below (Table 22).

Table 22. E2 ChIP-seq in DG75 cell lines - Perceived reads after different workflow steps and mapping to the human genome

DG75 Cell Line	Replicate - Sample Type	Read Count			Internal Designation
		Demultiplexed	Mappable (% of Demultiplxed)	Uniquely Mappable (% of Demultiplexed)	
CBF1 wt	E2-I-ChIP	17,455,101	94.81	69.06	LG620_wt_E2
	E2-I-input	19,901,128	97.69	70.21	LG620_wt_input
	E2-II-ChIP	34,613,332	94.78	68.89	LG625_wt_E2
	E2-II-input	27,927,224	97.93	70.56	LG625_wt_input
CBF1 ko	E2-I-ChIP	17,320,583	97.15	69.82	LG620_ko_E2
	E2-I-input	20,294,961	97.48	69.64	LG620_ko_input
	E2-II-ChIP	25,601,620	97.21	70.81	LG625_ko_E2
	E2-II-input	29,324,523	97.66	70.28	LG625_ko_input

Reads obtained after demultiplexing were directly subjected to Bowtie2 software for mapping to the human genome (hg19).

Significant E2 binding sites could well be detected in both DG75$^{doxHA\text{-}E2}$/CBF1 wt and ko cell lines (Table 23) applying the pipeline described in chapter 4.2.1.2 where biological ChIP-seq replicates were merged.

Table 23. E2 ChIP-seq in DG75 cell lines - Peaks identified in the human genome using MACS2

DG75 Cell LIne	Subjected to MACS2	Read Count		Allowed Duplicate Tags	Redundancy Rate (%)	E2 Peaks
		Merged Mapped Reads	Filtered			
CBF1 wt	E2-ChIP	49,354,861	48,526,538	2	1.68	1,937
	E2-input	46,790,793	45,788,697	2	2.14	
CBF1 ko	E2-ChIP	41,714,810	41,006,646	2	1.70	429
	E2-input	48,423,478	47,262,377	2	2.40	

Mapped reads of replicates were merged and subjected to MACS2 peak calling algorithm. Here reads were filtered for allowed duplicate tags, which represent maximum permitted reads mapping to the exact same position. This value is calculated by MACS2 in accordance with absolute read count and genome coverage. The redundancy rate is indicating the percentage of duplicate reads not allowed and displays a measurement for library complexity.

Also in this case identified peaks were submitted to further quality control steps (described in 4.2.1.2) before gaining a final peak list (Table 24) which then could be subjected to further bioinformatic analyses.

Table 24. E2 ChIP-seq in DG75 cell lines - Signal and mappability corrected peaks in the human genome

DG75 Cell Line	Identified by MACS2	Signal corrected	Blacklist corrected	GM12878 compatible	% of MACS2 peaks
CBF1 wt	1,937	1,818	1,793	1,789	92.4
CBF1 ko	429	286	271	271	63.2

Peaks identified by MACS2 were further filtered to exclude peaks which display a negative amplitude, fall on blacklisted regions or a chromosome not compatible with GM12878, the LCL used by ENCODE.

4.3.3 E2 binding sites in DG75 cell line differ from those identified in LCLs due to cell line specific enhancer signatures

Comparing the total of 1,789 detected E2 peaks in DG75$^{doxHA\text{-}E2}$/CBF1 wt with the 22,500 identified peaks in LCL described in chapter 4.2, a relevant difference in absolute numbers becomes evident. Several reasons could contribute to this finding. Low library complexity and poor read qualities could be excluded by FastQC quality control. Poor enrichment in the immunoprecipitation is not very like as predicted by qPCR analyses.

DG75 is an EBV negative Burkitt's lymphoma cell line harboring a t(8:14)(q24;q32) translocation bringing *MYC* under the control of the *IGH* gene locus and therefore driving proliferation (Ben-Bassat et al., 1977). In LCLs on the other hand EBV, with E2 as one of the most important factors, is the driving cause and indispensable for proliferation and immortalization. During immortalization the B cell changes its phenotype and becomes more similar to a activated B cell (reviewed in Thorley-Lawson, 2001). This process is accompanied by changes gene expression patterns and therefore also in the chromatin landscape, partly directly mediated by E3 proteins. This could be shown for several exemplary genomic regions (reviewed in Allday et al., 2015).

However, the DG75 cell line displays a completely different cellular system than LCL which is not in need of external pro proliferative signals and therefore very likely also exhibits a different chromatin landscape than LCL. Recently, a study on methylome analyses in different lymphomas, including information on important genome wide histone modifications in DG75, was published and sequencing data is now publically available (Kretzmer et al., 2015).

Comparing enhancer defining chromatin modifications at E2 binding sites in DG75$^{doxHA\text{-}E2}$/CBF1 wt with those in LCLs an interesting picture emerged where DG75 and LCL indeed differ significantly in their chromatin signatures (Fig. 31). To this end the available raw ChIP-seq data on histone modifications in DG75 and LCL was processed applying the self-generated bioinformatics pipeline for the identification of peaks and generation of signal tracks as described in chapter 3.6.1.

The majority of E2 binding sites in DG75$^{doxHA\text{-}E2}$/CBF1 wt is also present in LCL (1,325 *LCL/DG75 shared* sites = 74.1%) but both cell lines exhibit specific E2 binding sites (Fig. 31A). The E2 signal in LCLs was the highest at the LCL/DG75 shared binding sites, while E2 signal did only slightly differ between these shared and DG75 unique sites in DG75 (Fig. 31B).

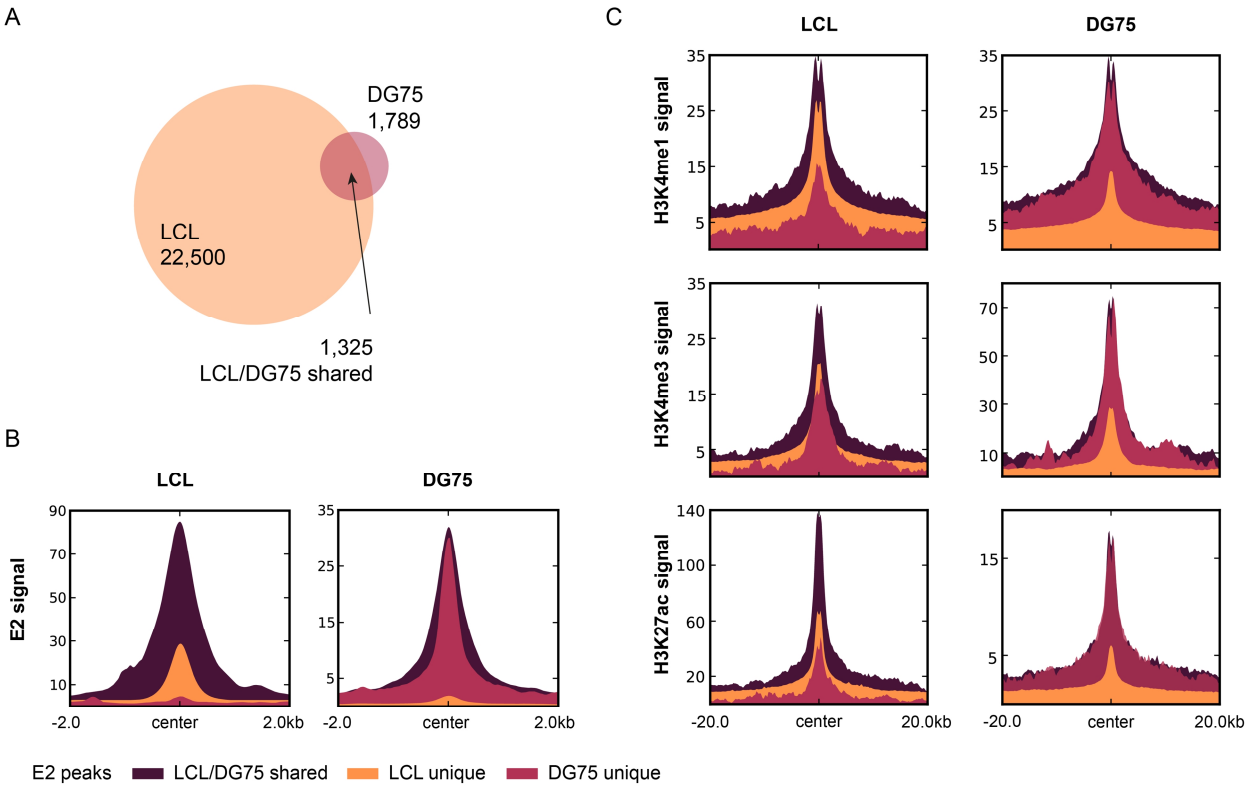


Figure 31. E2 binding sites in DG75 differ from those in LCL but are also located at cell line specific enhancers. (A) Intersection of E2 binding sites identified in LCLs and DG75$^{doxHA\text{-}E2}$/CBF1 wt. Anchor plots showing (B) E2 signals or (C) signals of histone modifications associated with active chromatin and enhancer state at E2 binding sites in LCL (ENCODE_Consortium, 2012) and DG75 (Kretzmer et al., 2015) cell line. Here, the mean normalized signal for each signal and peak subset was calculated for the region spanning 20 kb in each direction of E2 peak centers. Absolute numbers for signal intensities for the same histone modification cannot be compared between the two cell lines since the experiments were conducted at different laboratories also using different antibodies. Generation of the signal tracks for this analysis was performed applying the same data processing workflow for both data sets.

Investigating the three subgroups of E2 peaks, LCL/DG75 shared, LCL unique, and DG75 unique, separately for present histone modifications in the two different cell lines respectively, the LCL/DG75 shared E2 sites stand out as the subsets with the most prominent enrichment for all three investigated histone modifications (H3K4me1, H3K4me3, and H33K27ac) associated with active enhancers (Fig. 31C). Furthermore, it could be shown that the DG75 unique E2 binding sites display the lowest enrichment for all three investigated histone modifications in LCL, while the LCL unique E2 sites show the poorest enrichment in DG75.

Another point to be mentioned apart from different patterns of chromatin landscape between LCL and DG75 is the difference in expression of TFs. The DG75 cell lines used in the experiments conducted in this thesis are not expressing IRF4 and BATF only to a reduced extend (Fig. 29C). Evaluation of transcript levels detected by gene expression arrays in DG75, performed (by Sybille Thumann) to gain insights on E2 target genes, revealed several TFs which are only expressed at very low levels, while E2 associated factors EBF1 and CUX1 are transcribed at much higher levels (Fig. 32A). Among the TFs of the ENCODE ChIP-seq panel 7

factors, including IRF4 and BATF, were transcribed at very low levels with unsure protein expression status. MTA3 could be identified as part of the E3 cluster of positively correlating TFs on a genome wide level (Fig. 22). Since E2 and E3 clusters show also inter-connection, the depletion of MTA3 could also have a moderate impact on E2 binding in DG75. The other 4 TFs could not be identified to correlate with E2 signals in any of the conducted approaches and therefore are very unlikely to have an impact on E2 accession to chromatin in DG75 cell line.

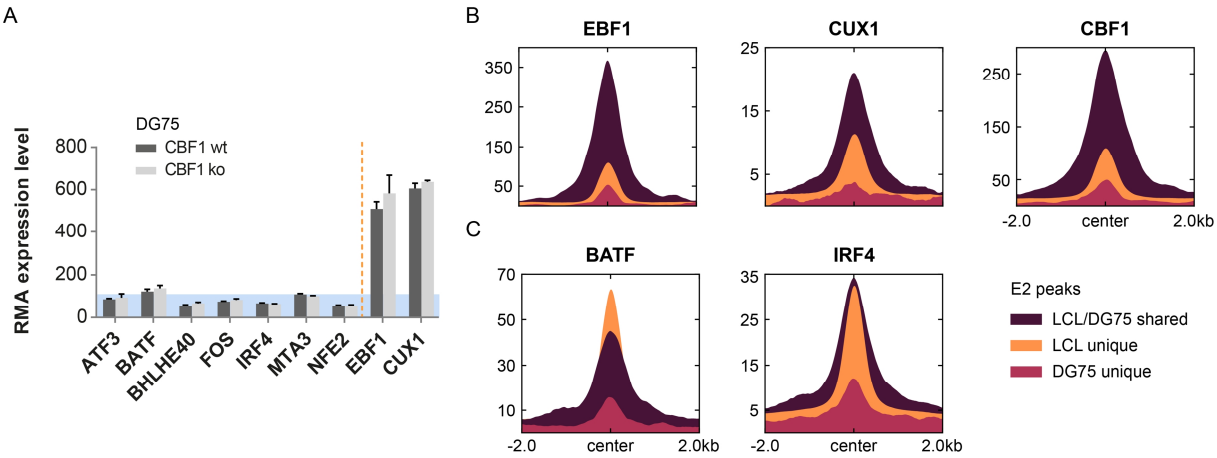


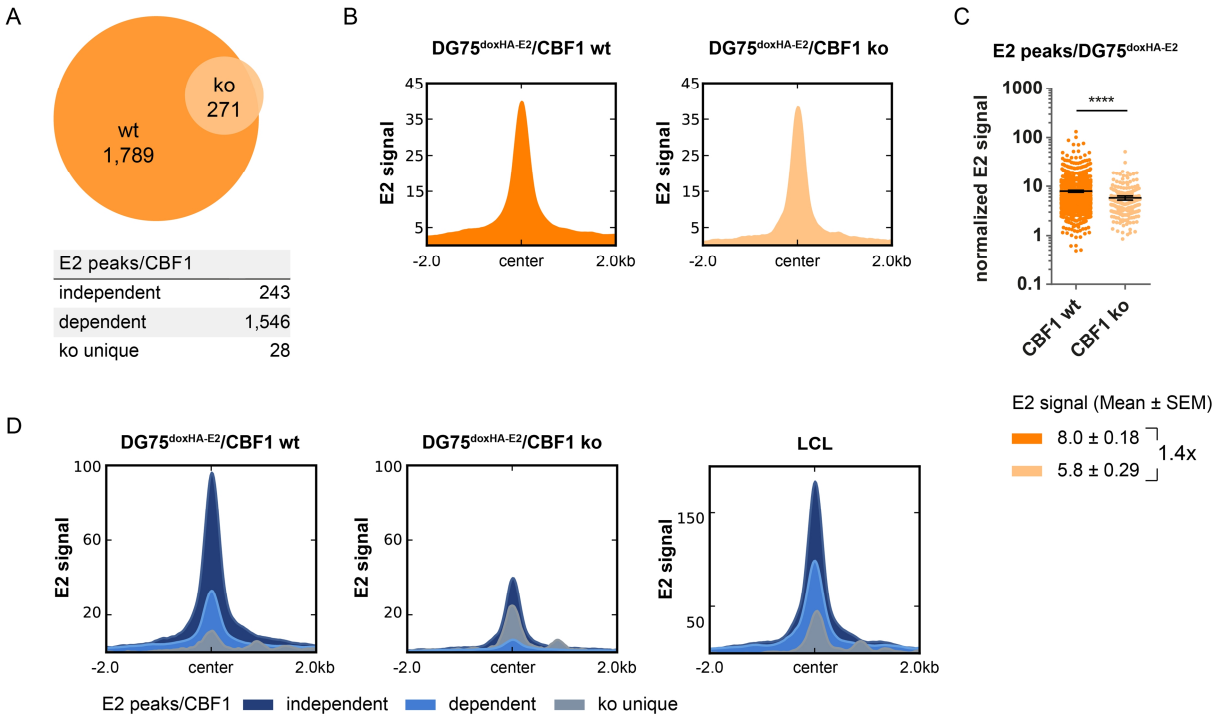
Figure 32. TFs expressed at very low levels in DG75 parental cell lines are enriched at LCL unique E2 binding sites. (A) RMA expression levels as received from gene expression analysis (GeneChIP® Human Gene 2.0 ST, PhD thesis Sybille Thuman) applying cDNA from parental DG75 cell lines, proficient and deficient for CBF1. An RMA value of 100 displays an approximate threshold for reliable detection of transcription. Only TFs of the ENCODE ChIP-seq panel were included in analyses for expression levels in DG75. (B and C) Anchorplots depicting TF signals at the three subsets of cell line specific E2 peaks. A region of 2 kb in each direction from peaks centers was analyzed. (B) TFs identified to be correlating with E2 signal in LCL and (C) TFs not or very low expressed in DG75 were included.

Information on TF enrichment, derived from LCL, at E2 binding sites was assessed as well and first a pattern very similar to the ones for histone modifications emerged for E2 associated TFs (Fig. 32B) but also for many other TFs (data not shown). Interestingly, BATF and IRF which are not expressed or only at very low levels in DG75 were showing very high enrichments at the LCL unique E2 binding sites indicating a role for those two TFs for E2 accession to these very binding sites.

In summary, chromatin landscape and TF expression in DG75 parental cell lines differ considerably from those present in LCL and therefore change accessibility of certain E2 binding sites. The E2 peaks which can be detected in both cell lines therefore display strong B cell enhancers which are most likely generally important for B cell identity.

4.3.4 E2 binding to chromatin is strongly but not exclusively dependent on CBF1

Within the DG75 cell system 271 E2 binding sites could be detected in the CBF1 ko situation accounting for 15.1% of the peaks in CBF1 wt (Fig. 33A). More specifically, 243 CBF1 independent E2 peaks could be identified which are present in both cell lines independent of CBF1 expression (*CBF1 independent*). 1,546 E2 sites could only be detected in the DG75$^{doxHA\text{-}E2}$/CBF1 wt cell line and therefore are *CBF1 dependent* E2 peaks. A small subset of 28 peaks could only be identified in DG75$^{doxHA\text{-}E2}$/CBF1 ko designated as *ko unique* E2 peaks.



E2 peaks/CBF1	
independent	243
dependent	1,546
ko unique	28

E2 signal (Mean ± SEM)	DG75$^{doxHA\text{-}E2}$/ CBF1 wt	DG75$^{doxHA\text{-}E2}$/ CBF1 ko	LCL
	27.0 ± 1.6	10.2 ± 0.6	49.6 ± 4.6
	10.6 ± 0.2	2.7 ± 0.1	29.4 ± 1.7
	4.0 ± 0.7	6.7 ± 1.0	11.1 ± 3.8

Figure 33. Chromatin binding of EBNA2 is mainly but not exclusively dependent on CBF1. Comparison of E2 binding sites detected in DG75$^{doxHA\text{-}E2}$/CBF1 wt and ko cell lines. (A) E2 peaks in DG75 cell line with CBF1 wt

or ko background were identified using MACS2. Peaks were subdivided according to their dependence on CBF1 expression. (B) Comparison of mean normalized ChIP-seq signal for E2 in DG75$^{doxHA\text{-}E2}$/CBF1 wt and ko background. (Here signal intensities are comparable since the same antibodies for the HA-E2 construct could be used in both cell lines and the expression levels are comparable). (C) The scatter plot shows the distribution of signal intensities and the mean with a 95% confidence interval of the mean normalized signal for E2 peaks in DG75/CBF1 wt or ko for a region flanking the peak center for 2 kb in each direction (Data underlying panel B). Signal means and SEMs are indicated below. (D) Anchorplots depicting signal intensities at CBF1 independent or dependent and ko unique EBNA2 peak subsets as defined in A. (E) Signal distribution of data underlying panel D, means and 95% confidence intervals are indicated. Statistical significance for differences of all means were assessed applying unpaired two-tailed t-test for log values with Welch's correction (**** $p<0.0001$); absolute means and SEMs are indicated below.

The mean E2 signal distribution and enrichment at E2 peaks in DG75$^{doxHA\text{-}E2}$/CBF1 wt is very similar to the one observed at E2 peaks in the CBF1 ko situation (Fig. 33B). However, taking a closer look at the mean signal distribution over all peaks it becomes evident that E2 signal is 1.4 fold higher in DG75$^{doxHA\text{-}E2}$/CBF1 wt than in the ko situation (Fig. 33C). Importantly, E2 signal intensities of the two different ChIP-seq experiments can be directly compared in this case, since the same protein, expressed in similar quantities, using the same antibodies was precipitated under the same experimental conditions and ChIP-seq data was analyzed and normalized applying the same pipeline.

Even more dramatic becomes this effect when observing E2 signal intensities at the three E2 peaks subsets, CBF1 independent, dependent, and ko unique, separately. E2 signal is most enriched at CBF1 independent E2 binding sites, in DG75$^{doxHA\text{-}E2}$/CBF1 wt as well as in CBF1 ko cell line but the total signal is the highest in the CBF1 wt situation (Fig. 33D and E). Hence, the strongest E2 binding sites in DG75 cell line are the ones that can still be detected in the CBF1 ko situation. Interestingly, the E2 binding sites detected in DG75 display a similar E2 signal distribution pattern in LCL as in the CBF1 wt situation. Here, CBF1 independent E2 peaks, detected in DG75, display the highest E2 enrichment as well, followed by CBF1 dependent and then ko unique peaks.

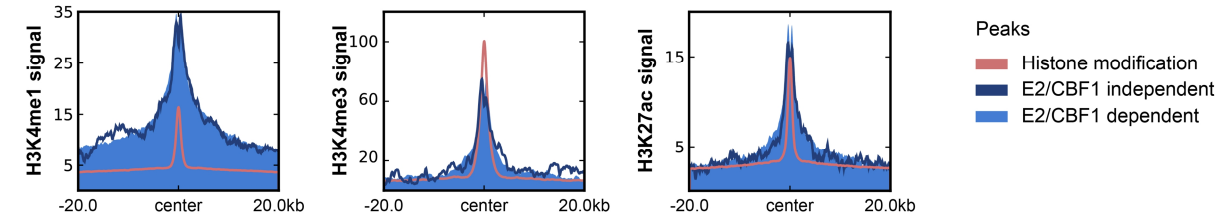
Taken together, significant E2 binding sites could be detected even in the absence of CBF1, but display a lower mean E2 enrichment and the strongest binding sites in the CBF1 wt situation. It seems to be very likely that other TFs contribute to the high enrichment of E2 at CBF1 independent peaks which allow accession to chromatin even in the absence of CBF1.

4.3.5 E2 targets strong enhancer in DG75 independent of CBF1 expression status

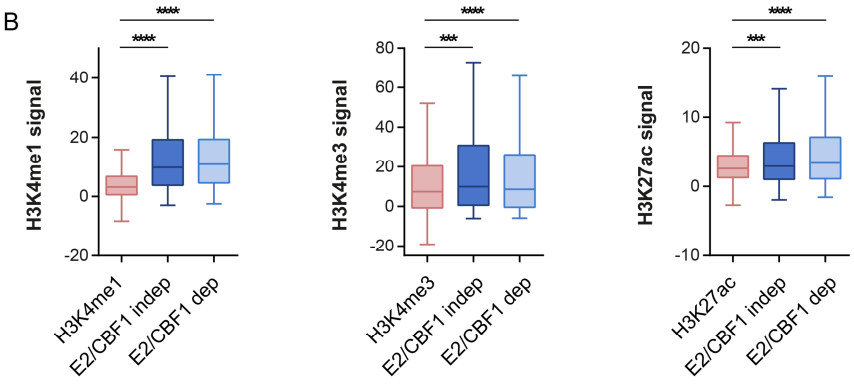
The identified CBF1 independent E2 peaks in DG75 cell line exhibit very high E2 signal intensities suggesting a potential involvement of strong enhancers as a prerequisite for high E2 signal intensities. To stress this idea CBF1 independent and dependent E2 peaks detected in DG75 were now analyzed for the prevalent histone modifications at those regions. To this end

ChIP-seq data on H3K4me1, H3K4me3, and H3K27ac (Kretzmer et al., 2015), characteristic for active enhancer elements were analyzed for their abundance at the different E2 peak subsets (Fig. 34).

A



B



C

Means	H3K4me1	H3K4me3	H3K27ac
	4.5 ± 0.02	12.3 ± 0.09	3.3 ± 0.01
	12.9 ± 0.74	17.0 ± 1.29	4.5 ± 0.32
	13.2 ± 0.28	16.6 ± 0.57	5.0 ± 0.15

Figure 34. CBF1 independent and dependent E2 binding sites in DG75 display almost identical histone modification patterns defining enhancer activity. CBF1 dependent and independent E2 binding sites identified in DG75 were analyzed for enhancer associated histone modifications. The normalized ChIP signals for the regions spanning 20 kb in each direction from peak center were used. (A) Anchor plots showing the histone modification signal profiles at E2 peak subsets. Peaks for each analyzed modification were used as references for average positive signals and model profiles. In (B) the data underlying panel (A) were used to generate boxplots depicting the signal distributions over the whole regions of 40 kb. Significances of differences of means were assessed applying unpaired two-tailed t-tests with Welch's correction (**** $p<0.0001$, *** $p<0.001$). The differences of means for CBF1 independent and dependent E2 peaks were not statistically significant ($p= 0.706$, 0.7595, and 0.1396 respectively). Boxplot whiskers extend to 1.5x interquartile range. (C) Table depicting means and SEMs of histone modification signals at peaks described in (A) and (B).

As a matter of fact, all three investigated enhancer defining histone modifications did not significantly differ in their signal intensities at CBF1 independent and dependent E2 peaks in DG75. Noticeably, H3K4me1 was 2.9x higher at both E2 peak subsets than at the average H3K4me1 peak in DG75 (Fig. 34A and B, left panels) indicating targeting of strong enhancers. Therefore a strong enhancer signature is not the defining feature of a CBF1 independent E2 binding site but rather as implicated in previous chapter, most likely the co-occurrence of specific TFs besides CBF1.

4.3.6 EBF1 as a potential chromatin anchor for E2 in the absence of CBF1

4.3.6.1 EBF1 is enriched at CBF1 independent E2 binding sites in LCL

In order to investigate how E2 accesses chromatin in the absence of CBF1 a de novo TF DNA sequence motifs search was performed to get a first idea about possible enriched motifs since no TF ChIP-seq experiments in DG75 were available. Applying MEME-ChIP (Machanick and Bailey, 2011) on CBF1 independent and 243 randomly selected CBF1 dependent E2 peaks, to account for the big differences in sample size, only two motifs could be discovered which match described TF sequence motifs: While both, the CBF1 and the EBF1 motif, could be discovered at CBF1 dependent E2 peaks, only the EBF1 motif was significantly enriched at CBF1 dependent E2 sites (Fig. 35A).

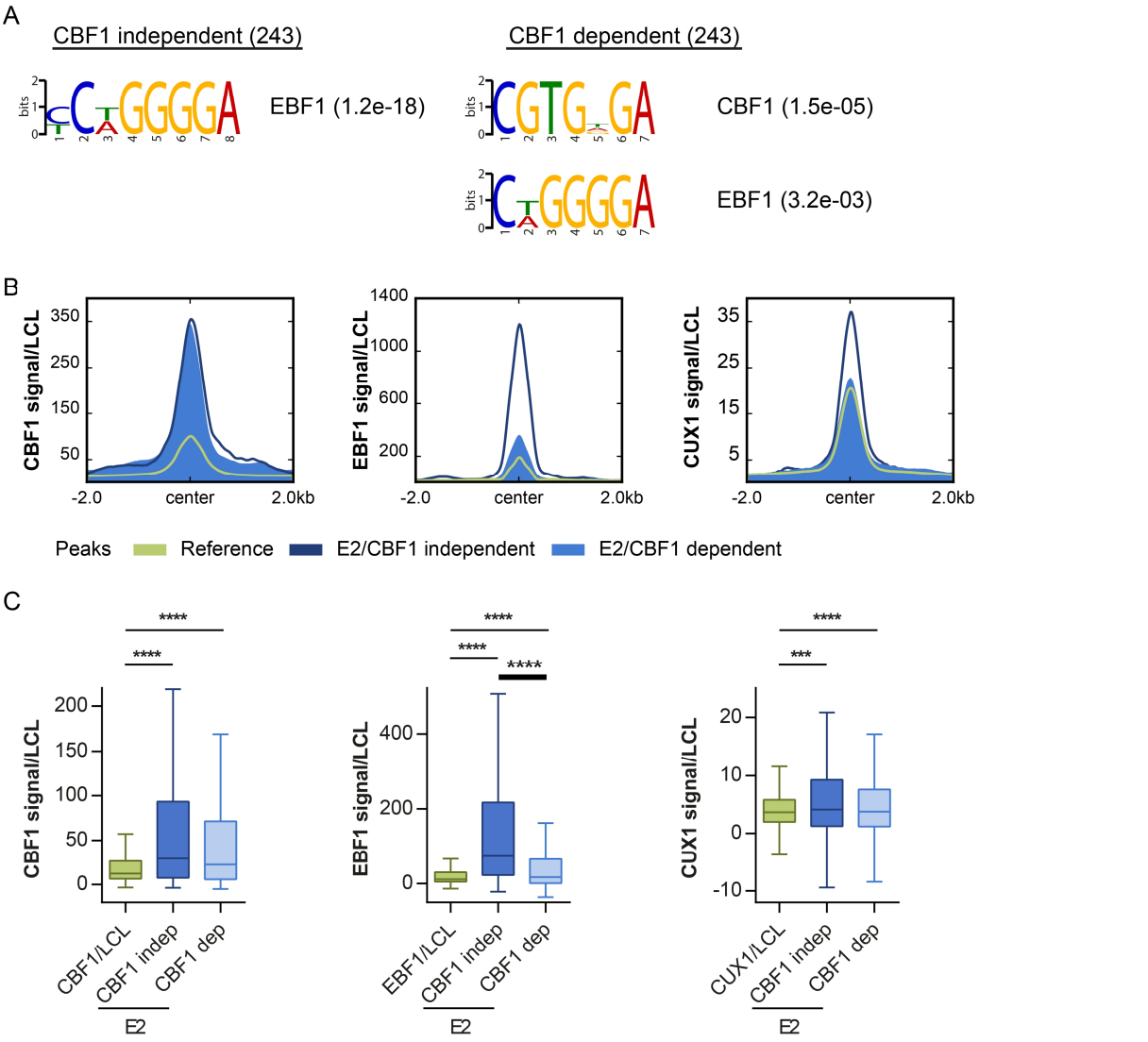


Figure 35. EBF1 is significantly enriched at CBF1 independent E2 peaks in LCLs. (A) Comparison of enriched DNA sequence motifs discovered at CBF1 independent and dependent E2 binding sites in DG75 cell lines using MEME-ChIP motif discovery tool (Machanick and Bailey, 2011). E-values for statistical significance of discovery and the TF most likely to recognize them as predicted by TOMTOM (Gupta et al., 2007) (scanning hocomoco v9 database) are shown. For this analysis 243 out of 1546 total CBF1 dependent E2 peaks were randomly

chosen for better comparison of E-values between two different sized populations. No significantly enriched motifs could be detected for CBF1 ko unique E2 peaks. (B) Anchor plots depicting mean normalized ChIP-seq signals for TFs derived from LCL at E2 binding sites identified in DG75 cell lines. Peaks of each investigated TF were used as references for average positive signals and model profiles. (C) The underlying data of panel (B) were used to generate boxplots depicting signal distributions. An unpaired two-tailed t-test with Welch's correction (**** $p<0.0001$) was performed to determine significant differences between means. The differences between the means for CBF1 independent and dependent E2 peaks were -2.892 ± 9.972, -93.82 ± 12.89, and -1.17 ± 0.622, for CBF1, EBF1, and CUX1 respectively and only statistically significant for the EBF1 signal ($p= 0.772$, 3.834E-12, and 0.069 respectively). Boxplot whiskers extend to 1.5x interquartile range.

Since EBF1 was also detected in the previous correlation analyses for E2 associated factors in LCL (4.2.5.5.1), this finding was particularly interesting. To get further insights on the contribution of TFs apart from CBF1 on E2 accession to chromatin, ChIP-seq information derived from ENCODE in LCL was analyzed for enrichment at E2 binding sites as detected in DG75, since those peaks show a very similar E2 signal distribution in DG75 and LCL (Fig. 33D). While CBF1 enrichment in LCL was not significantly different between CBF1 independent and dependent E2 binding sites, a highly significant enrichment of EBF1 at CBF1 independent over dependent E2 sites could be detected (Fig. 35B and C). CBF1 independent and dependent E2 peaks were also investigated for CUX1 signal in LCL, the second TF identified in the EBNA peak correlation analysis (4.2.5.5.1), but no significant enrichment of CBF1 independent over dependent E2 peaks could be identified. Moreover, the CUX1 sequence motif could not be identified in the MEME-ChIP motif enrichment analysis which indicates no important role for CUX1 in the presence of CBF1 independent chromatin binding of E2.

In summary, the enrichment analysis for sequence motifs at CBF1 independent and dependent E2 binding sites as well as the signal enrichment analyses for TFs in LCLs at those sites revealed a potential role of EBF1 in mediating chromatin access of E2 in the absence of CBF1.

Already the genome wide correlation analyses for TF binding patterns in LCL (4.2.5.3) showed a strong positive correlation of E2 and EBF1, almost in the range of E2 and CBF1 interaction. Signal correlation analyses restricted to E2 peaks as reference regions did show similar results (Fig. 36A). Here E2 showed an r_s of 0.46 and 0.40 in comparison to CBF1 and EBF1, respectively. For this analysis, also the correlation of E2 in comparison to PU.1 signal was examined, since, as described in the introduction of chapter 4.3, PU.1 was considered to be a potential adaptor protein for E2 several times in the literature. Nevertheless, neither correlation analyses on an E2 peaks wide or on a genome wide scale (Fig. 36A and B) nor motif enrichment analyses did show any potential involvement of PU.1 in CBF1 independent chromatin accession of E2. EBF1 on the other hand was revealed as a potential adaptor protein or chromatin access mediating factor for E2 action.

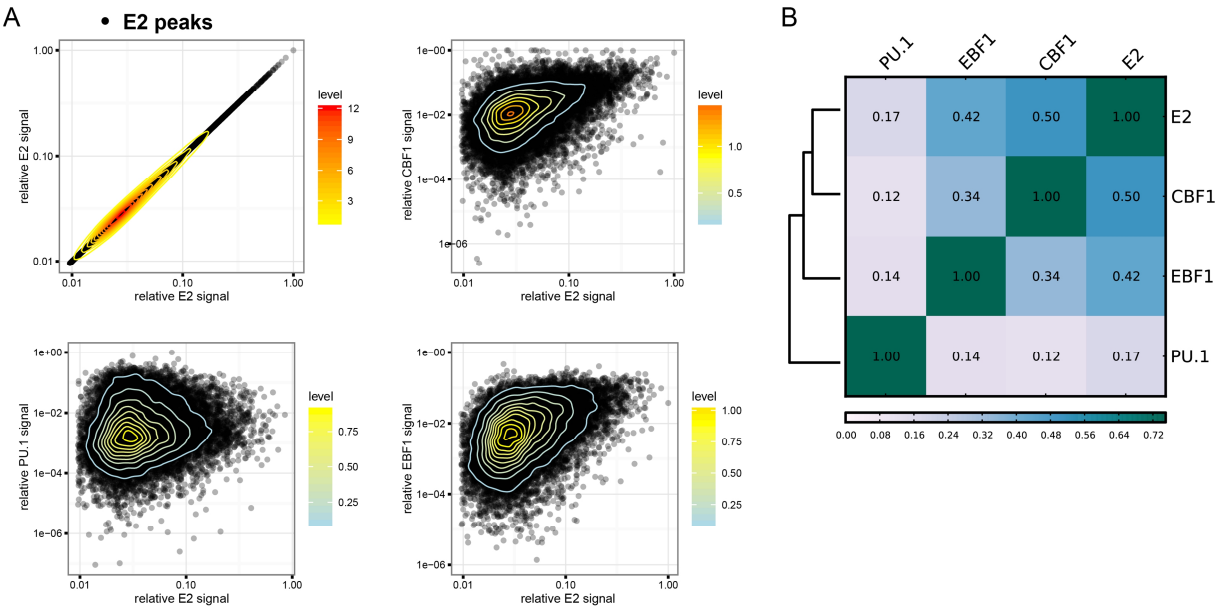


Figure 36. EBF1 shows a strong binding pattern correlation to E2, similar to known adaptor protein CBF1. In (A) E2 binding sites were investigated for signal intensities of other TFs. For every E2 peak the relative mean normalized E2 signal was plotted against the ones of E2 (perfect correlation), CBF1, PU.1 and EBF1 respectively. To obtain relative values, the highest peak signal was set to 1 and the other values were scaled accordingly. Each dot represents one E2 peak. Correlation analyses were performed and Spearman correlation coefficients (r_s) were calculated. E2 shows an r_s of 1.0, 0.46, 0.19, and 0.40 in comparison with E2 itself, CBF1, PU.1, and EBF1 respectively. (B) Correlation matrix showing signal pattern correlations for different ChIP-seq experiments on a genome wide scale. The human genome was divided in 100 bp bins and mapped reads for each experiment were counted for each bin. A correlation coefficient using Spearman correlation was calculated for each pair and is displayed and color coded in the matrix.

4.3.6.2 E2 and EBF1 protein-protein interaction in DG75 cell line

Since the bioinformatic analysis of CBF1 independent E2 binding sites in DG75 strongly indicated a functional role of EBF1 in chromatin accession of E2, the protein-protein interaction properties of E2 and EBF1 were assessed in Co-IP experiments (conducted by Cornelia Kuklik-Roos). Here, EBF1 was pulled down from DG75$^{doxHA\text{-}E2}$ cell lysates after transfection with an EBF1 expression plasmid (kindly provided by Prof. M. Sigvardsson, Lund University, Sweden; Mega et al. 2011) or corresponding empty plasmid and interaction with E2 was assessed in western blot experiments (Fig. 37). To this end, DG75$^{doxHA\text{-}E2}$ proficient but also deficient for CBF1 were used to restore the environment in which CBF1 independent E2 binding to chromatin could be detected. In a CBF1 competent DG75 background, a competition of CBF1 with EBF1 for E2 binding could be a possible scenario and a weak E2-EBF1 interaction might be unable to detect by this method.

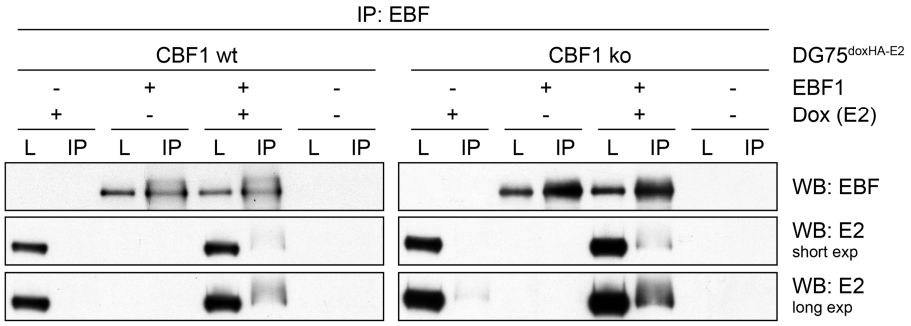


Figure 37. E2 and EBF1 protein-protein interaction could be detected in DG75doxHA-E2/CBF1 wt and ko cell lines. Co-IP experiments using EBF specific antibodies for IP were conducted 24 h after transfection of empty (pCDNA3) or EBF1-myc expression plasmid (pCDNA3.EBF1-5xmyc). Induction of HA-E2 expression by addition of Dox was performed directly after transfection. Total cell lysates (L) display 1% of the cells used for IP samples. One representative experiment is shown (n=2).

However, the Co-IP experiments in DG75$^{doxHA\text{-}E2}$/CBF1 wt and ko cells revealed a robust interaction of E2 with EBF1 upon EBF1 transfection and induction of E2 expression (Fig. 37) while the interaction between E2 with endogenously expressed EBF1 could only be detected in the CBF1 ko upon longer exposure times (Fig. 37, bottom right panel). This interaction could not be detected in the CBF1 wt situation (Fig. 37, bottom left panel), which is most probably due to overall lower E2 expression levels in the CBF1 wt cells (experiments are currently repeated by C. Kuklik-Roos with adjusted E2 expression levels). Furthermore, EBF1 is only expressed at very low endogenous levels and therefore only detectable in cell lysates by western blot upon very long exposure times, which lead to overexposure of other detected bands (data not shown).

Co-IP experiments pulling-down E2 from DG75$^{doxHA\text{-}E2}$ cell lysates with subsequent testing for EBF1 binding as well as confirmation of the interaction of endogenously expressed E2 and EBF1 in LCL is currently in preparation and could not be included in this thesis but certainly will shed light on the significance of E2-EBF1 protein-protein interaction.

5 Discussion

5.1 Epitope tagged E3A or E3C expressing LCLs as a versatile cellular system for studying chromatin interactions

The essential role of EBV encoded latent proteins E2, E3A, and E3C in infection and immortalization of primary B cells has been subject to many studies and the mechanisms by which they achieve specific gene regulation is still extensively researched (reviewed in Allday et al., 2015, and Kempkes and Ling, 2015). One important aspect of their distinct functions as TFs is the accession to chromatin and the targeting of certain functional elements. To gain information on binding sites of TFs in the human genome, ChIP-seq is the current standard method. However, the success of this method largely depends on the efficiency of the IP reaction. In the case of E2 highly specific antibodies suitable for IP were available and also other laboratories were able to perform successful E2 ChIP-seq experiments in LCL (Zhao et al., 2011b) and Mutu III, a Burkitt's lymphoma cell line showing type III latency expression pattern (McClellan et al., 2012).

The precipitation of E3A and also E3C proofed to be more difficult since no antibodies suitable for ChIP experiments were commercially available. Also other researchers were facing the same challenge and thus were using antibodies which are not specific for one distinct E3 protein but rather recognize all three E3 members (McClellan et al., 2012). In this study, individual E3 binding sites were further investigated by ChIP-qPCR in EBV negative Burkitt's lymphoma cell lines each ectopically expressing only one E3 protein. This approach could not give information on genome wide binding and therefore other strategies had to be developed to address this matter.

In the first part of this thesis the successful generation of LCLs, infected with recombinant EBV genomes, expressing Flag-tagged E3A or E3C could be demonstrated. Applying the recombineering technique (Warming et al., 2005), taking advantage of one selection marker which could be used for both, positive and negative selection of targeted genomes, it was possible to integrate the Flag-tag N-terminal and in frame to E3A or E3C respectively, without the permanent integration of selection markers and or recombination sites by established methods (as applied in Hertle et al., 2009, method adopted from Cherepanov and Wackernagel, 1995). Recombineering is a trending technique in manipulation of γ-herpesvirus genomes in general (reviewed in Warden et al., 2011) due to high targeting efficiencies (> 95%), the elegant usage of *galK* as positive and negative selection marker with its subsequent traceless elimination, and the

versatility of possible insertions, deletions, inversions, and point mutations. Furthermore, every BAC construct containing the *galK* gene at a position of interest can serve as starting point for the generation of various mutants e.g. for testing of different point mutations, and the cloning procedure does not have to be repeated for the first targeting step. Thus, this method was not only applied to generate recombinant EBV genomes (Seto et al., 2010, Jochum et al., 2012, Steinbruck et al., 2015) but was also used for manipulation of other γ-herpesvirus genomes such as KSHV (Wakeman et al., 2014, Bellare et al., 2015) and MHV68 (Rangaswamy et al., 2014, Rangaswamy and Speck, 2014).

The recent discovery and now commercial availability of *Clustered Regularly Interspaced Short Palindromic Repeats* (CRISPR)/Cas9 based genome editing strategies (Jinek et al., 2012), which shows dramatic improvement in efficiency and feasibility of precisely targeting genomic regions without leaving any traces, probably heralds a new age of genome editing possibilities. However, the EBV genome is relatively small (approx. 172 kb) compared to the human genome (approx. 3,234 Mb), is present in the cell as multiple copies, and involves the possibility to use the BACmid based EBV system. This includes the advantages of genome targeting and propagation in bacteria as well as the traceability by eGFP and selection marker expression. Therefore recombineering or related methods most probably will stay the method of choice for fast and effective generation of recombinant EBV strains.

The recombinant LCLs generated in this work did show latent EBV protein expression levels comparable to LCLs infected with wt EBV (Fig. 9C and D) and are not impaired in viability. Neither they are impaired in their protein-protein interaction with the DNA binding cellular protein CBF1, when pulled-down using Flag-tag specific antibodies (Fig. 10), nor in the repression of three well described target genes compared to wt LCLs (Fig. 11). Thus, these LCLs combine the advantage of ChIP specificity due to epitope-tagged proteins, as pointed out by the ENCODE project (Landt et al., 2012), with endogenous expression levels, since the epitope-tag coding sequence was integrated in the viral genome.

Simultaneously, other research groups were addressing the challenge of E3 protein precipitation in ChIP experiments with similar approaches of generating recombinant EBV genomes expressing epitope-tagged versions of E3 proteins. One group introduced a combination of Flag- and HA-tag C-terminally fused to E3C (E3C-F-HA) (Jiang et al., 2014) and E3A (E3A-F-HA) (Schmidt et al., 2015), respectively, in order to perform ChIP-seq experiments. Another study inserted a Strep-Flag-tag C-terminally fused to E3C within the EBV BAC system and subsequently infected BL31 EBV negative Burkitt's lymphoma cells with the derived recombinant viruses to perform E3C specific ChIP-qPCR experiments (Paschos et al., 2012). In contrast to these studies, the 3x Flag-tag used in this work was inserted N-terminally to E3A and

E3C, respectively. This choice was based on predicted secondary structures of the E3 proteins, where the N-terminal regions, covering the E3-family homology domain, are predicted to form α-helices and some β-strands while no ordered structures could be predicted for the C-terminal regions (Yenamandra et al., 2009). It has been described that disordered structures within proteins can display very important functions and might only form distinct secondary structures upon binding to interaction partners (Uversky, 2013). Also within the E3 proteins several functional *repressor* domains as well as many potential interacting proteins could be mapped to the C-terminal regions (reviewed in Allday et al., 2015). In order to preserve potential and not well understood functions of the C-terminal regions, the N-termini of E3A and E3C were targeted in this thesis.

As mentioned above, E3A and E3C were not impaired in their protein-protein interaction with CBF1 (Fig. 10), which have been described as very important for B cell transformation for both viral proteins (Maruo et al., 2005, Maruo et al., 2009). E3A and E3C interaction could also be detected by Flag-E3A as well as Flag-E3C pull-down in this thesis (Fig. 10). The formation of E3A and E3C complexes has been described previously by a Y2H screen (Calderwood et al., 2007) and could be confirmed by Co-IP experiments in B cells (Paschos et al., 2012). This interaction could already be identified in LCLs by the Kempkes group and furthermore the binding regions could be mapped to both N-terminal regions in HEK293 transfection experiments (dissertation S. Petermann, 2009).

Interestingly, in this thesis the protein-protein interaction of E3A and E3B could be demonstrated for the first time (Fig. 10) but was not further characterized.

Contrarily, a recent study to identify E3 protein interaction partners by *tandem affinity purification* (TAP) followed by mass-spec analysis of Flag-HA-tagged E3A, E3B, and E3C proteins did not reveal any E3 heterodimers but confirmed complexes of each investigated E3 with CBF1 and thus concluded the formation of distinct E3-CBF1 complexes (Ohashi et al., 2015). This finding could be due to experimental settings such as the integration of C-terminal tags as opposite to the N-terminal targeting in this thesis or harsher pull-down conditions. However, the study by Ohashi and colleagues underlines the strong binding of each E3 protein to CBF1 but the collective evidence by our group and Calderwood et al., collected in different cellular systems, investigating endogenous and transfected proteins, strongly indicates the formation of E3A and E3C as well as E3A and E3B heterodimers. The functional relevance and occurrence of those complexes in association with chromatin has yet to be determined. Some aspects of E3A and E3C binding within the human genome and their co-operation will be discussed in the following chapter.

5.2 EBNA transcription factors – exploiting enhancer elements

The overall aim of the second part of this thesis was the elaboration of the interplay of E2 as an activator and E3A and E3C as potential repressors of transcription by investigating chromatin binding and the associated prerequisites and co-occurring factors as the basis of EBNA protein mediated gene regulation. This question was addressed by performing and analyzing ChIP-seq experiments for E2, E3A, and E3C and the subsequent comparison with different data sets published by the ENCODE consortium.

5.2.1 Identification of E3, E3A, and E3C binding sites by ChIP-seq

At first the establishment of the ChIP-assay for Flag-tagged E3A and E3C was described and the successful deep-sequencing of the associated DNA fragments as well as the bioinformatic analysis and identification of significant binding sites could be shown. Here, each step from biochemistry to bioinformatics was controlled carefully and high quality results were obtained as discussed in the following.

Biochemistry

The cross-linking step of a standard ChIP-assay was successfully optimized to account for the indirect binding to DNA of the EBNA proteins applying disuccinimidyl glutarate (DSG) as an additional cross-linking reagent prior to formaldehyde (FA) treatment (Fig. 12). The beneficial impact of performing a dual cross-link using NHS-esters like DSG on ChIP efficiencies when precipitating TFs acting in complexes could already be described e.g. for NFκB (Nowak et al., 2008), STAT3, CDK9, PolII (Hou et al., 2007), and FOXM1 (Khongkow et al., 2014). Furthermore, the dual cross-linking procedure using DSG was applied for ChIP of the SWI/SNF chromatin-remodeling complex subunit SNF5 (Wilson et al., 2010). More recently, DSG dual cross-linking was applied in a genome wide approach to identify proteins which bind to enhancer or promoter elements in a cell specific manner by ChIP of distinct histone modifications defining functional chromatin elements and subsequent mass-spec analysis of associated proteins (Engelen et al., 2015).

Further steps of the ChIP protocol, specific for subsequent deep sequencing, were optimized and controlled mainly based on guidelines published by the ENCODE consortium (Landt et al., 2012).

Bioinformatic Analysis – Comparison to other studies on EBNA proteins

The bioinformatic analysis pipeline (Fig. 13) was constructed independently using the Galaxy platform (Giardine et al., 2005), which displayed the great advantage of traceability of each performed step, since complete workflows can be downloaded, shared, and recapitulated using

(the public) Galaxy server. Currently, more than 2,000 datasets are publicly available through Galaxy (Afgan et al., 2016).

The primary ChIP-seq data obtained in this thesis displayed very good quality features as assessed by FastQC (Table 16) and percentages of mapped reads (Table 17). Reads mapping to the EBV genome for all three performed ChIP-seq experiments could be detected as well, with a sequencing depth multiple times covering the entire EBV genome (Table 18). Applying MACS2 software significant binding sites for E2, E3A, and E3C could successfully be identified in the human (Table 19) and EBV (Table 20) genome. To this end, two biological replicates per ChIP were performed and sequenced as advised by the ENCODE project (Landt et al., 2012). However, there is no clear agreement in the field how to deal with data from replicate experiments. One way, also ENCODE suggests, is the application of *Irreproducible Discovery Rate* (IDR) analysis methodology, were peaks only count when significantly identified in both replicates (Li et al., 2011) but here the focus is drawn on highly enriched binding sites. Since in this thesis a quantitative analysis of binding sites enrichment should be performed as well, a different approach was chosen where mapped reads from biological replicates were merged and then subjected to peak calling in order to identify low enrichment reads as well.

However, detailed observations on the obtained binding sites in the human genome revealed the requirement for additional selection steps. To this end, "negative" peaks, which are wrongly detected by MACS2, peaks located on black-listed regions (Derrien et al., 2012), and finally peaks whose location were not compatible with the GM12878 genome were removed (Table 21). After this selection step 96.5, 90.8, and 69.8% of for E2, E3A, and E3C peaks, respectively identified by MACS2 remained in the final peak list. The peaks removed here, were clearly false positives or not adaptable for the cell line to compare. Therefore, this additional peak filtering strongly improves the overall significance of the final peaks lists.

The absolute numbers of detected peaks in the human genome, 22,500 E2, 13,490 E3A, and 8,733 E3C peaks are in general comparable to datasets published by other groups. For instance, the first E2 ChIP-seq study in IB4 cells, an LCL with two integrated EBV genomes, which initially was described not to be an ideal cell line for studying viral latency (Hurley et al., 1991), identified 5,151 E2 sites and also 10,529 CBF1 sites mapped to the human genome hg18 (Zhao et al., 2011b). It has to be noted, that in this study two biological replicates were reported to be analyzed but an independently performed reanalysis of the uploaded raw data, in order to update and compare these to data mapped to hg19, did not reproduce the results of the authors. After personal communication, the authors of the publication (Zhao et al., 2011b) admitted that only one replicate per experiment was used for the bioinformatic analysis respectively, since the quality of the second biological replicate was very low. Finally, 19,177 E2 and 38,063 CBF1 peaks

could be detected by applying the self-generated bioinformatics pipeline for peak calling to single experiments for E2 and CBF1 used by Zhao and colleagues. Later on, the same research group published different data sets which were compared with the E2 binding sites obtained by Zhao et al. but each time they were re-calculated: In a comparison with an EBNA-LP (E-LP) ChIP-seq 19,224 E2 binding sites (Portal et al., 2013) and subsequently in another study, addressing the potential occupation of super-enhancers by E2, even 42,251 peaks (Zhou et al., 2015) were calculated, each time applying different software and standards.

Simultaneously, a study performed in Mutu III, an EBV positive Burkitt's lymphoma cell line showing type III latency expression pattern, identified 21,605 E2 binding sites (McClellan et al., 2013).

However, the E2 ChIP-seq generated and analyzed in this thesis is based on two biological replicates, passing very high quality standards, performed in LCLs, the B cell line used by ENCODE for most of their experiments, and therefore displays the most reliable dataset for studying E2 binding properties in LCLs at the moment.

Chromatin binding properties of the E3 proteins was also investigated by several research groups. As already mentioned above, one study was conducted in Mutu III cells and revealed 7,044 E3 peaks but, due to antibody specificity issues, could not distinguish between E3A, E3B, and E3C peaks (McClellan et al., 2012). Furthermore, the Flag-HA-tagged E3A and E3C expressing LCLs were used for ChIP-seq experiments in LCLs, which could identify over 10,000 E3A (Schmidt et al., 2015) and over 13,000 E3C peaks (Jiang et al., 2014). These numbers very much resemble the findings presented in this thesis, while the Mutu III derived peak numbers differ noticeably. Even if Mutu III cells show a type III latency expression pattern, where all latent EBV proteins are expressed as in LCLs, not much is known about the chromatin landscape and TF expression pattern of these cells, which will turn out to be very important to chromatin accession by EBNA proteins (chapter 4.2.3), in comparison to LCLs. Therefore, it is not entirely clear how relevant the Mutu III derived data is when investigating E2 and E3 functions in immortalization and establishment of latency III.

5.2.2 Characterization of E2, E3A, and E3C binding sites in the EBV genome

Performing ChIP-seq experiments in LCLs infected with recombinant EBV strains made it also possible to investigate potential EBNA targeting of viral genomic sites. And indeed, 7 E2, 10 E3A, and 15 E3C binding sites could be identified (Table 20 and Fig. 14).

This finding was rather expected, since all three EBNA proteins were described to regulate viral genes. Early after infection E2 induces expression of viral genes *LMP1* and *LMAP2A/B* as

well as transcription from Cp, which gives rise to a polycistronic RNA coding for all six EBNAs (reviewed in Kempkes and Ling, 2015). Also the E3 proteins are known to regulate EBV transcription, but the current picture is still controversial: All three E3 proteins, E3A, E3B, and E3C, were described to repress E2 activated *LMP1* expression (Le Roux et al., 1994) but another study contrarily found E3C, in cooperation with E2, to induce *LMP1* expression (Lin et al., 2002). Furthermore, E3C was the only E3 protein so far, which was identified to bind to the *LMP1* promoter (Jimenez-Ramirez et al., 2006). E3A and E3C were also described to repress Cp derived transcription in reporter assays (Radkov et al., 1997, Waltzer et al., 1996), which could not be confirmed in EBV negative B cells with inducible E3C expression (Jimenez-Ramirez et al., 2006).

In this thesis, direct targeting of the bidirectional *LMP1/LMP2B* and the *LMP2A* promoter by E2 could be shown (Fig. 14, red columns). Also CBF1, the best described DNA adaptor, could be identified at both sites as well by re-analysis of published raw data (Zhao et al., 2011b). The B cell specific TF EBF1, which was described in this thesis to form complexes with E2 and was shown to be enriched at CBF1 independent E2 binding sites in LCL, was also found to bind to these promoters, with much higher enrichment at the *LMP2A* compared to the *LMP1/LMP2B* promoter (Fig. 38). Very recently a study on E2 binding, with the focus on comparing different latency states, identified E2, CBF1, and EBF1 binding sites in the viral genome as well (Lu et al., 2016). Here, only the two E2 binding sites at *LMP2A* and *LMP1/LMP2B* promoters could be identified, re-analyzing the data from Zhao et al., while the new EBF1 and CBF1 ChIP-seq experiments revealed a similar binding pattern as the ENCODE data shown in Fig. 38 (right panel). Furthermore, Lu and colleagues could show that E2 in fact recruits CBF1 and EBF1 to *LMP2A* and *LMP1/LMP2B* promoters as well as to Cp. Now, in this thesis it could be shown that also E3A and E3C are significantly enriched at the *LMP1/LMP2B* promoter while only E3C can be detected at the *LMP2A* promoter. Of the several TFs which were identified to correlate with E3 binding in the human genome (chapter 4.2.5), only BATF, BCL11A, and IRF4 were integrated in the EBV portal (Arvey et al., 2012) and therefore included in the IGV view (Fig. 38). BATF and BCL11A showed only moderate enrichment at these E3 peaks while IRF4 was not enriched. Due to the limited information on E3 associated TFs, predictions on E3 accession to DNA is very difficult in this case. The PU.1 signal track was also integrated in this comparison since PU.1 has been described to be important for E2 driven activation of the *LMP1* promoter (Laux et al., 1994b, Laux et al., 1994a, Johannsen et al., 1995), where E3C was characterized as a co-activating factor (Zhao and Sample, 2000, Lin et al., 2002) which targets the *LMP1* promoter as well (Jimenez-Ramirez et al., 2006). The data derived from ENCODE also showed PU.1 binding to the *LMP1/LMP2B* but not to the *LMP2A*

promoter. Therefore, a potential role of PU.1 in *LMP1* gene regulation seems very plausible and has to be further addressed in functional assays in the LCL system.

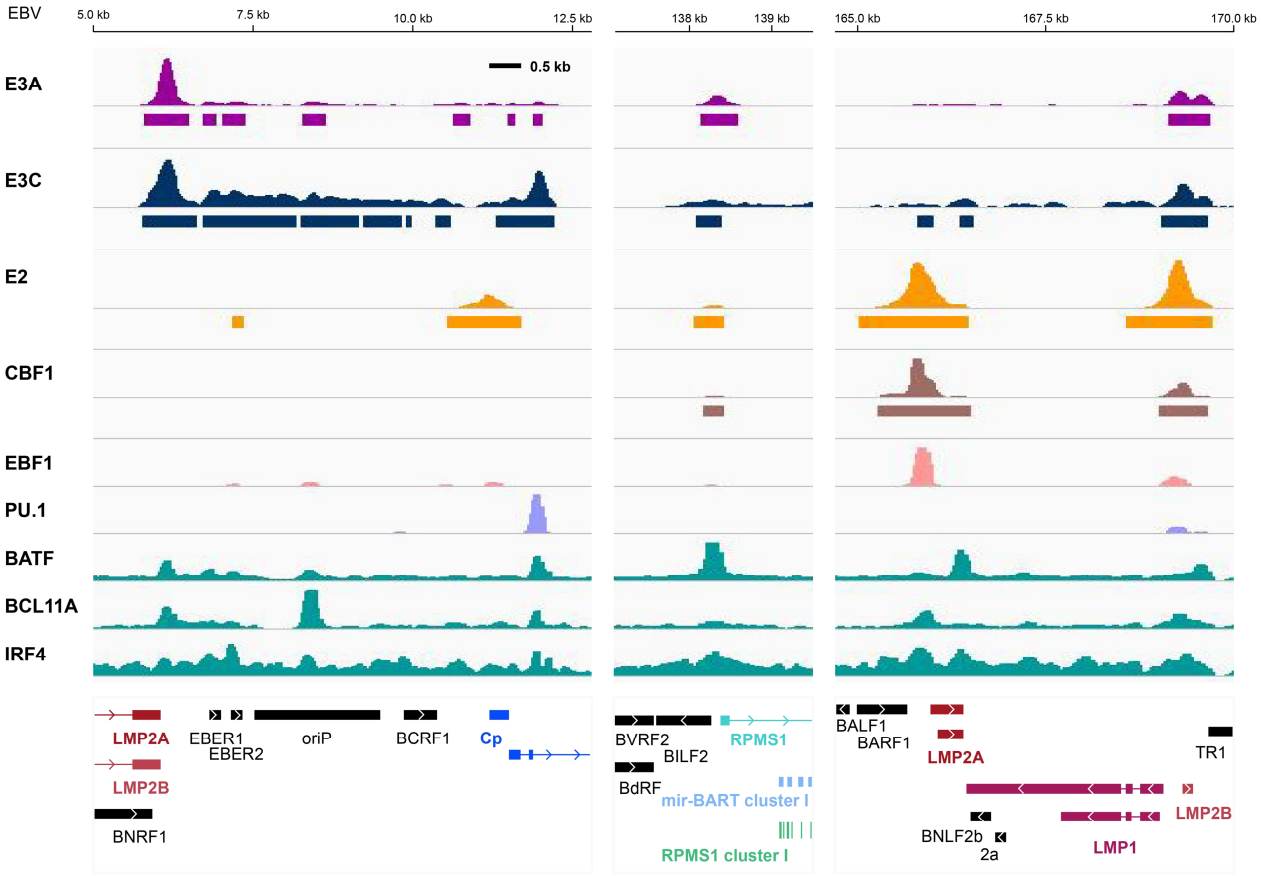


Figure 38. E2, E3A, and E3C binding sites in the EBV genome and co-occurrence of associated TFs. Schematic maps depicting three details of the EBV genome (HHV-4 type I, NC_007605.1, map provided by the EBV portal (Arvey et al., 2012)). Genes expressed during the lytic cycle are depicted in black and genes expressed during latency are highlighted in color. Also marked is the EBNA regulated Cp, which gives rise to different (polycistronic) splice variants coding for all EBNAs, including proteins of interest E2, E3A, and E3C. EBNA regulated *LMP1*, *LMP2A*, and *LMP2B* genes are shown in red. In the upper panels ChIP-seq signal profiles and underneath peaks called by MACS2 for E3A, E3C, E2, and CBF1 (Zhao et al., 2011b) are shown. Signal tracks of TFs EBF1, PU.1, BATF, BCL11A, and IRF4 were directly uploaded to IGV through the EBV portal server and are derived from ENCODE ChIP-seq experiments in GM12878, analyzed by the Lieberman group as described (Arvey et al., 2012). All signal tracks were set to show maximal intensities of the respective ChIP-seq signal within the genome.

Direct targeting of Cp in LCLs by E2 could be shown in this thesis (Fig. 38, left panel) as well as by Lu et al. using ChIP-qPCR. Re-analysis of CBF1 ChIP-seq data did not show significant binding sites in the Cp region. It was not possible to assess if this finding was due to low overall read coverage or has biological relevance. Since E2 activation of Cp has been studied and confirmed extensively, the former seems more likely. Also Lu and colleagues could detect significant binding of both, CBF1 and EBF1, at Cp in ChIP-qPCR experiments in LCL (Lu et al., 2016), which indicated CBF1 and/or EBF1 as E2 adaptors in this case. This time, E3A and E3C could be detected in close proximity, but not at the very same site as E2 and rather showed

enrichment over a larger genomic region of approx. 6.5 kb which is atypical for most TFs. The oriP region is located within this signal stretch and consists of highly repetitive sequences, which might lead to false positive repetitive ChIP signals. However, this was not the case for E2 or other TFs and therefore most likely displays E3 specific binding behavior. This finding might be due to higher order spatial organization of this genomic region mediated by E2 and the E3 proteins in order to tightly regulate Cp transcription but has to be further analyzed. The E3 associated factors BATF and BCL11A showed enrichment at the region boarders, while IRF4 shows no specific enrichment. PU.1 could be identified at only one boarder of the E3 stretch, 3' of Cp, and therefore might be involved in E3 but not E2 chromatin accession which would have to be verified by functional assays.

Also all four investigated TFs could be identified to bind at the promoter of the full length transcript of *RPMS1* (Fig. 38, middle panel), a putative ORF whose translation to a protein could not be confirmed to date but gives rise to *BART* ncRNAs and *BART* miRNAs. The *BART*s are forming three clusters in the *RPMS1* introns, which are largely deleted in the EBV B95.8 background. Only a few *BART*s of cluster I and two further downstream are still present. *BART* RNA can be detected during latent and lytic cycles of EBV infection but are found to be expressed at especially high levels in latency II. The molecular functions of *BART* ncRNAs is still to be determined but due to exclusive expression in the nucleus and no evidence of protein expression from several splicing variants, a role in viral or host gene regulation seems likely. The *BART* miRNAs show supporting functions in viral latency by targeting viral and cellular factors crucial in cell growth, survival and signaling pathways, but also cellular factors important in anti-viral immune responses (reviewed in Skalsky and Cullen, 2015). So far, no influence of E2 or E3 function on *BART* expression was reported to date. The signal enrichment for E2, E3A, E3C, and CBF1 is quite low compared to the other EBV genomic binding sites, but displays a significant peak for each factor at this co-occupied site. Also EBF1 was shown to be significantly enriched at this site, while PU.1 was not detected here. BATF was the only E3 correlating factor which was enriched here. In summary, a model where E2 is recruited to chromatin by CBF1 and EBF1 while E3A and E3C are recruited (i.a.) by BATF seems very likely.

5.2.3 E2, E3A, and E3C preferentially target enhancer modules in the human genome

In this thesis it could be demonstrated that E2 as well as E3A and E3C primarily target enhancer regions within the human genome in LCL and not, as suggested for some time in the past, mainly promoters. To this end the css analysis in the wt LCL GM12878 by ENCODE (Ernst et al., 2011), which segments the human genome into functional elements in a cell line specific manner,

was applied to assign functional states to binding sites. This analysis revealed that 65.9% of E2 and even 71.0 and 67.6 % of E3A and E3C binding sites, respectively, are located on enhancers (Fig. 15A). Interestingly, the percentage of E2 peaks at strong enhancers (47.0%) is higher than the ones of E3A or E3C (43.1 and 41.2% respectively). This phenomenon was further dissected and confirmed by enrichment analyses of enhancer defining histone modifications at E2 compared to E3A and E3C peaks (Fig. 16). All three investigated chromatin marks, H3K4me1 and H3K4me3, characteristic for enhancers, and H3K27ac, typical for active enhancer elements, as well as RNA polymerase II (PolII), indicating poised or actual transcription, were elevated at E2 peaks compared to E3A or E3C binding sites. Furthermore, this evaluation included the average signal distribution of the investigated factors at their respective peaks, allowing conclusions on the scale of the identified signal enrichments. Hence, it could be demonstrated that E2 peaks show higher and also broader H3K4me1 and H3K27ac signals as the average H3K4me1 or H3K27ac positive site respectively, which was far not as pronounced at E3 peaks.

Further examination of the EBNA binding sites at promoter regions, as predicted by css in GM12878, revealed a striking absence of annotated promoters by RefSeq in the majority of these peaks (Fig. 15B). According to RefSeq only 4.7% of E2 and 1.4 and 0.9% of E3A and E3C peaks, respectively, formerly annotated by css as promoter associated, are located within 1 kb upstream of a RefSeq gene. This can be explained by the criteria of css to annotate promoters, which are based on histone modifications and PolII occurrence, but not annotated genes of any kind. Therefore these “css only promoters” are most likely enhancers, which are frequently transcribed (reviewed in Plank and Dean, 2014, Kulic et al., 2015).

Taken together, these findings demonstrate that all three investigated EBNA proteins primarily target enhancer elements and E2 in particular is binding to strong enhancers, exhibiting high H3K4me1 and H3K27ac marks, while E3 proteins rather bind to regular enhancers. However, a significant percentage of E2 binding sites were also identified at RefSeq promoters, indicating a role of E2 in promoter targeted gene regulation for a subset of E2 peaks.

In relative distance analyses of E2, E3A, and E3C peaks and their respective induced or repressed genes, as identified by gene expression analyses in the Kempkes laboratory (Maier et al., 2006, Hertle et al., 2009, and diploma thesis A. Nowak, 2008), direct targeting of E2 induced genes by a small subset of E2 peaks could be revealed. This finding is consistent with a small subset of E2 peaks (4.7%) which are located at RefSeq gene promoters. This feature could not be observed towards E2 repressed genes and also E3A and E3C peaks seemed not to be shifted nearer towards regulated genes as expected by random distribution. Only when focusing on highly enriched E3A or E3C peaks a slight shift of shorter relative distances towards repressed target genes could be observed. Thus, a model emerges, in which the majority of E2 and also E3

proteins target enhancers and only a small subset of E2 and maybe E3A and E3C binding sites are located directly at the targeted gene (Fig. 15B).

E2 was already described to target enhancer elements, conjointly with CBF1, rather than promoters (Zhao et al., 2011b) and also genome wide studies on E3 binding sites in Mutu III revealed promoter distal binding (McClellan et al., 2012). Very recently, further studies of E3A and E3C binding sites in LCLs showed a similar picture, where enhancer targeting was outlined (Schmidt et al., 2015, Jiang et al., 2014, Wang et al., 2015). However, these studies did not include quantitative analyses and direct comparison of E2 versus E3A and E3C binding sites features which was only provided by this thesis.

In order to draw conclusions on the connections between binding sites of the single EBNAs and their target genes, experiments identifying three dimensional chromatin organization dependent on EBNA protein expression, should be performed. *Chromosome Conformation Capture* (3C), an assay to reveal chromatin interactions but only for distinct regions of interest, was already applied in order to reveal E3A or E3C mediated or inhibited chromatin loop formation at three model genomic loci harboring described target genes (McClellan et al., 2013). The authors described one promoter-enhancer interaction which is inhibited by E3A expression and presumable enhancer binding as well as two different E3C mediated repressive promoter-enhancer interactions. However, in that study many different cell lines were used, among those also Burkitts' lymphoma cell lines, which exhibit a chromatin landscapes very different from LCLs (discussed in chapter 5.3.1), and therefore are not displaying the ideal background for these interaction studies. Furthermore, many different more advanced methods have been developed in the recent past in order to study genome wide chromatin interactions. One example displays high-resolution capture Hi-C (Chi-C), which detects long range interactions preselected for promoter regions and was initially applied to investigate differences in promoter interactions between CD34+ hematopoietic progenitor cells and GM12878 LCLs demonstrating changes during differentiation processes (Mifsud et al., 2015). Therefore, information on promoter interactions in wt LCL are already available and have been used to formulate hypotheses on possible EBNA mediated interactions by combining these with E2 binding sites (Gunnell et al., 2016). However, a genome wide comparison of promoter-enhancer interactions dependent on the different EBNA proteins has not been performed yet and would certainly shed light on the connection between binding sites and target genes.

5.2.4 Enhancer signature is a prerequisite for accession of E2 to chromatin and is enriched upon E2 expression

The targeting of enhancers by all three EBNA proteins has been discussed extensively above but did not consider the presence of enhancer specific histone modifications as a prerequisite for or consequence of EBNA binding. In this thesis it could be demonstrated that enhancer signature is not only a prerequisite for E2 binding but also increases upon E2 expression. To this end ChIP-seq data on histone modification marks derived from CD19+ B cells (Bernstein et al., 2010) and LCLs (ENCODE_Consortium, 2012) were analyzed in the bioinformatic analysis pipeline designed and described in this thesis. Only E2 binding was studied in this context since E2 is the first latent EBV protein to be expressed upon infection and therefore able to access chromatin in the resting B cell prior to E3 protein expression which might interfere with E2 function. The enrichment of H3K4me1 at E2 peaks has been described previously (Zhao et al., 2011b) and the conclusion was drawn that E2 targets pre-existing enhancers in primary B cells. However, the mentioned study did not include any kind of quantitative assessment of the observed ChIP-seq signals, which have been derived from different laboratories applying different experimental features and antibodies and are provided by the ENCODE consortium.

The data presented in this thesis represent a profound and detailed examination of these two data sets with focus on E2 binding sites, including a normalization procedure which made it possible to compare CD19+ with LCL derived experiments. The enrichment of each histone modification and DNaseI HS was quantified relative to the absolute signal in the respective cell line and the increase of H3K4me3, H3K27ac, and DNaseI HS signals in LCLs compared to CD19+ cells could be demonstrated. Interestingly, H3K4me1, the enhancer hall mark, is not further enriched upon E2 expression in LCL and already shows are broad signal distribution. Therefore, E2 targets enhancers exhibiting broad H3K4me1 marks and subsequently might recruit factors which further open the chromatin and possibly even recruit PolII to initiate transcription.

5.2.5 Distinct combinations of cellular TFs characterize E2 versus E3 predominated chromatin regions

This thesis focused on the comparison between E2 and E3 modes of action and in particular the mechanisms and prerequisites for chromatin accession. Since E2, E3A, and E3C share a certain set of target genes, mostly in a counter-regulated manner, but also show uniquely regulated genes (Fig. 4) a co-occupation of binding sites in the human genome seemed very likely. It could be demonstrated that EBNA binding sites are shared to a certain degree, 27.4% of E2 sites are

positive for at least one E3 protein and vice versa 43.2% and 43.5% of E3A and E3C peaks, respectively, are E2 positive (Fig. 19). Furthermore, it was shown that the overlap of E2 and CBF1 binding sites (61.6% of E2 sites) was more significant than the overlap of E3A or E3C and CBF1 (45.4% and 45.3% respectively).

Different studies conducted in different cell lines and laboratories showed a partially similar picture: E2 binding sites were also shown to largely overlap with CBF1 sites (Zhao et al., 2011b), while the overlap between E3A or E3C with CBF1 binding sites was calculated to be smaller than shown here (16 and 16% respectively)(Jiang et al., 2014, Schmidt et al., 2015). This discrepancy is most likely due to the fact that for this analysis only the top 10,000 CBF1 binding sites as defined by enrichment were used for these analyses and low enrichment peaks were neglected. The intersection analysis of E2 and E3 proteins in MutuIII cells, which did not distinguish between the different E3s, revealed 25% of combined sites to be shared (McClellan et al., 2013). Studies in LCLs showed that only 9% of E3A and 9% of E3C sites were E2 positive and only 44% of E3A sites were described to be E3C positive (Jiang et al., 2014, Schmidt et al., 2015). Again, not the whole set of identified binding sites were used for this analysis in LCLs but rather the top enriched sites were analyzed.

5.2.5.1 Quantitative analysis of signal enrichment at binding sites as a novel strategy of determining possible interacting TFs

Now, the binding site co-occupation of E2 and E3 proteins was described and characterized in a quantitative and genome wide way for the first time. A new picture emerged when binding sites were not only compared for binding site overlaps, but sorted according to their signal enrichment and correlated with the signal enrichment of other factors. E2 and CBF1 showed very high signal intensity correlations at E2 peaks (Fig. 20A) as it could be shown for E3A and E3C at the respective other peak set (Fig. 20B and C). This analysis was extended to a genome wide scale including all TFs which were analyzed by ENCODE in LCL at that time (ENCODE_Consortium, 2012). A genome wide pattern of TF binding networks emerged (Fig. 21) which revealed TFs with high correlations to the EBNA proteins. Two subclusters could be identified; one included E2 and CBF1 and the other one E3A and E3C. TFs with the highest correlation values towards at least one EBNA protein were chosen for further analyses: CBF1, EBF1, and CUX1 showed very high correlation to E2, while 16 TFs correlated highly with both E3 proteins (Fig. 22). Also, the signal correlation between E3A and E3C was comparable to the ones of known dimers or members of the same protein complex, indicating a conjointly binding mechanism which could not be revealed by simple intersection analyses before.

Furthermore, the co-occupation of E2 and E3 binding sites could be characterized in more detail by including signal intensities for the analyses. An anti-correlation of E2 and E3A or E2 and E3C signals, together with E2 or E3 associated factors, at EBNA binding sites could be demonstrated and suggests reciprocal binding of most sites rather than actual shared sites (Fig. 23).

Some of the TFs identified to correlate with either E2 or E3 binding pattern have already been described as co-occurring TFs based on a panel of TFs defined by educated guesses and co-citations or motif enrichment analyses. In contrast, this thesis displays the first unbiased study including a very big data set on TF binding without any preselection of possible interacting factors. CBF1, the best described cellular protein to interact with all three EBNAs, could be assigned to correlate definitely with E2 over E3 signals in an unbiased quantitative approach for the first time. EBF1 has been suggested as co-occurring TF important for E2 binding by motif enrichment analysis at E2 peaks and subsequent peak overlap analysis (Zhao et al., 2011b). Very recently EBF1 has been proposed as recruiting factor for E2 (Lu et al., 2016). CUX1 on the other hand has previously not been described as TF related to E2 binding. Also, most factors of the E3 cluster have been discussed to be important for E3 accession to DNA or mediating specificity for E3 binding sites. These assumptions are mainly based on binding co-occurrences as determined by peak overlap analyses (McClellan et al., 2013, Jiang et al., 2014, Schmidt et al., 2015, Wang et al., 2015). However, only the interactions of E3A with BATF as well as E3C with IRF4 were studied in more detail: E3A and BATF binding at the same genomic region was confirmed by ChIP-re-ChIP-qPCR analysis (Schmidt et al., 2015), which only proofs the presence at the same chromatin fragment but not direct or indirect binding to each other. The direct interaction of IRF4 and E3C could be confirmed and mapped to E3C aas 130-159 (Banerjee et al., 2013). Only TFs CEBPB, MTA3, and PML have not been discussed previously to be important for chromatin accession of E3 proteins and display a novel piece of information.

5.2.5.2 Cluster analyses for E2 or E3 binding sites revealed subsets defined by combinatorial TF sets

After the quantitative approach described above to identify EBNA associated TFs, these were used for cluster searches of combinatorial TF co-occurrences. To this end E2 and E3 peaks were analyzed separately, peak intersection analyses including the previously identified EBNA correlating TFs were performed, and clusters were identified. For E2 and E3 peaks eight different clusters of defined TF compositions could be identified which are characterized by distinct histone modification patterns.

E2 peak clusters

The E2 peak clusters included combinations of TFs CBF1, EBF1, and CUX1 (Fig. 26). The highest E2 enrichment was observed for cluster I, which is positive for all three investigated TFs and shows histone modifications characteristic for active enhancers. This implies that the strongest enhancers with a combined composition of all three TFs display the most ideal E2 chromatin accession prerequisites. Cluster VII, which is devoid of EBF1 binding but shows the highest enrichments for H3K4me3 and PolII, indicating the presence of transcription, probably displays the E2 peaks subset comprised of mainly transcribed promoters and enhancers. Clusters II, III, VI, and VII are very similar in their chromatin signature and display regular enhancers bound by E2, while E2 sites of clusters IV and V are most likely poised enhancers, due to the low but present H3K4me1/3 enrichment and elevated H3K27me3 levels. However, cluster V also shows the highest percentage of sites located in heterochromatin, as defined by ENCODE css. These binding sites could be due to indirect chromatin interactions or might display actual targeting of heterochromatin by E2 for a subset of binding sites. Furthermore, clusters V and VI, which are both negative for CBF1 and EBF1, show the lowest E2 signal enrichment and outline the importance of these two factors on E2 binding and indicate an improving character. Recently, it was suggested that E2 in fact recruits these two factors to certain E2 target sites in order to access chromatin (Lu et al., 2016), a hypothesis which will be further discussed in the following chapter 5.3.

Furthermore, the whole set of TFs investigated by ENCODE was also assessed for the E2 peak clusters and revealed that indeed the majority of TF which were enriched at all E2 sites, showed the highest enrichment for clusters I and/or VII, while a depletion was apparent for clusters IV and V. This finding underlines the enhancer and promoter characteristics of clusters I and VII respectively, which are co-occupied by several TFs whose combinations of appearance probably determine the accessibility of discrete genomic loci for E2 binding.

Interestingly, the E2 clusters also differ in enriched sequence motifs (Fig. 27) which display another reference point for determining E2 binding sites. The enrichment of NFκB, EICE and ISRE motifs at cluster I peaks indicates a supportive role of these factors for E2 chromatin accession, since E2 showed the highest signal enrichment for this cluster. Interestingly, the NFκB motif is noticeably enriched at the four EBF1 positive clusters, which might be due to a potential interaction or support in mediating E2 specificity. Cluster VII which potentially depicts E2 accessible promoters reveals a role for TFs of the ETS family, like PU.1 and Spi-B, in this context. PU.1 was not included in the cluster search since it was already excluded after the genome wide (Fig. 22) and the EBNA peak wide (Fig. 23) correlation analyses due to a low or even anti-correlation with E2 signal. However, PU.1 was described to be important for mediating

E2 targeting of the LMP1 promoter (Laux et al., 1994a, Johannsen et al., 1995), and was also considered as E2 specificity mediating TF in genome wide analyses by peak comparison analyses which revealed an overlap of 22% (Zhao et al., 2011b). More recently the existence of EBNA controlled super-enhancers, characterized by disproportionately high enrichment of enhancer marks, PU.1, E2, E3A, and E3C was discussed (Zhou et al., 2015) but seems to be restricted to a very small set of binding sites (187), which also show extraordinarily high enrichment for the vast majority of TFs of the ENCODE set and therefore represents only a special case of chromatin accession by E2. Enrichment analyses for PU.1 at the eight different E2 peak clusters were performed as well (Fig. S4) and revealed the highest PU.1 enrichment at E2 peaks of clusters I and VII, like most investigated TFs. Thus PU.1, maybe also in combination with or substituted by other ETS TFs, might indeed be important for mediating E2 specificity of cluster VII sites, lacking EBF1. Also RUNX3, which was described as E2 co-occurring and potential factor for mediating chromatin accession (Zhao et al., 2011b, Portal et al., 2013) could not be confirmed to correlate with E2 binding in a general manner. It displays the highest enrichment at E2 clusters I and VII as well and therefore might indeed play a supporting role here.

The absence of the CUX1 sequence motif at all E2 binding sites indicates that either CUX1 does not only directly access DNA but is recruited to chromatin in an indirect fashion through other factors or CUX1 is able to bridge connections between distal chromatin regions and such indirect binding sites are also included in the binding site data set. The ability of CUX1 to regulate distant target genes was described previously (Vadnais et al., 2013) and could account for indirect chromatin binding sites.

E3 peak clusters

The E3 peaks could be divided into eight clusters as well by their combination of co-occurring TFs BATF, ATF2, BCL11A, FOXM1, NFIC, and IRF4 (Fig. 28), which exhibit the most similar binding patterns to E3 on EBNA peaks as identified in a first cluster analysis (Fig. 25). Cluster I, positive for both, E3A and E3C, as well as all six investigate TFs displays the highest enrichments for all enhancer defining histone modifications, E3A and E3C signals, implying a supporting function of the concerted presence of all factors on E3A and E3C binding. The clusters V, VI, VII, VIII consisting of E3A, E3C and different combinations lacking certain factors are associated with lower enrichment of enhancer marks. These clusters may represent weaker enhancers and E3 binding and might result in extenuated E3 mediated gene regulation. Cluster III, E3A and E3C co-occupied sites positive for only BATF and IRF4 display only very moderate enrichment of enhancer signatures. These sites could display a minimal prerequisite for E3 accessible enhancers. Clusters II and IV lack all six cellular TFs and exhibit the lowest E3A and E3C enrichment respectively. These sites might either represent indirect chromatin contacts

of the E3A/C complex or actual direct independent binding of E3A or E3C respectively at non-enhancer sites. However, the genome wide correlation analysis strikingly showed a very high positive correlation between E3A and E3C, which could otherwise only be observed for known dimers like BATF/IRF4 and MEF2A/MEF2C or members of the cohesion complex. In addition, the interaction of E3A and E3C could be demonstrated by Y2H (Calderwood et al., 2007) and Co-IP experiments in B cells (Paschos et al., 2012 and this thesis, Fig. 6). Therefore, a model in which E3A and E3C target chromatin as a heterodimer is favored here.

In Summary, these clusters of EBNA binding sites and associated TF combinations could represent a first list of prerequisites for E2 and E3 binding (Fig. 39) and a versatile starting point for further experiments to determine the functionality of the single involved factors. Finally, one might be able to describe such prerequisites of chromatin landscape and TFs combinations which determine E2 or E3 specificity.

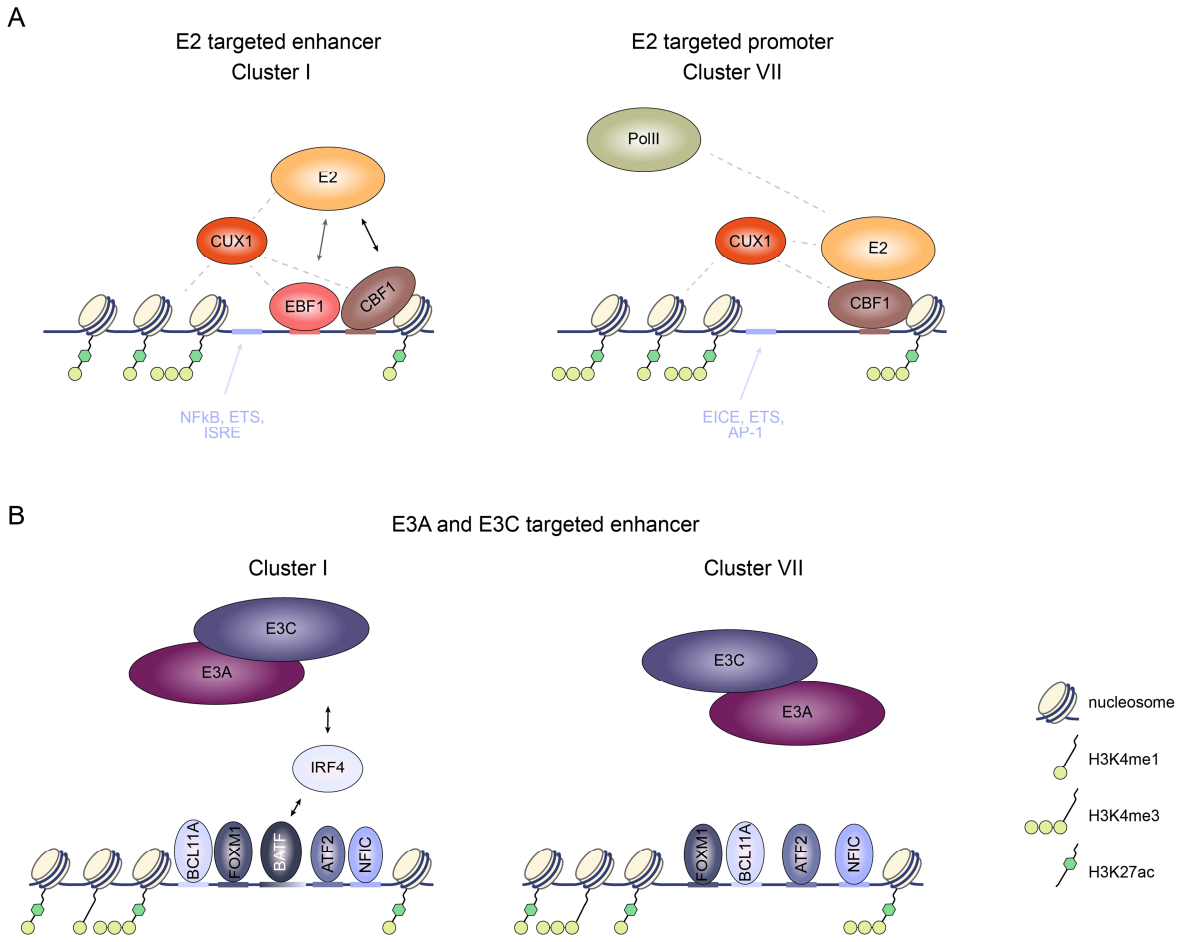


Figure 39. Hypothetical model of E2 and E3 targeted chromatin regions. (A) E2 is binding to enhancer regions, represented by cluster I (left panel), by the concerted action of CBF1, EBF1, and CUX1. CBF1 and EBF1 can directly access DNA and no interaction between these two TFs was described. The direct interaction of E2 and CBF1 has been demonstrated (Grossman et al., 1994, Henkel et al., 1994) (black arrow) and complex formation with EBF1 could be shown in this thesis (grey arrow). At promoter regions (right panel), represented by cluster VII, E2 is binding to DNA utilizing CBF1 and recruits PolII through yet unidentified factors. CUX1 is also important for

complexes of both clusters, yet the CUX1 binding motif was not found at E2 sites and indirect binding was already proposed in the past (Harada et al., 2008). Therefore it might bridge interactions between E2 and co-occurring factors. In both scenarios the co-occurring TFs as identified by the motif search also show signal enrichment at the respective clusters, yet they are not deterministic for cluster formation but might contribute to specificity. (B) E3A and E3C are binding to enhancers through combinations of the TFs ATF2, BATF, BCL11A, FOXM1, IRF4, and NFIC. The combined presence of all factors, as in cluster I, mediates the highest specificity for E3 proteins, while enhancers lacking several factors, like cluster VII, display lower E3 signals. Cluster VII shows that binding of E3 without BATF and IRF4 is still possible if substituted by the other four TFs. Yet, BATF and IRF4 display the minimal combination of TFs needed to mediate E3 specificity (cluster III). All six TFs specifying E3 clusters, can bind to DNA directly, mostly in (hetero) dimers and the conjointly binding of BATF and IRF4 to DNA has been shown (Glasmacher et al., 2012). However, it is not known in what combinations these TFs access DNA in this specific case.

5.2.6 B cell TF networks exploited by E2 and E3 proteins

In the previous section the E2 and E3 peak clusters and presence of distinct associated TFs were discussed and now shall be analyzed from a functional perspective.

CBF1 or RBPJ of the E2 cluster displays the most intensively studied and discussed TF and many functional aspects have been elaborated in the introduction (chapter 1.2.1.2). In this context it is very interesting that CBF1 was described to strongly correlate with dynamic NOTCH1 binding in T cells (Wang et al., 2014). Here, NOTCH1 function was induced and subsequently ChIP-seq for NOTCH1 and CBF1 performed and revealed approx. 10% of NOTCH1 peaks to be dynamic (only detectable upon induction) and predominantly located at enhancer sites. Interestingly, the CBF1 sites which correlate with NOTCH1 only appeared upon induction, thus indicating a stabilization of CBF1 binding to DNA by NOTCH1. Similar findings could be demonstrated in *drosophila* as well (Krejci and Bray, 2007).

EBF1 the *early B cell factor 1* is a sequence specific DNA binding TF, which plays a crucial role in defining B cell lineage specificity during differentiation and represses alternative cell fates. In concerted action with PAX5, PU.1, RUNX1, Ikaros, E2A, and FOXO1 the B cell specific transcription profile is established. EBF1 consists of a DBD, a helix-loop-helix dimerization domain, and a C-term transactivation domain and is highly conserved during metazoan evolution (reviewed in Boller and Grosschedl, 2014). The crystal structure of EBF1 which is binding to DNA as a dimer could be solved (Treiber et al., 2010a, Siponen et al., 2010). In gain-of-function and loss-of-function studies in pre-pro-B cells and pro-B cells, respectively, EBF1 was described to activate and repress genes associated with B cell function and EBF1 binding was associated with H3K4me2 (Treiber et al., 2010b). Furthermore, EBF1 was shown to induce DNA demethylation at the *CD79A* prom in plasmacytoma cells (Maier et al., 2004) and was linked to the chromatin remodeling complexes SWI/SNF and Mi-2/NuRD (Gao et al., 2009). Recently, EBF1 was also considered to act as a "pioneer factor" in order to establish B cell identity since its CTD, independent of its transactivating function, establishes chromatin accessibility and induces

DNA demethylation in previously naive chromatin (Boller et al., 2016). Therefore, E2 might well employ EBF1 to access important B cell lineage enhancers and drive B cell activation.
Moreover, a recent study reported E2 to recruit TFs CBF1 and EBF1 to its target sites rather than exploiting preexisting CBF1 and EBF1 positive enhancers to achieve target gene regulation (Lu et al., 2016). However, this study did show a significant decrease of both TFs, CBF1 and EBF1, at selected E2 binding sites upon E2 depletion but did not show an absolute abolishment of binding. Therefore, a second hypothesis, also supported by the dynamic NOTCH1 and CBF1 binding data, suggests complex stabilization of all factors involved by E2.

The third E2 correlating TF, CUX1, was described to act as activator and repressor of transcription, depending on the promoter context and expressed transcript variant. Several transcriptional roles of CUX1 in cell cycle progression, DNA damage response, and resistance to apoptotic signals could be demonstrated. Furthermore, several cancer links were described for CUX1, characterizing it as a haploinsufficient tumor suppressor gene (reviewed in Ramdzan and Nepveu, 2014). A consensus DNA binding motif for the CUX1 p110 variant was described to be enriched at genomic binding sites (ATCG/AAT) but also indirect DNA accession by protein-protein interactions was proposed (Harada et al., 2008) and promoter distal binding of target genes by CUX1, indicating enhancer binding, was described (Vadnais et al., 2013).

Interestingly, the CUX1 homologue in drosophila, Cut was described to be a downstream effector or target gene of Notch signaling, since Cut expression was lost in SuH (the drosophila homologue of CBF1) mutants and described to be activated or repressed by Notch function (Nepveu, 2001).

Of the E3 associated TFs, IRF4, a member of the *interferon regulatory factors* (IRF), is the best described so far. It is expressed in most cells of the immune system and during all developmental stages of B cell activation but during the germinal center (GC) reaction and plays a key role in late B cell differentiation. It was shown that IRF4 is upregulated by NFκB and represses BCL6, the master regulator of the GC reaction in GC B cells. Furthermore it was shown to play a role in class switch recombination and GC exit of centrocytes and aberrant IRF4 expression was linked to oncogenic pathologies like multiple myeloma, Hodgkin and Non-Hodgkin lymphomas. IRF4, which needs a cofactor to achieve DNA binding, was described as activator or repressor of transcription dependent on the interacting cofactor and context (reviewed in De Silva et al., 2012). The recruitment of IRF4 to DNA through ETS factors, like PU.1 and Spi-B, (Brass et al., 1999) or AP-1 family members, like BATF, (Glasmacher et al., 2012) could be demonstrated. Recently, the direct interaction of E3C and IRF4 was shown (Banerjee et al., 2013) and enrichment at E3C binding sites could be demonstrated (Jiang et al., 2014), underlining the

importance for E3C and E3A chromatin accession. Taken together, since IRF4 displays such a crucial TF in B cell development, again an essential B cell TF network is targeted by the EBNA proteins.

Also BATF (*basic leucine zipper* (bZIP) *TF ATF-like*) displays a TF with expression restricted to the hematopoietic system and belongs to the AP-1 family of TFs. Unlike AP-1 factors Fos or Jun, BATF is missing a transactivation domain and therefore is dependent on interacting factors for mediating transcriptional regulatory functions. BATF was described to heterodimerize with Jun and conjointly acts as repressors of transcription (Murphy et al., 2013). Recently, the recruitment of IRF4 to AICE composite sites by JUNB-BATF heterodimer could be shown and therefore allow an additional dimension of binding site specificity (Glasmacher et al., 2012). Unfortunately, JUNB was not included in the ENCODE ChIP-seq TF set used in this thesis and therefore it cannot be concluded what BATF heterodimer is recruiting IRF4 to composite sites. However, it was demonstrated that these BATF-IRF4 interactions display a crucial mechanism which is utilized by the E3 proteins to access specific regulatory regions and a mechanism by which E3A is tethered to DNA involving BATF was suggested analyzing binding data (Schmidt et al., 2015).

The remaining TFs of the E3 cluster were all described to co-occur at E3A or E3C binding sites by peak overlap analyses (Jiang et al., 2014, Schmidt et al., 2015, Wang et al., 2015).

ATF2, which is also a member of the AP-1 family of TFs, characterized by a bZIP domain, forms homo- or heterodimers with AP-1 members, like c-Jun, in order to specifically regulate target gene transcription. Furthermore, an oncogenic transformation potential was attributed to Jun-ATF dimers (reviewed in van Dam and Castellazzi, 2001). Jun-ATF activity is specifically enhanced by *Jun N-terminal Kinase* (JNK) members of the *Mitogen-Activated Protein Kinase* (MAPK) pathway, in contrast to Jun-Fos dimers, which are rather ERK targets (Karin et al., 1997, Davis, 1999).

BCL11A (*B cell chronic lymphocytic leukemia/lymphoma 11A*) displays a zinc-finger TF, which was identified as a protooncogene, frequently implicated in numerous B cell malignancies (Satterwhite et al., 2001). Initially, it was described as a crucial and specific factor for B cell lymphopoiesis (Liu et al., 2003, Yu et al., 2012). Later a deterministic role in plasmacytoid dendritic cell fate was shown and cell line specific binding sites were identified, which harbor the same consensus motif as GM12878 cells (determined by ENCODE, accessible via factorbook.org): EICE (Ippolito et al., 2014).

FOXM1 (*Forkhead (FKH) box protein M1*) is a member of the FOX family that consists of more than 50 proteins and was described to be important in cell cycle regulation and progression (reviewed in Carlsson and Mahlapuu, 2002) and therefore contributes to the pathogenesis of

several cancers (reviewed in Myatt and Lam, 2007). FOXM1 was described to act, in concert with MYB, as a master regulator of proliferation in germinal centers (Lefebvre et al., 2010). A recent study could show that FOXM1, which *in vitro* binds the FKH consensus motif, is specifically recruited to chromatin through co-factor interactions by direct binding to non-canonical DNA motifs (Sanders et al., 2015).

NFIC (*Nuclear Factor I C*), displays a member of the NFI family of site-specific TFs described to activate or repress transcription and bind DNA as dimers (reviewed in Gronostajski, 2000). The in vitro identified consensus DNA binding motif (Osada et al., 1996, Roulet et al., 2002) could be verified by analyses of ChIP-seq experiments (Bailey and Machanick, 2012).

In summary, 3 TFs could be described to define E2 clusters of peaks, while 6 distinct factors are responsible for E3 cluster formation. For both clusters distinct combinations of these TFs could be described. Recent advances in NGS methods and large data set comparisons made it possible to shed light on the importance of TF networks in mediating specific target gene regulation. For instance, combinatorial interactions of TFs, rather than the actions of single factors, were described to direct tissue-specific gene expression and determine cell fate (Ravasi et al., 2010). Here, the network structure was described to be dominated by facilitator TFs expressed broadly across tissues tended to interact with tissue restricted TF (specifiers) to result in specific functional consequences.

Also the formation of tissue specific enhancers was found to be dependent on distinct collaborative and hierarchical binding of TFs. A model was proposed where pioneer TFs, which are able to bind their recognition motif within compacted chromatin, already act in concert with further lineage specific TFs to select tissue specific enhancers and jointly displace nucleosomes. In a second step broadly expressed TFs mediate the actual enhancer function to activate distal target genes (reviewed in Heinz et al., 2015). One example for such a lineage defining and pioneer TF displays PU.1, which is required for the development of macrophages and B cells and influences the establishment of distinct gene expression programs in each cell type (Scott et al., 1994). However, PU.1 targets different sites in B cells compared to macrophages and it could be shown, that these differing binding sites were characterized by a set of B cell or macrophage specific TFs (Heinz et al., 2010). In the B cell EBF1, E2A, and OCT factors were found to be enriched at PU.1 sites, while in macrophages CEBP and AP1 factors could be identified. The corresponding motifs were found in close proximity to PU.1 motifs, indicating ternary protein-protein-DNA interactions and led to the conclusion that lineage defining TF composition might be a contributing factor to the formation of transcriptionally active and active genomic compartments (Pham et al., 2013).

In the LCL background PU.1 can be frequently identified to co-occur at E2 and E3 peaks, though it does not correlate significantly in signal intensity distribution Therefore, PU.1 could display a lineage restricted TF, or specifier, which collocates with more broadly expressed facilitator TFs, like EBF1 in the E2 or BATF, IRF4 or others in the E3 cluster, as it has been described for TF networks which determine lineage identity (Ravasi et al., 2010). Also, the fact that EBF1, BATF, IRF4, ATF2, and CBF1 are expressed at relatively high levels in GM12878 (Table S3) supports this argument. Interestingly, IRF4 displays the top expressed gene in GM12878 overall, as identified by CAGE (Fantom_Consortium et al., 2014) (Table S4). Actually, IRF4 was described to be expressed at low levels in the activated B cell and only re-expressed in the plasmablast stage at high levels (reviewed in Nutt et al., 2015). LCLs have been described to resemble activated B cells in their phenotype and expression pattern (Thorley-Lawson, 2001). However, this apparently does not apply to IRF4 expression, which is also induced by E2 action (DG75 expression data, S. Thumann) and IRF4 protein is stabilized by E3C (Banerjee et al., 2013). Hence, E2 might induce one of the cellular TFs mediating EBNA binding specificity.

Therefore, it seems very likely that E2 and later E3 proteins exploit B cell specific enhancers, which are already primed in CD19+ B cells, to achieve gene regulation of specific target genes usually triggered upon B cell activation, where specificity is mediated by the distinct sets of the co-occurring TFs. Finally, the clusters identified in this thesis display a first step on the way to identify deterministic features and prediction of E2 and E3 binding to chromatin.

5.3 CBF1 as a determining factor for E2 access to chromatin?

In the third part of this thesis, the dependency of E2 on CBF1 for binding chromatin was further elaborated. The usage of stable conditionally E2 expressing EBV negative DG75 B cell lines, which are proficient or deficient for CBF1 (Fig. 29), allowed a profound conclusion on CBF1 effects on E2 binding.

5.3.1 The effect of cell line specific chromatin signature and TF expression profile on E2 binding

Interestingly, the comparison of E2 peaks in DG75/CBF1 wt with the data from LCL revealed a difference in the chromatin landscape and therefore resulted in different E2 binding patterns between these two cell lines (Fig. 31). It could be show that the E2 binding sites present in both cell lines (*LCL/DG75 shared*) display the highest E2 enrichment in LCLs, while the E2 signal in DG75 is comparable between these LCL/DG75 shared and DG75 unique sites. However, the investigated enhancer characteristic histone modifications in both cell lines were enriched the

most at LCL/DG75 shared sites, followed by the unique E2 sites in the respective cell line. These findings showed that the E2 peaks detectable in both cell lines are in fact strong enhancers in the respective cell line. Also, this implicates that both, LCL and DG75 unique E2 sites, are indeed enhancer regions accessible for E2 binding only in the respective cell line. This conclusion could be one possible explanation for the different sets of E2 sites in those two cell lines. Furthermore it could be shown that the strong E2 binding sites from LCL which are associated with strong enhancer signatures are in fact the ones, which can still be identified in DG75. In contrast, the E2 binding intensities in DG75 are very similar between LCL/DG75 shared and DG75 unique sites which is most probably due to a very different enhancer distribution and intensities in this cell line. Moreover, expression patterns of E2 or E3 associated TFs in DG75, and in particular the impairment in BATF and IRF4 expression in DG75 cell line (Fig. 32), indicates a special role for these factors in mediating binding site specificity for E2. This might either be due to pioneering activity, in concert with other cellular factors, to ensure accessibility in the first place or they display cofactors in stabilizing chromatin binding per se. However, since the majority of E2 sites in DG75 are also present in LCL, and there they display the subset of strong binding sites associated with strong enhancer signatures, the ChIP-seq data derived from DG75 exhibits very important information.

5.3.2 CBF1 displays the key adaptor for E2 access to chromatin

The comparison of E2 binding sites in DG75/CBF1 wt and CBF1 ko revealed a strong dependency of E2 on CBF1 (Fig. 33), since 86.4% of E2 sites are lost in the ko situation. Also the analysis of E2 signal strength comparing the two situations showed a direct supportive effect of CBF1 on E2 binding for the first time. Furthermore, the examination of E2 peak subsets showed that the CBF1 independent E2 sites exhibit the strongest E2 signal in the CBF1 wt as well as in the CBF1 ko, although the mean signal in CBF1 wt is much higher than in the CBF1 ko. Investigation of these E2 peak subsets in LCL revealed a similar pattern, where the CBF1 independent peaks showed the highest signal. Even 28 CBF1 ko unique E2 peaks could be detected, which might be explained by peak detection thresholds, since they show a lower overall signal in the CBF1 ko line than the CBF1 independent peaks and still show a slight but not significant enrichment in the CBF1 wt cell line.

Since E2 signal strength was associated with strong enhancer signatures in LCLs, one explanation for the CBF1 independent peaks to exhibit stronger E2 signals in DG75/CBF1 wt might be due to a difference of enhancer signatures in DG75 for these two subsets. To stress this idea, data on histone modification patterns in DG75 (Kretzmer et al., 2015) were analyzed independently in the course of this thesis and E2 peak subsets were compared (Fig. 34).

Interestingly, the distribution of H3K4me1, H3K4me3, and H3K27ac was almost identical for CBF1 independent and dependent peaks, implying that not the DG75 cell line specific chromatin signature is responsible for higher E2 signals at the CBF1 independent peaks in the CBF1 wt line. Instead it becomes more evident that co-occurring factors besides CBF1 are responsible for strong E2 binding at enhancers.

5.3.3 EBF1 as a determining factor for E2 binding site specificity?

It could be shown that the consensus motif of the TF EBF1, which is essential for B cell lineage specification (reviewed in Hagman et al., 2012, and Boller and Grosschedl, 2014), was the only identified TF to be significantly enriched at CBF1 independent E2 binding sites (Fig. 35A). CBF1 dependent E2 sites were enriched for CBF1 and the EBF1 motif, but with less significance. In addition, the investigation of TF binding signal derived from LCL at these E2 peaks revealed a strong enrichment of EBF1 signal at CBF1 independent peaks (Fig. 35B and C). EBF1 already showed the highest correlation coefficient in the comparison with E2 signal distribution on a genome wide level (Fig. 21/22) and also when concentrating on EBNA peaks (Fig. 23). The correlation between E2 signal and EBF1 pattern became even more obvious when a special focus was directed on E2 peaks only (Fig. 36). The correlation between E2 and EBF1 (r_s = 0.42) almost scored the one of E2 and CBF1 (0.50), while the often discussed B cell lineage defining TF PU.1 only displayed moderate correlation with E2 signal distribution at E2 peaks. Combined these data strongly support the importance of EBF1 and CBF1 in mediating E2 accession of specific chromatin sites, while PU.1 might be important in priming enhancers in B cells to be accessibly for further TFs in the first place but does not recruit or stabilize E2 binding.

Eventually, the protein-protein interaction of E2 and EBF1 could be demonstrated for the first time in Co-IP experiments upon EBF1 transfection and E2 induction in DG75$^{doxHA\text{-}EE2}$ cells (Fig. 37). Complex formation could be detected in CBF1 wt and ko situations which demonstrated that CBF1 is not mediating E2-EBF1 interaction. These experiments now have to be repeated using E2 as IP target to confirm this interaction. Also, Co-IP experiments are not harboring information on direct interaction but interaction partners of complexes, directly or indirectly binding to each other, also mediated by DNA molecules can be identified. Therefore pull-down experiments applying heterologous expressed purified proteins need to be performed to distinguish between these scenarios. However, these data could demonstrate the complex formation of E2 and EBF1, also in the absence of CBF1, and underline the importance of this interaction.

In summary, it could be demonstrated that CBF1 displays the key adaptor for E2 in accessing chromatin and that EBF1 seems to support this interaction. EBF1 could not

completely replace CBF1s function as anchor to DNA but was sufficient for E2 binding at actual strong binding sites. Ongoing research in the Kempkes laboratory could already demonstrate a reduction in E2 binding intensity upon EBF1 knock-down (experiments conducted by S. Rieger) and support a hypothesis where both cellular TFs are needed to mediate accession to chromatin and specificity of binding sites.

The data obtained and analyzed in this thesis collectively point towards a B cell specific network of TFs and associated regulatory elements which are exploited by E2 and E3 proteins in order to regulate distinct target gene sets. PU.1, which does not correlate with E2 or E3 signals but frequently is found to co-occupy EBNA sites, was described as a pioneer factor for opening nucleosome occupied TF target sites (Barozzi et al., 2014), which is already expressed in hematopoietic precursor cells (reviewed in Choukrallah and Matthias, 2014). Recently, the pioneering activity of the C-terminal domain of EBF1 could be described in B cell fate decision (Boller et al., 2016). The additive and combinatorial effects of pioneer factors and rather broadly expressed TFs in the selection of cell type specific enhancers has been shown (Heinz et al., 2015) and together with the information on EBF1 expression and function displays the basis for E2 specificity.

6 References

AFGAN, E., BAKER, D., VAN DEN BEEK, M., BLANKENBERG, D., BOUVIER, D., CECH, M., CHILTON, J., CLEMENTS, D., CORAOR, N., EBERHARD, C., GRUNING, B., GUERLER, A., HILLMAN-JACKSON, J., VON KUSTER, G., RASCHE, E., SORANZO, N., TURAGA, N., TAYLOR, J., NEKRUTENKO, A. & GOECKS, J. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res,* 44, W3-W10.

ALLDAY, M. J., BAZOT, Q. & WHITE, R. E. 2015. The EBNA3 Family: Two Oncoproteins and a Tumour Suppressor that Are Central to the Biology of EBV in B Cells. *Curr Top Microbiol Immunol,* 391, 61-117.

ANDERTON, E., YEE, J., SMITH, P., CROOK, T., WHITE, R. E. & ALLDAY, M. J. 2008. Two Epstein-Barr virus (EBV) oncoproteins cooperate to repress expression of the proapoptotic tumour-suppressor Bim: clues to the pathogenesis of Burkitt's lymphoma. *Oncogene,* 27, 421-33.

ANDREWS, S. 2010. FastQC A Quality Control tool for High Throughput Sequence Data. 0.11.3 ed. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/: Babraham Institute for Bioinformatics

ARVEY, A., TEMPERA, I., TSAI, K., CHEN, H. S., TIKHMYANOVA, N., KLICHINSKY, M., LESLIE, C. & LIEBERMAN, P. M. 2012. An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. *Cell Host Microbe,* 12, 233-45.

BAER, R., BANKIER, A. T., BIGGIN, M. D., DEININGER, P. L., FARRELL, P. J., GIBSON, T. J., HATFULL, G., HUDSON, G. S., SATCHWELL, S. C., SEGUIN, C. & ET AL. 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature,* 310, 207-11.

BAILEY, T. L. & MACHANICK, P. 2012. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res,* 40, e128.

BAIN, M., WATSON, R. J., FARRELL, P. J. & ALLDAY, M. J. 1996. Epstein-Barr virus nuclear antigen 3C is a powerful repressor of transcription when tethered to DNA. *J Virol,* 70, 2481-9.

BANERJEE, S., LU, J., CAI, Q., SAHA, A., JHA, H. C., DZENG, R. K. & ROBERTSON, E. S. 2013. The EBV Latent Antigen 3C Inhibits Apoptosis through Targeted Regulation of Interferon Regulatory Factors 4 and 8. *PLoS Pathog,* 9, e1003314.

BAROZZI, I., SIMONATTO, M., BONIFACIO, S., YANG, L., ROHS, R., GHISLETTI, S. & NATOLI, G. 2014. Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol Cell,* 54, 844-57.

BELLARE, P., DUFRESNE, A. & GANEM, D. 2015. Inefficient Codon Usage Impairs mRNA Accumulation: the Case of the v-FLIP Gene of Kaposi's Sarcoma-Associated Herpesvirus. *J Virol,* 89, 7097-107.

BEN-BASSAT, H., GOLDBLUM, N., MITRANI, S., GOLDBLUM, T., YOFFEY, J. M., COHEN, M. M., BENTWICH, Z., RAMOT, B., KLEIN, E. & 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.

BERNSTEIN, B. E., STAMATOYANNOPOULOS, J. A., COSTELLO, J. F., REN, B., MILOSAVLJEVIC, A., MEISSNER, A., KELLIS, M., MARRA, M. A., BEAUDET, A. L., ECKER, J. R., FARNHAM, P. J., HIRST, M., LANDER, E. S., MIKKELSEN, T. S. & THOMSON, J. A. 2010. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol,* 28, 1045-8.

BOLLER, S. & GROSSCHEDL, R. 2014. The regulatory network of B-cell differentiation: a focused view of early B-cell factor 1 function. *Immunol Rev,* 261, 102-15.

BOLLER, S., RAMAMOORTHY, S., AKBAS, D., NECHANITZKY, R., BURGER, L., MURR, R., SCHUBELER, D. & GROSSCHEDL, R. 2016. Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming. *Immunity,* 44, 527-41.

BORNKAMM, G. W., BERENS, C., KUKLIK-ROOS, C., BECHET, J. M., LAUX, G., BACHL, J., KORNDOERFER, M., SCHLEE, M., HOLZEL, M., MALAMOUSSI, A., CHAPMAN, R. D., NIMMERJAHN, F., MAUTNER, J., HILLEN, W., BUJARD, H. & FEUILLARD, J. 2005. Stringent doxycycline-dependent control of gene activities using an episomal one-vector system. *Nucleic acids research,* 33, e137.

BOURILLOT, P. Y., WALTZER, L., SERGEANT, A. & MANET, E. 1998. Transcriptional repression by the Epstein-Barr virus EBNA3A protein tethered to DNA does not require RBP-Jkappa. *J Gen Virol,* 79, 363-70.

BRASS, A. L., KEHRLI, E., EISENBEIS, C. F., STORB, U. & SINGH, H. 1996. Pip, a lymphoid-restricted IRF, contains a regulatory domain that is important for autoinhibition and ternary complex formation with the Ets factor PU.1. *Genes Dev,* 10, 2335-47.

BRASS, A. L., ZHU, A. Q. & SINGH, H. 1999. Assembly requirements of PU.1-Pip (IRF-4) activator complexes: inhibiting function in vivo using fused dimers. *EMBO J,* 18, 977-91.

BURKITT, D. 1958. A sarcoma involving the jaws in African children. *Br J Surg,* 46, 218-23.

BURKITT, D. 1962a. A children's cancer dependent on climatic factors. *Nature,* 194, 232-4.

BURKITT, D. 1962b. Determining the climatic limitations of a children's cancer common in Africa. *Br Med J,* 2, 1019-23.

CALDERWOOD, M. A., VENKATESAN, K., XING, L., CHASE, M. R., VAZQUEZ, A., HOLTHAUS, A. M., EWENCE, A. E., LI, N., HIROZANE-KISHIKAWA, T., HILL, D. E., VIDAL, M., KIEFF, E. & JOHANNSEN, E. 2007. Epstein-Barr virus and virus human protein interaction maps. *Proc Natl Acad Sci U S A,* 104, 7606-11.

CARBONE, A., CESARMAN, E., SPINA, M., GLOGHINI, A. & SCHULZ, T. F. 2009. HIV-associated lymphomas and gamma-herpesviruses. *Blood,* 113, 1213-24.

CARBONE, A., GLOGHINI, A. & DOTTI, G. 2008. EBV-associated lymphoproliferative disorders: classification and treatment. *Oncologist,* 13, 577-85.

CARLSSON, P. & MAHLAPUU, M. 2002. Forkhead transcription factors: key players in development and metabolism. *Dev Biol,* 250, 1-23.

CEN, O. & LONGNECKER, R. 2015. Latent Membrane Protein 2 (LMP2). *Curr Top Microbiol Immunol,* 391, 151-80.

CHABOT, P. R., RAIOLA, L., LUSSIER-PRICE, M., MORSE, T., ARSENEAULT, G., ARCHAMBAULT, J. & 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.

CHEREPANOV, P. P. & WACKERNAGEL, W. 1995. Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene,* 158, 9-14.

CHOUKRALLAH, M. A. & MATTHIAS, P. 2014. The Interplay between Chromatin and Transcription Factor Networks during B Cell Development: Who Pulls the Trigger First? *Front Immunol,* 5, 156.

CICCONE, D. N., MORSHEAD, K. B. & OETTINGER, M. A. 2004. Chromatin immunoprecipitation in the analysis of large chromatin domains across murine antigen receptor loci. *Methods Enzymol,* 376, 334-48.

CLUDTS, I. & FARRELL, P. J. 1998. Multiple functions within the Epstein-Barr virus EBNA-3A protein. *J Virol,* 72, 1862-9.

COUNTRYMAN, J., JENSON, H., SEIBL, R., WOLF, H. & MILLER, G. 1987. Polymorphic proteins encoded within BZLF1 of defective and standard Epstein-Barr viruses disrupt latency. *J Virol,* 61, 3672-9.

DALBIES-TRAN, R., STIGGER-ROSSER, E., DOTSON, T. & SAMPLE, C. E. 2001. Amino acids of Epstein-Barr virus nuclear antigen 3A essential for repression of Jkappa-mediated transcription and their evolutionary conservation. *J Virol,* 75, 90-9.

DAVIS, R. J. 1999. Signal transduction by the c-Jun N-terminal kinase. *Biochem Soc Symp,* 64, 1-12.

DE SILVA, N. S., SIMONETTI, G., HEISE, N. & KLEIN, U. 2012. The diverse roles of IRF4 in late germinal center B-cell differentiation. *Immunol Rev,* 247, 73-92.

DELECLUSE, H. J., FEEDERLE, R., BEHRENDS, U. & MAUTNER, J. 2008. Contribution of viral recombinants to the study of the immune response against the Epstein-Barr virus. *Semin Cancer Biol,* 18, 409-15.

DELECLUSE, H. J., HILSENDEGEN, T., PICH, D., ZEIDLER, R. & 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, 8245-50.

DERRIEN, T., ESTELLE, J., MARCO SOLA, S., KNOWLES, D. G., RAINERI, E., GUIGO, R. & RIBECA, P. 2012. Fast computation and applications of genome mappability. *PLoS One,* 7, e30377.

DJEBALI, S., DAVIS, C. A., MERKEL, A., DOBIN, A., LASSMANN, T., MORTAZAVI, A., TANZER, A., LAGARDE, J., LIN, W., SCHLESINGER, F., XUE, C., MARINOV, G. K., KHATUN, J., WILLIAMS, B. A., ZALESKI, C., ROZOWSKY, J., RODER, M., KOKOCINSKI, F., ABDELHAMID, R. F., ALIOTO, T., ANTOSHECHKIN, I., BAER, M. T., BAR, N. S., BATUT, P., BELL, K., BELL, I., CHAKRABORTTY, S., CHEN, X., CHRAST, J., CURADO, J., DERRIEN, T., DRENKOW, J., DUMAIS, E., DUMAIS, J., DUTTAGUPTA, R., FALCONNET, E., FASTUCA, M., FEJES-TOTH, K., FERREIRA, P., FOISSAC, S., FULLWOOD, M. J., GAO, H., GONZALEZ, D., GORDON, A., GUNAWARDENA, H., HOWALD, C., JHA, S., JOHNSON, R., KAPRANOV, P., KING, B., KINGSWOOD, C., LUO, O. J., PARK, E., PERSAUD, K., PREALL, J. B., RIBECA, P., RISK, B., ROBYR, D., SAMMETH, M., SCHAFFER, L., SEE, L. H., SHAHAB, A., SKANCKE, J., SUZUKI, A. M., TAKAHASHI, H., TILGNER, H., TROUT, D., WALTERS, N., WANG, H., WROBEL, J., YU, Y., RUAN, X., HAYASHIZAKI, Y., HARROW, J., GERSTEIN, M., HUBBARD, T., REYMOND, A., ANTONARAKIS, S. E., HANNON, G., GIDDINGS, M. C., RUAN, Y., WOLD, B., CARNINCI, P., GUIGO, R. & GINGERAS, T. R. 2012. Landscape of transcription in human cells. *Nature,* 489, 101-8.

EISENBEIS, C. F., SINGH, H. & STORB, U. 1995. Pip, a novel IRF family member, is a lymphoid-specific, PU.1-dependent transcriptional activator. *Genes Dev,* 9, 1377-87.

ENCODE_CONSORTIUM 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol,* 9, e1001046.

ENCODE_CONSORTIUM 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489, 57-74.

ENGELEN, E., BRANDSMA, J. H., MOEN, M. J., SIGNORILE, L., DEKKERS, D. H., DEMMERS, J., KOCKX, C. E., OZGUR, Z., VAN, I. W. F., VAN DEN BERG, D. L. & POOT, R. A. 2015. Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. *Nat Commun,* 6, 7155.

EPSTEIN, M., ACHONG, B. & BARR, Y. 1964. Virus particles in cultured lymphoblasts from Burkitt's lymphoma. *Lancet,* 1, 702-703.

ERNST, J., KHERADPOUR, P., MIKKELSEN, T. S., SHORESH, N., WARD, L. D., EPSTEIN, C. B., ZHANG, X., WANG, L., ISSNER, R., COYNE, M., KU, M., DURHAM, T., KELLIS, M. & BERNSTEIN, B. E. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature,* 473, 43-9.

FANTOM_CONSORTIUM, THE, R. P., CLST, FORREST, A. R., KAWAJI, H., REHLI, M., BAILLIE, J. K., DE HOON, M. J., HABERLE, V., LASSMANN, T., KULAKOVSKIY, I. V., LIZIO, M., ITOH, M., ANDERSSON, R., MUNGALL, C. J., MEEHAN, T. F., SCHMEIER, S., BERTIN, N., JORGENSEN, M., DIMONT, E., ARNER, E., SCHMIDL, C., SCHAEFER, U., MEDVEDEVA, Y. A., PLESSY, C., VITEZIC, M., SEVERIN, J., SEMPLE, C., ISHIZU, Y., YOUNG, R. S., FRANCESCATTO, M., ALAM, I., ALBANESE, D., ALTSCHULER, G. M., ARAKAWA, T., ARCHER, J. A., ARNER, P., BABINA, M., RENNIE, S., BALWIERZ, P. J., BECKHOUSE, A. G., PRADHAN-BHATT, S., BLAKE, J. A., BLUMENTHAL, A., BODEGA, B., BONETTI, A., BRIGGS, J., BROMBACHER, F., BURROUGHS, A. M., CALIFANO, A., CANNISTRACI, C. V., CARBAJO, D., CHEN, Y., CHIERICI, M., CIANI, Y., CLEVERS, H. C., DALLA, E., DAVIS, C. A., DETMAR, M., DIEHL, A. D., DOHI, T., DRABLOS, F., EDGE, A. S., EDINGER, M., EKWALL, K., ENDOH, M., ENOMOTO, H., FAGIOLINI, M., FAIRBAIRN, L., FANG, H., FARACH-CARSON, M. C., FAULKNER, G. J., FAVOROV, A. V., FISHER, M. E., FRITH, M. C., FUJITA, R., FUKUDA, S., FURLANELLO, C., FURINO, M., FURUSAWA, J., GEIJTENBEEK, T. B., GIBSON, A. P., GINGERAS, T., GOLDOWITZ, D., GOUGH, J., GUHL, S., GULER, R., GUSTINCICH, S., HA, T. J., HAMAGUCHI, M., HARA, M., HARBERS, M., HARSHBARGER, J., HASEGAWA, A., HASEGAWA, Y., HASHIMOTO, T., HERLYN, M., HITCHENS, K. J., HO SUI, S. J., HOFMANN, O. M., et al. 2014. A promoter-level mammalian expression atlas. *Nature,* 507, 462-70.

FEEDERLE, R., BARTLETT, E. J. & DELECLUSE, H. J. 2010. Epstein-Barr virus genetics: talking about the BAC generation. *Herpesviridae,* 1, 6.

FRAPPIER, L. 2015. Ebna1. *Curr Top Microbiol Immunol,* 391, 3-34.

FRIBERG, A., THUMANN, S., HENNIG, J., ZOU, P., NOSSNER, E., LING, P. D., SATTLER, M. & KEMPKES, B. 2015. The EBNA-2 N-Terminal Transactivation Domain Folds into a Dimeric Structure Required for Target Gene Activation. *PLoS Pathog,* 11, e1004910.

FUCHS, K. P., BOMMER, G., DUMONT, E., CHRISTOPH, B., VIDAL, M., KREMMER, E. & 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, 4639-46.

FUJITA, N. & WADE, P. A. 2004. Use of bifunctional cross-linking reagents in mapping genomic distribution of chromatin remodeling complexes. *Methods,* 33, 81-5.

GAO, H., LUKIN, K., RAMIREZ, J., FIELDS, S., LOPEZ, D. & 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 U S A,* 106, 11258-63.

GIARDINE, B., RIEMER, C., HARDISON, R. C., BURHANS, R., ELNITSKI, L., SHAH, P., ZHANG, Y., BLANKENBERG, D., ALBERT, I., TAYLOR, J., MILLER, W., KENT, W. J. & NEKRUTENKO, A. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res,* 15, 1451-5.

GLASMACHER, E., AGRAWAL, S., CHANG, A. B., MURPHY, T. L., ZENG, W., VANDER LUGT, B., KHAN, A. A., CIOFANI, M., SPOONER, C. J., RUTZ, S., HACKNEY, J., NURIEVA, R., ESCALANTE, C. R., OUYANG, W., LITTMAN, D. R., MURPHY, K. M. & SINGH, H. 2012. A genomic regulatory element that directs assembly and function of immune-specific AP-1-IRF complexes. *Science,* 338, 975-80.

GORDADZE, A. V., ONUNWOR, C. W., PENG, R., POSTON, D., KREMMER, E. & LING, P. D. 2004. EBNA2 amino acids 3 to 30 are required for induction of LMP-1 and immortalization maintenance. *J Virol,* 78, 3919-29.

GOTTSCHALK, S., ROONEY, C. M. & HESLOP, H. E. 2005. Post-Transplant Lymphoproliferative Disorders. *Annu Rev Med,* 56, 29-44.

GRAHAM, F. L., SMILEY, J., RUSSELL, W. C. & NAIRN, R. 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol,* 36, 59-74.

GRONOSTAJSKI, R. M. 2000. Roles of the NFI/CTF gene family in transcription and development. *Gene,* 249, 31-45.

GROSSMAN, S. R., JOHANNSEN, E., TONG, X., YALAMANCHILI, R. & KIEFF, E. 1994. The Epstein-Barr virus nuclear antigen 2 transactivator is directed to response elements by the J kappa recombination signal binding protein. *Proc Natl Acad Sci U S A,* 91, 7568-72.

GUNNELL, A., WEBB, H. M., WOOD, C. D., MCCLELLAN, M. J., WICHAIDIT, B., KEMPKES, B., JENNER, R. G., OSBORNE, C., FARRELL, P. J. & WEST, M. J. 2016. RUNX super-enhancer control through the Notch pathway by Epstein-Barr virus transcription factors regulates B cell growth. *Nucleic Acids Res.*

GUPTA, S., STAMATOYANNOPOULOS, J. A., BAILEY, T. L. & NOBLE, W. S. 2007. Quantifying similarity between motifs. *Genome Biol,* 8, R24.

HAGMAN, J., RAMIREZ, J. & LUKIN, K. 2012. B lymphocyte lineage specification, commitment and epigenetic control of transcription by early B cell factor 1. *Curr Top Microbiol Immunol,* 356, 17-38.

HANAHAN, D. 1985. Techniques for transformation of E. coli. *In:* GLOVER, D. (ed.) *DNA cloning. A practical approach.* Oxford: IRL Press, Vol. 1: 109-135.

HARADA, R., VADNAIS, C., SANSREGRET, L., LEDUY, L., BERUBE, G., ROBERT, F. & NEPVEU, A. 2008. Genome-wide location analysis and expression studies reveal a role for p110 CUX1 in the activation of DNA replication genes. *Nucleic Acids Res,* 36, 189-202.

HARADA, S., YALAMANCHILI, R. & KIEFF, E. 2001. Epstein-Barr virus nuclear protein 2 has at least two N-terminal domains that mediate self-association. *J Virol,* 75, 2482-7.

HARTH-HERTLE, M. L., SCHOLZ, B. A., ERHARD, F., GLASER, L. V., DOLKEN, L., ZIMMER, R. & 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, e1003638.

HAYWARD, S. D., LIU, J. & FUJIMURO, M. 2006. Notch and Wnt signaling: mimicry and manipulation by gamma herpesviruses. *Sci STKE,* 2006, re4.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell,* 38, 576-89.

HEINZ, S., ROMANOSKI, C. E., BENNER, C. & GLASS, C. K. 2015. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol,* 16, 144-54.

HENKEL, T., LING, P. D., HAYWARD, S. D. & PETERSON, M. G. 1994. Mediation of Epstein-Barr virus EBNA2 transactivation by recombination signal-binding protein J kappa. *Science,* 265, 92-5.

HENLE, G., HENLE, W. & DIEHL, V. 1968. Relation of Burkitt's tumor-associated herpes-ytpe virus to infectious mononucleosis. *Proc Natl Acad Sci U S A,* 59, 94-101.

HENLE, W., DIEHL, V., KOHN, G., ZUR HAUSEN, H. & HENLE, G. 1967. Herpes-type virus and chromosome marker in normal leukocytes after growth with irradiated Burkitt cells. *Science,* 157, 1064-5.

HERTLE, M. L., POPP, C., PETERMANN, S., MAIER, S., KREMMER, E., LANG, R., MAGES, J. & KEMPKES, B. 2009. Differential gene expression patterns of EBV infected EBNA-3A positive and negative human B lymphocytes. *PLoS pathogens,* 5, e1000506.

HICKABOTTOM, M., PARKER, G. A., FREEMONT, P., CROOK, T. & ALLDAY, M. J. 2002. Two nonconsensus sites in the Epstein-Barr virus oncoprotein EBNA3A cooperate to bind the co-repressor carboxyl-terminal-binding protein (CtBP). *J Biol Chem,* 277, 47197-204.

HOU, T., RAY, S. & BRASIER, A. R. 2007. The functional role of an interleukin 6-inducible CDK9.STAT3 complex in human gamma-fibrinogen gene expression. *J Biol Chem,* 282, 37091-102.

HSIEH, J. J. & HAYWARD, S. D. 1995. Masking of the CBF1/RBPJ kappa transcriptional repression domain by Epstein-Barr virus EBNA2. *Science,* 268, 560-3.

HUMME, S., REISBACH, G., FEEDERLE, R., DELECLUSE, H. J., BOUSSET, K., HAMMERSCHMIDT, W. & SCHEPERS, A. 2003. The EBV nuclear antigen 1 (EBNA1) enhances B cell immortalization several thousandfold. *Proc Natl Acad Sci U S A,* 100, 10989-94. Epub 2003 Aug 28.

HUMMELER, K., HENLE, G. & HENLE, W. 1966. Fine structure of a virus in cultured lymphoblasts from Burkitt lymphoma. *J Bacteriol,* 91, 1366-8.

HURLEY, E. A., KLAMAN, L. D., AGGER, S., LAWRENCE, J. B. & THORLEY-LAWSON, D. A. 1991. The prototypical Epstein-Barr virus-transformed lymphoblastoid cell line IB4 is an unusual variant containing integrated but no episomal viral DNA. *J Virol,* 65, 3958-63.

HUTT-FLETCHER, L. M. 2007. Epstein-Barr virus entry. *J Virol,* 81, 7825-32.

IPPOLITO, G. C., DEKKER, J. D., WANG, Y. H., LEE, B. K., SHAFFER, A. L., 3RD, LIN, J., WALL, J. K., LEE, B. S., STAUDT, L. M., LIU, Y. J., IYER, V. R. & TUCKER, H. O. 2014. Dendritic cell fate is determined by BCL11A. *Proc Natl Acad Sci U S A,* 111, E998-1006.

JACKSON, V. 1999. Formaldehyde cross-linking for studying nucleosomal dynamics. *Methods,* 17, 125-39.

JACKSTADT, R., ROH, S., NEUMANN, J., JUNG, P., HOFFMANN, R., HORST, D., BERENS, C., BORNKAMM, G. W., KIRCHNER, T., MENSSEN, A. & HERMEKING, H. 2013. AP4 is a mediator of epithelial-mesenchymal transition and metastasis in colorectal cancer. *J Exp Med,* 210, 1331-50.

JHA, H. C., BANERJEE, S. & ROBERTSON, E. S. 2016. The Role of Gammaherpesviruses in Cancer Pathogenesis. *Pathogens,* 5.

JIANG, S., WILLOX, B., ZHOU, H., HOLTHAUS, A. M., WANG, A., SHI, T. T., MARUO, S., KHARCHENKO, P. V., JOHANNSEN, E. C., KIEFF, E. & ZHAO, B. 2014. Epstein-Barr virus nuclear antigen 3C binds to BATF/IRF4 or SPI1/IRF4 composite sites and recruits Sin3A to repress CDKN2A. *Proc Natl Acad Sci U S A,* 111, 421-6.

JIMENEZ-RAMIREZ, C., BROOKS, A. J., FORSHELL, L. P., YAKIMCHUK, K., ZHAO, B., FULGHAM, T. Z. & SAMPLE, C. E. 2006. Epstein-Barr virus EBNA-3C is targeted to and regulates expression from the bidirectional LMP-1/2B promoter. *J Virol,* 80, 11200-8.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337, 816-21.

JOCHUM, S., MOOSMANN, A., LANG, S., HAMMERSCHMIDT, W. & ZEIDLER, R. 2012. The EBV immunoevasins vIL-10 and BNLF2a protect newly infected B cells from immune recognition and elimination. *PLoS Pathog,* 8, e1002704.

JOHANNSEN, E., KOH, E., MOSIALOS, G., TONG, X., KIEFF, E. & 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, 253-62.

KARIN, M., LIU, Z. & ZANDI, E. 1997. AP-1 function and regulation. *Curr Opin Cell Biol,* 9, 240-6.

KAVATHAS, P., BACH, F. H. & DEMARS, R. 1980. Gamma ray-induced loss of expression of HLA and glyoxalase I alleles in lymphoblastoid cells. *Proc Natl Acad Sci U S A,* 77, 4251-5.

KEMPKES, B. & LING, P. D. 2015. EBNA2 and Its Coactivator EBNA-LP. *Curr Top Microbiol Immunol,* 391, 35-59.

KEMPKES, B. & ROBERTSON, E. S. 2015. Epstein-Barr virus latency: current and future perspectives. *Curr Opin Virol,* 14, 138-44.

KHONGKOW, P., KARUNARATHNA, U., KHONGKOW, M., GONG, C., GOMES, A. R., YAGUE, E., MONTEIRO, L. J., KONGSEMA, M., ZONA, S., MAN, E. P., TSANG, J. W., COOMBES, R. C., WU, K. J., KHOO, U. S., MEDEMA, R. H., FREIRE, R. & LAM, E. W. 2014. FOXM1 targets NBS1 to regulate DNA damage-induced senescence and epirubicin resistance. *Oncogene,* 33, 4144-55.

KIEFF, E. & RICKINSON, A. B. 2007. Epstein-Barr virus and its replication. *In:* KNIPE, D. M. & HOWLEY, P. M. (eds.) *Fields Virology.* Philadelphia: Lippincott - Williams & Wilkins, pp. 2603-2654.

KIESER, A. & STERZ, K. R. 2015. The Latent Membrane Protein 1 (LMP1). *Curr Top Microbiol Immunol,* 391, 119-49.

KNIGHT, J. S., LAN, K., SUBRAMANIAN, C. & ROBERTSON, E. S. 2003. Epstein-Barr virus nuclear antigen 3C recruits histone deacetylase activity and associates with the corepressors mSin3A and NCoR in human B-cell lines. *J Virol,* 77, 4261-72.

KOUSKOUTI, A., SCHEER, E., STAUB, A., TORA, L. & TALIANIDIS, I. 2004. Gene-specific modulation of TAF10 function by SET9-mediated methylation. *Mol Cell,* 14, 175-82.

KOVALL, R. A. & HENDRICKSON, W. A. 2004. Crystal structure of the nuclear effector of Notch signaling, CSL, bound to DNA. *Embo J,* 23, 3441-51.

KREJCI, A. & BRAY, S. 2007. Notch activation stimulates transient and selective binding of Su(H)/CSL to target enhancers. *Genes Dev,* 21, 1322-7.

KRETZMER, H., BERNHART, S. H., WANG, W., HAAKE, A., WENIGER, M. A., BERGMANN, A. K., BETTS, M. J., CARRILLO-DE-SANTA-PAU, E., DOOSE, G., GUTWEIN, J., RICHTER, J., HOVESTADT, V., HUANG, B., RICO, D., JUHLING, F., KOLAROVA, J., LU, Q., OTTO, C., WAGENER, R., ARNOLDS, J., BURKHARDT, B., CLAVIEZ, A., DREXLER, H. G., EBERTH, S., EILS, R., FLICEK, P., HAAS, S., HUMMEL, M., KARSCH, D., KERSTENS, H. H., KLAPPER, W., KREUZ, M., LAWERENZ, C., LENZE, D., LOEFFLER, M., LOPEZ, C., MACLEOD, R. A., MARTENS, J. H., KULIS, M., MARTIN-SUBERO, J. I., MOLLER, P., NAGEL, I., PICELLI, S., VATER, I., ROHDE, M., ROSENSTIEL, P., ROSOLOWSKI, M., RUSSELL, R. B., SCHILHABEL, M., SCHLESNER, M., STADLER, P. F., SZCZEPANOWSKI, M., TRUMPER, L., STUNNENBERG, H. G., PROJECT, I. M.-S., PROJECT, B., KUPPERS, R., AMMERPOHL, O., LICHTER, P., SIEBERT, R., HOFFMANN, S. & RADLWIMMER, B. 2015. DNA methylome analysis in Burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control. *Nat Genet,* 47, 1316-25.

KRUEGER, F. 2012. Trim Galore! http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/: Babraham Institute for Bioinformatics

KU, M., JAFFE, J. D., KOCHE, R. P., RHEINBAY, E., ENDOH, M., KOSEKI, H., CARR, S. A. & BERNSTEIN, B. E. 2012. H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions. *Genome Biol,* 13, R85.

KULIC, I., ROBERTSON, G., CHANG, L., BAKER, J. H., LOCKWOOD, W. W., MOK, W., FULLER, M., FOURNIER, M., WONG, N., CHOU, V., ROBINSON, M. D., CHUN, H. J., GILKS, B., KEMPKES, B., THOMSON, T. A., HIRST, M., MINCHINTON, A. I., LAM, W. L., JONES, S., MARRA, M. & KARSAN, A. 2015. Loss of the Notch effector RBPJ promotes tumorigenesis. *J Exp Med,* 212, 37-52.

LAI, E. C. 2002. Keeping a good pathway down: transcriptional repression of Notch pathway target genes by CSL proteins. *EMBO Rep,* 3, 840-5.

LAICHALK, L. L. & THORLEY-LAWSON, D. A. 2005. Terminal differentiation into plasma cells initiates the replicative cycle of Epstein-Barr virus in vivo. *J Virol,* 79, 1296-307.

LANDT, S. G., MARINOV, G. K., KUNDAJE, A., KHERADPOUR, P., PAULI, F., BATZOGLOU, S., BERNSTEIN, B. E., BICKEL, P., BROWN, J. B., CAYTING, P., CHEN, Y., DESALVO, G., EPSTEIN, C., FISHER-AYLOR, K. I., EUSKIRCHEN, G., GERSTEIN, M., GERTZ, J., HARTEMINK, A. J., HOFFMAN, M. M., IYER, V. R., JUNG, Y. L., KARMAKAR, S., KELLIS, M., KHARCHENKO, P. V., LI, Q., LIU, T., LIU, X. S., MA, L., MILOSAVLJEVIC, A., MYERS, R. M., PARK, P. J., PAZIN, M. J., PERRY, M. D., RAHA, D., REDDY, T. E., ROZOWSKY, J., SHORESH, N., SIDOW, A., SLATTERY, M., STAMATOYANNOPOULOS, J. A., TOLSTORUKOV, M. Y., WHITE, K. P., XI, S., FARNHAM, P. J., LIEB, J. D., WOLD, B. J. & SNYDER, M. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res,* 22, 1813-31.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9, 357-9.

LAUX, G., ADAM, B., STROBL, L. J. & MOREAU-GACHELIN, F. 1994a. 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, 5624-32.

LAUX, G., DUGRILLON, F., ECKERT, C., ADAM, B., ZIMBER-STROBL, U. & BORNKAMM, G. W. 1994b. Identification and characterization of an Epstein-Barr virus nuclear antigen 2-responsive cis element in the bidirectional promoter region of latent membrane protein and terminal protein 2 genes. *J Virol,* 68, 6947-58.

LE ROUX, A., KERDILES, B., WALLS, D., DEDIEU, J. F. & PERRICAUDET, M. 1994. The Epstein-Barr virus determined nuclear antigens EBNA-3A, -3B, and -3C repress EBNA-2-mediated transactivation of the viral terminal protein 1 gene promoter. *Virology,* 205, 596-602.

LEE, S., SAKAKIBARA, S., MARUO, S., ZHAO, B., CALDERWOOD, M. A., HOLTHAUS, A. M., LAI, C. Y., TAKADA, K., KIEFF, E. & JOHANNSEN, E. 2009. Epstein-Barr virus nuclear protein 3C domains necessary for lymphoblastoid cell growth: interaction with RBP-Jkappa regulates TCL1. *J Virol,* 83, 12368-77.

LEFEBVRE, C., RAJBHANDARI, P., ALVAREZ, M. J., BANDARU, P., LIM, W. K., SATO, M., WANG, K., SUMAZIN, P., KUSTAGI, M., BISIKIRSKA, B. C., BASSO, K., BELTRAO, P., KROGAN, N., GAUTIER, J., DALLA-FAVERA, R. & CALIFANO, A. 2010. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol Syst Biol,* 6, 377.

LI, Q., BROWN, J. B., HUANG, H. & BICKEL, P. J. 2011. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.,* 5, 1752-1779.

LI, X., WANG, W., WANG, J., MALOVANNAYA, A., XI, Y., LI, W., GUERRA, R., HAWKE, D. H., QIN, J. & CHEN, J. 2015. Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. *Mol Syst Biol,* 11, 775.

LIN, J., JOHANNSEN, E., ROBERTSON, E. & KIEFF, E. 2002. Epstein-Barr virus nuclear antigen 3C putative repression domain mediates coactivation of the LMP1 promoter with EBNA-2. *J Virol,* 76, 232-42.

LING, P. D. & HAYWARD, S. D. 1995. Contribution of conserved amino acids in mediating the interaction between EBNA2 and CBF1/RBPJk. *J Virol,* 69, 1944-50.

LING, P. D., HSIEH, J. J., RUF, I. K., RAWLINS, D. R. & HAYWARD, S. D. 1994. EBNA-2 upregulation of Epstein-Barr virus latency promoters and the cellular CD23 promoter utilizes a common targeting intermediate, CBF1. *J Virol,* 68, 5375-83.

LIU, P., KELLER, J. R., ORTIZ, M., TESSAROLLO, L., RACHEL, R. A., NAKAMURA, T., JENKINS, N. A. & COPELAND, N. G. 2003. Bcl11a is essential for normal lymphoid development. *Nat Immunol,* 4, 525-32.

LU, F., CHEN, H. S., KOSSENKOV, A. V., DEWISPELEARE, K., WON, K. J. & LIEBERMAN, P. M. 2016. EBNA2 Drives Formation of New Chromosome Binding Sites and Target Genes for B-Cell Master Regulatory Transcription Factors RBP-jkappa and EBF1. *PLoS Pathog,* 12, e1005339.

LUCCHESI, W., BRADY, G., DITTRICH-BREIHOLZ, O., KRACHT, M., RUSS, R. & FARRELL, P. J. 2008. Differential gene regulation by Epstein-Barr virus type 1 and type 2 EBNA2. *J Virol,* 82, 7456-66.

MACHANICK, P. & BAILEY, T. L. 2011. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics,* 27, 1696-7.

MAIER, H., OSTRAAT, R., GAO, H., FIELDS, S., SHINTON, S. A., MEDINA, K. L., IKAWA, T., MURRE, C., SINGH, H., HARDY, R. R. & HAGMAN, J. 2004. Early B cell factor cooperates with Runx1 and mediates epigenetic changes associated with mb-1 transcription. *Nat Immunol,* 5, 1069-77.

MAIER, S., SANTAK, M., MANTIK, A., GRABUSIC, K., KREMMER, E., HAMMERSCHMIDT, W. & KEMPKES, B. 2005. A somatic knockout of CBF1 in a human B-cell line reveals that induction of CD21 and CCR7 by EBNA-2 is strictly CBF1 dependent and that downregulation of immunoglobulin M is partially CBF1 independent. *J Virol,* 79, 8784-92.

MAIER, S., STAFFLER, G., HARTMANN, A., HOCK, J., HENNING, K., GRABUSIC, K., MAILHAMMER, R., HOFFMANN, R., WILMANNS, M., LANG, R., MAGES, J. & KEMPKES, B. 2006. Cellular target genes of epstein-barr virus nuclear antigen 2. *J Virol,* 80, 9761-71.

MARSHALL, D. & SAMPLE, C. 1995. Epstein-Barr virus nuclear antigen 3C is a transcriptional regulator. *J Virol,* 69, 3624-30.

MARUO, S., JOHANNSEN, E., ILLANES, D., COOPER, A., ZHAO, B. & KIEFF, E. 2005. Epstein-Barr virus nuclear protein 3A domains essential for growth of lymphoblasts: transcriptional regulation through RBP-Jkappa/CBF1 is critical. *J Virol,* 79, 10171-9.

MARUO, S., WU, Y., ITO, T., KANDA, T., KIEFF, E. D. & TAKADA, K. 2009. Epstein-Barr virus nuclear protein EBNA3C residues critical for maintaining lymphoblastoid cell growth. *Proc Natl Acad Sci U S A,* 106, 4419-24.

MCCLELLAN, M. J., KHASNIS, S., WOOD, C. D., PALERMO, R. D., SCHLICK, S. N., KANHERE, A. S., JENNER, R. G. & WEST, M. J. 2012. Downregulation of integrin receptor-signaling genes by Epstein-Barr virus EBNA 3C via promoter-proximal and -distal binding elements. *J Virol,* 86, 5165-78.

MCCLELLAN, M. J., WOOD, C. D., OJENIYI, O., COOPER, T. J., KANHERE, A., ARVEY, A., WEBB, H. M., PALERMO, R. D., HARTH-HERTLE, M. L., KEMPKES, B., JENNER, R. G. & WEST, M. J. 2013. Modulation of enhancer looping and differential gene targeting by Epstein-Barr virus transcription factors directs cellular reprogramming. *PLoS Pathog,* 9, e1003636.

MEGA, T., LUPIA, M., AMODIO, N., HORTON, S. J., MESURACA, M., PELAGGI, D., AGOSTI, V., GRIECO, M., CHIARELLA, E., SPINA, R., MOORE, M. A., SCHURINGA, J. J., BOND, H. M. & MORRONE, G. 2011. Zinc finger protein 521 antagonizes early B-cell factor 1 and modulates the B-lymphoid differentiation of primary hematopoietic progenitors. *Cell Cycle,* 10, 2129-39.

MERKENSCHLAGER, M. & ODOM, D. T. 2013. CTCF and cohesin: linking gene regulatory elements with their targets. *Cell,* 152, 1285-97.

MIFSUD, B., TAVARES-CADETE, F., YOUNG, A. N., SUGAR, R., SCHOENFELDER, S., FERREIRA, L., WINGETT, S. W., ANDREWS, S., GREY, W., EWELS, P. A., HERMAN, B., HAPPE, S., HIGGS, A., LEPROUST, E., FOLLOWS, G. A., FRASER, P., LUSCOMBE, N. M. & OSBORNE, C. S. 2015. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet,* 47, 598-606.

MURPHY, T. L., TUSSIWAND, R. & MURPHY, K. M. 2013. Specificity through cooperation: BATF-IRF interactions control immune-regulatory networks. *Nat Rev Immunol,* 13, 499-509.

MURRAY, P. & BELL, A. 2015. Contribution of the Epstein-Barr Virus to the Pathogenesis of Hodgkin Lymphoma. *Curr Top Microbiol Immunol,* 390, 287-313.

MYATT, S. S. & LAM, E. W. 2007. The emerging roles of forkhead box (Fox) proteins in cancer. *Nat Rev Cancer,* 7, 847-59.

NEMEROW, G. R., MOLD, C., SCHWEND, V. K., TOLLEFSON, V. & 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, 1416-20.

NEPVEU, A. 2001. Role of the multifunctional CDP/Cut/Cux homeodomain transcription factor in regulating differentiation, cell growth and development. *Gene,* 270, 1-15.

NEUHIERL, B., FEEDERLE, R., HAMMERSCHMIDT, W. & 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, 15036-41. Epub 2002 Oct 30.

NGUYEN, A. T. & ZHANG, Y. 2011. The diverse functions of Dot1 and H3K79 methylation. *Genes Dev,* 25, 1345-58.

NOWAK, D. E., TIAN, B. & BRASIER, A. R. 2005. Two-step cross-linking method for identification of NF-kappaB gene network by chromatin immunoprecipitation. *Biotechniques,* 39, 715-25.

NOWAK, D. E., TIAN, B., JAMALUDDIN, M., BOLDOGH, I., VERGARA, L. A., CHOUDHARY, S. & BRASIER, A. R. 2008. RelA Ser276 phosphorylation is required for activation of a subset of NF-kappaB-dependent genes by recruiting cyclin-dependent kinase 9/cyclin T1 complexes. *Mol Cell Biol,* 28, 3623-38.

NUTT, S. L., HODGKIN, P. D., TARLINTON, D. M. & CORCORAN, L. M. 2015. The generation of antibody-secreting plasma cells. *Nat Rev Immunol,* 15, 160-71.

O'NIONS, J. & ALLDAY, M. J. 2004. Deregulation of the cell cycle by the Epstein-Barr virus. *Adv Cancer Res,* 92, 119-86.

OHASHI, M., HOLTHAUS, A. M., CALDERWOOD, M. A., LAI, C. Y., KRASTINS, B., SARRACINO, D. & JOHANNSEN, E. 2015. The EBNA3 family of Epstein-Barr virus nuclear proteins associates with the USP46/USP12 deubiquitination complexes to regulate lymphoblastoid cell line growth. *PLoS Pathog,* 11, e1004822.

OSADA, S., DAIMON, S., NISHIHARA, T. & IMAGAWA, M. 1996. Identification of DNA binding-site preferences for nuclear factor I-A. *FEBS Lett,* 390, 44-6.

PASCHOS, K., PARKER, G. A., WATANATANASUP, E., WHITE, R. E. & ALLDAY, M. J. 2012. BIM promoter directly targeted by EBNA3C in polycomb-mediated repression by EBV. *Nucleic Acids Res,* 40, 7233-46.

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M. & DUCHESNAY, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research,* 12, pp. 2825--2830.

PENG, C. W., XUE, Y., ZHAO, B., JOHANNSEN, E., KIEFF, E. & 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, 1033-8.

PHAM, T. H., MINDERJAHN, J., SCHMIDL, C., HOFFMEISTER, H., SCHMIDHOFER, S., CHEN, W., LANGST, G., BENNER, C. & REHLI, M. 2013. Mechanisms of in vivo binding site selection of the hematopoietic master transcription factor PU.1. *Nucleic Acids Res,* 41, 6391-402.

PLANK, J. L. & DEAN, A. 2014. Enhancer function: mechanistic and genome-wide insights come together. *Mol Cell,* 55, 5-14.

POPE, J. H., HORNE, M. K. & SCOTT, W. 1968. Transformation of foetal human keukocytes in vitro by filtrates of a human leukaemic cell line containing herpes-like virus. *Int J Cancer,* 3, 857-66.

PORTAL, D., ZHAO, B., CALDERWOOD, M. A., SOMMERMANN, T., JOHANNSEN, E. & KIEFF, E. 2011. EBV nuclear antigen EBNALP dismisses transcription repressors NCoR and RBPJ from enhancers and EBNA2 increases NCoR-deficient RBPJ DNA binding. *Proceedings of the National Academy of Sciences of the United States of America,* 108, 7808-13.

PORTAL, D., ZHOU, H., ZHAO, B., KHARCHENKO, P. V., LOWRY, E., WONG, L., QUACKENBUSH, J., HOLLOWAY, D., JIANG, S., LU, Y. & KIEFF, E. 2013. Epstein-Barr virus nuclear antigen leader protein localizes to promoters and enhancers with cell transcription factors and EBNA2. *Proc Natl Acad Sci U S A,* 110, 18537-42.

PRUITT, K. D., TATUSOVA, T. & MAGLOTT, D. R. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res,* 33, D501-4.

PULVERTAFT, J. V. 1964. Cytology of Burkitt's Tumour (African Lymphoma). *Lancet,* 39, 238-40.

RAAB-TRAUB, N. 2015. Nasopharyngeal Carcinoma: An Evolving Role for the Epstein-Barr Virus. *Curr Top Microbiol Immunol,* 390, 339-63.

RADKOV, S. A., BAIN, M., FARRELL, P. J., WEST, M., ROWE, M. & ALLDAY, M. J. 1997. Epstein-Barr virus EBNA3C represses Cp, the major promoter for EBNA expression, but has no effect on the promoter of the cell gene CD21. *J Virol,* 71, 8552-62.

RADKOV, S. A., TOUITOU, R., BREHM, A., ROWE, M., WEST, M., KOUZARIDES, T. & ALLDAY, M. J. 1999. Epstein-Barr virus nuclear antigen 3C interacts with histone deacetylase to repress transcription. *J Virol,* 73, 5688-97.

RAMDZAN, Z. M. & NEPVEU, A. 2014. CUX1, a haploinsufficient tumour suppressor gene overexpressed in advanced cancers. *Nat Rev Cancer,* 14, 673-82.

RAMIREZ, F., DUNDAR, F., DIEHL, S., GRUNING, B. A. & MANKE, T. 2014a. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res,* 42, W187-91.

RAMIREZ, F., DUNDAR, F., DIEHL, S., GRUNING, B. A. & MANKE, T. 2014b. *deepTools: github wiki page* [Online]. Available: https://github.com/fidelram/deepTools/wiki/QC#wiki-bamCorrelate [Accessed 21.04.2016 2016].

RANGASWAMY, U. S., O'FLAHERTY, B. M. & SPECK, S. H. 2014. Tyrosine 129 of the murine gammaherpesvirus M2 protein is critical for M2 function in vivo. *PLoS One,* 9, e105197.

RANGASWAMY, U. S. & SPECK, S. H. 2014. Murine gammaherpesvirus M2 protein induction of IRF4 via the NFAT pathway leads to IL-10 expression in B cells. *PLoS Pathog,* 10, e1003858.

RAVASI, T., SUZUKI, H., CANNISTRACI, C. V., KATAYAMA, S., BAJIC, V. B., TAN, K., AKALIN, A., SCHMEIER, S., KANAMORI-KATAYAMA, M., BERTIN, N., CARNINCI, P., DAUB, C. O., FORREST, A. R., GOUGH, J., GRIMMOND, S., HAN, J. H., HASHIMOTO, T., HIDE, W., HOFMANN, O., KAMBUROV, A., KAUR, M., KAWAJI, H., KUBOSAKI, A., LASSMANN, T., VAN NIMWEGEN, E., MACPHERSON, C. R., OGAWA, C., RADOVANOVIC, A., SCHWARTZ, A., TEASDALE, R. D., TEGNER, J., LENHARD, B., TEICHMANN, S. A., ARAKAWA, T., NINOMIYA, N., MURAKAMI, K., TAGAMI, M., FUKUDA, S., IMAMURA, K., KAI, C., ISHIHARA, R., KITAZUME, Y., KAWAI, J., HUME, D. A., IDEKER, T. & HAYASHIZAKI, Y. 2010. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell,* 140, 744-52.

RICKINSON, A. B. & KIEFF, E. 2007. Epstein-Barr Virus. *In:* KNIPE, D. M. & HOWLEY, P. M. (eds.) *Fields Virology.* 5 ed. Philadelphia: Lippincott-Williams & Wilkins, pp. 2656-2700.

ROBERTSON, E. S., GROSSMAN, S., JOHANNSEN, E., MILLER, C., LIN, J., TOMKINSON, B. & KIEFF, E. 1995. Epstein-Barr virus nuclear protein 3C modulates transcription through interaction with the sequence-specific DNA-binding protein J kappa. *J Virol,* 69, 3108-16.

ROBERTSON, E. S., LIN, J. & KIEFF, E. 1996. The amino-terminal domains of Epstein-Barr virus nuclear proteins 3A, 3B, and 3C interact with RBPJ(kappa). *J Virol,* 70, 3068-74.

ROCHFORD, R. & MOORMANN, A. M. 2015. Burkitt's Lymphoma. *Curr Top Microbiol Immunol,* 390, 267-85.

ROULET, E., BUSSO, S., CAMARGO, A. A., SIMPSON, A. J., MERMOD, N. & BUCHER, P. 2002. High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites. *Nat Biotechnol,* 20, 831-5.

ROZOWSKY, J., EUSKIRCHEN, G., AUERBACH, R. K., ZHANG, Z. D., GIBSON, T., BJORNSON, R., CARRIERO, N., SNYDER, M. & GERSTEIN, M. B. 2009. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat Biotechnol,* 27, 66-75.

SAMBROOK, J. & GETHING, M. J. 1989. Protein structure. Chaperones, paperones. *Nature,* 342, 224-5.

SANDERS, D. A., GORMALLY, M. V., MARSICO, G., BERALDI, D., TANNAHILL, D. & BALASUBRAMANIAN, S. 2015. FOXM1 binds directly to non-consensus sequences in the human genome. *Genome Biol,* 16, 130.

SATTERWHITE, E., SONOKI, T., WILLIS, T. G., HARDER, L., NOWAK, R., ARRIOLA, E. L., LIU, H., PRICE, H. P., GESK, S., STEINEMANN, D., SCHLEGELBERGER, B., OSCIER, D. G., SIEBERT, R., TUCKER, P. W. & DYER, M. J. 2001. The BCL11 gene family: involvement of BCL11A in lymphoid malignancies. *Blood,* 98, 3413-20.

SCHMIDT, S. C., JIANG, S., ZHOU, H., WILLOX, B., HOLTHAUS, A. M., KHARCHENKO, P. V., JOHANNSEN, E. C., KIEFF, E. & ZHAO, B. 2015. Epstein-Barr virus nuclear antigen 3A partially coincides with EBNA3C genome-wide and is tethered to DNA through BATF complexes. *Proc Natl Acad Sci U S A,* 112, 554-9.

SCOTT, E. W., SIMON, M. C., ANASTASI, J. & SINGH, H. 1994. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science,* 265, 1573-7.

SETO, E., MOOSMANN, A., GROMMINGER, S., WALZ, N., GRUNDHOFF, A. & HAMMERSCHMIDT, W. 2010. Micro RNAs of Epstein-Barr virus promote cell cycle progression and prevent apoptosis of primary human B cells. *PLoS Pathog,* 6, e1001063.

SILVA, A. L., OMEROVIC, J., JARDETZKY, T. S. & LONGNECKER, R. 2004. Mutational analyses of Epstein-Barr virus glycoprotein 42 reveal functional domains not involved in receptor binding but required for membrane fusion. *J Virol,* 78, 5946-56.

SIPONEN, M. I., WISNIEWSKA, M., LEHTIO, L., JOHANSSON, I., SVENSSON, L., RASZEWSKI, G., NILSSON, L., SIGVARDSSON, M. & 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, 25875-9.

SKALSKA, L., WHITE, R. E., PARKER, G. A., TURRO, E., SINCLAIR, A. J., PASCHOS, K. & ALLDAY, M. J. 2013. Induction of p16(INK4a) is the major barrier to proliferation when Epstein-Barr virus (EBV) transforms primary B cells into lymphoblastoid cell lines. *PLoS Pathog,* 9, e1003187.

SKALSKY, R. L. & CULLEN, B. R. 2015. EBV Noncoding RNAs. *Curr Top Microbiol Immunol,* 391, 181-217.

SKARE, J., EDSON, C., FARLEY, J. & STROMINGER, J. L. 1982. The B95-8 isolate of Epstein-Barr virus arose from an isolate with a standard genome. *J Virol,* 44, 1088-91.

SPENDER, L. C., LUCCHESI, W., BODELON, G., BILANCIO, A., KARSTEGL, C. E., ASANO, T., DITTRICH-BREIHOLZ, O., KRACHT, M., VANHAESEBROECK, B. & FARRELL, P. J. 2006. Cell target genes of Epstein-Barr virus transcription factor EBNA-2: induction of the p55alpha regulatory subunit of PI3-kinase and its role in survival of EREB2.5 cells. *J Gen Virol,* 87, 2859-67.

STEINBRUCK, L., GUSTEMS, M., MEDELE, S., SCHULZ, T. F., LUTTER, D. & HAMMERSCHMIDT, W. 2015. K1 and K15 of Kaposi's Sarcoma-Associated Herpesvirus Are Partial Functional Homologues of Latent Membrane Protein 2A of Epstein-Barr Virus. *J Virol,* 89, 7248-61.

TANNER, J., WEIS, J., FEARON, D., WHANG, Y. & KIEFF, E. 1987. Epstein-Barr virus gp350/220 binding to the B lymphocyte C3d receptor mediates adsorption, capping, and endocytosis. *Cell,* 50, 203-13.

THORLEY-LAWSON, A. D., DUNMIRE, S. K., HOGQUIST, K. A., BALFOUR, H. H., COHEN, J. I., ROCHFORD, R., MOORMANN, A. M., MURRAY, P., BELL, A., HEALEY, J. A., SANDEEP, S. D., RAAB-TRAUB, N., ASCHERIO, A. & MUNGER, K. L. 2015. Part III Viral Infection and Associated Diseases. *In:* MUENZ, C. (ed.) *Epstein-Barr Virus.* Springer International Publishing Switzerland, pp. 151-386.

THORLEY-LAWSON, D. A. 2001. Epstein-Barr virus: exploiting the immune system. *Nature Rev Immunol,* 1, 75-82.

THORLEY-LAWSON, D. A. & ALLDAY, M. J. 2008. The curious case of the tumour virus: 50 years of Burkitt's lymphoma. *Nat Rev Microbiol,* 6, 913-24.

TONG, X., DRAPKIN, R., REINBERG, D. & KIEFF, E. 1995a. The 62- and 80-kDa subunits of transcription factor IIH mediate the interaction with Epstein-Barr virus nuclear protein 2. *Proc Natl Acad Sci U S A,* 92, 3259-63.

TONG, X., WANG, F., THUT, C. J. & 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, 585-8.

TOUITOU, R., HICKABOTTOM, M., PARKER, G., CROOK, T. & ALLDAY, M. J. 2001. Physical and functional interactions between the corepressor CtBP and the Epstein-Barr virus nuclear antigen EBNA3C. *J Virol,* 75, 7749-55.

TREIBER, N., TREIBER, T., ZOCHER, G. & GROSSCHEDL, R. 2010a. Structure of an Ebf1:DNA complex reveals unusual DNA recognition and structural homology with Rel proteins. *Genes Dev,* 24, 2270-5.

TREIBER, T., MANDEL, E. M., POTT, S., GYORY, I., FIRNER, S., LIU, E. T. & GROSSCHEDL, R. 2010b. Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. *Immunity,* 32, 714-25.

UVERSKY, V. N. 2013. The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini. *FEBS Lett,* 587, 1891-901.

VADNAIS, C., AWAN, A. A., HARADA, R., CLERMONT, P. L., LEDUY, L., BERUBE, G. & NEPVEU, A. 2013. Long-range transcriptional regulation by the p110 CUX1 homeodomain protein on the ENCODE array. *BMC Genomics,* 14, 258.

VAN DAM, H. & CASTELLAZZI, M. 2001. Distinct roles of Jun : Fos and Jun : ATF dimers in oncogenesis. *Oncogene,* 20, 2453-64.

WAKEMAN, B. S., JOHNSON, L. S., PADEN, C. R., GRAY, K. S., VIRGIN, H. W. & SPECK, S. H. 2014. Identification of alternative transcripts encoding the essential murine gammaherpesvirus lytic transactivator RTA. *J Virol,* 88, 5474-90.

WALTZER, L., PERRICAUDET, M., SERGEANT, A. & MANET, E. 1996. Epstein-Barr virus EBNA3A and EBNA3C proteins both repress RBP-J kappa-EBNA2-activated transcription by inhibiting the binding of RBP-J kappa to DNA. *J Virol,* 70, 5909-15.

WANG, A., WELCH, R., ZHAO, B., TA, T., KELES, S. & JOHANNSEN, E. 2015. Epstein-Barr Virus Nuclear Antigen 3 (EBNA3) Proteins Regulate EBNA2 Binding to Distinct RBPJ Genomic Sites. *J Virol,* 90, 2906-19.

WANG, F., KIKUTANI, H., TSANG, S. F., KISHIMOTO, T. & KIEFF, E. 1991. Epstein-Barr virus nuclear protein 2 transactivates a cis-acting CD23 DNA element. *J Virol,* 65, 4101-6.

WANG, H., ZANG, C., TAING, L., ARNETT, K. L., WONG, Y. J., PEAR, W. S., BLACKLOW, S. C., LIU, X. S. & ASTER, J. C. 2014. NOTCH1-RBPJ complexes drive target gene expression through dynamic interactions with superenhancers. *Proc Natl Acad Sci U S A,* 111, 705-10.

WANG, L., GROSSMAN, S. R. & 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, 430-5.

WARDEN, C., TANG, Q. & ZHU, H. 2011. Herpesvirus BACs: past, present, and future. *J Biomed Biotechnol,* 2011, 124595.

WARMING, S., COSTANTINO, N., COURT, D. L., JENKINS, N. A. & COPELAND, N. G. 2005. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res,* 33, e36.

WEST, M. J., WEBB, H. M., SINCLAIR, A. J. & WOOLFSON, D. N. 2004. Biophysical and mutational analysis of the putative bZIP domain of Epstein-Barr virus EBNA 3C. *J Virol,* 78, 9431-45.

WHITE, R. E., GROVES, I. J., TURRO, E., YEE, J., KREMMER, E. & ALLDAY, M. J. 2010. Extensive co-operation between the Epstein-Barr virus EBNA3 proteins in the manipulation of host gene expression and epigenetic chromatin modification. *PLoS One,* 5, e13979.

WILSON, B. G., WANG, X., SHEN, X., MCKENNA, E. S., LEMIEUX, M. E., CHO, Y. J., KOELLHOFFER, E. C., POMEROY, S. L., ORKIN, S. H. & ROBERTS, C. W. 2010. Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer Cell,* 18, 316-28.

YENAMANDRA, S. P., SOMPALLAE, R., KLEIN, G. & KASHUBA, E. 2009. Comparative analysis of the Epstein-Barr virus encoded nuclear proteins of EBNA-3 family. *Comput Biol Med,* 39, 1036-42.

YOUNG, L. S. & MURRAY, P. G. 2003. Epstein-Barr virus and oncogenesis: from latent genes to tumours. *Oncogene,* 22, 5108-21.

YU, Y., WANG, J., KHALED, W., BURKE, S., LI, P., CHEN, X., YANG, W., JENKINS, N. A., COPELAND, N. G., ZHANG, S. & LIU, P. 2012. Bcl11a is essential for lymphoid development and negatively regulates p53. *J Exp Med,* 209, 2467-83.

YUE, W., DAVENPORT, M. G., SHACKELFORD, J. & PAGANO, J. S. 2004. Mitosis-specific hyperphosphorylation of Epstein-Barr virus nuclear antigen 2 suppresses its function. *J Virol,* 78, 3542-52.

ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol,* 9, R137.

ZHAO, B., MAR, J. C., MARUO, S., LEE, S., GEWURZ, B. E., JOHANNSEN, E., HOLTON, K., RUBIO, R., TAKADA, K., QUACKENBUSH, J. & KIEFF, E. 2011a. Epstein-Barr virus nuclear antigen 3C regulated genes in lymphoblastoid cell lines. *Proc Natl Acad Sci U S A,* 108, 337-42.

ZHAO, B., MARSHALL, D. R. & SAMPLE, C. E. 1996. A conserved domain of the Epstein-Barr virus nuclear antigens 3A and 3C binds to a discrete domain of Jkappa. *J Virol,* 70, 4228-36.

ZHAO, B., MARUO, S., COOPER, A., M, R. C., JOHANNSEN, E., KIEFF, E. & CAHIR-MCFARLAND, E. 2006. RNAs induced by Epstein-Barr virus nuclear antigen 2 in lymphoblastoid cell lines. *Proc Natl Acad Sci U S A,* 103, 1900-5.

ZHAO, B. & SAMPLE, C. E. 2000. Epstein-barr virus nuclear antigen 3C activates the latent membrane protein 1 promoter in the presence of Epstein-Barr virus nuclear antigen 2 through sequences encompassing an spi-1/Spi-B binding site. *J Virol,* 74, 5151-60.

ZHAO, B., ZOU, J., WANG, H., JOHANNSEN, E., PENG, C. W., QUACKENBUSH, J., MAR, J. C., MORTON, C. C., FREEDMAN, M. L., BLACKLOW, S. C., ASTER, J. C., BERNSTEIN, B. E. & KIEFF, E. 2011b. Epstein-Barr virus exploits intrinsic B-lymphocyte transcription programs to achieve immortal cell growth. *Proc Natl Acad Sci U S A,* 108, 14902-7.

ZHOU, H., SCHMIDT, S. C., JIANG, S., WILLOX, B., BERNHARDT, K., LIANG, J., JOHANNSEN, E. C., KHARCHENKO, P., GEWURZ, B. E., KIEFF, E. & ZHAO, B. 2015. Epstein-Barr virus oncoprotein super-enhancers control B cell growth. *Cell Host Microbe,* 17, 205-16.

ZUR HAUSEN, A., VAN REES, B. P., VAN BEEK, J., CRAANEN, M. E., BLOEMENA, E., OFFERHAUS, G. J., MEIJER, C. J. & VAN DEN BRULE, A. J. 2004. Epstein-Barr virus in gastric carcinomas and gastric stump carcinomas: a late event in gastric carcinogenesis. *J Clin Pathol,* 57, 487-91.

7 Appendices

7.1 Supplementary figures

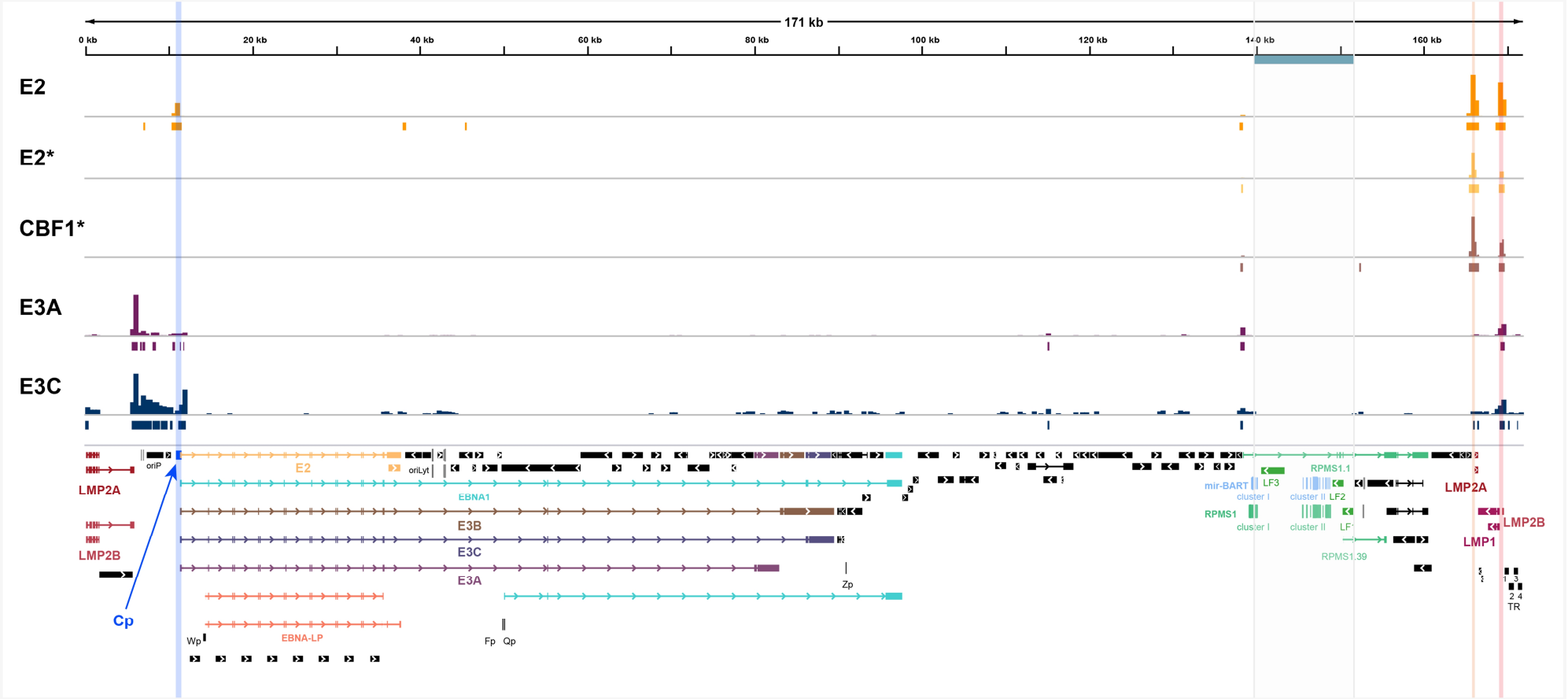


Figure S1. Identification of E2, E3A, and E3C binding sites in the EBV genome. Schematic map of the EBV genome (HHV-4 type I, NC_007605.1) as provided by the EBV portal (Arvey et al., 2012) depicting gene positions (lower panel). Genes expressed during the lytic cycle are depicted in black, microRNAs in grey and genes expressed during latency are highlighted in color. Also marked is the EBNA regulated Cp, which gives rise to different (polycistronic) splice variants coding for all EBNAs, including proteins of interest E2, E3A, and E3C. The light grey box to the right encompasses a region, which is deleted in the B95.8 EBV genome used for EBV BACmid generation compared to HHV-4 type I reference genome. Genes affected by the B95.8 deletion are highlighted in blue and green. Thus it is not possible that reads from ChIP-seq analysis derive from this genomic region. EBNA regulated *LMP1*, *LMP2A*, and *LMP2B* genes are shown in red, with the bidirectional promoter controlling *LMP1* and *LMP2B* expression as well as the *LMP2A* promoter highlighted with light red columns. In the upper panels ChIP-seq signal profiles and underneath peaks called by MACS2 for E2, E3A, and E3C are shown. (*) Additionally published data for E2 and CBF1 (Zhao et al., 2011b) is shown for comparison.

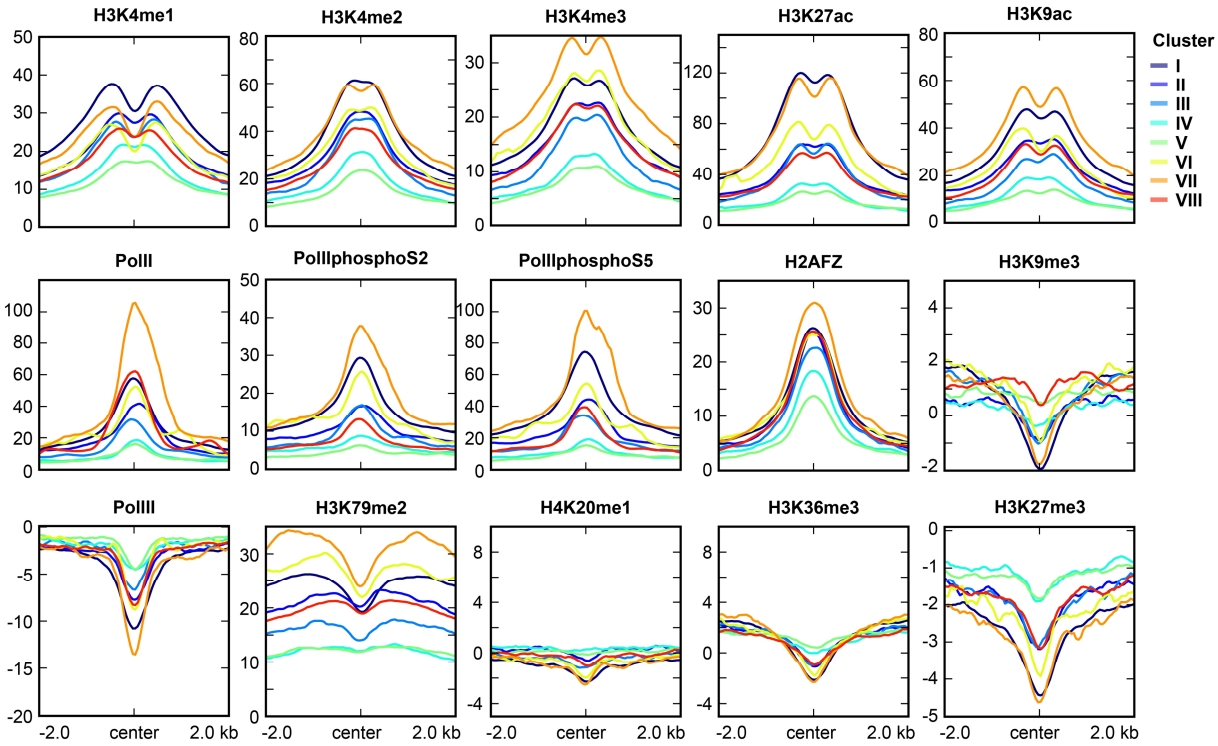


Figure S2. Signal distribution of histone modifications, histone variant H2AFZ, and RNA polymerases at E2 peak clusters. Anchor plots depicting mean signal distributions of available histone modifications, histone variants, and RNA polymerases (variants) at the 8 different E2 peak clusters. A region of 2 kb in each direction of the peak center was analyzed. ChIP-seq signals from ENCODE were normalized to their respective input samples and RPKM (see chapter 3.6.1).

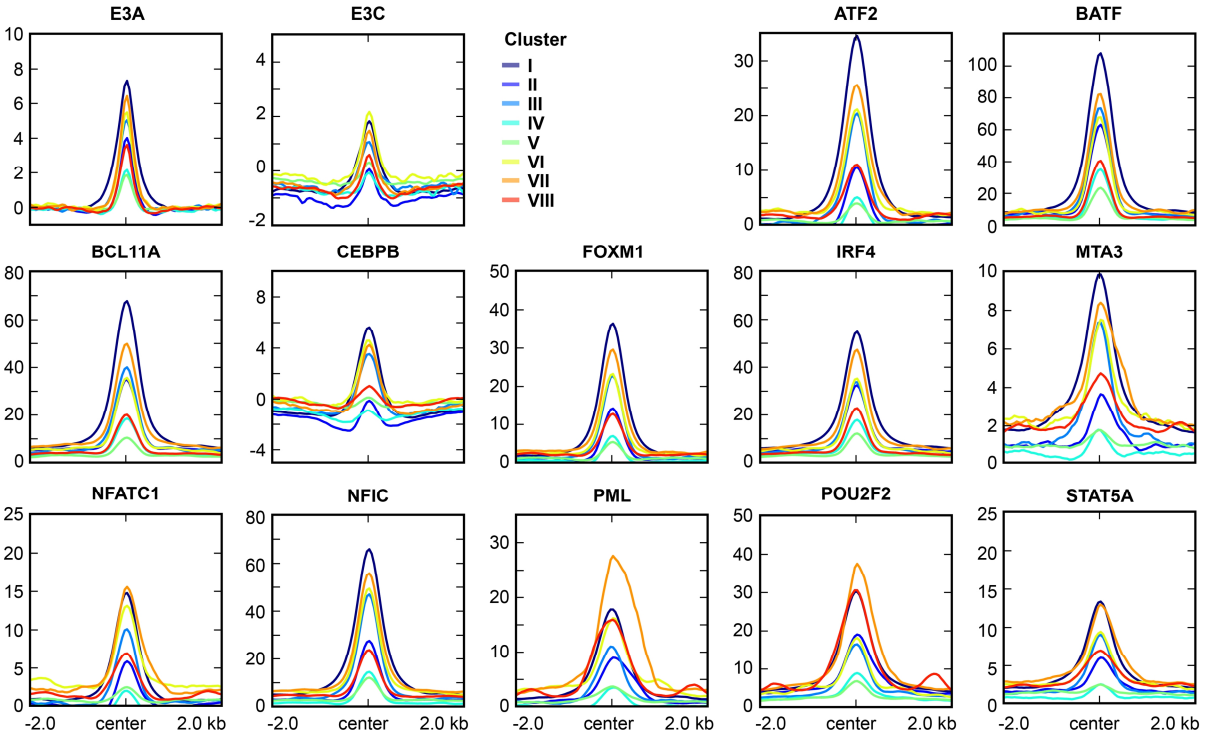


Figure S3. Signal distribution of TFs found to cluster with E3A and E3C in EBNA peak wide correlation analysis at E2 peak clusters. Anchor plots depicting mean signal distributions of TFs, which are positively correlating with E3 signals (as detected by correlation analysis using EBNA peaks as reference) at the 8 different E2 peak clusters. A region of 2 kb in each direction of the peak center was analyzed. ChIP-seq signals from ENCODE were normalized to their respective input samples and RPKM (see chapter 3.6.1).

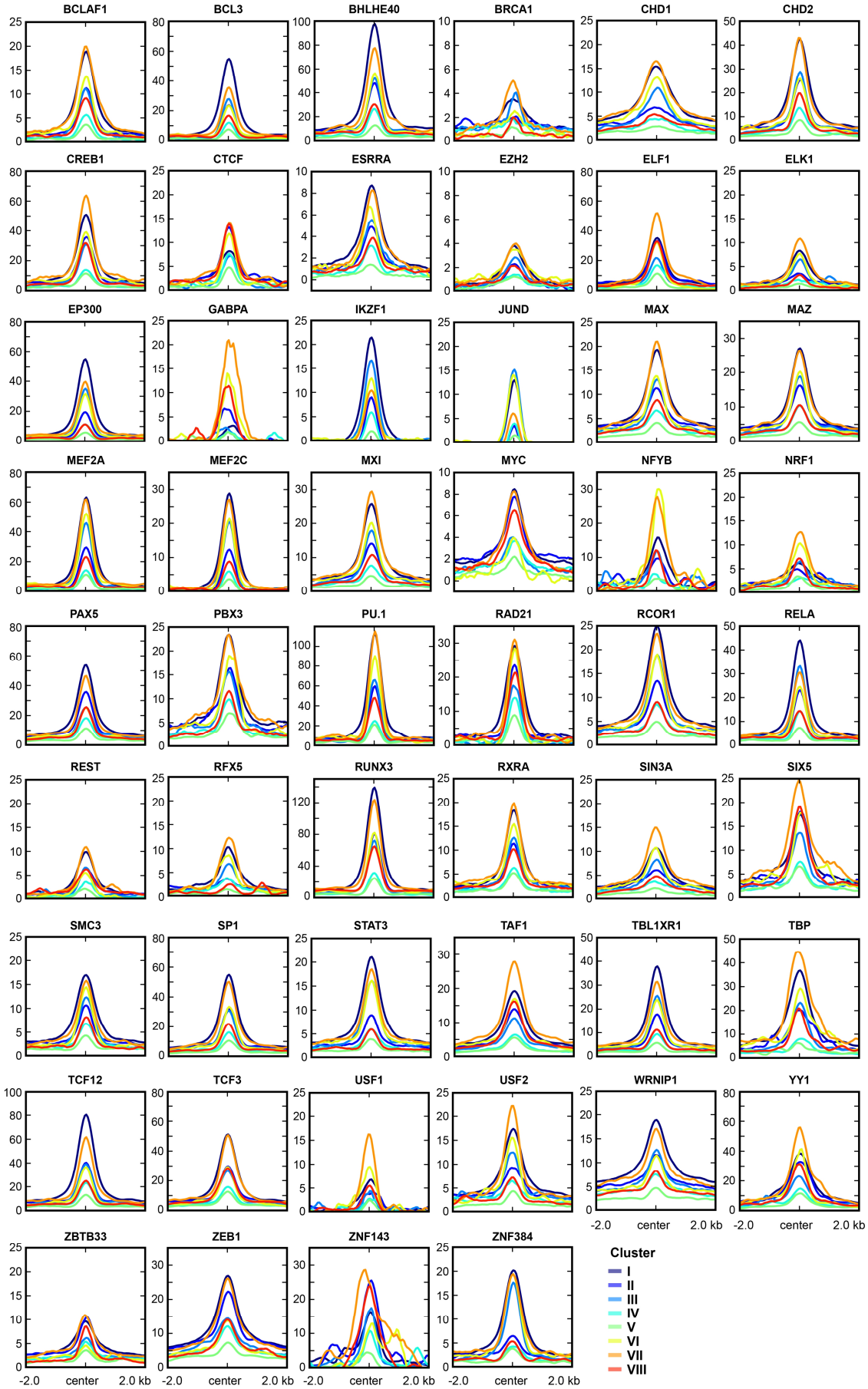


Figure S4. Signal distribution of TFs analyzed by ENCODE, which were not identified in E2 or E3 clusters in EBNA peak wide correlation analysis but showed enrichment at E2 peak clusters. Anchor plots depicting

mean signal distributions of TFs, which are positively correlating with E3 signals (as detected by correlation analysis using EBNA peaks as reference) at the 8 different E2 peak clusters. A region of 2 kb in each direction of the peak center was analyzed. ChIP-seq signals from ENCODE were normalized to their respective input samples and RPKM (see chapter 3.6.1).

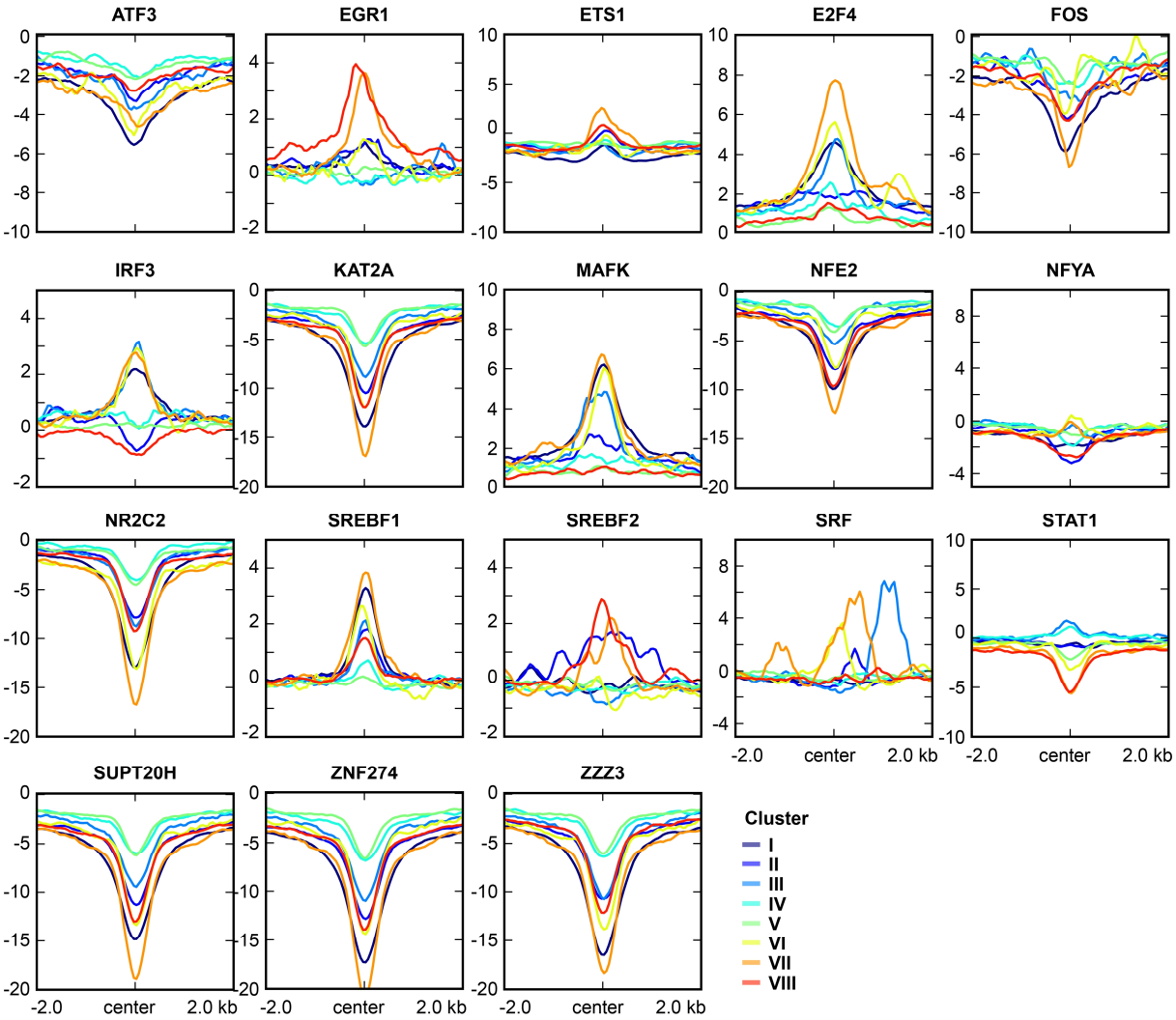


Figure S5. Signal distribution of TFs analyzed by ENCODE, which were not identified in E2 or E3 clusters in EBNA peak wide correlation analysis and were not enriched at E2 peak clusters. Anchor plots depicting mean signal distributions of TFs, which are positively correlating with E3 signals (as detected by correlation analysis using EBNA peaks as reference) at the 8 different E2 peak clusters. A region of 2 kb in each direction of the peak center was analyzed. ChIP-seq signals from ENCODE were normalized to their respective input samples and RPKM (see chapter 3.6.1).

7.2 Supplementary Tables

Table S1. TF and histone modification ChIP-seq experiments by the ENCODE project used in this study

ChIP	Experiment Name	Narrow Peaks (bed)	Broad Peaks (bed)	ENCODE Reanalyzed	Lab	Release Date	Antibody
ATF2	ENCSR000BQK	-	ENCFF001TWB, ENCFF001TWC	ENCFF002CGO	R. Myers, HAIB	29.02.12	ENCAB000ASU
ATF3	ENCSR000BJY	-	ENCFF001TWD, ENCFF001TWE	ENCFF002CGP	R. Myers, HAIB	18.07.11	ENCAB000ADZ
BATF	ENCSR000BGT	-	ENCFF001TWF, ENCFF001TWG	ENCFF002CGQ	R. Myers, HAIB	18.07.11	ENCAB000AED
BCL11A	ENCSR000BHA	-	ENCFF001TWH, ENCFF001TWI	ENCFF002CGR	R. Myers, HAIB	18.07.11	ENCAB000AEE
BCL3	ENCSR000BNQ	-	ENCFF001TWJ, ENCFF001TWK	ENCFF002CGS	R. Myers, HAIB	18.07.11	ENCAB000AEG
BCLAF1	ENCSR000BJZ	-	ENCFF001TWL, ENCFF001TWM	ENCFF002CGT	R. Myers, HAIB	18.07.11	ENCAB000AEH
BHLHE40	ENCSR000DZJ	ENCFF001VDW	-	ENCFF002COK	M. Snyder, Stanford	29.10.11	ENCAB000AEK
BRCA1	ENCSR000DZS	ENCFF001VDX	-	ENCFF002COL	M. Snyder, Stanford	29.10.11	ENCAB000AEL
CEBPB	ENCSR000BRX	-	ENCFF001TWN, ENCFF001TWO	ENCFF002CGU	R. Myers, HAIB	10.09.12	ENCAB000AFB
CHD1	ENCSR000DZE	ENCFF001VEA	-	ENCFF002CON	M. Snyder, Stanford	14.05.12	ENCAB000AFE
CHD2	ENCSR000DZR	ENCFF001VEB	-	ENCFF002COO	M. Snyder, Stanford	29.10.11	ENCAB000AFG
CREB1	ENCSR000BUF	-	ENCFF001TWP, ENCFF001TWQ	-	R. Myers, HAIB	10.09.12	ENCAB000AFN
CTCF	ENCSR000AKB	-	ENCFF001SUB	ENCFF002CDP	B. Bernstein, Broad	10.02.11	ENCAB000AXY
CUX1	ENCSR000DYR	ENCFF001VDY	-	-	M. Snyder, Stanford	20.08.12	ENCAB000AFA
E2F4	ENCSR000DYY	ENCFF001VEE	-	ENCFF002COR	M. Snyder, Stanford	14.05.12	ENCAB000AFV
EBF1	ENCSR000DZQ	ENCFF001VEF	-	ENCFF002COS	M. Snyder, Stanford	29.10.11	ENCAB000AFX
EGR1	ENCSR000BRG	-	ENCFF001TWS	ENCFF002CGW	R. Myers, HAIB	29.02.12	ENCAB000ASX
ELF1	ENCSR000BMB	-	ENCFF000NVE, ENCFF001TWX	ENCFF002CGX	R. Myers, HAIB	18.07.11	ENCAB000AGA
ELK1	ENCSR000DZB	ENCFF001VEG	-	ENCFF002COT	M. Snyder, Stanford	14.05.12	ENCAB000AGB
EP300	ENCSR000DZD	ENCFF001VEX	-	ENCFF002CPE	M. Snyder, Stanford	14.05.12	ENCAB000AJM
ESRRA	ENCSR000DYQ	ENCFF001VEH	-	-	M. Snyder, Stanford	20.08.12	ENCAB000AGE
ETS1	ENCSR000BKA	-	ENCFF001TWZ, ENCFF001TXB	ENCFF002CGY	R. Myers, HAIB	18.07.11	ENCAB000AGG
EZH2	ENCSR000ARD	-	ENCFF001SUC	ENCFF002CDQ	B. Bernstein, Broad	06.03.12	ENCAB000AGH
FOS	ENCSR000EYZ	ENCFF001VDZ	-	ENCFF002COM	S. Weissman, Yale	29.10.11	ENCAB000AEQ
FOXM1	ENCSR000BRU	-	ENCFF001TXC, ENCFF001TXD	ENCFF002CGZ	R. Myers, HAIB	29.02.12	ENCAB000AGP
GABPA	ENCSR000BGC	-	ENCFF001TXE, ENCFF001TXF	ENCFF002CHA	R. Myers, HAIB	18.07.11	ENCAB000AGR
H2AFZ	ENCSR000AOV	-	ENCFF001SUD	-	B. Bernstein, Broad	10.02.11	ENCAB000ASY
H3K27ac	ENCSR000AKC	-	ENCFF001SUG	-	B. Bernstein, Broad	10.02.11	ENCAB000ANA
H3K27me3	ENCSR000AKD	-	ENCFF001SUI	-	B. Bernstein, Broad	10.02.11	ENCAB000ANB
H3K36me3	ENCSR000AKE	-	ENCFF001SUJ	-	B. Bernstein, Broad	10.02.11	ENCAB000ADU
H3K4me1	ENCSR000AKF	-	ENCFF001SUE	-	B. Bernstein, Broad	10.02.11	ENCAB000ADW
H3K4me2	ENCSR000AKG	-	ENCFF001SUL	-	B. Bernstein, Broad	10.02.11	ENCAB000ANF
H3K4me3	ENCSR000AKA	-	ENCFF001SUF	-	B. Bernstein, Broad	10.02.11	ENCAB000BLJ
H3K79me2	ENCSR000AOW	-	ENCFF001SUN	-	B. Bernstein, Broad	10.02.11	ENCAB000ANH
H3K9ac	ENCSR000AKH	-	ENCFF001SUO	-	B. Bernstein, Broad	10.02.11	ENCAB000ANK
H3K9me3	ENCSR000AOX	-	ENCFF001SUP	-	B. Bernstein, Broad	10.02.11	ENCAB000ANX
H4K20me1	ENCSR000AKI	-	ENCFF001SUQ	-	B. Bernstein, Broad	10.02.11	ENCAB000ANZ
IKZF1	ENCSR000EUJ	ENCFF001VEJ	-	ENCFF002COU	P.Farnham, USC	14.05.12	ENCAB000AHV

IRF3	ENCSR000DZX	ENCFF001VEK	-	-	M.I Snyder, Stanford	29.10.11	ENCAB000AHY
IRF4	ENCSR000BGY	-	ENCFF001TXG, ENCFF001TXH	ENCFF002CHB	R. Myers, HAIB	18.07.11	ENCAB000AHZ
JUND	ENCSR000EYV	ENCFF001VEM	-	ENCFF002COV	M. Snyder, Stanford	29.10.11	ENCAB000AID
KAT2A	ENCSR000DNO	ENCFF001VEI	-	-	K. Struhl, HMS	29.10.11	ENCAB000AHA
MAFK	ENCSR000DYV	ENCFF001VEN	-	-	M. Snyder, Stanford	20.08.12	ENCAB000AIJ
MAX	ENCSR000DZF	ENCFF001VEO	-	ENCFF002COW	M. Snyder, Stanford	14.05.12	ENCAB000AIL
MAZ	ENCSR000DZA	ENCFF001VEQ	-	ENCFF002COX	M. Snyder, Stanford	14.05.12	ENCAB000AIM
MEF2A	ENCSR000BKB	-	ENCFF001TXI, ENCFF001TXJ	ENCFF002CHC	R. Myers, HAIB	18.07.11	ENCAB000AIQ
MEF2C	ENCSR000BNG	-	ENCFF001TXK, ENCFF001TXL	ENCFF002CHD	R. Myers, HAIB	18.07.11	ENCAB000AIR
MTA3	ENCSR000BRH	-	ENCFF001TXM, ENCFF001TXN	ENCFF002CHE	R. Myers, HAIB	29.02.12	ENCAB000AIS
MXI1	ENCSR000DZI	ENCFF001VER	-	ENCFF002COY	M. Snyder, Stanford	29.10.11	ENCAB000AIT
MYC	ENCSR000DKU	ENCFF001USG	-	ENCFF002DAI	V. Iyer, UTA	17.03.11	ENCAB000AET
NFATC1	ENCSR000BQL	-	ENCFF001TXO; ENCFF001TXP	ENCFF002CHF	R. Myers, HAIB	29.02.12	ENCAB000AJE
NFE2	ENCSR000DZY	ENCFF001VES	-	ENCFF002COZ	M. Snyder, Stanford	29.10.11	ENCAB000AJB
NFIC	ENCSR000BRN	-	ENCFF001TXQ, ENCFF001TXR	ENCFF002CHG	R. Myers, HAIB	29.02.12	ENCAB000AJF
NFYA	ENCSR000DNN	ENCFF001VEU	-	ENCFF002CPB	K. Struhl, HMS	29.10.11	
NFYB	ENCSR000DNM	ENCFF001VEV	-	ENCFF002CPC	K. Struhl, HMS	29.10.11	ENCAB000AJD
NR2C2	ENCSR000EUL	ENCFF001VFP	-	ENCFF002CPS	P.Farnham, USC	29.10.11	ENCAB000AMA
NRF1	ENCSR000DZO	ENCFF001VEW	-	ENCFF002CPD	M. Snyder, Stanford	29.10.11	ENCAB000AJI
PAX5	ENCSR000BHD	-	ENCFF001TXY, ENCFF001TXZ	ENCFF002CHJ	R. Myers, HAIB	18.07.11	ENCAB000AJS
PBX3	ENCSR000BGR	-	ENCFF001TYC, ENCFF001TYD	ENCFF002CHL	R. Myers, HAIB	18.07.11	ENCAB000AJU
PML	ENCSR000BQM	-	ENCFF001TYE, ENCFF001TYF	ENCFF002CHM	R. Myers, HAIB	29.02.12	ENCAB000AKA
POLR2A	ENCSR000EAD	ENCFF001VFA	-	ENCFF002CPG	M. Snyder, Stanford	14.05.12	ENCAB000AOC
POLR3G	ENCSR000EYU	ENCFF001VFC	-	ENCFF002CPJ	S. Weissman, Yale	29.10.11	ENCAB000AKB
POU2F2	ENCSR000BGP	-	ENCFF001TYK, ENCFF001TYL; ENCFF001TYM	ENCFF002CHP	R. Myers, HAIB	18.07.11	ENCAB000AKC
RAD21	ENCSR000BMY	-	ENCFF001TYQ, ENCFF001TYR	ENCFF002CHR	R. Myers, HAIB	18.07.11	ENCAB000AKG
RCOR1	ENCSR000DZC	ENCFF001VEC	-	ENCFF002COP	M. Snyder, Stanford	14.05.12	ENCAB000AFK
RELA	ENCSR000EAG	ENCFF001VET	-	ENCFF002CPA	M. Snyder, Stanford	14.05.12	ENCAB000AJG
REST	ENCSR000BQS	-	ENCFF001TXS, ENCFF001TXT	ENCFF002CHH	R. Myers, HAIB	29.02.12	ENCAB000AJK
RFX5	ENCSR000DZW	ENCFF001VFF	-	ENCFF002CPL	M. Snyder, Stanford	29.10.11	ENCAB000AKJ
RUNX3	ENCSR000BRI		ENCFF001TYS, ENCFF001TYU	ENCFF002CHS	R. Myers, HAIB	29.02.12	ENCAB000AKM
RXRA	ENCSR000BJD	-	ENCFF001TYT, ENCFF001TYV	ENCFF002CHT	R. Myers, HAIB	18.07.11	ENCAB000AKN
SIN3A	ENCSR000DYX	ENCFF001VFG	-	ENCFF002CPM	M. Snyder, Stanford	14.05.12	ENCAB000AKR
SIX5	ENCSR000BJE	-	ENCFF001TYW, ENCFF001TYX	ENCFF002CHU	R. Myers, HAIB	18.07.11	ENCAB000AKV
SMC3	ENCSR000DZP	ENCFF001VFH	-	ENCFF002CPN	M. Snyder, Stanford	29.10.11	ENCAB000AKX
SP1	ENCSR000BHK	-	ENCFF001TYZ, ENCFF001TYY	ENCFF002CHV	R. Myers, HAIB	18.07.11	ENCAB000AKY
SPI1	ENCSR000BGQ	-	ENCFF001TYN, ENCFF001TYO, ENCFF001TYP	ENCFF002CHQ	R. Myers, HAIB	18.07.11	ENCAB000AKF
SREBF1	ENCSR000DYU	ENCFF001VFJ	-	-	M. Snyder, Stanford	20.08.12	ENCAB000ALC
SREBF2	ENCSR000DYT	ENCFF001VFK	-	-	M. Snyder, Stanford	20.08.12	ENCAB000ALD
SRF	ENCSR000BGE	-	ENCFF001TZA, ENCFF001TZB	ENCFF002CHW	R. Myers, HAIB	18.07.11	ENCAB000ALE

STAT1	ENCSR000DZM	ENCFF001VFL	-	ENCFF002CPO	M. Snyder, Stanford	29.10.11	ENCAB000ALF
STAT3	ENCSR000DZV	ENCFF001VFM	-	ENCFF002CPP	M. Snyder, Stanford	29.10.11	ENCAB000ALH
STAT5A	ENCSR000BQZ	-	ENCFF001TZE, ENCFF001TZF	ENCFF002CHX	R. Myers, HAIB	29.02.12	ENCAB000ALI
SUPT20H	ENCSR000DNP	ENCFF001VFI	-	-	K. Struhl, HMS	29.10.11	ENCAB000ALB
TAF1	ENCSR000BGS	-	ENCFF001TZG, ENCFF001TZH	ENCFF002CHY	R. Myers, HAIB	18.07.11	ENCAB000ALM
TBL1XR1	ENCSR000DYZ	ENCFF001VFN	-	ENCFF002CPQ	M. Snyder, Stanford	14.05.12	ENCAB000ALP
TBP	ENCSR000DZZ	ENCFF001VFO	-	ENCFF002CPR	M. Snyder, Stanford	29.10.11	ENCAB000ALR
TCF12	ENCSR000BGZ	-	ENCFF001TZI, ENCFF001TZJ	ENCFF002CHZ	R. Myers, HAIB	18.07.11	ENCAB000ALT
TCF3	ENCSR000BQT	-	ENCFF001TZK, ENCFF001TZL	ENCFF002CIA	R. Myers, HAIB	29.02.12	ENCAB000ALU
USF1	ENCSR000BGI	-	ENCFF001TZM, ENCFF001TZN	ENCFF002CIB	R. Myers, HAIB	18.07.11	ENCAB000AMF
USF2	ENCSR000DZU	ENCFF001VFQ	-	ENCFF002CPT	M. Snyder, Stanford	29.10.11	ENCAB000AMH
WRNIP1	ENCSR000EAA	ENCFF001VFS	-	ENCFF002CPU	M. Snyder, Stanford	29.10.11	ENCAB000AMJ
YY1	ENCSR000BNP	-	ENCFF001TZO, ENCFF001TZP	ENCFF002CIC	R. Myers, HAIB	18.07.11	ENCAB000ANT
ZBTB33	ENCSR000BHC	-	ENCFF001TZQ, ENCFF001TZR	ENCFF002CID	R. Myers, HAIB	18.07.11	ENCAB000AML
ZEB1	ENCSR000BND	-	ENCFF001TZS, ENCFF001TZT	ENCFF002CIE	R. Myers, HAIB	18.07.11	ENCAB000AMO
ZNF143	ENCSR000DZL	ENCFF001VFU	-	ENCFF002CPW	M. Snyder, Stanford	29.10.11	ENCAB000AMR
ZNF274	ENCSR000EUK	ENCFF001VFT	-	ENCFF002CPX	P. Farnham, USC	29.10.11	ENCAB000AMU
ZNF384	ENCSR000DYP	ENCFF001VFW, ENCFF000WGY (bigbed)		-	M. Snyder, Stanford	20.08.12	ENCAB000AMW
ZZZ3	ENCSR000DNQ	ENCFF001VFV	-	ENCFF002CPY	K. Struhl, HMS	29.10.11	ENCAB000AMX

ChIP-seq experiments for TFs and histone modifications used in this thesis are listed. All supplied information including experimental procedures and all submitted files can be found at www.encodeproject.org. Peak files (bed) which were used for e.g. cluster analyses are highlighted in bold letters. Signal tracks published by the ENCODE project were not used in this thesis, since they are not normalized to input samples. To correct for input reads signal tracks from aligned reads (bam files) were generated using a Galaxy workflow described in 3.6.1.

Table S2. Accession numbers for data published by other laboratories used in this thesis

Name	Cells Line	Data Deposit Platform	Accession No. (Experiment)	Accession No. (Sample)	ENCODE experiment	Description	Publication
DNaseI HS	GM12878	ENCODE/GEO	GSE29692	GSM736620	ENCSR000EMT	DNase-seq	-
H3K4me1	CD19+ primary cells	GEO	GSE18927	GSM1027296	-	ChIP-seq	Bernstein et al. (2010)
H3K4me3				GSM1027300	-	ChIP-seq	
H3K27ac				GSM1027287	-	ChIP-seq	
Input				GSM1027304	-	ChIP-seq control	
DNaseI HS				GSM701507	-	DNase-seq	
H3K4me1	DG75	EMBL-EBI European Nucleotide Archive (ENA)	PRJEB1912 (study) ERS333899 (sample = DG75)	ERX297414	-	ChIP-seq	Kretzmer et al. (2015)
H3K4me3				ERX297407	-	ChIP-seq	
H3K27ac				ERX297417	-	ChIP-seq	
Input				ERX297450	-	ChIP-seq control	

ChIP-seq and DNase-seq experiments used in this thesis are listed including accession details. For all listed experiments fastqsanger files were downloaded and read mapping as well as further down-stream processing was conducted as explained in chapter 3.6.

Table S3. Expression levels of the TFs included in the ENCODE ChIP-seq data set used in this thesis

Gene ID	Gene Name	Expression Level H1-hESC	CD20+ B cells	GM12878	Cluster
ENSG00000115966	ATF2	8	10	26	
ENSG00000162772	ATF3	4	22	4	
ENSG00000156127	BATF		6	32	
	BCL11A				
	BCL3				
ENSG00000029363	BCLAF1	8	27	36	
ENSG00000134107	BHLHE40	0.9	18	39	
ENSG00000012048	BRCA1	3	0.6	12	
ENSG00000172216	CEBPB	4	10	0.8	
ENSG00000153922	CHD1	3	7	12	
ENSG00000173575	CHD2	4	16	9	
ENSG00000118260	CREB1	4	9	10	
ENSG00000102974	CTCF	18	25	17	
ENSG00000257923	CUX1	3	2	1	
ENSG00000205250	E2F4	65	54	16	
ENSG00000164330	EBF1		22	6	
ENSG00000120738	EGR1	18	98	1	
ENSG00000120690	ELF1	2	40	34	
ENSG00000126767	ELK1	15	8	8	
ENSG00000100393	EP300	22	24	3	
ENSG00000173153	ESRRA	10	9	3	
ENSG00000134954	ETS1	4	39	21	
ENSG00000106462	EZH2	12	6	31	
	FOS				
ENSG00000111206	FOXM1	23	1	5	
ENSG00000154727	GABPA	2	6	16	
ENSG00000185811	IKZF1		36	17	
ENSG00000126456	IRF3	23	43	7	
ENSG00000137265	IRF4		35	148	
ENSG00000140968	IRF8		155	70	
ENSG00000130522	JUND	16	1432	4	
ENSG00000108773	KAT2A	22	71	13	
	MAFK				
ENSG00000125952	MAX	9	32	17	
ENSG00000103495	MAZ	26	20	4	
ENSG00000068305	MEF2A	6	19	17	
ENSG00000081189	MEF2C		18	21	
ENSG00000057935	MTA3	16	3	6	
ENSG00000119950	MXI1	2	6	3	
ENSG00000136997	MYC	15	40	24	
	NFATC1				
	NFE2				
	NFIC				
ENSG00000001167	NFYA	8	10	7	
ENSG00000120837	NFYB	6	2	10	
ENSG00000177463	NR2C2	7	13	9	
ENSG00000106459	NRF1	8	10	6	
ENSG00000196092	PAX5		92	8	
ENSG00000167081	PBX3	8	6	49	
ENSG00000140464	PML	3	9	1	
ENSG00000028277	POU2F2	1	45	7	
ENSG00000164754	RAD21	16	18	58	
ENSG00000168214	RBPJ	21	4	32	
ENSG00000089902	RCOR1	7	7	6	
ENSG00000173039	RELA	6	35	4	
ENSG00000084093	REST	5	6	10	
ENSG00000143390	RFX5	13	71	38	
ENSG00000020633	RUNX3		35	9	
ENSG00000186350	RXRA	3	0.6	0.9	
ENSG00000169375	SIN3A	13	10	5	
	SIX5				
ENSG00000108055	SMC3	9	7	24	
ENSG00000185591	SP1	20	15	21	
ENSG00000066336	SPI1		55	2	
ENSG00000072310	SREBF1	10	9	1	
ENSG00000198911	SREBF2	70	43	27	
ENSG00000112658	SRF	57	20	18	
ENSG00000115415	STAT1	10	9	168	
ENSG00000168610	STAT3	14	14	17	
ENSG00000126561	STAT5A	2	17	8	
ENSG00000102710	SUPT20H	8	6	8	
ENSG00000147133	TAF1	2	5	5	
ENSG00000177565	TBL1XR1	7	8	18	
ENSG00000112592	TBP	10	16	13	
ENSG00000140262	TCF12	11	7	10	
ENSG00000071564	TCF3	64	98	7	
ENSG00000158773	USF1	16	98	11	
ENSG00000105698	USF2	16	62	12	
ENSG00000124535	WRNIP1	12	14	7	
ENSG00000100811	YY1	13	14	13	
ENSG00000177485	ZBTB33	2	4	16	
ENSG00000148516	ZEB1		5	12	
ENSG00000166478	ZNF143	7	8	8	
ENSG00000171606	ZNF274	3	12	3	
ENSG00000126746	ZNF384	12	13	7	
ENSG00000036549	ZZZ3	3	4	8	

List of all the TFs used by ENCODE for ChIP-seq in GM12878 at the time of this analysis and their expression levels as determined and quantified by RNA-seq analysis of long poly adenylated RNA in different ENCODE cell lines (E-GEOD-26284) (Djebali et al., 2012). Data was downloaded via EMBL-EBI Expression Atlas. Expression levels cannot be directly compared between factors since e.g. RNA stability and translation rates are not represented by this analysis. Expression levels are color coded from very low (white $\leq$ 10% of max. observed) to high (blue $\geq$ 90% of max. observed) expression. TFs which were not included in this analysis are labeled grey, while IRF8 is highlighted because it is not represented by the ENCODE TF set. TFs which were used for the E2 or E3 peak cluster analysis are marked in orange or dark blue, respectively.

Table S4. Highly expressed TFs in GM12878 as identified by CAGE

Expression value (log10)	TPM	TF	Expression value (log10)	TPM	TF
2.6	395.59	IRF4	1.24	38.59	TCF4
2.46	288.23	PLEK	1.24	16.28	CREB5
2.37	235.43	POU2AF1	1.21	68.7	TCF7
2.07	117.22	IRF8	1.21	21.24	NFKBIZ
2.05	111.79	SPIB	1.2	26.21	IKZF2
2.04	107.59	RUNX3	1.2	17.04	MSC
1.89	77.41	ASCL1	1.18	88.56	ELF1
1.84	68.31	IKZF3	1.18	22.24	MSC
1.83	67.01	SP140	1.16	26.52	IRF7
1.77	57.46	EOMES	1.16	13.45	BATF
1.72	94.52	RUNX3	1.15	15.97	HOXC4
1.71	50.2	SP140	1.14	12.91	DMRT2
1.69	47.53	PLEK	1.13	12.61	PLEK
1.69	47.53	CREB5	1.13	12.38	IKZF2
1.59	38.28	ZBTB32	1.11	159.7	HMGA1
1.56	71.68	EGR2	1.11	12	NR3C1
1.53	33.09	IFI16	1.11	11.92	LHX2
1.52	32.4	HNF4G	1.1	15.44	SOX18
1.5	30.72	HNF1B	1.09	87.26	MEF2C
1.46	27.81	ZBED1	1.09	51.66	ZNF296
1.45	35.76	BATF	1.09	18.03	IRF5
1.43	25.67	EOMES	1.08	11.16	ZFAT
1.42	25.52	MEF2C	1.08	11.16	IRF8
1.41	50.97	C11orf9	1.08	11.08	PRRX1
1.4	65.94	POU2F2	1.08	11	IKZF1
1.4	34.77	NFATC2	1.07	48.22	TOX2
1.4	24.38	SPI1	1.05	28.43	AKNA
1.4	24.15	TP73	1.04	9.86	E2F8
1.38	23.08	ASCL1	1.03	9.63	IKZF2
1.34	20.94	MEF2B	1.02	19.33	MSC
1.32	24.61	IRF2	1.02	9.55	IKZF1
1.32	19.71	IKZF1	1.02	9.55	LHX2
1.31	19.64	NR3C1	1.02	9.48	SP110
1.31	19.64	TP63	1.01	60.75	NFKB2
1.3	19.1	PAX5	1	11	ZBTB38
1.27	17.8	TBX21	1	9.02	TFDP2
1.26	42.26	EBF1	1	9.02	RUNX3
1.26	17.12	HNF4G			

Ranked list of TF promoter expression in GM12878 relative to the median expression in the FANTOM5 collection is shown as determined by *Cap Analysis of Gene Expression* (CAGE, library ID: CNhs12331, experiment accession ID: DRX007776) (Fantom_Consortium et al., 2014). TPM = tags per million. TFs assigned to the E2 cluster or the E3 cluster of TFs are highlighted in orange or blue, respectively.

7.3 Affirmation

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorliegende Dissertation mit dem Titel

"Target Gene Regulation by EBV Latent Transcription Factors –
Exploiting Cellular Enhancer Elements"

von mir selbstständig und ohne unerlaubte Hilfe angefertigt ist.

München, August 2016 Laura Glaser

Erklärung

Hiermit erkläre ich,

dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.

dass ich mich anderweitig einer Doktorprüfung ohne Erfolg nicht unterzogen habe.

München, August 2016 Laura Glaser

7.4 Curriculum Vitae

Personal Data

Name	Laura Viola Glaser
Nationality	German

Education

09/11 - 11/15	Ph.D. student	Helmholtz Zentrum München
	Ph.D. thesis "Target Gene Regulation by EBV Latent Transcription Factors – Exploiting Cellular Enhancer Modules"	
	Supervisor: Prof. Dr. Bettina Kempkes *Helmholtz Zentrum München,* Research Unit Gene Vectors	
	Participation and successful completion of the *Helmholtz Graduate School Environmental Health (HELENA)*	
10/06 - 07/11	Diploma in Biology (Dipl.-Biol. Univ.)	Ludwig-Maximilians-Universität München
	Major subject Human Genetics, Final grade 1.0 (very good, equals A+)	
	Diploma thesis "Identification and Characterization of Alternative Transcripts of the KSHV Gene K10/vIRF4"	
	Supervisor: Prof. Dr. Bettina Kempkes, *Helmholtz Zentrum München*	
09/97 - 07/06	A-levels	Gymnasium Starnberg
	General qualification for university entrance, Final grade 1.8 (good, equals A-)	

Work Experience

09/12 - 11/14	Helmholtz Zentrum München, Working Group of Prof. Dr. Bettina Kempkes *Supervisor for several undergraduate students*
09/13 - 12/13	Temple University, Philadelphia, PA, Fels Institute for Cancer Research & Molecular Biology, Dr. Italo Tempera *Visiting Researcher*, Training: Chromosome Conformation Capture method Recipient of HELENA (Helmholtz Graduate School Environmental Health) travel grant
09/09 - 12/09 06/09 - 07/09	Helmholtz Zentrum München, Working Group of Prof. Dr. Bettina Kempkes *Student Research Assistant*
11/09 - 12/09	Ludwig-Maximilians-Universität München, Institute for Anthropology and Human Genetics Junior Research Group Chromobodies, Dr. Ulrich Rothbauer *Internship* "Evaluation of Potential LC3 Chromobodies in U2OS Cells with Induced Autophagy"
08/09	Tulane University, New Orleans, LA, Xavier Center for Bioenvironmental Research, Environmental Endocrinology Laboratory, Prof. Dr. John McLachlan/Dr. Chasity Coleman *Internship* "Estrogenic Effect of Bisphenol A in Human Smooth Muscle and Leiomyoma Cells"
02/09 - 03/09	Ludwig-Maximilians-Universität München, Institute of Genetics *Tutor for undergraduate students* Supervision of experiments and correction of protocols