

Erklärung

Hiermit erkläre ich, dass die vorliegende Arbeit mit dem Titel

„THE IMPACT OF EPSTEIN-BARR VIRUS PROTEIN BZLF1 ON THE CELLULAR TRANSCRIPTOME AND GENOME STRUCTURE“

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

München, 20. Dezember 2016

Alexander Buschle

Dissertation eingereicht am 20. Dezember 2016

Erstgutachter: Prof. Dr. Dirk Eick
Zweitgutachter: Prof. Dr. Heinrich Leonhardt

Tag der mündlichen Prüfung: 29. Juni 2017

Table of Contents

| | | |
|----------|---|-----------|
| 1 | SUMMARY | 9 |
| 2 | INTRODUCTION | 10 |
| 2.1 | DNA METHYLATION | 11 |
| 2.2 | HISTONES AND THEIR MODIFICATIONS | 12 |
| 2.3 | CHROMATIN STRUCTURES AND ORGANIZATION | 13 |
| 2.4 | MRNA TRANSCRIPTS AND TECHNIQUES TO QUANTIFY THEM | 15 |
| 2.5 | VIRUSES HIJACK EPIGENETIC MECHANISMS | 16 |
| 2.6 | EBV | 17 |
| 2.7 | EPIGENETICS IN EBV'S LIFE CYCLE | 18 |
| 2.8 | BZIP PROTEINS | 20 |
| 2.9 | THE STRUCTURE OF BZLF1 | 21 |
| 2.10 | SCOPE AND AIM OF THIS THESIS | 23 |
| 3 | RESULTS | 25 |
| 3.1 | THE CELLULAR AND VIRAL MODEL SYSTEMS – AN EXPERIMENTAL OVERVIEW | 25 |
| 3.2 | THE EXPRESSION OF BZLF1 IN NON-INDUCED CELLS ORIGINATES FROM THE LEAKY PRTR 4816 PLASMID | 28 |
| 3.3 | BZLF1 IS EXPRESSED PRIOR TO DOXYCYCLINE INDUCTION BUT ITS LEVEL DOES NOT INDUCE THE VIRAL LYTIC CYCLE | 30 |
| 3.4 | BZLF1 BINDS CELLULAR CHROMATIN WITH HIGH FREQUENCY AND WITH A CLEAR PREFERENCE FOR CpG MOTIFS | 32 |
| 3.4.1 | THE EXPERIMENTAL SETUP OF CHIP-SEQ EXPERIMENTS | 32 |
| 3.4.2 | BZLF1 BINDING SITES IN TWO B-CELL LINES | 33 |
| 3.5 | THE EXPRESSION OF BZLF1 CHANGES THE EXPRESSION OF CELLULAR GENES | 41 |
| 3.5.1 | THE EXPERIMENTAL SETUP OF RNA-SEQ EXPERIMENTS | 41 |
| 3.5.2 | INDUCED EXPRESSION OF BZLF1 IN DG75 4816 CELLS REGULATES ONLY FEW CELLULAR GENES | 42 |

| | | |
|------------|---|------------|
| 3.5.3 | THE EXPRESSION OF BZLF1 IN RAJI 4816 CELLS RESULTS IN MASSIVE DOWN-REGULATION OF CELLULAR TRANSCRIPTS. | 48 |
| 3.5.4 | KEGG PATHWAY ANALYSIS | 58 |
| 3.6 | BZLF1 BINDING SITES IN PROMOTER REGIONS OF REGULATED GENES ONLY EXPLAIN A MINORITY OF CELLULAR GENE REGULATION | 59 |
| 3.6.1 | A MINORITY OF BZLF1-REGULATED GENES HAS PROMOTER REGIONS WITH BZLF1 BINDING SITES | 60 |
| 3.7 | CHROMATIN INTERACTIONS OF SELECTED GENES PRIOR TO AND AFTER BZLF1 EXPRESSION | 63 |
| 3.7.1 | THE EXPERIMENTAL SETUP OF THE CAPTURE-C APPROACH | 63 |
| 3.7.2 | DOWN-REGULATED GENES | 66 |
| 3.7.3 | UP-REGULATED GENES | 71 |
| 3.7.4 | NON-REGULATED GENES | 74 |
| 3.7.5 | A REGION WITH INCREASED NUMBERS OF INTERACTIONS 15 H POST BZLF1 INDUCTION | 78 |
| 3.7.6 | COMPARTMENTALIZATION OF INTERACTIONS | 80 |
| 3.7.7 | THE PATTERN OF CHROMATIN INTERACTIONS IN ADJACENT GENE LOCI | 83 |
| 3.7.8 | GENES WITH LONG-RANGE INTERACTIONS | 89 |
| 4 | <u>DISCUSSION</u> | 93 |
| 4.1 | BZLF1 EXPRESSION PRIOR TO DOXYCYCLINE INDUCTION | 94 |
| 4.2 | BZLF1 BINDING SITES IN THE CELLULAR GENOME | 95 |
| 4.3 | DIFFERENTIAL GENE EXPRESSION UPON INDUCTION OF BZLF1 | 98 |
| 4.4 | REGULATED CELLULAR GENES WITH BZLF1 BINDING SITES | 101 |
| 4.5 | CHROMATIN INTERACTIONS OF STRONGLY REGULATED CELLULAR GENES | 103 |
| 4.6 | ACHIEVEMENTS AND OPEN QUESTIONS | 110 |
| 5 | <u>MATERIAL</u> | 112 |
| 5.1 | OLIGONUCLEOTIDES | 112 |
| 5.2 | PLASMIDS | 112 |
| 5.3 | ANTIBODIES | 113 |
| 5.4 | EUKARYOTIC CELL LINES | 113 |
| 5.5 | CELL CULTURE MEDIA AND ADDITIVES FOR EUKARYOTIC CELLS | 114 |
| 5.6 | CHEMICALS AND ENZYMES | 114 |
| 5.7 | BUFFERS AND SOLUTIONS | 115 |

| | | |
|-------------|---|------------|
| 5.8 | COMMERCIAL KITS | 116 |
| 5.9 | SOFTWARE | 117 |
| 5.10 | SPECIAL DEVICES AND CONSUMABLES | 120 |
| 6 | METHODS | 122 |
| 6.1 | EUKARYOTIC CELL CULTURE | 122 |
| 6.1.1 | CELL CULTURE CONDITIONS | 122 |
| 6.1.2 | THAWING EUKARYOTIC CELLS | 122 |
| 6.1.3 | FREEZING EUKARYOTIC CELLS | 122 |
| 6.1.4 | TRANSFECTION OF EUKARYOTIC CELLS | 123 |
| 6.1.5 | ESTABLISHMENT OF STABLE CELL LINES | 123 |
| 6.2 | FLOW CYTOMETRY | 123 |
| 6.3 | CHIP-SEQUENCING (NGS) | 124 |
| 6.3.1 | CELL PREPARATION | 124 |
| 6.3.2 | CHROMATIN IMMUNOPRECIPITATION | 124 |
| 6.3.3 | CHIP PULL-DOWN CONTROL | 125 |
| 6.3.4 | LIBRARY PREPARATION AND ILLUMINA SEQUENCING | 125 |
| 6.3.5 | BIOINFORMATIC ANALYSIS | 127 |
| 6.4 | RNA-SEQUENCING (NGS) | 129 |
| 6.4.1 | CELL PREPARATION | 129 |
| 6.4.2 | MAGNETIC BEADS CELL SORTING | 129 |
| 6.4.3 | LIBRARY PREPARATION AND SEQUENCING | 131 |
| 6.4.4 | BIOINFORMATIC ANALYSIS | 133 |
| 6.5 | LINKAGE OF CHIP-SEQ AND RNA-SEQ RESULTS | 135 |
| 6.6 | CAPTURE-C (NGS) | 135 |
| 6.6.1 | CELL PREPARATION | 135 |
| 6.6.2 | CHROMATIN CONFORMATION CAPTURE (3C) | 135 |
| 6.6.3 | LIBRARY PREPARATION, SAMPLE CAPTURE, AND SEQUENCING | 136 |
| 6.6.4 | BIOINFORMATIC ANALYSIS | 138 |
| 6.7 | QUANTITATIVE REAL TIME PCR | 139 |
| 6.8 | PROTEIN ANALYSIS | 140 |
| 6.8.1 | WESTERN BLOT | 140 |

| | | |
|-----------|---|------------|
| 6.8.2 | WESTERN BLOT PROTEIN QUANTIFICATION | 143 |
| 7 | <u>LITERATURE</u> | 144 |
| 8 | <u>ABBREVIATIONS</u> | 176 |
| 9 | <u>APPENDIX</u> | 180 |
| 9.1 | RAJI EBV'S GENE EXPRESSION | 180 |
| 9.2 | ERCC RNAs | 181 |
| 9.3 | BOXPLOTS | 182 |
| 9.4 | TOP 25 UP- AND DOWN-REGULATED GENES OF DG75 4816 AND RAJI 4816 CELLS 6 H AFTER INDUCTION OF BZLF1 | 183 |
| 9.5 | TOP 25 UP- AND DOWN-REGULATED GENES OF DG75 4816 AND RAJI 4816 CELLS 6 H AFTER INDUCTION OF BZLF1 WITH AT LEAST ONE PEAK WITHIN THEIR PROMOTERS | 185 |
| 9.6 | LIST OF GENES AND POSITIONS CAPTURED DURING CAPTURE-C ANALYSIS | 187 |
| 10 | <u>CURRICULUM VITAE</u> | 189 |
| 11 | <u>DANKSAGUNGEN</u> | 191 |

1 Summary

The BZLF1 protein of the Epstein-Barr virus is a viral transcription factor, which binds the viral genome sequence specifically to activate viral gene transcription of lytic genes initiating virus *de novo* synthesis. In the viral genome BZLF1 binds two classes of DNA motifs: a canonical AP-1 like motif and an alternative DNA motif with a CpG di-nucleotide, which must be methylated for BZLF1 to bind. In mammalian cells CpG methylation in the promoter region is generally associated with transcriptionally repressed and silenced genes, because 5'-methylcytosines attract family members of nuclear proteins with methyl-CpG-binding domains, which interfere with transcription factor binding. BZLF1 is a homolog of members of the cellular AP-1 protein family, suggesting that BZLF1 not only regulates viral but also cellular genes. The aim of this thesis was to decipher BZLF1's role in cellular transcription and chromatin architecture.

In human B-cell lines BZLF1 was expressed from an inducible plasmid to compare the transcriptomes of EBV-positive and EBV-negative cells upon BZLF1 expression. BZLF1 binding sites were identified in the cellular genome as well as genes regulated at different time points after induced BZLF1 expression. A Capture-C approach was used to study BZLF1's role in modulating the architecture of chromatin concentrating on promoter-enhancer interactions.

Chromatin immunoprecipitation followed by next generation sequencing and bioinformatic analysis identified two major binding motifs in up to 230,000 sites bound by BZLF1. Interestingly, BZLF1 exclusively binds motifs containing a CpG di-nucleotide at lower concentrations. RNA expression profiling (RNA-seq) experiments that used artificial spike-in RNAs as normalization control revealed that cellular genes were mostly down-regulated after BZLF1 induction. Because 50 % of the genes regulated after BZLF1 expression did not show BZLF1 binding sites within their promoters, I investigated changes of the three-dimensional chromatin structures of selected genes upon expression of BZLF1 in a Capture-C approach. The data show that the expression of BZLF1 resulted in a massive loss of chromatin interactions, which may contribute to the down-regulation of cellular genes upon BZLF1 expression.

2 Introduction

In 1928 Fred Griffith found that genetic information is transferable by the transformation of different *Streptococcus pneumoniae* strains (Griffith, 1928). The DNA was identified as the molecule transferring the genetic information by Oswald Avery and colleagues in 1944 (Avery et al., 1944). After Erwin Chargaff and Gerald Wyatt discovered that the DNA bases adenosine and thymine as well as guanine and cytosine occur in about the same quantities (Zamenhof et al., 1952; Wyatt, 1952) and X-ray crystallography data of the DNA were provided by Rosalind Franklin (Franklin and Gosling, 1953) and Maurice Wilkins (Wilkins et al., 1953), James Watson and Francis Crick could propose the model of the double strand helix in 1953 (Watson and Crick, 1953). Between 145-147 base pairs of DNA are wrapped around an octamer of each two histone proteins H2A, H2B, H3, and H4 (Luger et al., 1997; Kornberg, 1974) to build a nucleosome (Kornberg, 1977). With the help of nucleosomes the DNA strand of approximately 2 meter length and 2 nm diameter in a human cell can be compacted to 46 chromosomes, which fit to the cell's nucleus with a diameter of about 6 μm (Alberts et al., 2002).

With few exceptions all cells of multicellular organisms contain the same genetic information stored in the DNA sequence. In humans the differential usage of the approximately 20,000 genes (Clamp et al., 2007; Ezkurdia et al., 2014) allows cells to differentiate to several hundred cell types. The change in expression profiles stems from epigenetic modifications, which govern gene expression like software governs the hardware of a computer. Epigenetic information can be encoded by several means: e.g. methylation of cytosine in a cytosine-phosphatidyl-guanine (CpG) context (Jones, 2012), incorporation of histone variants or the modification of histone tails (Bannister and Kouzarides, 2011; Venkatesh and Workman, 2015). These epigenetic marks can support or suppress the interaction of chromatin fibers with promoters regulating gene expression (Mora et al., 2015). Another way to influence gene regulation epigenetically is the expression non-coding RNAs, which are not translated to a protein. They involve micro RNAs, long non-coding RNAs, small interfering RNAs, enhancer RNAs, and piwi-interacting RNAs (Morris and Mattick, 2014), which generally interfere with gene expression using different modes of action.

In general, epigenetic modifications, which are set by writers and interpreted by readers are reversible and can be erased by either active erasers or passive dilution (Allis and Jenuwein, 2016).

2.1 DNA methylation

In the human genome the fraction of methylated versus non-methylated CpGs differs between cell lines and depending on the analyzed loci (Ehrlich et al., 1982). By the deamination of cytosines in a CpG context these di-nucleotides occur less frequently than expected by statistical distribution (Bird, 1980; Chahwan et al., 2010). The methylation of cytosines is mediated by the three DNA methyltransferases (DNMTs) DNMT1, DNMT3a, and DNMT3b in mammalian cells. All DNMTs flip the cytosine base out of the DNA helix to transfer the methyl group from the donor molecule S-adenosyl methionine (SAM) to the fifth carbon of cytosine (Cheng and Blumenthal, 2008, Figure 2-1).

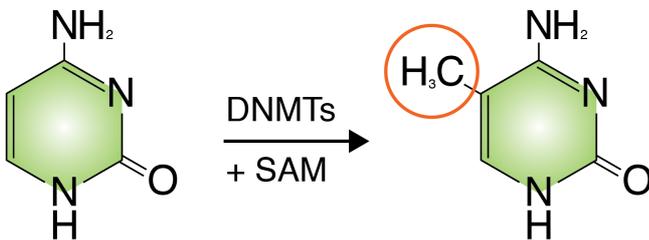


Figure 2-1: Cytosine methylation by DNMTs

The DNMT proteins transfer the methyl group of the donor molecule SAM to the fifth carbon of the cytosine molecule.

DNMT1 is associated with replication foci (Leonhardt et al., 1992) to maintain the methylation pattern in dividing cells and preferentially binds to hemi-methylated DNA (Gruenbaum et al., 1982). Both, DNMT3a and DNMT3b are mainly associated with de novo DNA methylation, but do not show a preference for non-methylated or hemi-methylated DNA (Okano et al., 1998). Nevertheless, under certain circumstances DNMT1 can also perform de novo methylation and DNMT3a/b can maintain a pre-existing methylation pattern (Lei et al., 1996; Chen et al., 2003).

Methylated CpG di-nucleotides are associated with different epigenetic mechanisms like imprinting (Shemer et al., 2000; Tremblay et al., 1995), gene and retrotransposon silencing (Boyes and Bird, 1992; Hsieh, 1994; Kaneko-Ishino and Ishino, 2010), and polycomb repression (Viré et al., 2006). High densities of methylated CpG di-nucleotides can be found in promoters and in CpG-islands (Deaton and Bird, 2011; Saxonov et al., 2006).

Cytosine methylation can be erased passively upon cell division by preventing methylation maintenance (Rougier et al., 1998; Valinluck and Sowers, 2007; Hashimoto et al., 2012). For an active de-methylation the ten-eleven-translocation (TET) proteins are required to oxidize 5-methylcytosine to 5-hydroxymethylcytosine (Tahiliani et al., 2009) and further over 5-formylcytosine (5fC) to 5-carboxylcytosine (5caC) (Ito et al., 2011; He et al., 2011a). Subsequently, both the 5fC or 5caC modification can be recognized by the thymine DNA glycosylase and removed by base excision repair (He et al., 2011b; Maiti and Drohat, 2011).

Methods of different complexity have been developed to detect cytosine methylation. In early days a chemical approach was used to detect cytosine methylation (Ohmori et al., 1978). Later assays based on methylation-sensitive endonucleases (Huang et al., 1997; Ushijima et al., 1997) and affinity based precipitation methods (Cross et al., 1994; Weber et al., 2005; Keshet et al., 2006) were developed to identify sites with methylated cytosines. Bisulfite conversion (Hayatsu et al., 1970) in combination with Sanger-sequencing (Sanger et al., 1977) resulted in the resolution of cytosine methylation at the level of single bases (Frommer et al., 1992), which was extended to whole genome analysis (Cokus et al., 2008; Lister et al., 2008) after immunoprecipitation (Brinkman et al., 2012; Statham et al., 2012).

2.2 Histones and their modifications

DNA is wrapped around nucleosomes for a better compaction and histone modifications classify DNA sequences for chromatin interacting proteins. Besides the linker histone H1 and the four canonical core histone proteins H2A, H2B, H3, and H4 several isoforms exist but for H4 (Bernstein and Hake, 2006; Maze et al., 2014). Additionally, post-translational modifications are preferentially introduced in the histone tails N-terminally. Two studies identified 130 posttranscriptional modifications (PTM) on histones and 708 unambiguous histone isoforms (Tan et al., 2011; Tian et al., 2012).

Immunoprecipitation is commonly used to pull down the DNA sequences associated with histone variants or modifications and microarrays or next generation sequencing are employed to identify the underlying DNA sequences (Barski et al., 2007; Mikkelsen et al., 2007; Robertson et al., 2007). The limited resolution caused by the imperfect chromatin

fragmentation can be increased with the ChIP-exo technique, in which the DNA overhangs are chopped off from the 5' end with an exonuclease (Rhee and Pugh, 2011).

Histone modifications serve to attract or repel other proteins that bind to and interpret the histone tails or their modifications (Kouzarides, 2007). Like DNA methylation the histone marks H3K27me3 (Boyer et al., 2006; Roh et al., 2006; Mikkelsen et al., 2007; Young et al., 2011), H3K9me3 (Bannister et al., 2001; Volpe, 2002), and H3K79me3 are commonly associated with gene repression (Barski et al., 2007). Active chromatin is defined by mono-methylated H3K9, H3K27, H3K79 (Barski et al., 2007), and other marks such as H3K27ac.

The histone marks are associated with discrete tasks: H3K4me3 is precipitated with promoters (Roh et al., 2006; Barski et al., 2007; Birney et al., 2007) and the H3K36me3 mark is associated with the gene body of actively transcribed genes (Bannister et al., 2005; Barski et al., 2007).

Also enhancers can be identified with the help histone marks. Both, the H3K4me1 and H3K27ac marks are associated with active enhancers (Creighton et al., 2010), while the combination of H3K4me1 and H3K27me3 is found in poised enhancers (Rada-Iglesias et al., 2011). To promote gene transcription enhancers have to get in contact with more or less distant chromatin region, which the 3D chromatin architecture provides.

2.3 Chromatin structures and organization

For a long time DNA sequences, which were not associated with coding genes, were thought to be useless (Pennisi, 2012). Gradually more and more features like non-coding RNAs, structural components and enhancer functions could be assigned to the “junk” DNA.

DNA sequences that act as enhancers hold the potency to increase the basal transcription of a promoter (Banerji et al., 1981). It is generally accepted that enhancers that are often far distant elements, must interact with promoters to enhance transcription. The interactions between chromatin regions such as promoters and enhancers mainly occur in close spatial proximity but can also span an entire chromosome. Even inter-chromosomal interactions occur especially between small, gene rich chromosomes (Lieberman-Aiden et al., 2009).

The interactions between different chromatin regions occur cell type specifically (Shen et al., 2012; Sanyal et al., 2012; Phillips-Cremins et al., 2013). In enhancer sequences the

histone double variant H3.3/ H2A.Z is enriched (Jin et al., 2009), specific histone marks are common, and the transcriptional mediator p300 is frequently bound (chapter 2.3; Heintzman et al., 2007, 2009; Visel et al., 2009). Additionally, many transcription factors bind cooperatively to enhancer sites (Chen et al., 2008) to activate enhancers or to enhance gene expression (Ghisletti et al., 2010). The identification of enhancers can be supported by DNaseI digestion since the chromatin associated with enhancers is open and therefore easy accessible for endonucleases (Song et al., 2011).

Chromatin organizers like CTCF and cohesin keep chromatin loops in place to stabilize the interactions between enhancers and promoters (Mishiro et al., 2009). Nevertheless, only few enhancers target their nearest promoter and can span several chromatin organizers (Sanyal et al., 2012). Enhancers are transcribed to further stabilize cis-regulatory interactions and the resulting RNAs support the transcription of the targeted genes (Kim et al., 2010; Ørom et al., 2010; Rinn et al., 2007).

A special type of enhancer is made up of clusters of individual enhancers spanning up to 50 kbps to form super-enhancers. These are commonly associated with genes defining the identity of a given cell (Whyte et al., 2013).

Over the last decades more and more sophisticated approaches have been developed to identify the interactions between chromatin regions. In early approaches radioactive probe-hybridization techniques were used to identify interacting chromatin visually in fixed cells (Gall and Pardue, 1969). The introduction of fluorescent in situ hybridization (FISH) allowed to visualize the results under a microscope (Pinkel et al., 1986). A modern version of FISH marks DNA sequences in living cells using a modified version of the CRISPR/Cas9 technology (Anton et al., 2014). The chromosome conformation capture (3C) technique allows to analyze chromatin interactions of given loci (Dekker et al., 2002) and was the starting point for a number of techniques elucidating the chromatin architecture in high throughput. To perform 3C, the chromatin gets crosslinked, digested with a restriction enzyme and religated under conditions of very low DNA concentration to only link DNA that is hold in close proximity by DNA-protein complexation. After reversing the crosslink a common PCR with locus specific primers is used to test if two fragments interact with each other (Dekker et al., 2002). The combination of the 3C approach with ChIP (ChIP-loop) reduces the background and allows to identify protein mediated interactions (Horike et al., 2005; Cai et al., 2006; Ren et al., 2012). The chromosome conformation capture carbon

copy (5C) technique allows to test multiple interactions in parallel by a clever primer design but still requires, like 3C, the previous selection of primers at regions to be tested for interaction (Dostie et al., 2006). The chromosome conformation capture-on-chip (4C) was developed to find all sequences interacting with one defined locus. Therefore the 3C approach was extended for an additional restriction enzyme digestion and ligation step to form small DNA circles, which can be amplified by an inverse PCR from the locus of interest (Simonis et al., 2006). With the Hi-C technique all chromatin interactions in the nucleus can be monitored. The technique starts like 3C, but prior to ligation the crosslinked and restriction enzyme digested DNA fragments are filled up with a biotin-linked nucleoside triphosphate. After sheering the biotin-associated DNA fragments are precipitated with streptavidin beads and sequenced with high throughput (Lieberman-Aiden et al., 2009). The chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) combines Hi-C with ChIP and allows to identify DNA interactions involving proteins of interest like the chromatin organizer CTCF (Fullwood et al., 2009; Li et al., 2012; DeMare et al., 2013). The Capture-C technique combines the 3C approach with a DNA-RNA hybridization based and sequence-specific pull-down of regions of interest (ROI). All regions interacting with the ROI are pulled down with the biotinylated RNA probes and identified with an increased resolution compared with other genome-wide techniques (Hughes et al., 2014).

2.4 mRNA transcripts and techniques to quantify them

The interplay of epigenetic chromatin marks defines the resulting gene expression conducted by RNA polymerase II (Pol II). First Pol II forms the pre-initiation complex together with other proteins including the TATA-box-binding protein (TBP) and several general transcription factors. In a second step the productive transcription starts with Pol II's promoter escape and the capping of the early transcript. The process of gene transcription is called elongation and is associated with the splicing machinery. The final step is the termination of the transcription after which the poly A chain is attached to complete the mRNA (Shandilya and Roberts, 2012).

The resulting mRNAs can be analyzed by RNA-sequencing. An early but still frequently use method is the quantitative real-time fluorescence-based reverse transcription polymerase chain reaction. Messenger RNA gets transcribed to cDNA and is then multiplied.

After each duplication the fluorescence of a DNA intercalating dye gets measured to determine the amount of double stranded DNA (Morrison et al., 1998; Bustin et al., 2005; Bustin, 2000).

The first possibility to analyze differential gene expression between two different samples on a high throughput scale was the microarray. To do so the mRNA gets transcribed to cDNA and each sample is fluorescently labeled in a different color. Upon hybridization to a sequence-specific DNA microarray the color of the more frequent sample predominates for each analyzed sequence (Brown and Botstein, 1999). Nowadays, sequence specific-sequencing methods like Illumina sequencing are available (Bentley et al., 2008). High throughput techniques have to deal with the problem of huge amounts of ribosomal RNAs (rRNA), which represent up to 98 % of all RNAs in RNA preparations (Benes et al., 2011). Several different techniques are used to overcome this predominance. Ribosomal RNAs are either depleted by not-so-random hexamer primers, polyA-specific primers or by a PolyA-specific pull-down and depletion (Armour et al., 2009; Zhao et al., 2014). Another problem of RNA-sequencing was that the direction of the sequenced strand was unknown. The introduction of strand-specific RNA-seq allowed to identify the strand the detected RNA is originating from (Parkhomchuk et al., 2009; Levin et al., 2010; Zhao et al., 2015). A common method is to amplify the first DNA strand on the RNA template and degrade the RNA with RNase H. For the amplification of the second strand deoxyuridine triphosphate is used. After library preparation the uracil bases can be excised by the uracil-DNA glycosylase and heat dissociates the remaining bases. The sequencing reads map to the strand, which coded for the mRNA (Parkhomchuk et al., 2009).

2.5 Viruses hijack epigenetic mechanisms

Initially the chromatinization of viral DNA by the infected host cell was used as a defense mechanism to silence the viral DNA (Knipe, 2015; Mäntylä et al., 2016). Over time viruses learnt how to deal with this mechanism and to use epigenetic reprogramming to their own favor. Like computer viruses can take control over hard- or software, real viruses take control of a cell by mimicking or abusing cellular epigenetic principles controlling the cellular DNA. To do so viruses introduce new players, which manipulate the cell's epigenetic program.

For example, the influenza A H3N2 virus mimics parts of the histone protein H3 tail to interfere with the host's gene expression, reducing the cellular immune response (Marazzi et al., 2012). The human cytomegalovirus is another example for the manipulation of an epigenetic mechanism. Viral DNA is occupied by cellular histones upon infection (Nitzsche et al., 2008), but with the help of the viral IE1 protein the virus redistributes the histones that have been assembled on its genome to promote its own gene expression (Zalckvar et al., 2013). Viruses also use non-coding RNAs to influence the host cell's activities. The human immunodeficiency virus (HIV) harbors sequences in its envelope encoding genes, which perfectly match the sequences of human micro RNAs presumably to inactivate them and to evade the host's immune response (Holland et al., 2013). The Epstein-Barr virus (EBV) uses micro RNAs to evade the surveillance by CD4⁺ and CD8⁺ T cells by multiple mechanisms like suppression of antigen presentation by MHC class I and II or suppression of pro-inflammatory cytokines (Tagawa et al., 2016; Albanese et al., 2016).

Viruses can either attack a cell directly upon infection to turn it into a virus factory or establish latency to escape the immune system to be able to replicate and synthesize new virions during an opportune moment. Several viruses use CpG methylation to govern their own latency. The transcriptional start site of HIV is flanked by methylated CpG islands to presume the latent state of infection (Kauder et al., 2009), while the methylation of several of EBV's lytic promoters is required to escape from latency (Kalla et al., 2010; Bergbauer et al., 2010). The intricate processes that EBV has evolved will be explained in greater detail in the next chapters.

2.6 EBV

In 1958 the Irish physician Denis Burkitt described malignant tumors in the jaws of African children (Burkitt, 1958). Some years later two groups of scientists, involving the team of Michael Epstein and Yvonne Barr, could cultivate cells of this malignancy independently (Epstein and Barr, 1964; Pulvertaft, 1964). Additionally, virus particles were visualized in these cells by electron microscopy (Epstein et al., 1964). In 1967 the virus was named after its discoverers as Epstein-Barr virus and could be associated with infectious mononucleosis (Henle et al., 1967). Two years later a study in Africa revealed that EBV infected about 85-

95 % of all examined children older than two years (Henle et al., 1969). This indicated that EBV is one of the most successful viruses infecting humans.

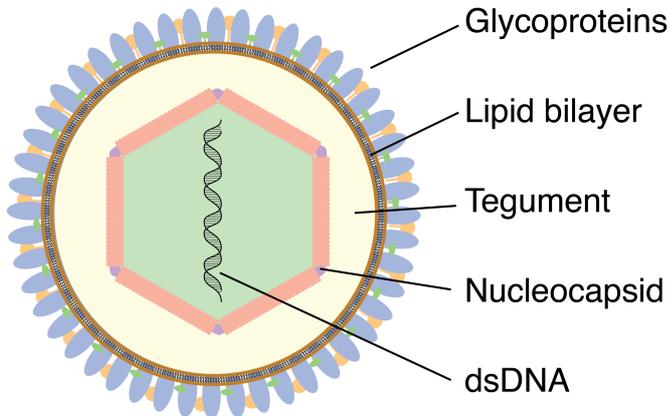


Figure 2-2: Schematic structure of an EBV virion.

A lipid bilayer, which comprises glycoproteins, encompasses the tegument of the virion. The nucleocapsid harbors the epigenetically naïve, double stranded DNA.

EBV belongs to the family of herpesviridae, which harbor a unique virion structure (Figure 2-2). The outer layer is a lipid bilayer envelope

comprising glycoproteins and encompassing the tegument. The double stranded DNA is stored in the most inner nucleocapsid. There are three subfamilies of herpesviridae: alpha-, beta-, and gamma-herpesviridae (Baron and Whitley, 1996). EBV belongs to the gamma-herpesviridae (Davison, 2007) and was recognized by the WHO as a class I carcinogen because it can cause Burkitt lymphoma in its main target, the human B-cells, is involved in the formation of nasopharyngeal carcinoma (zur Hausen et al., 1970), and associated with other human tumors.

2.7 Epigenetics in EBV's life cycle

The key of EBV's success lies in its life long persistence in B-cells coupled with occasional viral reactivations. To achieve this EBV's life cycle is organized in three phases (Figure 2-3), which start when the epigenetically naïve and linear viral DNA enters the host cell (Kintner and Sugden, 1981; Johannsen et al., 2004; Kalla et al., 2010):

I. The pre-latent phase is defined by the expression of latent genes and certain lytic genes such as BZLF1, which induce the proliferation of quiescent naïve or memory B-cells (Wen et al., 2007; Kalla et al., 2010) together with BCRF1, BNLF2a and viral micro RNAs, which inhibit the host's immune responses directed against EBV (Jochum et al., 2012; Albanese et al., 2016; Tagawa et al., 2016). In the pre-latent phase structural proteins are not expressed and an amplification of viral DNA does not take place (Kalla et al., 2010), despite the expression of BZLF1. During the pre-latent phase the viral DNA acquires

histones, histone marks, chromatin organizers, and methylation of CpGs (Kalla et al., 2010; Schmeinck, 2011; Shaw et al., 1979; Dyson and Farrell, 1985).

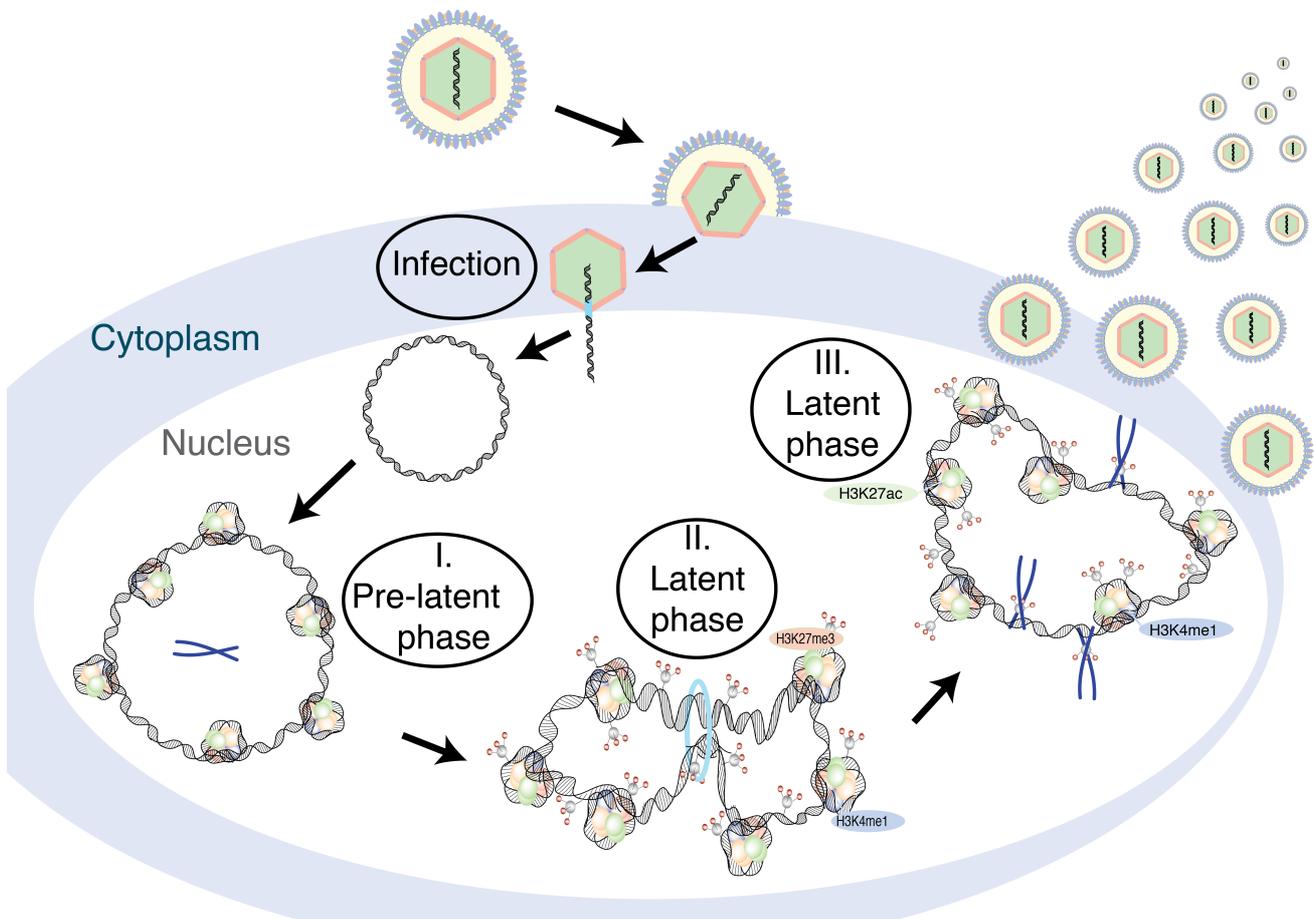


Figure 2-3: EBV's life cycle

EBV's life cycle starts after the epigenetically naïve and linear viral DNA enters the nucleus of the host cell. During the pre-latent phase the viral genome acquires cellular histones, histone marks, chromatin organizers and CpG methylation. In the latent phase the chromatin is densely compacted and histone marks like H3K27me3 denote the chromatin as silent. Upon the expression of BZLF1 repressive chromatin marks are exchanged to active marks, the viral chromatin opens up, and viral structural proteins are expressed to release progeny virions.

II. In the latent phase the EBV genome is fully chromatinized, latent genes are expressed exclusively to maintain the viral episome during proliferation of the host cell but all lytic viral genes are silenced. With each replication of the host cell multiple copies of the EBV episome are replicated as well (Adams, 1987; Kirchmaier and Sugden, 1995). For the latent replication of EBV's episome both the origin of replication and the nuclear antigen 1 (EBNA1) are required and sufficient (Yates et al., 1985).

During the latent phase viral lytic genes have to be suppressed to hide the virus from the immune system. Therefore the degree of CpG methylation reaches a very high level on the viral DNA and the nucleosomes of the repressed lytic genes are densely packed and

covered with Polycomb-group proteins to prevent immune recognition of the coded proteins (Woellmer et al., 2012).

III. The lytic phase is characterized by the expression of BZLF1, which functions as the molecular switch from EBV's latent to its lytic phase (Countryman and Miller, 1985; Takada et al., 1986). The promoters of repressed lytic genes remain methylated during the lytic phase (Schmeinck, 2011). Nevertheless BZLF1 binds its DNA motifs in methylated repressed chromatin, recruits chromatin remodelers such as INO80 and SNF2H and induces the loss of repressive Polycomb-group proteins to open the viral chromatin (Schmeinck, 2011; Schaeffner, 2015). Additionally, activation marks such as H3K4me3 get established to activate the expression of viral, lytic genes (Woellmer et al., 2012) driving the expression of viral, structural proteins and de novo synthesis of viral progeny DNA. CpG methylation of lytic viral promoters is the prerequisite to activate early lytic viral genes (Kalla et al., 2010; Bergbauer et al., 2010) contrary to the commonly accepted rule that links CpG methylation of promoters with repressed genes.

2.8 bZIP proteins

BZLF1 has to get in contact with viral DNA to activate the lytic cycle. Evolution yielded multiple classes of proteins to achieve sequence-specific DNA-protein contacts. The most famous representatives are basic leucine zipper (bZIP), helix-turn-helix (Ellenberger, 1994), and zinc finger DNA binding proteins (Laity et al., 2001). BZIP proteins contain a C-terminal dimerization domain with a hepta-repeat of hydrophobic leucines (Landschulz et al., 1988), which form coiled-coils and mediate dimerization. The N-terminal DNA-binding domain (Talanian et al., 1990) binds to the major groove of DNA (Shuman et al., 1990), but only upon DNA binding (Shuman et al., 1990) the basic N-terminal domain adopts the alpha-helical structure (O'Neil et al., 1991) to get in contact with the DNA by hydrogen bonds and van der Waal contacts (Nakabeppu and Nathans, 1989; O'Neil et al., 1990). The DNA binding of bZIP proteins is also described as 'scissors-grip' (Vinson et al., 1989).

Many bZIP proteins have already been identified in different species including *Arabidopsis thaliana*, *Drosophila melanogaster*, Human herpesvirus 4 and 8, human simian T-lymphotropic virus, Gallid herpesvirus 2, *Homo sapiens*, and *Saccharomyces cerevisiae* (Deppmann et al., 2006; Reinke et al., 2010). In yeast the bZIP protein Gnc4 (König and

Richmond, 1993) is an activator of metabolic pathways under nearly any kind of amino acid starvation (Natarajan et al., 2001; Hinnebusch and Natarajan, 2002). The protein sequence homology in the DNA binding domains of Gcn4 and several human bZIP proteins of the AP-1 family including c-Jun, Jun-B, c-Fos, and Fra-1 is impressive (Farrell et al., 1989). The AP-1 protein family consists of three subfamilies: (i) Jun (c-Jun, JunB, and JunD), Fos (c-Fos, FosB, Fra-1, and Fra-2), and the activating transcription factor (ATFa, ATF2, ATF3, ATF4, and B-ATF). Different dimerization products of these proteins either bind the DNA sequence of the 12-O-tetradecanoylphorbol-12-acetate (TPA) response elements (TRE) or cyclic adenosine monophosphate (cAMP) response elements (CRE) with different affinities (Eferl and Wagner, 2003; Gustems et al., 2014; Halazonetis et al., 1988). Upon binding members of the AP-1 protein family are involved in the regulation of cellular differentiation, apoptosis, and proliferation (Shaulian and Karin, 2001, 2002; Eferl and Wagner, 2003; Ameyar et al., 2003), which involves the transactivation domain that most AP-1 protein family members contain (Cohen and Curran, 1988; Cohen et al., 1989; Angel et al., 1988; Bohmann and Tjian, 1989; Sutherland et al., 1992).

The viral BZLF1 protein is another member of the bZIP proteins with a strong sequence homology at protein level with Gcn4 and other members of the AP-1 family (Farrell et al., 1989). The next chapter will provide a detailed description of BZLF1.

2.9 The structure of BZLF1

EBV's life cycle has three major phases: (i) the infection of the cell and the establishment of latency during the pre-latent phase; (ii) the latent phase, during which the virus only maintains itself and attempts to escape the immune system; (iii) the lytic phase, which supports the synthesis of new virion particles to infect further cells. The immediate-early viral protein BZLF1 mediates the switch from EBV's latent to lytic phase (Countryman and Miller, 1985; Takada et al., 1986; Chevallier-Greco et al., 1986). There are several ways to induce the lytic cycle artificially. BZLF1 can be ectopically overexpressed (Countryman and Miller, 1985; Chevallier-Greco et al., 1986; Kenney et al., 1989) or induced by certain reagents. These can either be the phorbol ester TPA (zur Hausen et al., 1978; Baumann et al., 1998) or an antibody directed against the immunoglobulin of an EBV infected B-cell (Takada, 1984; Takada and Ono, 1989).

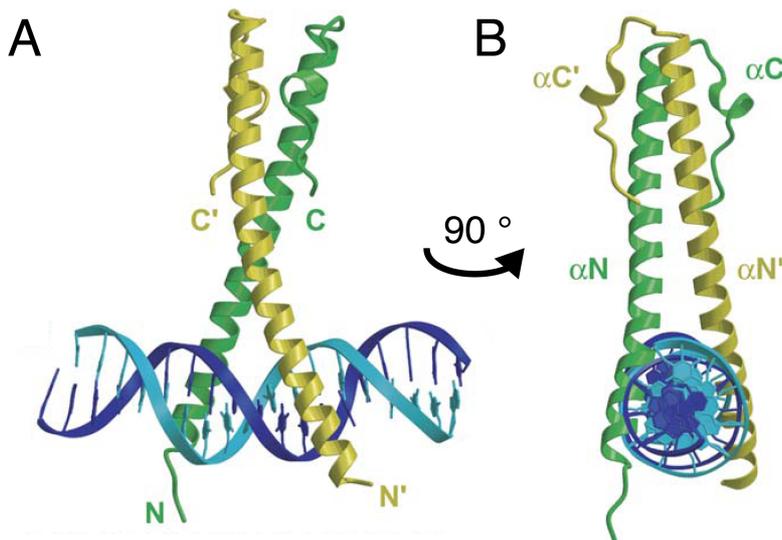


Figure 2-4: The crystal structure of a truncated BZLF1 dimer bound to DNA (modified from Petosa et al., 2006)

A: Two N-terminal truncated BZLF1 dimers (yellow and green) form a homodimer and bind the DNA double-helix (light and dark blue). **B:** The same BZLF1-DNA complex like depicted in panel A but turned for 90° around the y-axis.

The BZLF1 gene harbors three exons and results in a 245 amino acids spanning protein

(Biggin et al., 1987) of about 35 kDa, which contains five domains. The N-terminus harbors the transactivation domain, which is the largest part of the protein (amino acids (aa): 1-167), followed by the regulatory domain (aa: 168-177). The protein also encompasses the DNA binding domain (aa: 178-194) followed by the dimerization domain (aa: 195-227). The most C-terminal domain stabilizes the dimerization of BZLF1 (aa: 228-245) (Countryman and Miller, 1985; Farrell et al., 1989; Chang et al., 1990; Flemington and Speck, 1990a; Taylor et al., 1991; Flemington et al., 1992; Chi and Carey, 1993; Petosa et al., 2006; El-Guindy et al., 2006; McDonald et al., 2009). The BZLF1 protein dimerizes like other proteins of the bZIP family members (Farrell et al., 1989; Lieberman and Berk, 1990; Chang et al., 1990) to form a coiled coil structure (Chang et al., 1990; Hicks et al., 2001). Nevertheless, BZLF1 lacks the typical heptad-repeat of leucine residues, which mediates the dimerization in other bZIP protein members (Chang et al., 1990; Flemington and Speck, 1990a). As a consequence BZLF1 does not form heterodimers with other bZIP family members (Chang et al., 1990; Petosa et al., 2006; Reinke et al., 2010) but uses its C-terminus to stabilize its homodimerization (Hicks et al., 2003). Only after dimerization BZLF1 responsive elements (ZREs) are bound as well as AP-1 and C/EBP associated binding motifs (Farrell et al., 1989; Lieberman et al., 1990; Kouzarides et al., 1991). The basic region of BZLF1 contacts the major groove of the DNA (Petosa et al., 2006, Figure 2-4). Two major binding motifs were identified in the viral genome, one containing and one lacking a CpG di-nucleotide in the binding motif (Kalla et al., 2010; Bergbauer et al., 2010). The motif comprising the CpG pair is only bound efficiently when methylated (Bhende et al., 2004; Bergbauer et al., 2010). By phosphorylation of serine 173 in BZLF1's regulatory domain the binding affinity to DNA can

be increased, which is required to activate replication (El-Guindy et al., 2007). The transactivation domain accomplishes several tasks. It is required to evict nucleosomes at viral genes essential for lytic DNA replication (Woellmer et al., 2012). The activation domain was also found to stimulate the formation of the pre-initiation complex (Chi and Carey, 1993) and to interact with the CREB-binding protein (CBP) and the TATA-binding protein TFIID, which are involved in transcriptional mediation themselves (Adamson and Kenney, 1999; Lieberman and Berk, 1991).

2.10 Scope and aim of this thesis

BZLF1 is expressed in newly infected cells but it cannot activate the lytic cycle initially. This feature relies on BZLF1's ability to bind two different DNA sequence motifs. One class of ZREs is reminiscent of AP-1 binding motifs (Lieberman et al., 1990; Chang et al., 1990) and does not contain a CpG pair (nonCpG ZRE). The function of this motif is not affected by DNA methylation in the latent phase. The second class contains a CpG di-nucleotide (CpG ZRE) within BZLF1's binding motif and can only be stably bound when methylated (Bergbauer et al., 2010). In fact BZLF1 preferentially binds methylated CpG ZREs compared with nonCpG ZREs (Bhende et al., 2004; Dickerson et al., 2009; Bergbauer et al., 2010). Since EBV DNA is non-methylated upon B-cell infection, the early expression of BZLF1 does not induce relevant lytic viral genes. As a consequence a strictly latent infection ensues.

The ability of BZLF1 to bind methylated DNA was also described for the c-JUN/c-FOS hetero-dimer, a member of to the AP-1 protein family (Gustems et al., 2014). Additionally, BZLF1 shows sequence homology in its DNA binding domain with different AP-1 family members (Farrell et al., 1989; Taylor et al., 1991). In contrast to other bZIP proteins BZLF1 forms homo-dimers only (Reinke et al., 2010) presumably because it lacks the typical heptad-repeat of leucine residues (Lieberman and Berk, 1990; Flemington and Speck, 1990a; Chang et al., 1990). The similarity to the binding sites of AP-1 family proteins sparked the idea that BZLF1 might influence the global expression of cellular genes to EBV's favor. Additionally, individual cellular genes were reported to be regulated upon expression of BZLF1 (Dreyfus et al., 1999; Lu et al., 2000; Morrison et al., 2001, 2004; Chang et al., 2006; Jones et al., 2007; Hsu et al., 2008; Bristol et al., 2010; Zuo et al., 2011; Tsai et al., 2009; Lan et al., 2013). BZLF1 is a known transcriptional activator in the viral genome, but

surprisingly cellular genes were identified, which were up- or down-regulated upon expression of BZLF1. For example the transforming growth factor β 1 (*TGFB1*), interleukin 10 (*IL10*), and the AP-1 family member *FOS* are up-regulated (Cayrol and Flemington, 1995; Mahot et al., 2003; Flemington and Speck, 1990b) whereas the class II major histocompatibility complex (*CIITA*), the TATA-box binding protein (*TBP*), and *AP-1* family member *JUN* (Li et al., 2009a; Mauser et al., 2002; Sato et al., 1992) are down-regulated.

Two B-cell lines, which represent two different situations in B-cells, were used to clarify BZLF1's role in host cells. The DG75 cell line is EBV-negative lacking the EBV episome. This cell line mimics the pre-latent phase, when EBV infects primary B-cells and expresses BZLF1 initially but transiently (Kalla et al., 2010). Very much in contrast to this situation, only BZLF1 but no other viral proteins are expressed in DG75 cells. The Raji cell line is latently infected with EBV, expresses a number of viral latent proteins and non-coding RNAs, but efficiently enters the lytic phase upon BZLF1 expression.

The experimental thesis work is divided into three major parts: (i) ChIP-sequencing was performed to clarify if BZLF1 binds the cellular genome and the identified binding sites were searched for BZLF1 binding motifs. BZLF1 is a transcriptional activator in the viral genome but its impact on cellular genes is unclear. (ii) Therefore differentially expressed genes should be identified upon BZLF1 induction in transcriptome analysis. Both parts were combined bioinformatically in an attempt to link BZLF1 binding sites with the promoters of regulated cellular genes. (iii) After BZLF1-regulated cellular candidates genes were identified I asked whether BZLF1 might act as an enhancer factor and influences the chromatin architecture. Towards this end Capture-C experiments were performed with selected candidate genes and their chromosomal loci.

3 Results

3.1 The cellular and viral model systems – an experimental overview

The life cycle of EBV comprises three different phases, in which the expression of BZLF1 differs: the pre-latent, latent and lytic phase. In the pre-latent phase, shortly after infection, BZLF1 expression peaks but becomes non-detectable later in the latent phase. During latency, the virus persists in the cell, maintains itself during cell proliferation and avoids to be detected by the immune system. Thus, only a few viral genes, but not BZLF1, are expressed during the latent phase. *In vivo*, the virus establishes latency in memory B-cells. When an infected B-cell encounters a cognate antigen the activated B-cell receptor signaling pathway includes the expression of BZLF1. Upon its expression the virus enters the lytic phase and expresses all lytic viral proteins in an ordered fashion to manipulate the host cell, synthesize and massively amplify viral DNA, express all viral structural proteins, and assemble new virions.

The methylation status of viral DNA governs the onset of EBV's lytic productive cycle. During the pre-latent phase non-methylated DNA prevents the switch to the lytic cycle, but in the latent phase, after the methylation of the viral DNA is complete, the lytic switch is supported. BZLF1 can bind two classes of DNA motifs, one of which contains a CpG base pair within its binding motif. BZLF1 binds this motif only if methylated. This class of binding motif is called CpG ZRE, while the class that BZLF1 binds methylation-independently does not contain a CpG pair and is called nonCpG ZRE. Upon infection the viral DNA is epigenetically naïve, contains non-functional CpG ZREs, which BZLF1 cannot bind, but which are mandatory for the completion of the lytic cycle. CpG ZREs predominantly occur in the promoters of viral, lytic genes (Bergbauer et al., 2010; Kalla et al., 2010). Only after the CpG ZREs are methylated within a few weeks after infection, BZLF1 can bind lytic promoters and activate all lytic, viral genes.

DNA methylation is commonly associated with gene silencing, but the virus uses this epigenetic mechanism to avoid the premature expression of lytic genes immediately after infection and subsequently during the pre-latent phase. With this strategy the virus always

establishes a latent infection and hides from the immune system, because viral, lytic proteins are prominent targets of the host's cellular immune response. In order to escape from latency the transcriptional activator BZLF1 allows the virus to overcome the epigenetic repression associated with CpG methylation.

BZLF1 is the viral homolog to the cellular bZIP protein family AP-1 (Farrell et al., 1989) suggesting that BZLF1 can also control cellular genes to support viral infection and proliferation indirectly. Several publications describe cellular genes to be regulated after induced expression of BZLF (Cayrol and Flemington, 1995; Mahot et al., 2003; Li et al., 2009a; Sato et al., 1992), but a global, systematic study to identify BZLF1 binding sites and its regulated cellular genes in B-cells is missing.

To address this question, two B-cell lines were employed. The DG75 and Raji cells are both Burkitt Lymphomas but differ in an important detail. While DG75 cells are free of EBV, Raji cells are EBV positive and contain dozens of EBV episomes per cell. Raji EBV supports the lytic viral phase, but the endogenous Raji genome lacks the BALF2 gene (Zhang et al., 1988). A BALF2 deficiency makes the virus incapable of synthesizing viral DNA in the lytic phase. As a consequence, expression of viral genes responsible for the synthesis of viral DNA in the lytic phase is blocked and the Raji cells do not release virions (Decaussin et al., 1995). The usage of both cell lines allows studying BZLF1 functions in the context of EBV negative and EBV positive cells and in the absence or presence of other viral proteins (Bhende et al., 2004; Woellmer et al., 2012). Additionally, the DG75 cell line represents a B-cell, which never had contact with BZLF1 before, while the Raji cell line is trapped in the latent phase and gets pushed into the lytic phase by the induced expression of BZLF1. A conditional BZLF1 allele was employed to compare both cell lines in the absence of the BZLF1 protein and upon induced expression of BZLF1.

Prior to my PhD work the laboratory had established an inducible system, in which BZLF1 is expressed from the 4816 plasmid only upon doxycycline addition, as seen in Figure 3-1. The plot shows the transcript levels of BZLF1 and other lytic genes with BZLF1 binding sites in their promoters and their regulation upon doxycycline-induced BZLF1 expression. Transcripts of the viral capsid gene BcLF1 and cellular Cytochrome C (cytC) gene were not affected by the induction of BZLF1, as expected.

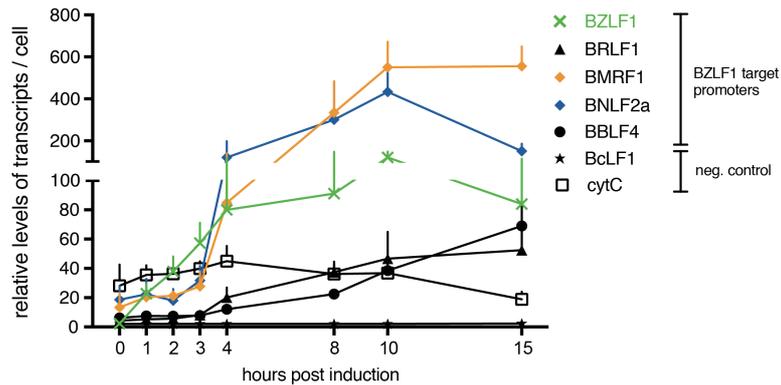


Figure 3-1: qPCR analysis of viral gene transcripts after doxycycline-induced BZLF1 expression. (n = 3)

The figure shows the expression of several viral genes together with a cellular gene as a reference prior to (0 h) and at several time points after doxycycline-induced expression of BZLF1 in Raji 4816 cells. The promoters of all viral genes but BcLF1 contain BZLF1 binding motifs, which BZLF1 binds in the lytic phase. The x-axis shows the hours post BZLF1 induction. The y-axis depicts the relative levels of transcripts per cell. Prior to doxycycline addition, the BZLF1 transcript (green) is barely detectable. After induction, the transcript level rises within one hour and peaks at round 10 h. Genes regulated by BZLF1 respond with the depicted kinetics. Neither the viral BcLF1 gene, nor the cellular reference gene cytC are regulated by BZLF1. Paulina Mrozek-Gorska performed the experiment summarized here. (n = 3, mean and SD are shown)

The inducible 4816 plasmid does not only contain BZLF1 but comprises a sequence coding for the truncated nerve growth factor (NGF) receptor and the green fluorescent protein (GFP), which are, besides BZLF1, induced by doxycycline. NGF receptor and GFP were used for cell sorting and visual evaluation by FACS, respectively. Both proteins are co-expressed from a bicistronic mRNA and separated by an IRES sequence.

Chromatin immunoprecipitation sequencing (ChIP-seq) and ribonucleic acid sequencing (RNA-seq) were used to identify BZLF1 binding sites within cellular DNA and regulated cellular genes, respectively.

BZLF1 is known as a promoter factor in EBV and therefore the two sets of ChIP-seq and RNA-seq results were combined in a bioinformatic approach with the initial idea to discover BZLF1 regulated genes with its cognate binding sites located within their promoters. Surprisingly, my results did not suggest this role of BZLF1 in regulating cellular genes. In cellular DNA very many BZLF1 binding sites were identified, but only few BZLF1-regulated genes contained BZLF1 binding sites within their promoters. These results were insufficient to provide a molecular explanation of the RNA-seq data. As a consequence, Capture-C experiments were performed to detect potential enhancer, which BZLF1 might engage to regulate cellular target genes.

3.2 The expression of BZLF1 in non-induced cells originates from the leaky pRTR 4816 plasmid

Prior to this thesis the pRTR 4816 (4816) plasmid had been in use in our lab to study the consequences of induced BZLF1 expression with respect to viral gene regulation. As seen in Figure 3-1 only upon doxycycline-induced (100 ng/ml) expression of BZLF1, other viral lytic genes containing BZLF1 binding sites within their promoters became activated and detectably transcribed.

During my thesis preliminary results clearly indicated that BZLF1 is expressed at low levels even in the absence of doxycycline. To measure BZLF1 protein levels in the non-induced state Western blot immunodetections were performed with different cell lines prior to and after induction (Figure 3-2, panel A). Parental DG75 and Raji cell lines were used as negative controls. Parental Raji cells do not express BZLF1, the cells are strictly latently infected, while the DG75 cell line is EBV negative and therefore lacks the BZLF1 gene. As expected both cell lines did not show a BZLF1 signal (Figure 3-2, panel A, lane 1, 2) but DG75 4816 (lane 3) and Raji 4816 (lane 4, 5, 6) cells did prior to doxycycline addition. All protein lysates from non-induced cells (lane 3-6) showed a clear signal of the expected mass of BZLF1 (35-38 kDa, Marschall et al., 1989). Fetal calf serum (FCS) from BioSell is currently in use in our lab to cultivate DG75 4816 cells (lane 3) and Raji 4816 cells (lane 4). Raji 4816 cells were additionally tested with two different FCS batches: FCS from PAA (lane 6), which had been used for the results shown in Figure 3-1, and the tetracycline-free FCS from Clontech Laboratories (lane 5) were tested to exclude that traces of tetracycline or its derivatives, which FCS might contain, induce basal levels of BZLF1. All FCS samples gave similar results in Western blot experiments documenting BZLF1 expression prior to addition of doxycycline were independent of the batch of FCS. Lanes 7, 8, and 9 show Raji 4816 cells induced for 15 h. While lanes 1-6 were loaded with protein lysates from 35,000 cells (100 %), lanes 7, 8, and 9 contained protein lysate from 88 (0.25 %), 176 (0.5 %), or 264 doxycycline-treated cells (0.75 %) indicating a very strong induced expression of the BZLF1 protein.

A bacterially produced part to the BZLF1 protein of 24.4 kDa (3 ng, 2 ng, 1 ng purified protein) was added for the absolute quantification of BZLF1 dimers per cell in this and similar Western blots. Panel B of Figure 3-2 shows the quantification of the data after scanning the Western blot images. The BZLF1 protein signals were quantified with ImageJ and used to

calculate the BZLF1 dimers per cell. No BZLF1 was present in parental DG75 or Raji cells. Prior to doxycycline-induced BZLF1 expression 2.9×10^5 BZLF1 dimers were present per DG75 4816 cell. The maximum of BZLF1 dimers (7.5×10^5) was found in Raji 4816 cells prior to induction (-TET, column 5). After doxycycline-induced expression of BZLF1, 8.8×10^7 BZLF1 dimers could be identified, which exceeds the number of BZLF1 dimers in non-induced cells by a factor of about 100. In essence, Western blot quantifications revealed, that BZLF1 is already expressed from the 4816 plasmid prior to doxycycline induction, but BZLF1 level increased considerably after induction. Nevertheless, a protein level close to 8.8×10^7 dimers per cell seems extremely high even after 15 h of induction when compared with other natural transcription factor levels (about 10^5 MYC proteins/ cell, Biggin, 2011). Therefore, this number should be considered with reservation.

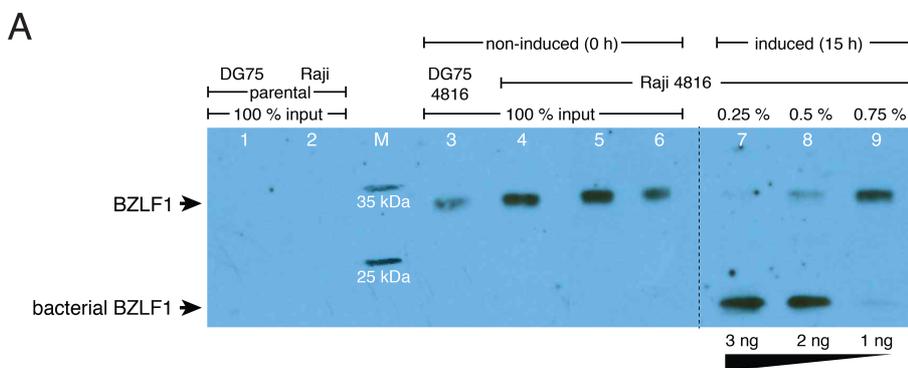
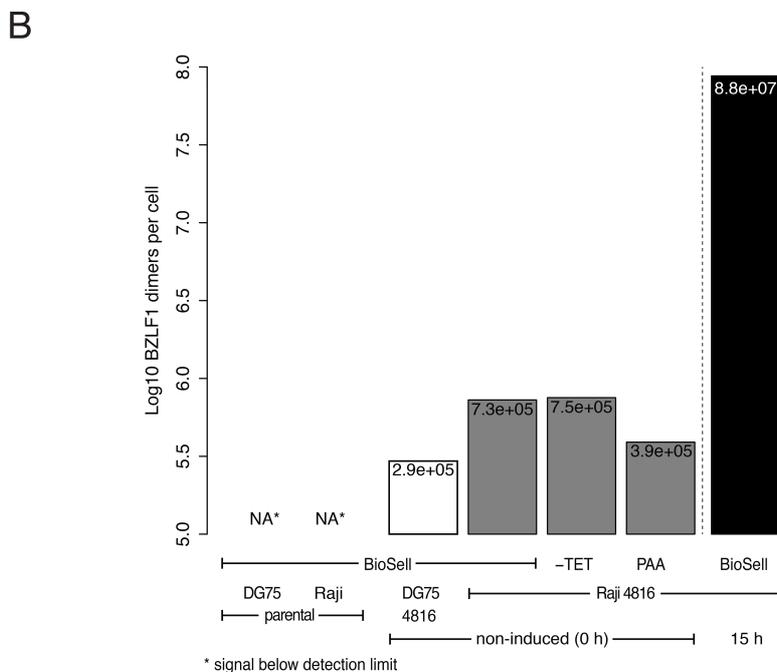


Figure 3-2: BZLF1 protein levels prior to and after doxycycline-induced expression in DG75 and Raji cell lines.

Panel A shows the quantification of BZLF1 protein in DG75 and Raji cells together with a bacterially expressed and purified partial BZLF1 protein. ImageJ was used to quantify the signals of Western blots and the results are summarized in panel B. **A:** Parental DG75 and Raji cells and cells stably transfected with the 4816 plasmid were cultured in medium with different FCS. Protein extracts from 35,000 cells were loaded per lane indicated by “100 % input”. EBV-negative DG75 (lane 1) and Raji (2) parental cells were used as controls. No BZLF1 signals were detected in the two parental cell lines. Non-induced (0 h) DG75 4816 (lane 3) and Raji 4816 (lane 4) cells were analyzed for BZLF1 expression, when no doxycycline was added to cell culture media with FCS from BioCell. Both cell lines showed BZLF1 signals in cell culture media supplemented with



different commercially available batches of FCS (lane 5: tetracycline-free FCS from Clontech Laboratories; lane 6: FCS from PAA), suggesting that tetracycline or its derivatives that FCS might contain are not responsible for non-induced expression of BZLF1 prior addition of doxycycline. Raji 4816 cells induced for 15 h were used as a

positive control. In lanes 7, 8, and 9 protein extracts were loaded, which equal 0.25, 0.5 and 0.75 % of 35,000 cells, respectively. A bacterially produced and purified BZLF1 protein with known mass (24.4 kDa) was used in different amounts (3, 2, and 1 ng in lanes 7, 8, and 9, respectively) to provide a standard for absolute protein quantification. **B**: The graph shows the quantification of the western blot data obtained from panel **A**. The signal intensities were measured with ImageJ. The y-axis indicates the calculated numbers of dimeric BZLF1 molecules per cell on a log₁₀ scale. In parental DG75 (1) and Raji (2) cell lines, BZLF1 molecules could not be detected, but non-induced DG75 4816 and Raji 4816 cells contained up to 7.5×10^5 BZLF1 dimers per cell. Far more BZLF1 dimers (8.8×10^7) were found in doxycycline-induced Raji 4816 cells. The quantification showed that even prior to BZLF1 induction detectable levels of BZLF1 dimers were present in the cells with the 4816 plasmid. Upon induction the level of BZLF1 molecules increased about 100-fold.

In parallel with the Western blot experiment, the Tiergesundheitsdienst Bayern e.V. analyzed a sample of the BioCell FCS for the concentrations of tetracycline and its derivatives. Table 3-1 summarizes the results from the analyzed samples. The detection limit for all tetracycline derivatives in BioCell FCS was 25 ng/ ml, but none of the substrates reached this level, indicating a concentration below detection limit. In cell culture medium FCS makes up 10 % (v/ v), thus the concentration of tetracycline derivatives cannot possibly exceed 2.5 ng/ ml in cell media used throughout my experiments.

Table 3-1: **Concentration of tetracycline and its derivatives in BioSell's FCS.**

The table summarizes the tetracycline derivatives tested by the Tiergesundheitsdienst Bayer e.V. by high performance liquid chromatography (HPLC). None of the compounds reached concentrations higher than the detection limit of 25 ng/ ml.

| Tetracycline derivatives | Detection limit [ng/ ml] | Detected [ng/ ml] |
|----------------------------------|--------------------------|-------------------|
| Chlortetracycline (incl. epimer) | 25 | n.n |
| Doxycycline | 25 | n.n |
| Oxytetracycline (incl. epimer) | 25 | n.n |
| Tetracycline (incl. epimer) | 25 | n.n |

3.3 BZLF1 is expressed prior to doxycycline induction but its level does not induce the viral lytic cycle

Next generation RNA-seq experiments were employed to analyze viral gene expression prior to (0 h) and after doxycycline-induced BZLF1 expression (3 h / 6 h), to judge the effect in non-induced cells comprising the p4816 plasmid. The results are summarized and visualized in Figure 3-3, which shows RNA-seq reads mapped to the entire Raji EBV genome (coordinates #1 to #166,184). In tracks 1-3 the read coverage (reads per base) is indicated for each time point (0 h: 0-250, 3 h: 0-10,000, 6 h: 0-50,000).

Additionally, each row indicates a single selected viral gene and its fold induction 6 h after doxycycline-induced BZLF1 expression. The reads per gene were counted with the

bioinformatic tool “HTSeq count” and differential gene expression was calculated with “DESeq2”. The positions of the genes are linked to the RNA expression profiles by colored columns, which indicate latent (orange) and lytic genes (blue). While the latent genes EBNA1 (x 0.4) and EBNA2 (x 0.3) were down-regulated 6 h after doxycycline-induced BZLF1 expression, other latent genes were hardly affected by BZLF1 and their transcript levels stayed constant. Contrary to EBV latent genes, all lytic genes were strongly up-regulated 6 h after doxycycline-induced BZLF1 expression. A moderate up-regulation was detected for a late lytic gene lacking a BZLF1 binding site within its promoter (BcLF1, x 4.9). The strongest induction of a viral gene comprising BZLF1 binding sites could be seen with the BBLF3/2 gene (x 257).

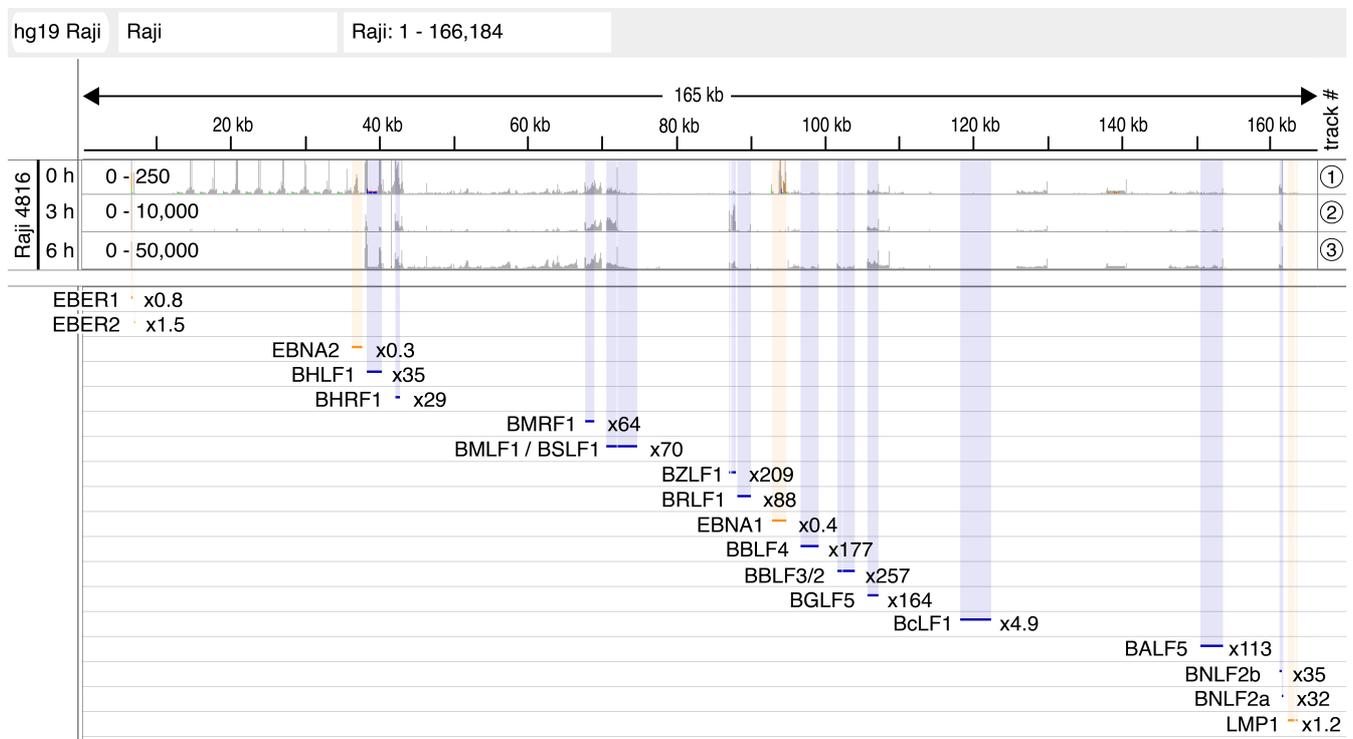


Figure 3-3: Viral gene regulation after doxycycline-induced BZLF1 expression in next generation RNA-sequencing.

The figure shows the complete Raji EBV genome (nucleotide coordinates #1 to #166,184) in an IGV browser window. On top, the raw reads 0, 3, and 6 h after BZLF1 induction are shown. They originate from next generation RNA-sequencing, which was performed with Raji 4816 RNA isolated from non-induced and induced cells. Prior to BZLF1 induction (track 1), reads reached a maximum coverage of about 250 reads per base. Three and six hours post induction (track 2 / 3) the counts reached up to 10,000 and 50,000 respectively, indicating a substantial induction of viral transcription. The lower part shows the positions of selected viral genes, their acronyms and their x-fold induction six hours after induced expression of BZLF1. The gene reads were counted with HTSeq count and the fold change was calculated and normalized with DESeq2. Vertical, semi-transparent columns connect the gene positions with the raw read counts. Viral latent genes are shown in orange, lytic genes are shown in blue. The latent genes EBER1, EBER2, and LMP1 were not regulated by BZLF1, but EBNA1 and EBNA2 are down-regulated by a factor of 0.4 and 0.3, respectively. Expression of BZLF1 led to a strong up-regulation of all selected lytic viral genes except BcLF1, which lacks a BZLF1 binding site within its promoter. The range of x-fold induction reaches from a factor of 4.9 (BcLF1) to 257 (BBLF3/2). The expression level and exact position of every gene can be found in Table 9-1 in the attachment.

Six viral genes BMRF1 (x 64), BSLF1 (x 70), BBLF4 (x 177), BBLF2 / 3 (x 257), and BALF5 (x 113) are essential for lytic amplification of viral DNA (Fixman et al., 1992). The sixth gene on this list, BALF2, is deleted in Raji EBV DNA (Zhang et al., 1988). Additionally, four genes support lytic viral replication, mediated by the origin of lytic replication (oriLyt) in cis:

BHLF1 (x 35) (Rennekamp and Lieberman, 2011), BMLF1 (x 70), BZLF1 (x 209) and BRLF1 (x 88) (Fixman et al., 1992).

The very strong induction of viral gene expression (track 2: 3 h, 40 fold; track 3: 6 h 200 fold), upon doxycycline-induced BZLF1 expression revealed that the doxycycline-independent expression of BZLF1 (track 1) is insufficient to trigger the lytic phase of EBV.

3.4 BZLF1 binds cellular chromatin with high frequency and with a clear preference for CpG motifs

3.4.1 The experimental setup of ChIP-Seq experiments

The viral, lytic cycle in Raji 4816 cells is not activated despite a constitutive expression of BZLF1 in the absence of doxycycline. This situation should provide a model, which is suitable to identify the binding sites of BZLF1 in the cellular genome.

EBV's transcription factor BZLF1 binds DNA sequence motifs of two classes in the viral genome: CpG ZREs and nonCpG ZREs. While BZLF1 binds CpG ZREs only when methylated (Bergbauer et al., 2010), it binds nonCpG ZREs unconditionally and independently of any base modification. In primary B-lymphocytes BZLF1 is a critical activation factor (Kalla et al., 2010); it was thus conceivable to postulate that BZLF1 does not only bind and regulate viral genes but cellular genes as well.

BZLF1 is the viral relative and homolog of members of the cellular transcription factor family AP-1. They are important regulators of cellular functions and it was therefore suspected that BZLF1 also controls cellular genes. To locate the BZLF1 binding sites in cellular DNA, a ChIP-seq approach was used. Cells were either left non-induced (0 h) or were induced with doxycycline (15 h) to trigger BZLF1 expression. No crosslinking was performed. The DNA was sheared prior to precipitation of BZLF1-associated chromatin. Prior to library preparation and Illumina next generation sequencing (NGS) a quantitative PCR (qPCR) was performed to control for a successful pull-down of BZLF1 bound DNA.

The ChIP-seq data were further analyzed *in silico* to identify the binding motifs of BZLF1 within the cellular genome. Data received from the sequencer were mapped to the human genome 19 (Hg19) and peak calling was performed to identify regions with enriched sequencing reads (peaks). The DREME algorithm was used to identify BZLF1 binding motifs within the peaks found in cellular DNA.

3.4.2 BZLF1 binding sites in two B-cell lines

3.4.2.1 Analysis of the conditional B-cell line DG75 4816

In DG75 4816 cells BZLF1 is the only EBV protein because this Burkitt lymphoma cell line is EBV-negative. DNA sequences bound by BZLF1 were pulled down by ChIP to identify BZLF1 binding sites within the cellular genome of DG75 4816 cells.

Figure 3-4 depicts the results of a qPCR documenting the successful pull-down of BZLF1 associated chromatin in two experiments. For each experiment the parental DG75 cell line (white) was tested as well as the DG75 4816 cell line prior to (black, 0 h) and post BZLF1 induction (grey, 15 h). The early growth response 1 gene (*EGR1*) served as a negative control (unpublished observation), the interleukin 8 gene (*IL8*) as a positive control (Hsu et al., 2008).

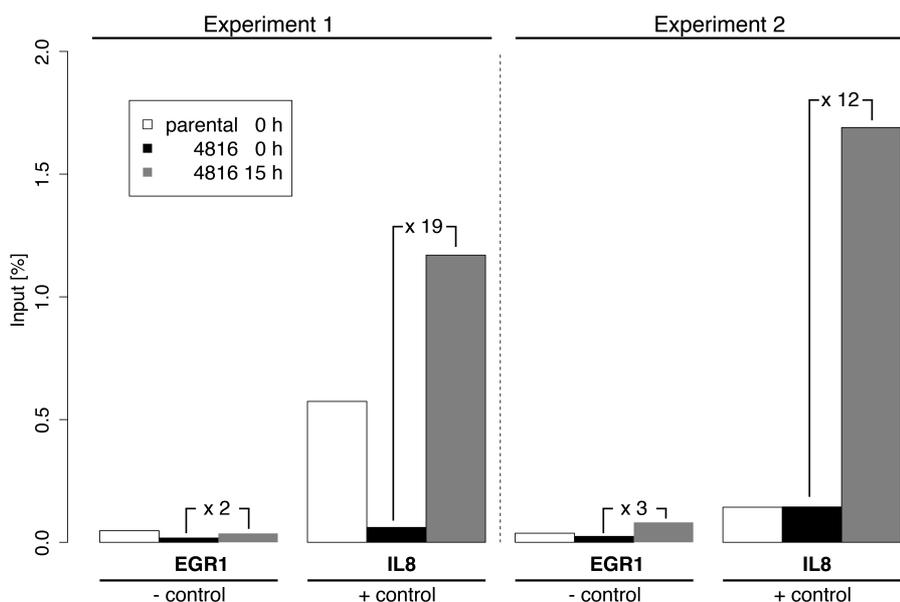


Figure 3-4: Evaluation of ChIP experiments with qPCR.

The figure shows the qPCR results of two independent pilot experiments after chromatin immunoprecipitations (ChIP) with parental DG75 and DG75 4816 cells. The y-axis shows the percentage of the input material, which was precipitated with the monoclonal BZ1 antibody directed against BZLF1. Two cellular loci were tested: *EGR1* served as a negative control (unpublished observation); *IL8* as a positive control (Hsu et al., 2008). The white, black and grey columns show the results obtained with parental DG75 cells, non-

induced (0 h) and induced (15 h) DG75 4816 cells, respectively. The results at the *IL8* locus indicated that BZLF1 is present at this locus only after induced expression. BZLF1 levels were low at the control locus *EGR1* in

parental DG75 cells, in non-induced, and induced DG75 4816 cells, indicating that the ChIP experiments are selective and informative.

The negative control was hardly detectable in parental DG75 cells or in non-induced or BZLF1 induced DG75 4816 cells. The positive control was strongly bound 15 h after BZLF1 induction. The results documented that BZLF1-bound DNA sequences were precipitated selectively and specifically in the chosen experimental setting indicating that the remaining DNA of both experiments was well suited for next-generation-sequencing.

The data obtained from Illumina sequencing (single-end) were mapped on Hg19 and processed for peak calling with the MACS2 algorithm. Only peaks consistently identified in both experiments were further analyzed. The results depicted in Figure 3-5 originate from the motifs identified by the DREME algorithm from sequences identified with the MACS2 algorithm. The figure shows binding motifs identified prior to (0 h) and after (15 h) doxycycline-induced expression of BZLF1.

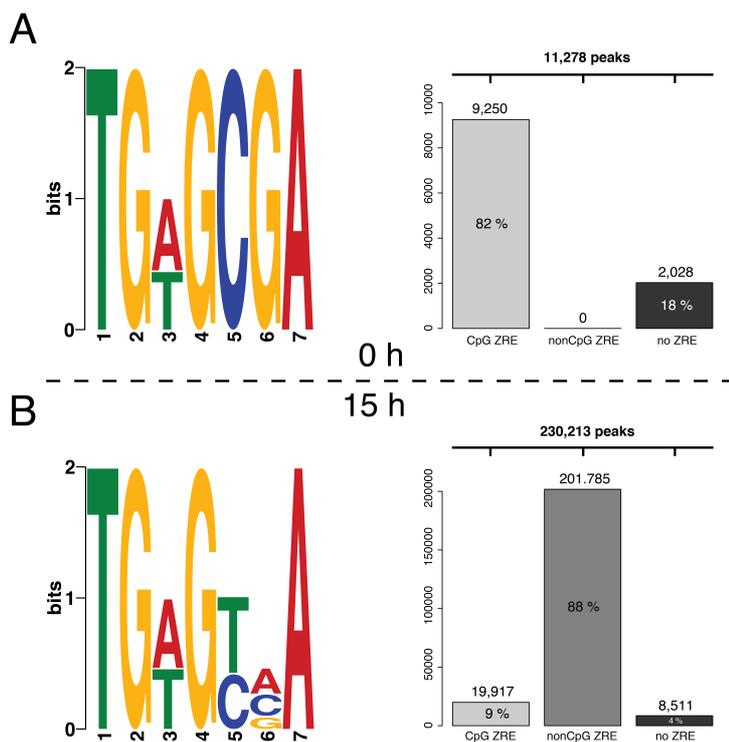


Figure 3-5: Identification of BZLF1 consensus motifs from next generation ChIP-seq data with the MACS2 peak caller and the DREME algorithm.

Chromatin immunoprecipitations with the BZLF1 antibody BZ1 were performed with non-induced DG75 4816 cells and cells induced for 15 h. NGS data were analyzed for consensus binding motifs of BZLF1. The analysis revealed two consensus motifs that differ in non-induced cells (0 h) and induced cells 15 h post induction. **A**: In non-induced cells the peak caller MACS2 identified 11,278 peaks, which were further analyzed with the DREME algorithm. About 82 % of the identified peaks contained a ZRE motif with a CpG ZRE. No nonCpG ZREs could be identified in any of the peaks by DREME. In the remaining 18 % of the peaks DREME did not identify a discrete third ZRE motif (no ZRE). **B**: After induction for 15 h the peak caller identified 230,213 peaks, 96 % of which contained one of the two classes of ZRE motifs. About 9 % of identified peaks belong to the CpG ZRE class, 88 % belong to the class of nonCpG ZREs, and

4 % of the peaks did not encompass a known ZRE motif sequence. At lower levels BZLF1 preferentially bound to CpG ZREs. After induced expression, BZLF1 bound about 20 times more sites in DG75 4816 cellular DNA. Among the identified binding motifs nonCpG ZREs predominated by a factor of about 10 over the class of CpG ZREs.

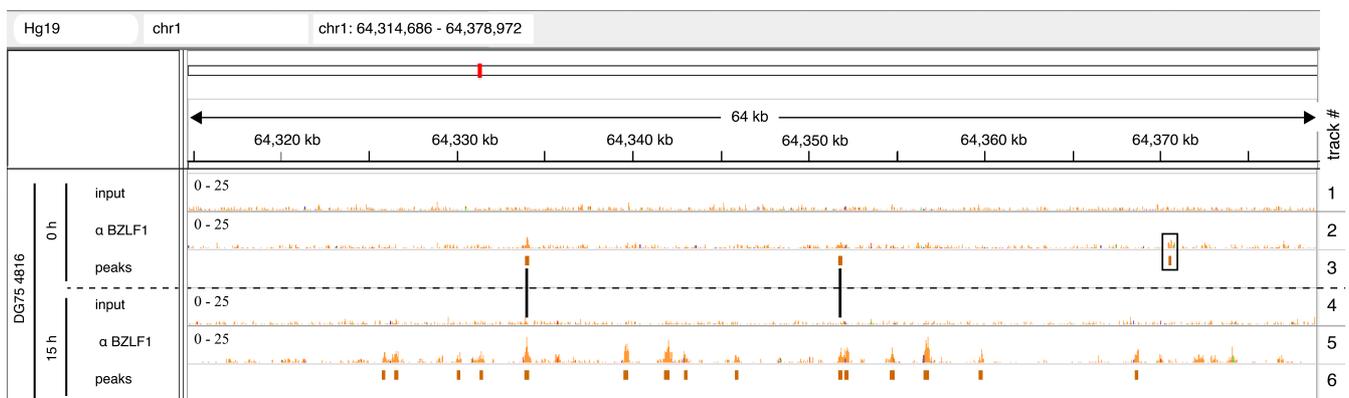
Prior to doxycycline-induced BZLF1 expression 11,278 peaks could be identified by MACS2. About 82 % percent of the peaks contained the known CpG ZRE (TGWGC GA) motif, while the nonCpG ZRE (TGWGTMA) motif did not show up in the list of the top 10 most

probable motifs identified by DREME. Eighteen percent of the peak sequences identified with the help of MACS2 did not contain a known BZLF1 binding motif.

After BZLF1 was induced for 15 h, a completely different distribution of ZREs was found. Now, within the 230,213 peaks both classes of ZREs were identified with MACS2, a 21-fold increase compared with non-induced cells. The majority of about 88 % of all peaks belong to the class of nonCpG ZREs. Fifteen hours after BZLF1 induction the number of identified CpG ZREs increased only about two fold (from 9,250 to 19,917) and then made up about nine percent of all peaks. Among the called peaks, only four percent did not fall into any of the two classes. These 8,511 peaks did not contain a clear alternative motif in the list provided by the DREME algorithm indicative of sequence-specific binding of BZLF1.

During the visual inspection and comparison of peaks identified at 0 h and 15 h, a certain pattern was noticed, which is exemplary visualized in Figure 3-6. The figure shows the raw data of the input and the α -BZLF1 ChIP data prior to and post BZLF1 induction in a snapshot generated by the integrated genome viewer (IGV). In addition to the raw and ChIP data peaks identified by MACS2 are depicted for both time points (panel A).

A



B

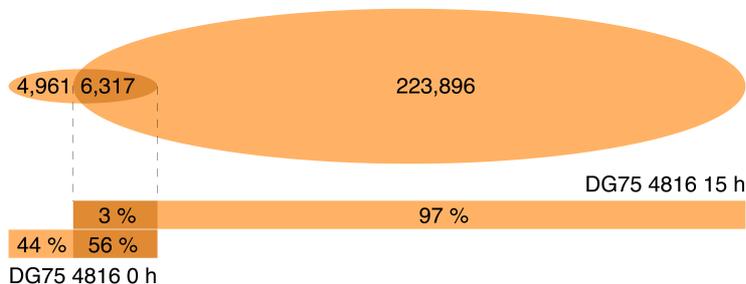


Figure 3-6: Visualization of BZLF1 peaks in the IGV browser prior to and after induced expression of BZLF1 in DG75 4816 cells.

Three categories of peaks were identified with the MACS2 algorithm. Panel A visualizes these peaks and their identified categories; panel B summarizes the categories genome-wide in a Venn diagram. **A:** The visualization is an example of a snapshot of the IGV browser comprising nucleotide coordinates #64,314,686 – #64,378,972 of chromosome 1 of Hg19. All six tracks show results obtained with the DG75 4816 cell line. Tracks 1 to 3 show results of non-induced DG75 4816 cells (0 h), tracks 4 to 6 show data of the induced DG75 4816 cells (15 h).

Tracks 1, 2, 4, and 5 show next generation ChIP-seq raw data. ChIP input data are depicted in tracks 1 and 4, ChIP data obtained with an α -BZLF1 antibody are shown in track 2 and 5. The track height represents the read counts 0-25 for DG75 4816 cells for raw and input data. Track 3 and 6 depict the peaks, which MACS2 identified comparing track 2 with track 1 (0 h) and track 5 with track 4 (15 h), respectively. Three categories of peaks become evident in this experiment. One category comprises peaks, which are identified in both non-induced and induced cells. Connecting black vertical lines highlight these peaks. The second and largest category comprises peaks shown in track 6 that only appear in DG75 4816 cells after induction of BZLF1. A third category shown in track 3 indicates peaks that were identified in non-induced cells, only, but were lost 15 h after induced expression of BZLF1. **B:** The plot shows a Venn diagram (top), which provides the absolute numbers of BZLF1 peaks and peaks that were present both prior to (0 h) and post BZLF1 induction (15 h). The bar diagram (bottom) depicts the relative abundance of the peak categories.

Panel B summarizes the peaks identified prior to and post BZLF1 induction and their overlaps. The most striking finding is that many more peaks were identified 15 h post induction (about 230,000), when compared with peaks prior to doxycycline-induced BZLF1 induction (about 11,000). About 97 % of these peaks appear only 15 h post BZLF1 induction (about 224,000) as expected, but about 6,000 peaks were present both prior to and were maintained 15 h post BZLF1 induction. They are depicted in Figure 3-6 (panel A) by connecting black vertical lines. Another category of peaks was detected prior to doxycycline-induced BZLF1 expression, only, but lost 15 h post induction of BZLF1 (about 5,000, marked by a black square in Figure 3-6).

3.4.2.2 Analysis of the conditional cell line Raji 4816

In contrast to DG75 cells, Raji cells are EBV-positive and thus BZLF1 is not the only viral protein potentially expressed in Raji cells. Many target genes of BZLF1 are known in the

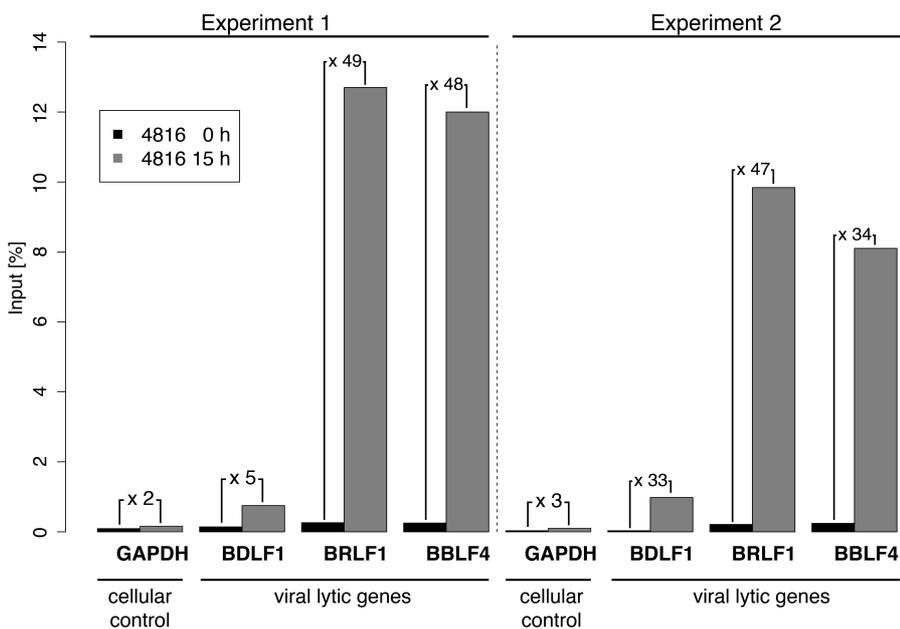


Figure 3-7: Evaluation of ChIP experiments with qPCR

The figure shows qPCR results of two independent pilot experiments with Raji 4816 cells prior to and post induction of BZLF1. The y-axis indicates the percentage of chromatin recovery after chromatin immunoprecipitation as shown in Figure 3-4. Three viral lytic genes (BDLF1, BRLF1, and BBLF4) and a cellular control gene (*GAPDH*), which were used as positive and negative controls, respectively, are indicated on the x-axis (Bergbauer et al., 2010). The black and grey columns show results prior to and post induction of BZLF1, respectively.

EBV DNA sequence and were used as positive controls. In a pilot experiment, a qPCR approach evaluated successful pull-down of promoter sequences of three viral, lytic genes (positive controls) and the cellular *GAPDH* gene (negative control). The experiment was performed two times and the results are visualized in Figure 3-7. The % input of the pulled down promoter DNA prior to BZLF1 induction is shown in black, post BZLF1 induction in grey. While the cellular control is hardly enriched, the promoters of the lytic genes known to be bound by BZLF1 are strongly enriched (up to 49 fold) after ChIP. Since the pull-down of BZLF1 target sequences was successful the samples of both experiments were used for library preparation and single-end sequencing on an Illumina machine.

The reads obtained from the sequencer were then mapped to Hg19 and the Raji EBV genome. After further processing, the MACS2 peak caller was used to identify peaks in sequences pulled down prior to (30,155) and after (144,877) doxycycline-induced BZLF1 expression (Figure 3-8). Prior to induction of BZLF1 the CpG ZRE (TGWGCGA) was identified by DREME in about 70 % of all identified peaks, nonCpG ZREs were absent and 30 % of the remaining peaks did not contain a discrete, additional binding motif of BZLF1 in the top 10 list of motifs identified by the DREME algorithm.

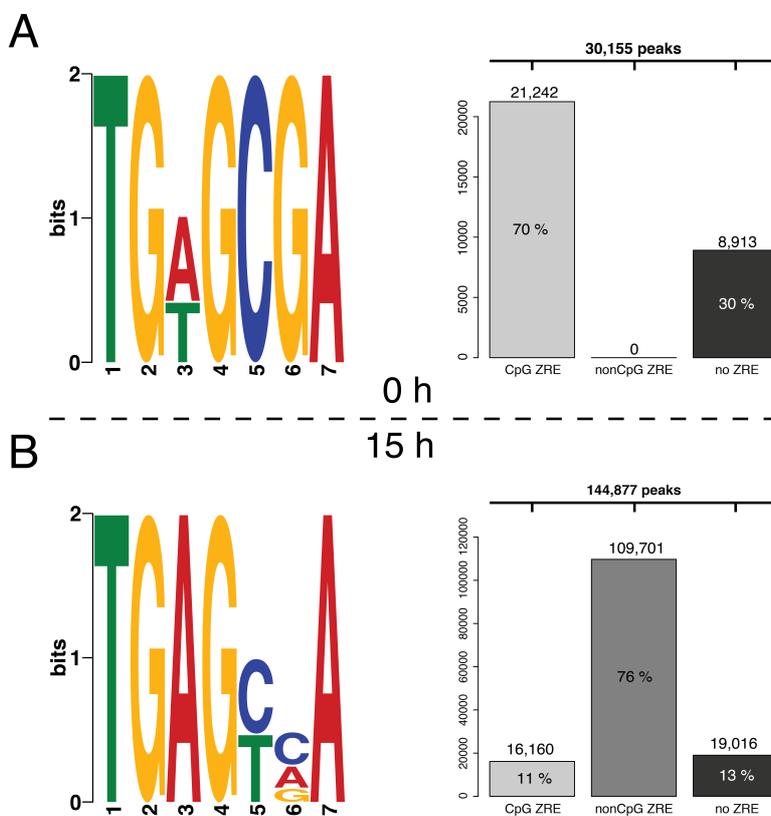


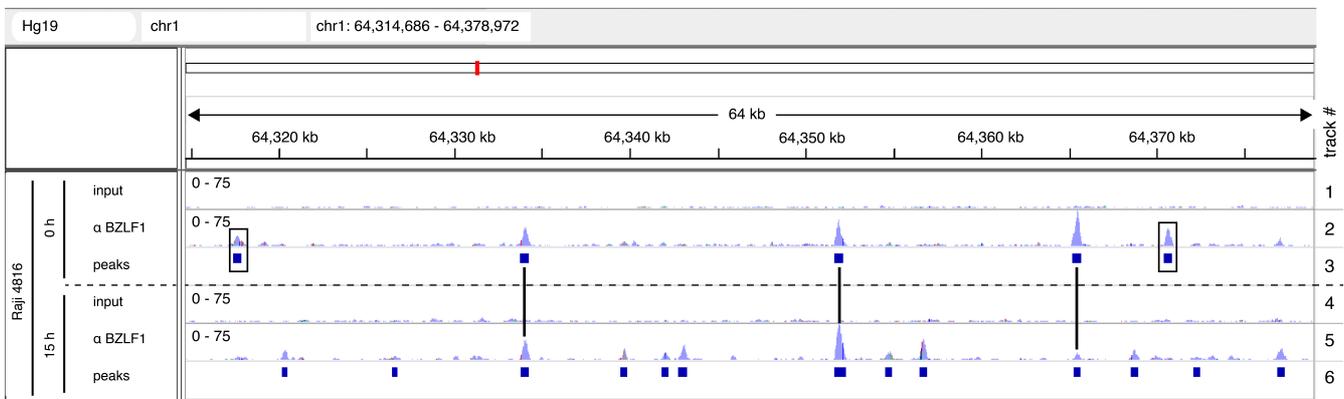
Figure 3-8: Identification of BZLF1 consensus motifs from next generation ChIP-seq data with the MACS2 peak caller and the DREME algorithm.

The experiment was performed and NGS data were analyzed as described for DG75 4816 cells in Figure 3-5. In Raji 4816 cells two consensus motifs could be identified. Their relative abundance differs between non-induced cells (0 h) and induced cells (15 h). **A:** In non-induced cells (0 h) 30,155 peaks were identified. About 70 % of the identified peaks contained a CpG ZRE, while nonCpG ZREs were absent. The remaining fraction of peaks (30 %) did not contain an additional third ZRE motif (no ZRE). **B:** 144,877 peaks were identified 15 h post BZLF1 induction, 87 % of which contained two classes of ZRE motifs. The CpG ZRE class represented 11 % of the peaks, while the nonCpG ZRE class covered 76 %. No additional ZRE motifs could be identified in the remaining 13 %. In Raji 4816 cells BZLF1 exclusively bound CpG ZRE

motifs when expressed at lower level. After induced expression, the number of BZLF1-associated peaks increased about five-fold. In induced cells, the class with nonCpG ZRE binding motifs predominated by a factor of about seven over the class with the CpG ZRE motif.

Fifteen hours after doxycycline-induced BZLF1 expression MACS2 identified about 5 times more peaks and the distribution of the two classes of ZREs changed considerably. Around 87 % of all peaks contained one of the two ZRE motifs; the majority (76 % were nonCpG ZREs, the minority (11 %) were CpG ZREs. While nonCpG ZRE motifs were absent prior to BZLF1 induction the nonCpG ZRE class surpassed the class of CpG ZREs by about a factor of seven after BZLF1 induction.

A



B

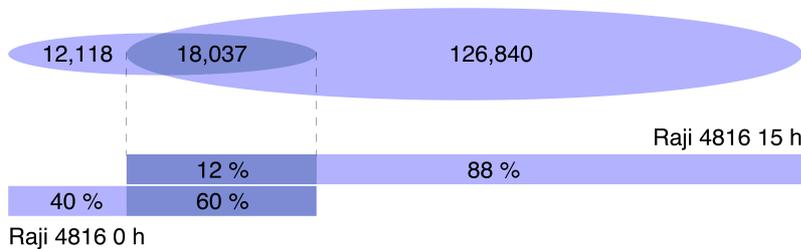


Figure 3-9: Visualization of BZLF1 peaks in the IGV browser prior to and after induced expression of BZLF1 in Raji 4816 cells.

A: Shown is an example of a snapshot of the IGV browser comprising nucleotide coordinates #64,314,686 – #64,378,972 of chromosome 1 of Hg19. All six tracks show results obtained with the Raji 4816 cell line. The setup of the figure is identical to Figure 3-6. **B:** A Venn diagram shows the absolute numbers and relative abundance of identified BZLF1 binding sites.

In addition to the bioinformatic and statistical analysis the data, which were similar to the data obtained with DG75 cells in Figure 3-6, were also investigated in the IGV browser. In Raji 4816 cells (Figure 3-9), the three previously identified categories of peaks re-emerged:

peaks, which appeared after induced BZLF1 expression (about 145,000); peaks, which were identified after BZLF1 induction and were already present prior to BZLF1 induction (about 18,000, panel B; connected by black vertical lines, panel A); and peaks, which were

identified prior to but lost after BZLF1 induction (about 12,000 in panel B; marked by a black square in panel A). Additionally, peaks are shown, which were already present prior to induction (about 30,000 in panel B; black squares in panel A).

To get an impression whether the identified binding sites are common in both cell lines, the two sets of results prior to and after BZLF1 induction were merged and analyzed together. An IGV snapshot shown in Figure 3-10 provides an example. Data obtained from the parental DG75 cell line were added to assess the specificity of the α -BZLF1 antibody (BZ1) and the ChIP protocol.

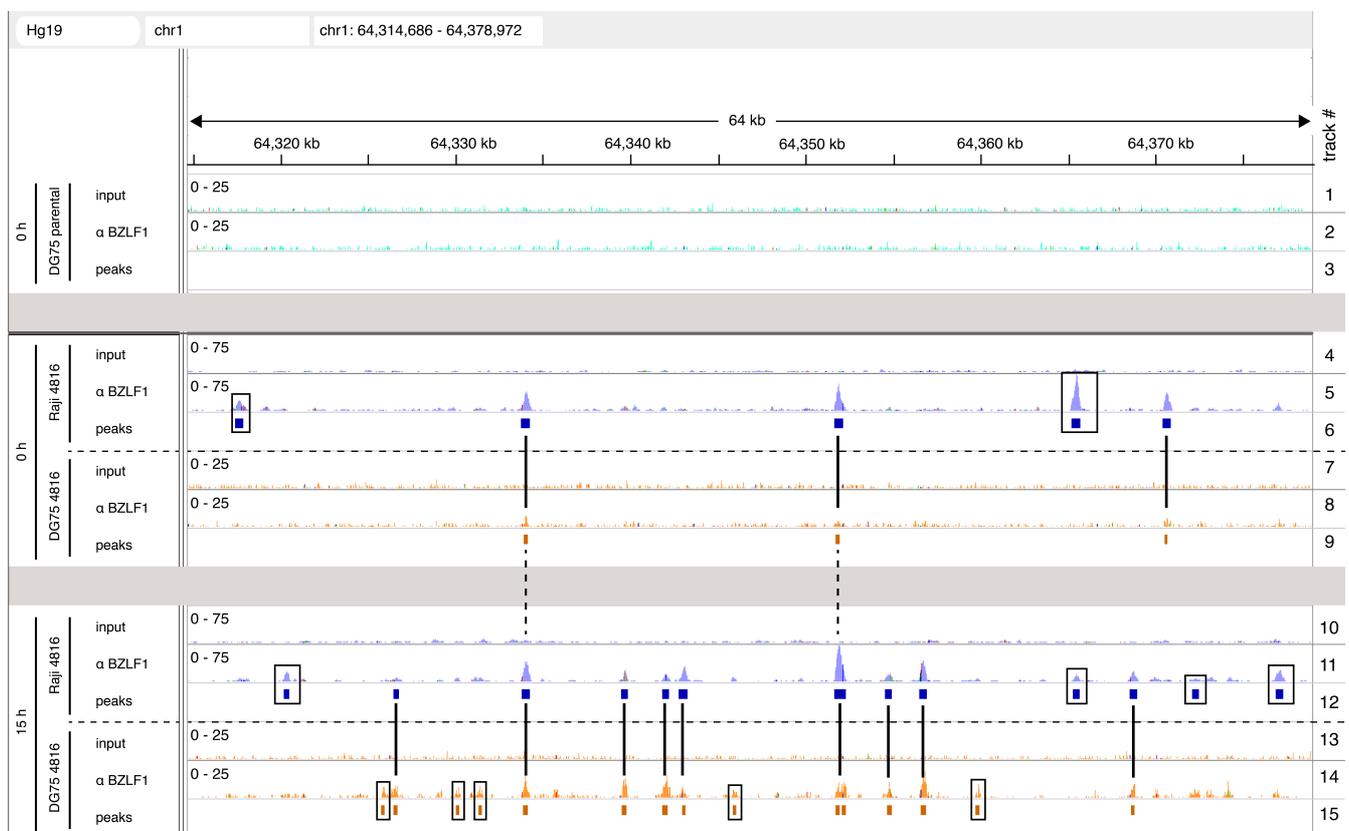


Figure 3-10: Visualization of BZLF1 peaks in the IGV browser prior to and after induced expression of BZLF1 comparing the Raji 4816 with the DG75 4816 cell line.

The visualization is an example of a snapshot of the IGV browser comprising nucleotide coordinates #64,314,686 – #64,378,972 of chromosome 1 of Hg19. On the right side of the figure 15 tracks are indicated. The first three tracks contain data of the parental cell line DG75. The first two tracks show the raw reads of NGS. Track 1 depicts the input and track 2 the results obtained after chromatin immunoprecipitations with an α -BZLF1 antibody. The 3rd track indicates the peaks identified with the MACS2 algorithm. No peaks were found here. Tracks 4 to 9 show the non-induced state (0 h) in Raji 4816 (tracks 4-6) and DG75 4816 (tracks 7-9) cells. Tracks 4 and 7 show the input; tracks 5 and 8 data of the α -BZLF1 ChIP; and tracks 6 and 9 the identified MACS2 peaks. The track height represents the read counts from 0-75 and 0-25 for Raji 4816 and for DG75 4816 cells, respectively. Solid black vertical lines connect peaks, which could be identified in both cell lines in the non-induced and induced state. Peaks that show up in one of the two cell lines, only, are indicated by black squares. Tracks 10 to 15 show both cell lines in their induced state (15 h). The assembly of tracks is identical to data

described for the non-induced state as shown above. Individual peaks are marked by black squares; vertical dashed lines connect peaks that appear in both cell lines at all time points.

Three tracks (input, α-BZLF1 ChIP, and MACS2 peaks) are shown for each cell line prior to and after induction of BZLF1. No peaks were identified in the parental DG75 cells (tracks 1-3), but in the two cell lines Raji 4816 and DG75 4816 three categories of peaks were found: peaks, which were identified in only one of the two cell lines are indicated by black squares; peaks, which were identified in both cell lines in either non-induced cells or cells induced for 15 h are indicated by black vertical lines; and peaks, which were identified in both cell lines and during both time points are connected by dashed black lines.

Three Venn diagrams summarize the numbers of peaks of the two cell lines and the shared or overlapping peaks between different time points (Figure 3-11). Panel A shows the overlap (about 9,500) between Raji 4816 peaks and DG75 4816 peaks prior to BZLF1 induction (0 h). Panel B shows that in Raji 4816 and DG75 4816 cells about 90,000 peaks overlap 15 h after BZLF1 induction. Panel C shows the intersection of all peaks identified both in DG75 4816 and Raji 4816 cells prior to and after BZLF1 induction (about 4,600).

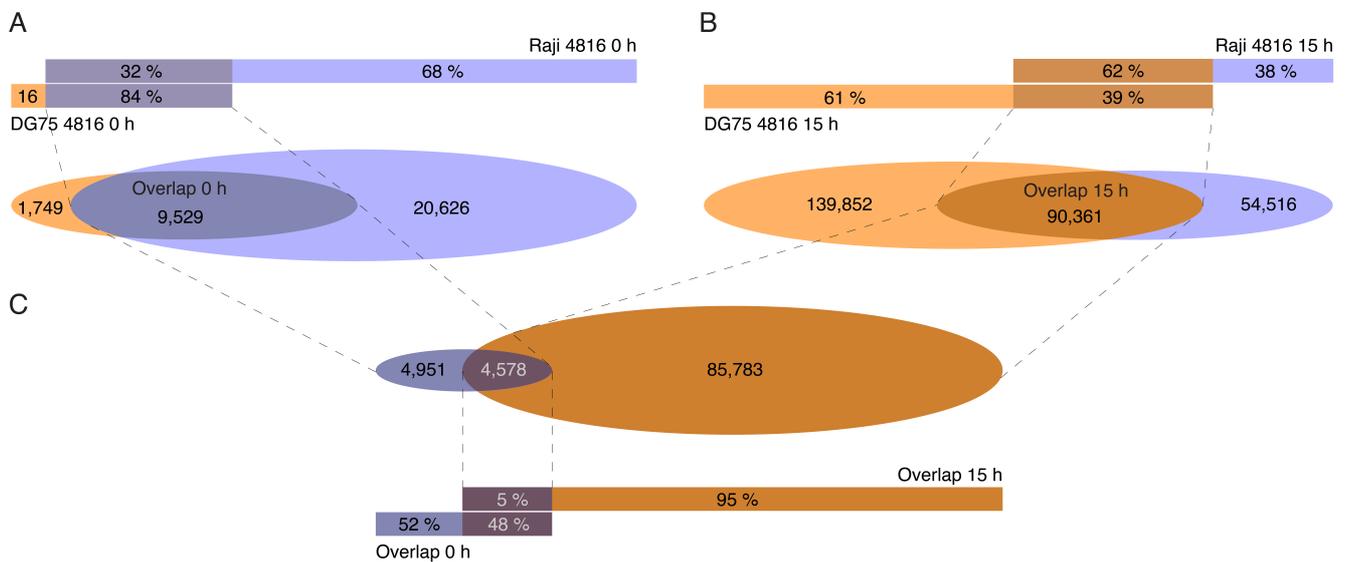


Figure 3-11: Peaks identified after α-BZLF1 ChIP in Raji 4816 and DG75 4816 cells.

Shown are Venn diagrams summarizing the results after peak identification in non-induced cells (panel A) and after BZLF1 induction with doxycycline for 15 h (panel B) in two cell lines. The overlaps between the two cell lines identified at different time points were investigated again (panel C). The results represent the number of peaks identified in both cell lines at both time points, which are indicated in Figure 3-10 with dashed, vertical lines. The absolute numbers of peaks in each cell line and the number of identical peaks, which were found in both cell lines (“overlap”) are indicated together with their relative abundance (%).

3.5 The expression of BZLF1 changes the expression of cellular genes

3.5.1 The experimental setup of RNA-seq experiments

The ChIP-seq experiments investigated the binding sites of BZLF1 in the DNA of two B-cell lines. BZLF1 does not only bind viral DNA, but is a transcriptional activator of viral genes as well. Upon expression during the viral latent phase, BZLF1 binds the promoters of several lytic, viral genes (Bergbauer et al., 2010) and induces their expression dramatically.

Since BZLF1 is the viral homologue of the cellular AP-1 family, it was expected to bind cellular promoters and regulate their gene expression. Several publications dealt with selected BZLF1 binding and locus specific regulation of single cellular target genes (Cayrol and Flemington, 1995; Mahot et al., 2003; Li et al., 2009a; Sato et al., 1992). To address the role of BZLF1 in regulating cellular gene expression globally, several RNA-seq experiments with full-length BZLF1 and a BZLF1 mutant lacking the transactivation domain (AD-tBZLF1) were performed in the two different B-cell lines used previously in the ChIP-seq experiments. DG75 4816 and Raji 4816 cells express BZLF1 upon doxycycline induction. DG75 cells are a model of EBV-negative B-cells but the Raji cell line is latently infected and efficiently enters the lytic phase of EBV's life cycle upon induced BZLF1 expression. The AD-tBZLF1 mutant is used as a control and is also only expressed upon doxycycline induction. The parental DG75 and Raji cell lines do not harbor an inducible plasmid and were not expected to be affected by the addition of doxycycline.

The protein should be expressed in a high fraction of the cells to analyze the consequences of BZLF1 on cellular gene expression. BZLF1 might regulate the expression of cellular transactivators or repressors, which themselves might have an impact on cellular gene regulation. It was mandatory to analyze early time points (3 and 6 h) after BZLF1 expression to avoid secondary effects on gene expression. Doxycycline-mediated expression of BZLF1 is relatively fast but not efficient enough to obtain a uniform high expression level in all doxycycline-treated cells. Therefore the analysis had to be restricted to cells expressing BZLF1. The doxycycline regulated and co-expressed NGF-receptor was used to sort BZLF1 or AD-tBZLF1 positive cells with magnetic beads. After sorting, the cells were used for RNA-isolation, followed by library preparation and strand-specific, single-end Illumina sequencing. In certain experiments an additional control was introduced because it was unclear if BZLF1

altered the entire cellular transcriptome, which would make the data unsuitable for data normalization. Therefore, experiments in Raji 4816 cells with full-length BZLF1 were conducted with additional External RNA Controls Consortium (ERCC) spike-in RNAs to be able to normalize the RNA expression with a stable external reference. The commercially available set of ERCC spike-ins consists of 92 artificial RNAs of different concentration and serves as a reference if added to an equal number of cells prior to RNA extraction.

The results obtained with DG75 4816 cells were visualized as relative log expression (RLE) and mean average (MA) plots, while the data from of the Raji 4816 cells also include a principal component analysis (PCA) and a KEGG pathway analysis.

3.5.2 Induced expression of BZLF1 in DG75 4816 cells regulates only few cellular genes

After addition of doxycycline for 3 and 6 h, NGF-receptor-positive cells were enriched with an antibody coupled to magnetic beads.

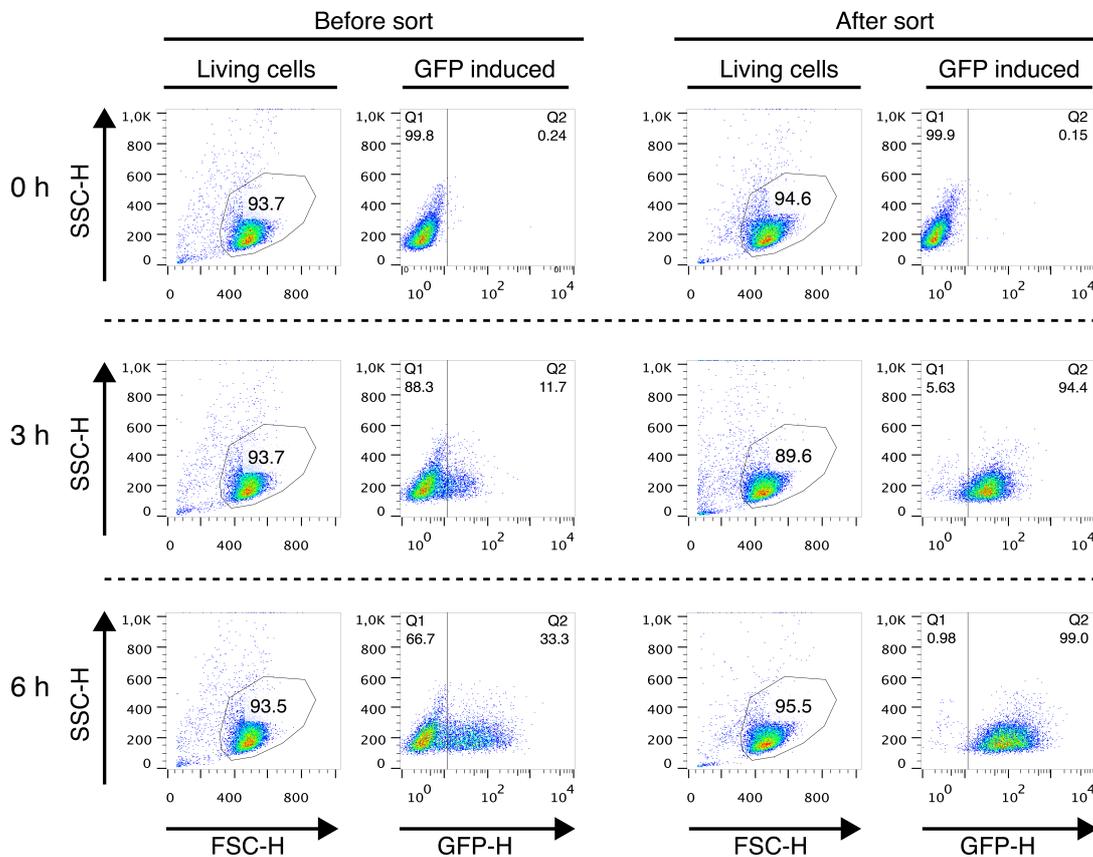


Figure 3-12: Enrichment of doxycycline-induced cells with magnetic beads and an α -NGF-receptor antibody.

The figure shows FACS data of DG75 4816 cells before (left) and after (right) sorting of NGF-receptor positive cells. The NGF-receptor and the GFP gene are co-regulated with BZLF1 upon addition of doxycycline. 3 and 6 h post induction the cells were incubated with an α -NGF-receptor antibody and a secondary antibody coupled to magnetic beads and selected on a suitable column. Non-induced cells were sorted for NGF-receptor negative cells. Three rows show FACS data 0, 3, and 6 h after doxycycline induction. Living cells were identified by forward and sideward scatter criteria and their percentages in the gates are indicated. GFP signals are shown on the x-axis in combination with sideward scatter data. Population statistics of GFP-positive and -negative cells are provided. Prior to addition of doxycycline, GFP expressing cells were scarce. Upon induction for 3 and 6 h, GFP-positive cells constituted about 12 and 33 % of all cells, respectively. Upon sorting the majority of cells were GFP-positive. Cell sorting did not detectably affect cell viability.

The aim was to isolate RNA only from cells, which expressed BZLF1 (Figure 3-12). The fraction of GFP positive cells increased after sorting from 12 and 33 % to 94 and 99 % when induced for 3 and 6 h, respectively.

Subsequently, RNA was extracted from the sorted cells and reverse transcribed to cDNA. After library preparation, single-end sequence reads were obtained. For data visualization the RLE was calculated and the data were depicted as MA plots. To calculate the RLE, the medians of the natural-logarithm-transformed (ln-transformed) read counts of the 0 h probe set were subtracted from the ln-transformed read counts of every gene of the six samples. The results were depicted as boxplots, whose setup is described in detail in Figure 9-1 (page 182). The MA plots show the log₂ fold change over the mean of normalized read counts between the two analyzed time points of each individual gene.

3.5.2.1 Comparison of the transcriptomes of DG75 cells expressing full-length BZLF1 (4816): 3 hours versus non-induced cells

The DG75 4816 cells were analyzed 3 h after induction to restrict the analysis to cellular genes that are directly regulated by BZLF1. The results are shown in Figure 3-13.

When compared with the raw data, the DESeq2 normalization centers the median of the boxplots around zero as expected (panel A). It is notable that the spread of the boxplot's interquartile range (IQR) is wider at 3 h post induction than in non-induced cells (0 h).

The MA plot in panel B distinguishes between non-significantly regulated genes (grey dots) and significantly regulated genes (red dots) according to the adjusted p-value < 0.05. Red triangles represent values, which exceed the figure margins and can therefore not be visualized as a dot. Additionally, only genes with more than 20 read counts per gene were

taken into consideration for analysis. Furthermore, genes up-regulated < 2.5 fold or down-regulated > 0.4 fold were excluded from the analysis to make sure that only strongly regulated genes are used for further data processing. Using these criteria only 2 genes were up-regulated and 5 genes down-regulated in DG75 cells 3 h post BZLF1 induction.

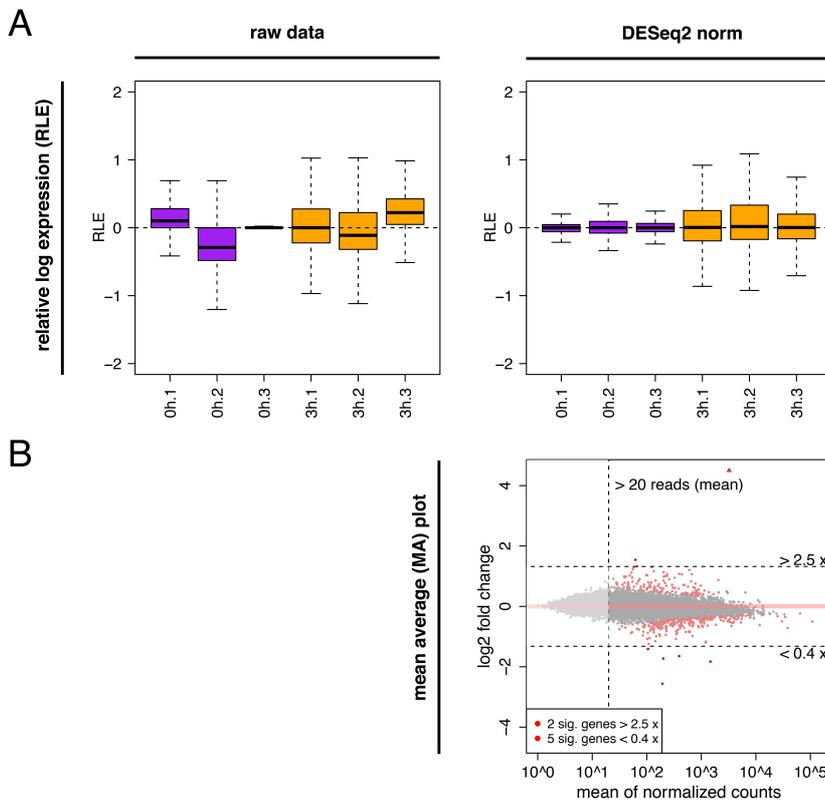


Figure 3-13: Data analysis of differentially expressed genes in DG75 4816 cells 3 h after BZLF1 induction

The depicted data from DG75 4816 cells are based on next generation RNA-sequencing prior to and 3 h after doxycycline-induced BZLF1 expression. The figure visualizes the data in two distinct modes comparing raw with normalized data. The two categories of data visualization are “relative log expression” (RLE) and “mean average (MA) plot”. “Raw data” (left panel) is compared with “DESeq2 normalization” (right panel). **A:** For RLE visualization, the medians of the ln-transformed read counts of the 0 h probe set is subtracted from the ln-transformed read counts of every gene of the six samples (0 and 3 h). The results are depicted as boxplots, their setup is explained in detail in Figure 9-1 (page 182). The purple boxplots show the 0 h samples, the orange boxplots show the samples, which were induced for 3 h. The

medians of the boxplots based on raw reads are chaotically distributed around zero, while the medians of the boxplots after DESeq2 normalization are centered on zero. The boxplots depicting the values 3 h after induction show a wider spread of the interquartile range (IQR is the colored box of the boxplot: for an accurate explanation see Figure 9-1) than the 0 h boxplots. **B:** While the RLE plots show six individual data sets, the MA plot summarizes all samples in a single plot. The MA plot visualizes the log2 fold change (y-axis) of gene expression between two time points (0 and 3 h) for every single gene. The x-axis shows the mean of normalized read counts per gene from three experiments. The grey dots represent genes, which were not significantly regulated, while the red dots indicate genes with an adjusted p-value below 0.05. The vertical, dashed line indicates genes with fewer than 20 read counts, which were excluded from the analysis. The two horizontal, dashed lines separate genes, which are regulated < 2.5 and > 0.4-fold and which were excluded from the analysis. The legend indicates the number of genes, which exceed these two limits and are up- or down-regulated 3 h after BZLF1 induction.

3.5.2.2 Comparison of the transcriptomes of DG75 cells expressing full-length BZLF1 (4816): 6 hours versus non-induced cells

An analogous analysis was performed with cells harvested 6 h after induced BZLF1 expression. The results are summarized in Figure 3-14.

Similar to the results obtained in the RLE analysis 3 h after BZLF1 induction the medians of the boxplots at 6 h are centered at around zero after DESeq2 normalization (panel A). The boxplots of the 6 h samples showed a wider spread of the IQR than the 0 h boxplots but also the slightly wider spread of the IQR than seen for the 3 h boxplots in Figure 3-13.

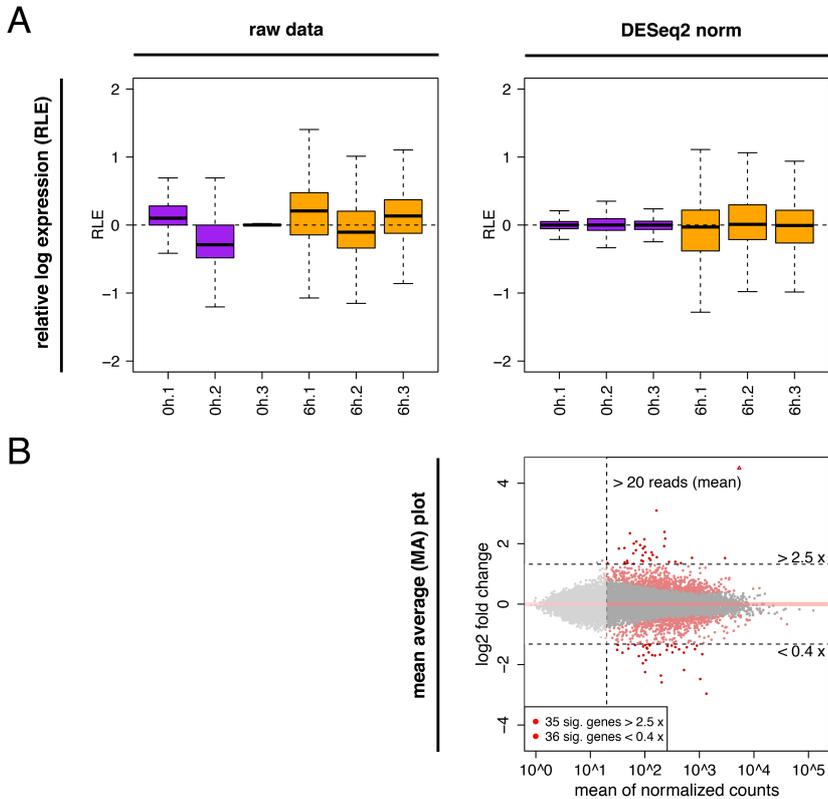


Figure 3-14: Data analysis of differentially expressed genes in DG75 4816 cells 6 h after BZLF1 induction

The depicted data from DG75 4816 cells are based on next generation RNA-sequencing prior to and 6 h after doxycycline-induced BZLF1 expression. The figure is composed identically as Figure 3-13. **A:** 6 h after induction the medians of the boxplots based on raw reads were unevenly distributed around zero in the RLE plots. After DESeq2 normalization the medians of the boxplots were centered closely around zero. A wider spread of the IQRs was visible in the boxplots 6 h after BZLF1 induction when compared with 0 h boxplots. This pattern is very similar to the distribution seen in Figure 3-13 when BZLF1 was expressed for 3 h. **B:** In the MA plot analysis 35 genes were significantly up-regulated (> 2.5 fold) and 36 genes were down-regulated (< 0.4 fold) in DG75 4816 cells 6 h after BZLF1 induction.

In the MA plot (panel B) about 70 genes were identified, which comply with the set criteria explained in (Figure 3-13). About half of the identified genes were up- (35) or down-regulated (36). After 6 h of induction many more genes were significantly regulated and the average distance from zero of the majority of genes increased when compared with the MA plot shown in Figure 3-13 (panel B).

Table 9-3 in the appendix provides a list of the 25 strongest up- and down-regulated genes 6 h after doxycycline-induced BZLF1 expression in DG75 4816 cells.

3.5.2.3 Comparison of the transcriptomes of doxycycline-induced parental DG75 cells: 6 hours versus non-induced cells

Doxycycline could have adverse effects on the cellular transcriptome. To analyze if the addition of doxycycline has such consequences, the parental DG75 cells were treated with doxycycline for 6 h and the transcriptome was analyzed as described in the two previous chapters (3.5.2.1 and 3.5.2.2). The results are summarized in Figure 3-15.

As seen before in the RLE analysis, the medians of the boxplots centered at around zero after DESeq2 normalization (panel A). The spread of the boxplot’s IQR 6 h after doxycycline addition was wider when compared to the 0 h sample but small when compared with the IQR’s spread observed 6 h after full-length BZLF1 expression (Figure 3-14).

As expected the MA plot (panel B) showed no significantly up- or down-regulated genes with the set criteria explained in Figure 3-13 suggesting that doxycycline did not have a measureable effect on the transcriptome of parental DG75 cells.

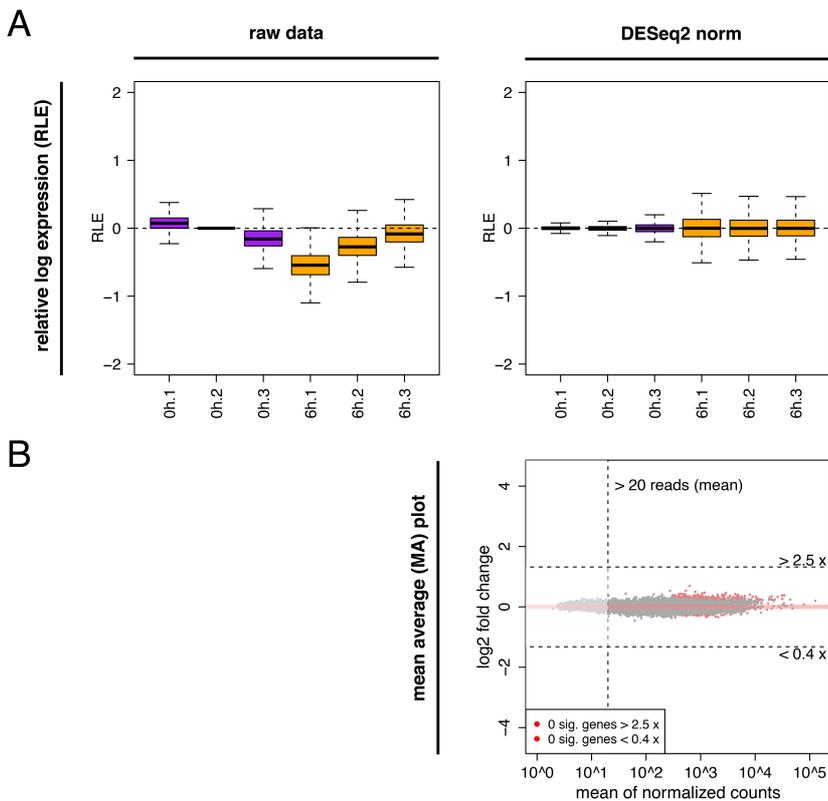


Figure 3-15: Data analysis of differentially expressed genes 6 h after doxycycline addition in parental DG75 cells.

The depicted data are based on next generation RNA-sequencing with parental DG75 cells prior to and 6 h after doxycycline addition. The parental DG75 cell line was used as a negative control to exclude adverse effects on cellular gene regulation upon addition of doxycycline. Parental DG75 cells do not harbor a doxycycline regulated expression plasmid. The figure is composed like Figure 3-13. **A:** Prior to doxycycline addition, the medians of the boxplots representing the raw counts arrange around zero. 6 h after doxycycline was added, the medians of all three samples are located below zero. After DESeq2 normalization, the medians of all six samples center on zero. The spread of the IQR from 6 h samples is wider compared with samples prior to addition of doxycycline (0 h values). **B:** No

cellular gene was induced or inhibited 6 h after the addition of doxycycline with the set criteria explained in Figure 3-13.

3.5.2.4 Comparison of the transcriptomes of DG75 cells expressing an activation domain truncated BZLF1 (5694): 6 hours versus non-induced cells

BZLF1 is a known transcriptional activator in the viral genome and thus it was unexpected to identify down-regulated genes in RNA-seq experiments with the full-length BZLF1 protein. BZLF1 could compete for binding sites with other transcription factors such as AP-1 family members and thus down-regulate certain genes. To investigate this hypothesis an AD-tBZLF1 mutant was employed, which binds to all BZLF1's cognate motifs but does not transactivate target genes in EBV's genome (Flemington et al., 1992). Apart from using the mutant BZLF1, the experimental setup was the same as for the full-length BZLF1 protein induced for 6 h. The results are summarized in Figure 3-16.

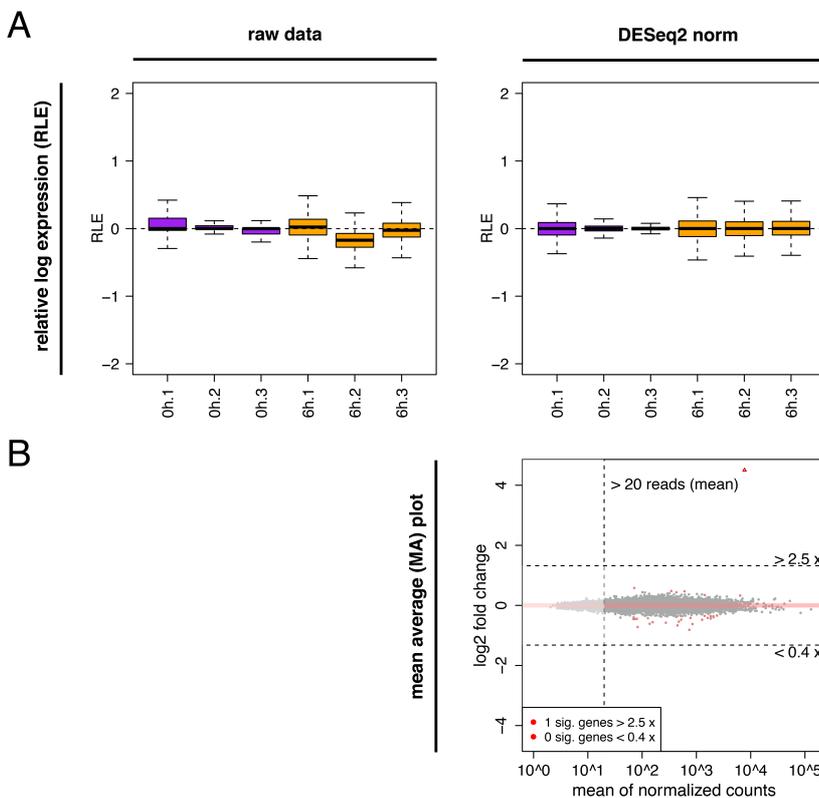


Figure 3-16: Data analysis of differentially expressed genes prior to and 6 h after induced expression of a truncated BZLF1 lacking its transactivation domain in DG75 5694 cells.

The depicted data result from DG75 5694 cells after next generation RNA-sequencing prior to and 6 h after doxycycline-induced expression of the BZLF1 mutant lacking the transactivation domain (AD-tBZLF1). The figure is composed like Figure 3-13. **A**: Prior to induced expression of AD-tBZLF1, the medians of the boxplot representing the raw reads center on zero. Six h post induction, the medians of two of the three samples do not deviate much from zero, but the spread of the IQRs is more narrow than in Figure 3-14, in which full-length BZLF1 was expressed for 6 h. Following DESeq2 normalization, all medians of the boxplots center on zero with an about uniform and narrow spread of the

IQRs. **B**: The single up-regulated gene identified in the MA plot visualization is the NGF-receptor, whose expression reflects its induction from the 5694 plasmid upon addition of doxycycline. It is depicted as a triangle, which indicates that it exceeds the figure margins.

Similar to all previously shown RLE data the DESeq2 normalization centered the medians of the boxplots on zero (panel A). The IQR's spread of the 6 h samples with the AD-tBZLF1 mutant was much smaller than the one observed for full-length BZLF1 (Figure 3-14). Very few genes were significantly up- or down-regulated in the MA plot analysis (panel B) and

only one gene is up-regulated more than 2.5 fold. This gene is the NGF-receptor and its expression originates from the plasmid, which also expressed the AD-tBZLF1 mutant. None of the genes exceeded the threshold defining down-regulated genes (< 0.4 fold).

The experiments with the AD-tBZLF1 mutant showed that BZLF1’s activation domain is essential and indispensable to induce measurable changes in cellular gene transcription.

3.5.3 The expression of BZLF1 in Raji 4816 cells results in massive down-regulation of cellular transcripts.

Identical to the experiments with DG75 4816 cells described in chapter 3.5.2, Raji 4816 cells were induced for 3 and 6 h and NGF-receptor positive cells were enriched by cell sorting. As a reference non-induced Raji cells were sorted for NGF-receptor negative cells.

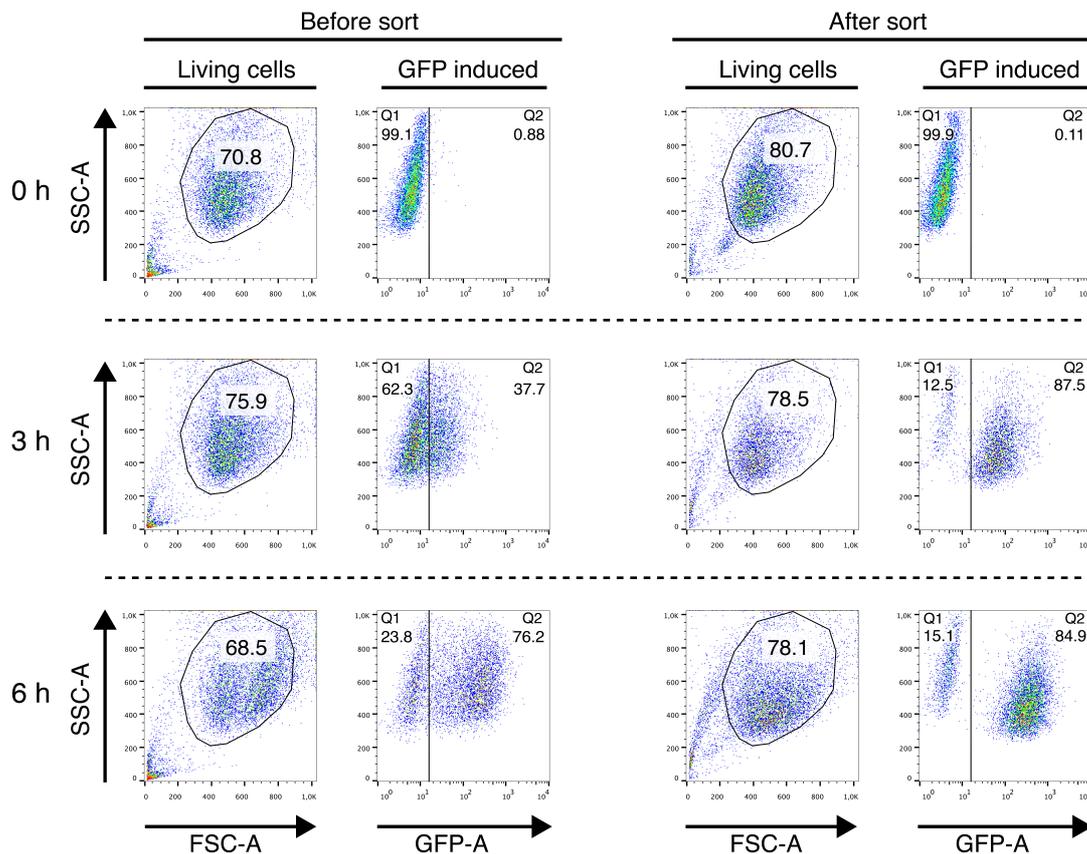


Figure 3-17: Enrichment of doxycycline-induced cells with magnetic beads and an α-NGF-receptor antibody.

The figure shows FACS data of Raji 4816 cells prior to (left) and after (right) sorting of NGF-receptor positive cells. The experimental setup and the composition of the figure are explained in the legend of Figure 3-12. Prior to doxycycline addition, hardly any cells expressed GFP. Three and 6 h after BZLF1 induction about 38 and 76 % of the living cells were GFP positive. After sorting, the majority of the cells (88 and 85 % after 3 and 6 h, respectively) showed a GFP-positive signal.

The FACS data prior to and after BZLF1 induction are depicted in Figure 3-17. After the magnetic beads sort at least 78 % of the cells were alive and at least 85 % GFP-positive.

3.5.3.1 ERCC spike-in RNAs for monitoring differential RNA expression experiments

In preliminary experiments not included here, Raji cells were analyzed after induction with BZLF1 for 3 and 6 h. The aim was to identify the cellular genes, which BZLF1 was expected to induce, but only few up-regulated cellular genes were identified. Several reasons could account for this failure, but we hypothesized that BZLF1 might alter the majority of cellular genes such that standard data normalization would eliminate such drastic changes in the transcriptomic data.

In the following experiments, an additional RNA-based control was added to address this potential problem and to analyze the data properly. ERCC RNAs are commercially available and ideally suited for data normalization to control for RNA-sequencing experiments. Exactly 92 polyadenylated artificial RNAs of different concentrations, which are added prior to RNA extraction and library preparation, serve as internal standard for subsequent data normalization. Figure 3-18 illustrates why data normalization can be problematic and how spike-in RNAs solve this problem.

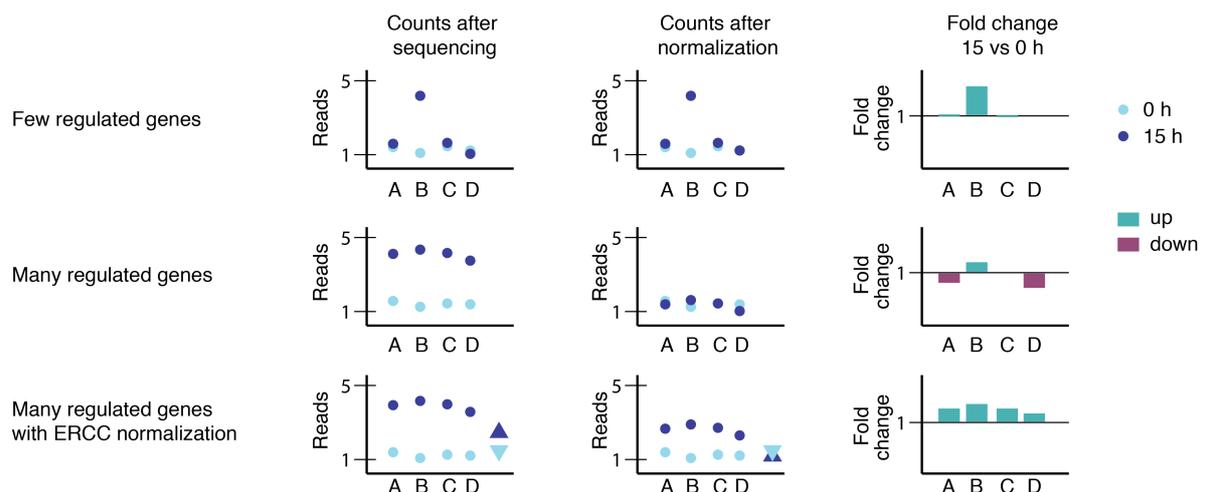


Figure 3-18: Normalization without and with ERCC control RNAs on samples with global gene regulation. The figure shows different extremes of gene regulation to explain why ERCC normalization is required. Bioinformatics programs can easily detect regulated genes, if only few genes are regulated (“Few regulated genes”). If the majority of genes is affected (“Many regulated genes”) bioinformatic programs are misled and “normalize” the regulated genes eradicating the effects of drastic gene regulation. With the help of ERCC spike-ins the normalization is achieved independently of cellular genes. Only the ERCC RNAs (“Many regulated genes with ERCC normalization”), which are equally added to all samples, are used for normalization. Cellular genes

are repositioned relative to the control RNAs and reveal their real regulation. (Figure adapted from Lovén et al., 2012).

In many biological models, only few genes change their expression levels upon a certain regulating event. Data normalization relies on this assumption because the global transcript level is considered to be stable on average when comparing one data set with another. Few, strongly regulated genes can be reliably detected with this approach. If the signal causes a global change in transcript levels, conventional data normalization cannot cope with this situation but introduces a severe bias. After data normalization the two data sets appear to be similar because the algorithm eliminates the global change. As a consequence the global change will go unnoticed and up- and down-regulated genes will not be detected.

In contrast, with externally added ERCC RNAs, the bioinformatic scripts use the ERCC control RNAs to normalize and correct cellular gene reads accordingly. The cellular genes are normalized relative to the ERCC RNAs and reveal their real level or regulation.

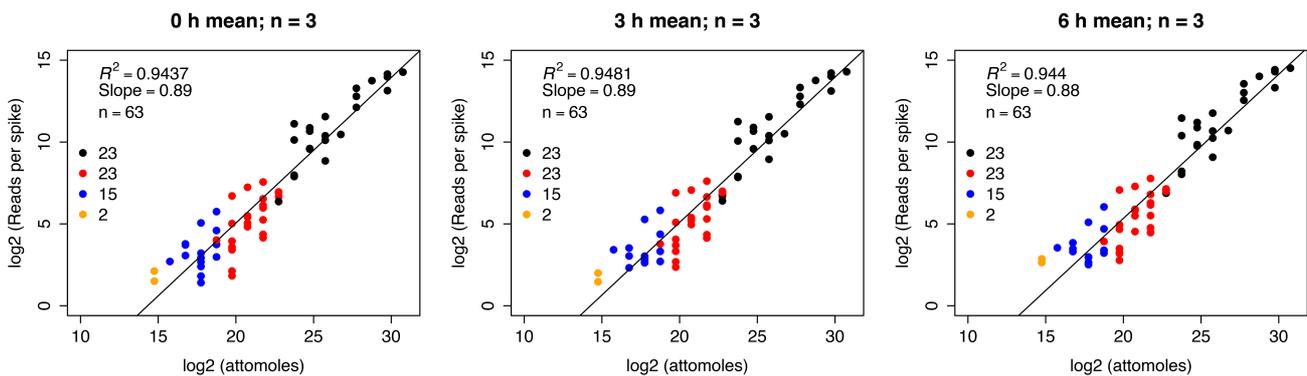


Figure 3-19: Visualization of ERCC spike-in control RNAs at different time points

Ninety-two synthetic ERCC RNAs of different concentrations were added to RNA samples obtained from the identical numbers of cells to control global changes of RNA levels in the Raji 4816 cell line. Based on the reads from the ERCC spike-in controls, reads obtained from cellular and viral genes were normalized accordingly. The figure shows the means of the reads of each of the detected and mapped ERCC spike-in control RNAs from three experiments. The three graphs show the mean of ERCC reads from cells prior to BZLF1 induction (0 h) and cells induced for 3 and 6 h, respectively. The x-axis shows the amount of attomoles RNAs added to the samples on a log2 scale. Mean reads per individual ERCC spike-in RNA after next generation RNA-sequencing are displayed on the y-axis on a log2 scale. In the legend three values are given. The coefficient of determination is shown as R^2 , the slope of the regression line is indicated, and the third line indicates numbers of individual spike-in RNAs out of 92 total RNAs, which could be identified. The 92 ERCC spike-in RNAs are shown in four groups with 23 RNAs each, which differ in their concentration level. The four colored dots are ordered by black-red-blue-yellow indicating the group of concentration, in which a yellow dot indicates control RNAs with the lowest concentrations. All RNAs of the two groups with the highest concentrations could be identified in all samples, but only 15 and 2 RNAs were found in the third (blue) and fourth (yellow) class of RNAs, respectively, with lower and lowest concentrations. The figure indicates that ERCC spike-in controls were detected and can be quantified according to their individual molar concentration and in a dose-dependent fashion. The results indicate that the technical settings justify data normalization according to the chosen criteria. The data also provide the lower levels of detection, which reflects the overall read coverage.

In the experiments with Raji cells the same amount of control RNAs was added to the same number of non-treated cells and cells induced for 3 and 6 h with doxycycline. Following the ERCCs addition, RNAs were extracted from the different samples, processed to libraries and sequenced. Figure 3-19 shows the distribution of the ERCCs in samples induced with BZLF1 for 0, 3, and 6 h. The spiked attomoles of ERCC RNAs were plotted against the reads per spike detected during sequencing. In all three replicas 63 of the 92 spike-in ERCC RNAs could be identified. In Figure 3-19 the 92 ERCCs are shown in four groups comprising 23 RNAs each. Black dots represent RNAs with the highest concentration, while yellow shows the RNAs with the lowest concentration. Red and blue represent intermediate states. All ERCC RNAs of the two groups comprising RNAs with the highest concentration could be detected. Still 15 control RNAs could be detected for the blue group and two RNAs for the group, which contains the RNAs with the lowest concentrations. The plots show that the chosen sequencing depth detected a sufficiently large number of ERCC RNAs in all three replicas indicating that the ERCC RNAs can be used for read normalization.

As described for the DG75 cell lines, normalization was performed with DESeq2. Instead of all genes, the ERCC spike-in reads were used for normalization and the reads of the cellular genes were corrected accordingly following the normalized ERCC spike-in RNAs. [Tamas Schauer](#) provided the basic script, which contained the code used for ERCC-based normalization (personal communication) with the DESeq2 algorithm.

3.5.3.2 Comparison of Raji 4816 transcriptomes after ERCC controlled expression of full-length BZLF1: 3 hours versus non-induced cells

Figure 3-20 summarizes the results 3 and 0 h after full-length BZLF1 expression in three categories: PCA, RLE, and MA plots. Each of the categories shows three different conditions: raw data, global normalization via DESeq2, and ERCC-based normalization. The PCA is a method to reduce the amount of dimensions for data visualization and uses the first principal component (PC 1) and the second principal component (PC 2) here to visualize the variance between the six samples. PC 1 shows the largest variance between samples, PC 2 the second largest. Prior to normalization the 3 h samples cluster in a group in PCA (panel A), while 0 h samples were widely distributed. After DESeq2 and ERCC normalizations the samples of two different time points cluster in two groups on the axis of the PC 1.

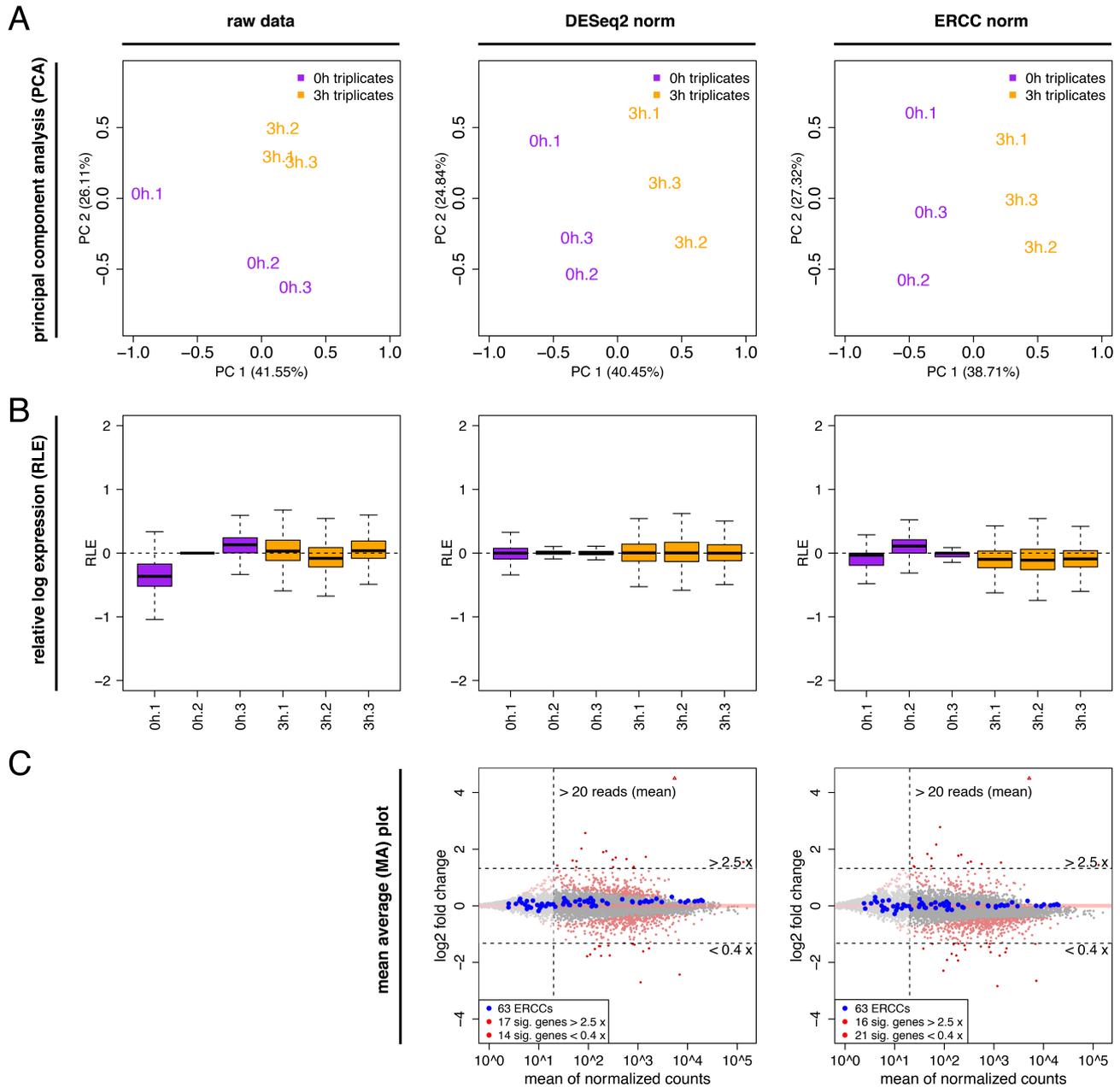


Figure 3-20: Data analysis of differentially expressed genes 3 h post induction of BZLF1.

The depicted data from Raji 4816 cells are based on next generation RNA-sequencing prior to and 3 h after doxycycline-induced BZLF1 expression. The figure visualizes the data in three different ways and compares raw data with DESeq2 and ERCC spike-in normalized (“ERCC norm”) data. The three categories of data visualization are “principle component analysis” (PCA), RLE and MA plots. “Raw data” (left panel) is compared with “DESeq2 normalization” (middle panel) and “ERCC normalized” data (right panel). **A**: The PCA summarizes the data of the six samples in a 2D plot, which is defined by the first (PC 1) and second (PC 2) principal component (PC). PC 1 comprises the largest, PC 2 the second largest variance between the samples. The axes of the plots are unitless, but the denomination of the axes shows the percentage of variance between the samples explained by the PCs. Prior to normalization, the 3 h samples cluster closely together, while the 0 h samples are distantly located. After DESeq2 and ERCC normalization, the 0 and 3 h samples build clusters, which are clearly separated from each other. The origin and setup of the RLE and MA plots are explained in detail in Figure 3-13. **B**: In the RLE plot of the raw data the medians of 0 h samples were located at around zero, but are widely distributed. 3 h after BZLF1 induction the medians were located much closer at zero. Post DESeq2 normalization the medians of all samples were located on zero and the spreads of the boxplots’ IQRs of the 3 h samples were wider. Post ERCC normalization the 0 h samples were closely distributed around zero but

the medians of all 3 h boxplots were located slightly below the zero line. **C**: ERCC RNAs are depicted as blue dots. In MA plot analysis and after DESeq2 normalization the ERCC control RNAs were clearly located above zero. Seventeen and 14 genes were up-regulated more than 2.5 fold and down-regulated less than 0.4 fold, respectively. After ERCC normalization the control RNAs centered at zero. As a consequence all cellular genes were shifted slightly downwards. Subsequently, 16 and 21 genes were identified to be up- and down-regulated, respectively.

After DESeq2 normalization the medians of the boxplots centered on zero in RLE analysis (panel B). After ERCC normalization the 0 h samples positioned around zero, while the medians of all three 3 h sample boxplots position slightly, but clearly below zero.

The DESeq2 normalization handles the transcriptome data from Raji cells as described previously in chapter 3.5.2.1 with data obtained from DG75 cell lines. The ERCC RNAs were ignored during the normalization procedure, but still included in the plot (Figure 3-20, panel C). They showed a slight tendency to be up-regulated 3 h after induction. The identical molar mass of ERCC RNAs was added to all samples indicating that the cellular genes were down-regulated on average. This bias becomes obvious in the MA plot after ERCC normalization. ERCC RNAs were used for data normalization and became centered on zero on average (Figure 3-20, panel C, right). As a consequence, cellular genes were shifted during the process of data normalization. Comparing genes, which were normalized via DESeq2 or via the ERCC spike-in RNAs revealed this shift in data points (Figure 3-20, panel C, right). This shift is noticeable when up- and down-regulated genes were compared after 3 h of BZLF1 expression: Seventeen and 14 genes were up- and down-regulated after DESeq2 normalization, while 16 and 21 genes were up- and down-regulated after ERCC normalization with the set criteria explained in Figure 3-13.

3.5.3.3 Comparison of Raji 4816 transcriptomes after the ERCC controlled expression of full-length BZLF1: 6 hours versus non-induced cells

Three hours after addition of doxycycline and induction of BZLF1 the effect of ERCC RNAs on adequate data normalization was barely evident (Figure 3-20) but 6 h after doxycycline-induced BZLF1 expression the advantage of ERCC spike-in RNAs as an external standard became impressive (Figure 3-21).

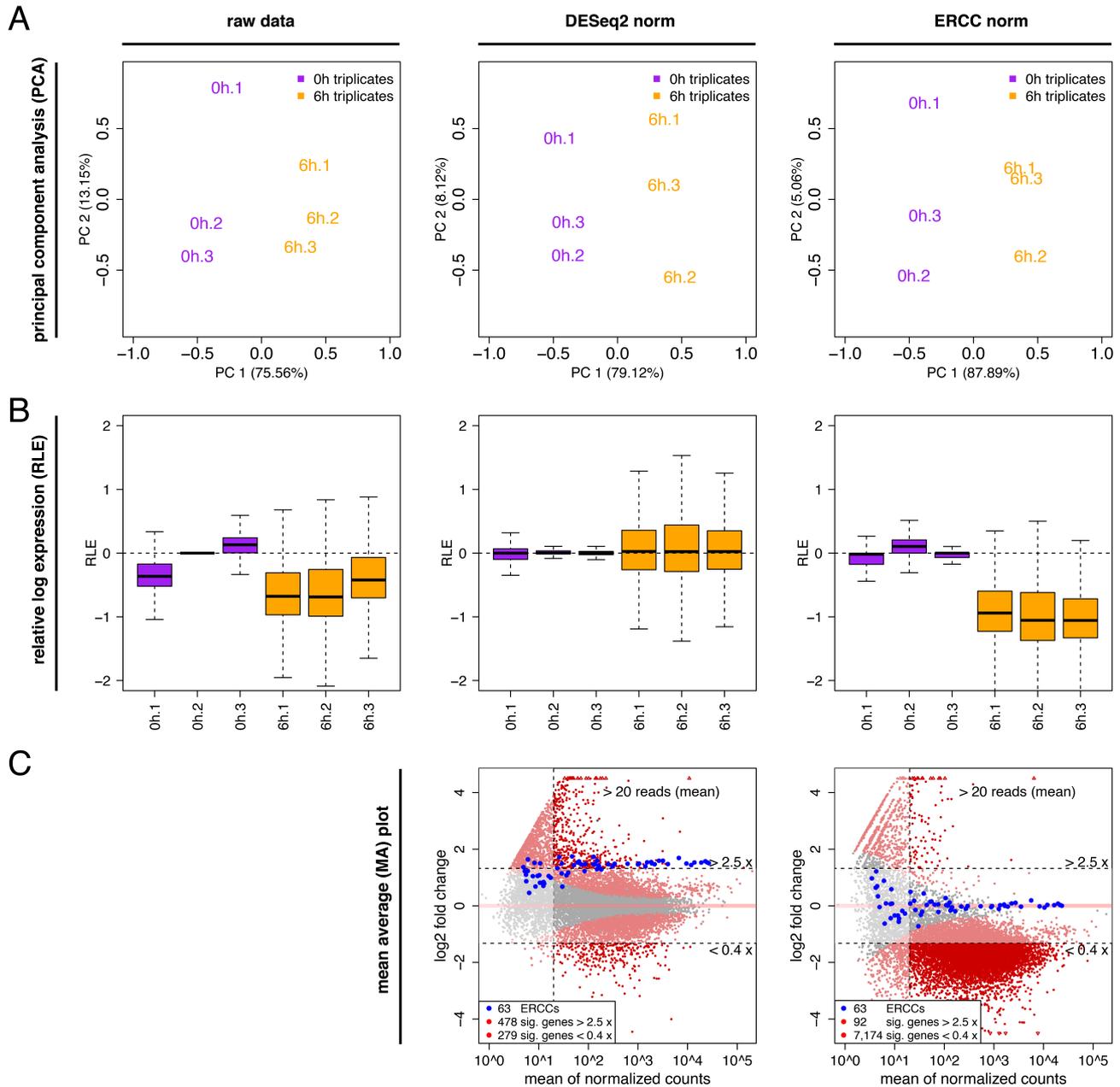


Figure 3-21: Data analysis of differentially expressed genes 6 h post induction of BZLF1

The depicted data from Raji 4816 cells are based on next generation RNA-sequencing prior to and 6 h after doxycycline-induced BZLF1 expression. The overall composition of the figure was explained in Figure 3-20. **A**: Prior to and after both normalization steps, the distribution of 0 h and 6 h samples was very similar in PCA. In all three plots the time points clustered in two groups. Up to 88 % of the variance between the samples could be explained by PC 1. PC 2, showing the differences within the same time points, explained only 5 % of the variance after ERCC normalization. **B**: Prior to normalization the medians of the boxplots of the 0 h samples were widely distributed around zero, while the medians of the 6 h samples were located below zero. After DESeq2 normalization, the medians of all samples were on or close to zero. After ERCC-bases data normalization the medians of the 0 h boxplots were centered closely at around zero, while the medians of 6 h samples were all close to -1. The 6 h data of all samples depicted far wider IQR spreads than 0 h samples. **C**: After DESeq2 normalization the ERCC control RNAs centered just above the dashed line in the MA plot indicating that ERCC RNA reads appeared to be up-regulated more than 2.5 fold. 478 genes appeared up- and 279 genes down-regulated 6 h after BZLF1 induction. Post ERCC normalization control RNAs were centered on zero while the cellular genes were shifted downwards and adjusted accordingly. After ERCC normalization 92 and 7,174 genes were up- and down-regulated, respectively.

Already prior to normalization the 0 and 6 h samples clustered to two groups in the PCA. Even though the positions of the samples changed slightly, the two groups persisted. Prior to normalization the boxplots of the 0 h samples positioned at around zero in the RLE analysis, while the 6 h boxplots were positioned obviously deeper. The DESeq2 normalization centered the medians of the boxplot on zero as expected, but data normalization on the basis of ERCC RNAs yielded a very different outcome: The medians of the 0 h boxplots were slightly shifted and positioned at around zero but the medians of the 6 h boxplots centered at about minus one, which illustrates a dramatic down-shift. In all conditions the spread of the IQRs was wider at 6 h post induction as compared with data of non-induced cells (0 h) (Figure 3-21, panel B).

In the MA plot analysis and after DESeq2 normalization about 500 genes were up- and 300 genes down-regulated. The ERCC RNAs seemed to be about 2.5 fold up-regulated indicating that DESeq2 normalization is inappropriate with this set of data. In contrast, ERCC normalization showed a realistic distribution of cellular genes. After the ERCCs RNAs were centered on zero the cellular genes were down-shifted according to the ERCC RNAs as expected. After normalization only 92 genes were up-regulated 6 h post induction of BZLF1 but more than 7,000 genes were identified to be down-regulated.

Table 9-4 in the appendix provides a list of the 25 strongest up- and down-regulated genes 6 h after doxycycline-induced BZLF1 expression in Raji 4816 cells.

3.5.3.4 Comparison of parental Raji transcriptomes after addition of doxycycline: 6 hours versus non-induced cells

Parental Raji cells were induced with doxycycline only to exclude gene regulation by the addition of doxycycline. No ERCC control RNAs were used in this experiment, which is summarized in Figure 3-22.

In PCA raw, unadjusted data and data after DESeq2 or ERCC normalization showed no obvious clustering (Figure 3-22, panel A). Independent of the normalization method used in RLE analysis, all the samples clustered at around or on zero (panel B). None of the cellular genes were found to be significantly regulated beyond a significance level of 0.05 in the MA plot analysis (panel C).

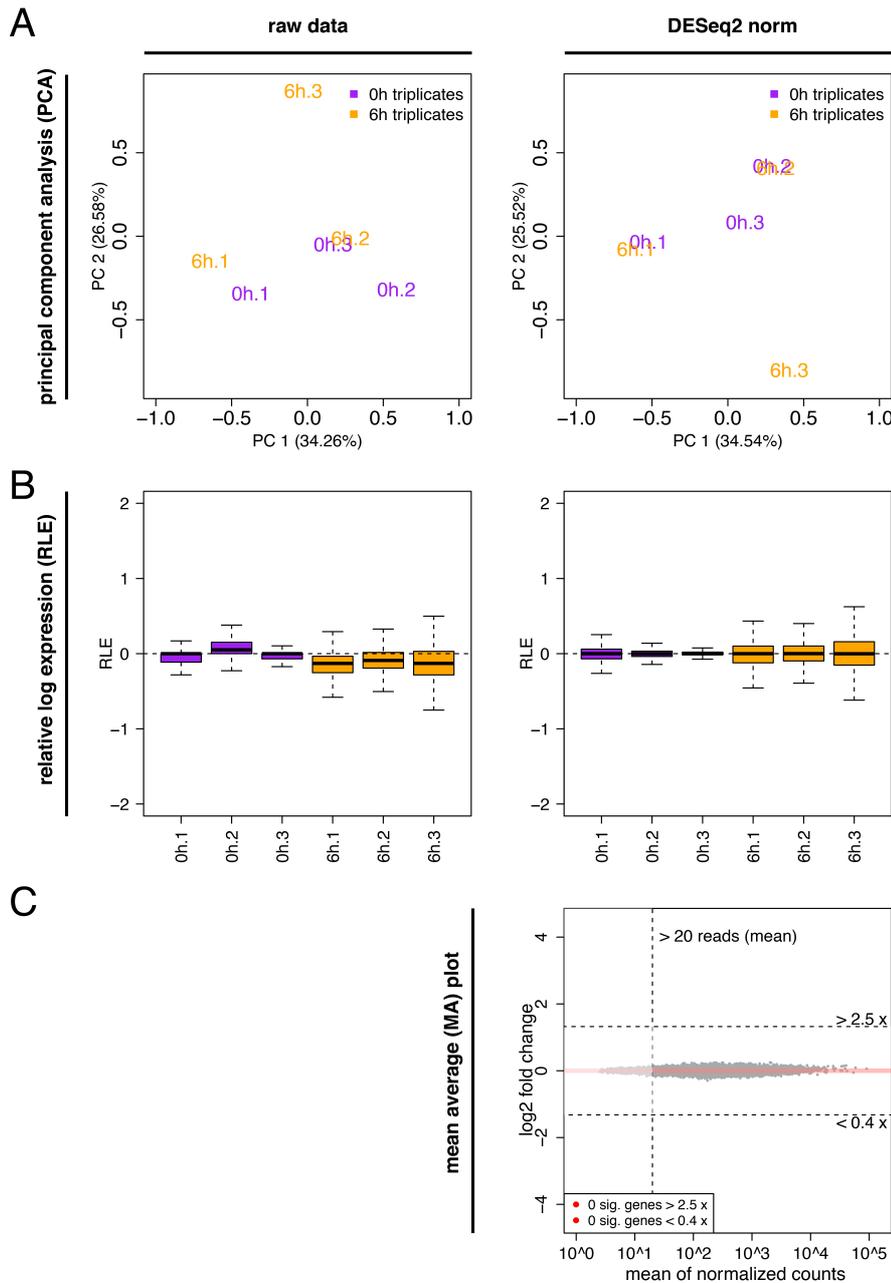


Figure 3-22: Data analysis of differentially expressed genes 6 h after doxycycline addition

The depicted data are based on next generation RNA-sequencing with parental Raji cells prior to and 6 h after doxycycline addition. The parental Raji cell line does not harbor a doxycycline regulated expression plasmid and serves as a control to exclude adverse effects on cellular gene regulation upon addition of doxycycline. The figure is composed like Figure 3-20 excluding the ERCC normalization. **A**: The samples in the PCA appeared randomly distributed except that samples of the same experiment frequently clustered together. **B**: In the RLE analysis the 0 h samples of the raw data clustered closely around zero, while the medians of the 6 h samples were all closely below zero. After DESeq2 normalization the medians of all boxplots centered on zero. **C**: In the MA plot analysis no regulated genes were identified by the given criteria.

3.5.3.5 Comparison of Raji 5694 cells expressing a BZLF1 mutant lacking the transactivation domain: 6 hours versus non-induced cells

The AD-tBZLF1 mutant was expressed for 6 h to exclude that the down-regulation of cellular genes was caused by BZLF1 competing for binding sites of cellular transcription factors. No ERCCs control RNAs were used in this experiment, which is summarized in Figure 3-23.

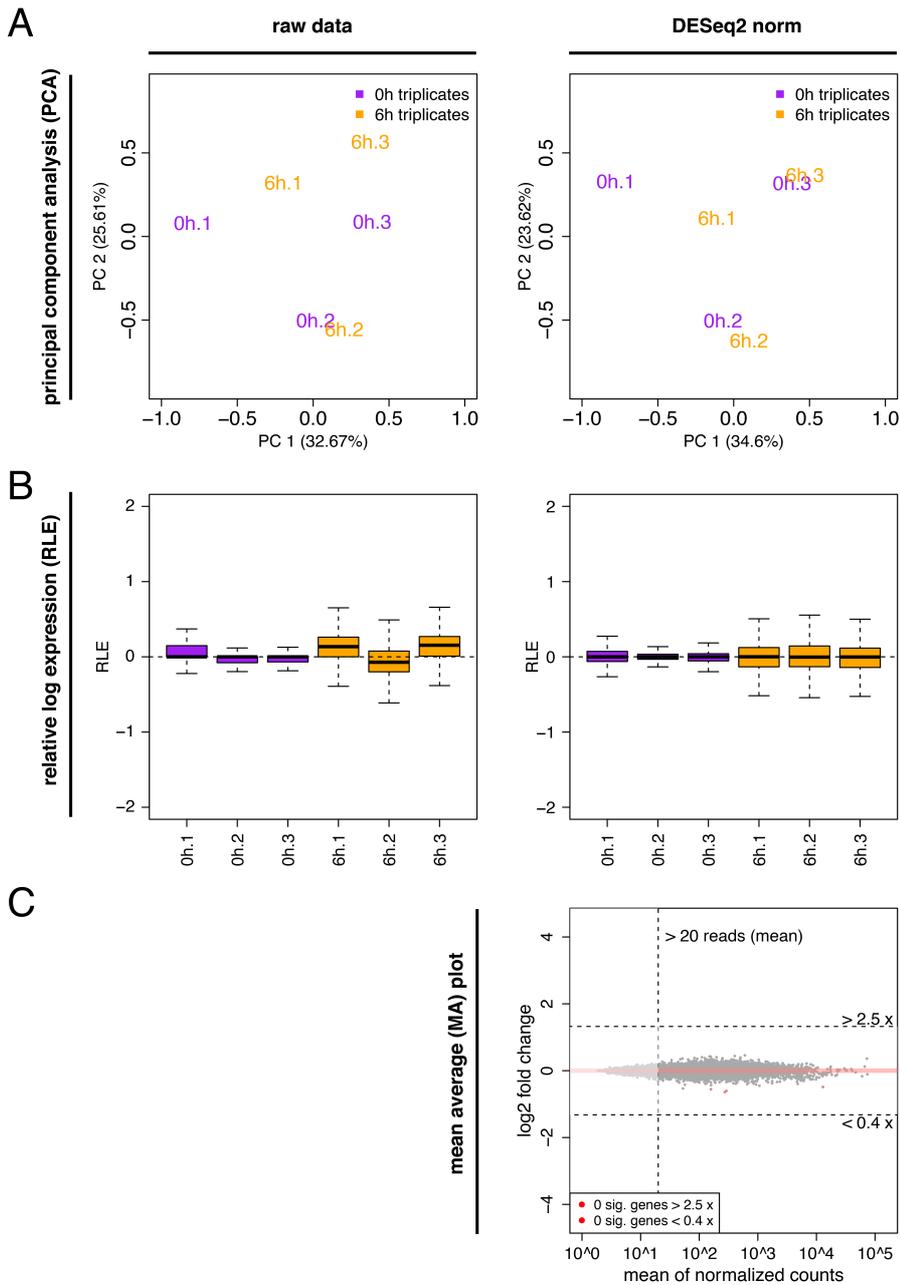


Figure 3-23: Data analysis of differentially expressed genes prior to and 6 h after induced expression of BZLF1 with a truncated activation domain.

The depicted data result from Raji 5694 cells based on next generation RNA-sequencing prior to and 6 h after doxycycline-induced expression of the BZLF1 mutants lacking the transactivation domain (AD-tBZLF1). The figure is composed like Figure 3-20, but lacks the ERCC normalization, since no ERCC RNAs were used. **A:** During all phases of normalization the samples were distributed without clear directions in PCA. **B:** In RLE analysis and prior to normalization 0 h samples clustered closely around zero, while 6 h samples were distributed a little further around zero. The DESeq2 normalized samples were all centered on zero. **C:** The DESeq2 normalization visualized as MA plot showed that none of the cellular genes were regulated by the set criteria explained in Figure 3-13.

In PCA prior to and after DESeq2 normalization the different time points of the same experiments mostly clustered together (Figure 3-23, panel A). In the RLE analysis the medians representing the raw data centered around zero (panel B) and after DESeq2 normalization the medians of the boxplot clustered on zero. After the DESeq2 analysis only very few genes were significantly down-regulated, but none exceeded the previously defined threshold described for up- and down-regulated genes in the MA plot analysis (panel C).

3.5.4 KEGG pathway analysis

The ERCC controlled experiments in Raji cells expressing the full-length BZLF1 showed that the RNA-seq results can only be trusted when normalized with the help of an external standard, the ERCC spike-in RNAs. A Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was conducted, which allows assigning regulated genes to cellular pathways. This was done after ERCC RNA-based normalization with data from Raji 4816 cells after 6 h of induced BZLF1 expression. A minimal numbers of regulated genes are required to run the KEGG analysis, which is neither reached for the 3 h samples nor for up-regulated genes 6 h after doxycycline addition. Therefore, the 3,300 most stable genes (Figure 3-24) and the 1,000 most strongly down-regulated genes (Figure 3-25) 6 h after BZLF1 induction were analyzed for KEGG pathway contribution. The two figures show gene numbers contained within the defined pathways and the numbers of genes of these pathways regulated in Raji 4816 cell after induction of 6 h.

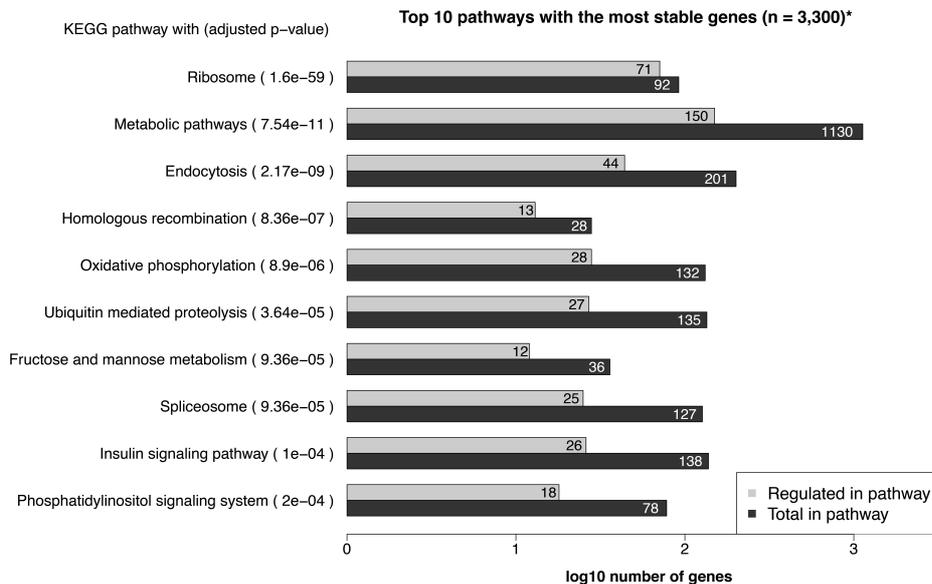


Figure 3-24: KEGG pathway analysis of the 3,300 most stable genes after induced expression of BZLF1.

The figure shows the top ten KEGG pathways, which were identified with the 3,300 most stable genes in Raji 4816 cells induced for 6 h. The genes originate from ERCC spike-in controlled RNA-seq data after normalization and are obtained from the Raji 4816 cell line shown in Figure 3-21. The pathways were identified with the help of the “webgestalt” tool. The x-axis depicts the number of genes on a log10 scale. The names of the pathways and the adjusted p-values are indicated. The number of genes not affected after BZLF1’s induced expression is indicated together with the total number of genes, which the given pathway comprises. The ten KEGG pathways shown in the figure suggest that BZLF1 does not regulate basic metabolic functions of the cells.

* Pathways related to distinct diseases were excluded from analysis.

The top ten pathways identified with the 3,300 most stable genes suggested that the basic metabolic functions of the cell were not affected after BZLF1 expression (Figure 3-24), whereas the 1,000 most strongly down-regulated genes comprise pathways related to immune defense or pathways important in immune cells (Figure 3-25).

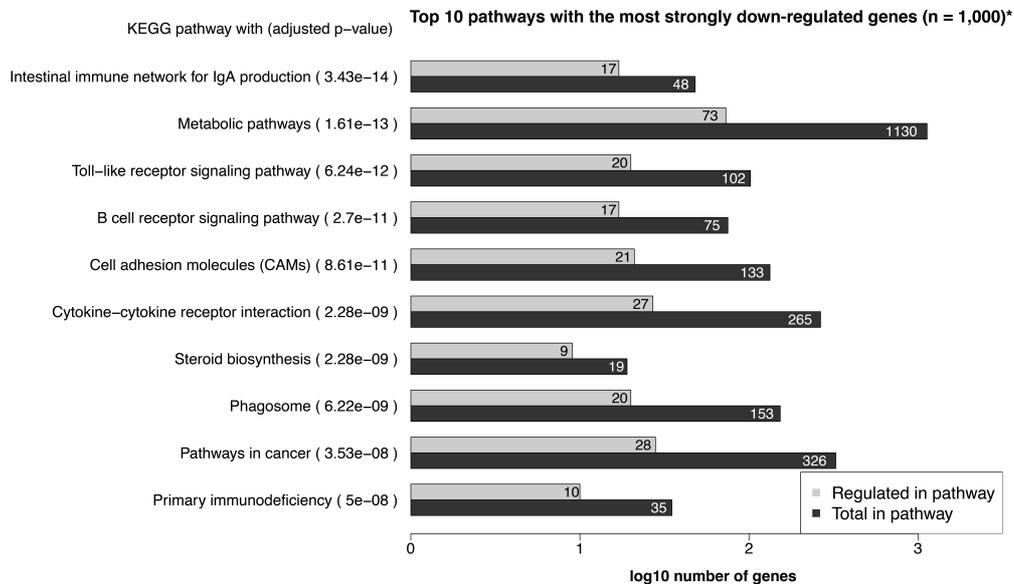


Figure 3-25: KEGG pathway analysis of the 1000 most strongly down-regulated genes

The figure shows the top ten KEGG pathways, which comprised the 1000 most strongly down-regulated genes in Raji 4816 cells 6 h after induction with doxycycline. Affected pathways included important functions in immune cells and host immune responses. * Pathways related to distinct diseases were excluded from analysis.

3.6 BZLF1 binding sites in promoter regions of regulated genes only explain a minority of cellular gene regulation

The analysis of the viral genome revealed that BZLF1 binds within the promoter region of those viral genes, which are up-regulated after induction of BZLF1 (Bergbauer et al., 2010). A similar mechanism was expected to exist in BZLF1 controlled cellular genes.

To test this hypothesis, ChIP-seq and RNA-seq data were searched for BZLF1 binding sites in cellular genes regulated after BZLF1 induction. Regulated genes were defined as genes with a mean expression > 20, an adjusted p-value < 0.05, and a log₂ fold change > 2.5 fold or < 0.4 fold. The promoters (-5/ +1 kbps from the TSS) of up- and down-regulated genes were searched for one or more BZLF1 peaks. The linkage of BZLF1 regulated genes in DG75 4816 and Raji 4816 cell lines and already identified BZLF1 binding sites was investigated with the help of the programming software R.

3.6.1 A minority of BZLF1-regulated genes has promoter regions with BZLF1 binding sites

Figure 3-26 shows the results after combining ChIP-seq and RNA-seq data. The plot highlights the significantly regulated genes ($p_{adj} < 0.05$) with a mean expression > 20 and the thresholds defining up-regulated (> 2.5 fold) and down-regulated (< 0.4 fold) genes. Genes to the left of the vertical dashed line have no identified BZLF1 peaks within their promoter region (-5/ +1 kbps from the TSS); genes with one or more BZLF1 binding sites are shown to the right of the vertical dashed line in Figure 3-26. Genes located in each one of the four areas are summarized in this figure. Cyan and magenta dots stand for up-regulated and down-regulated genes, respectively.

Panel A links genes regulated 3 h after BZLF1 expression with 230,213 peaks identified 15 h after doxycycline-induced BZLF1 expression. Only seven genes were defined as regulated and four of these contained a single peak within their promoter region. Panel B links the genes regulated 6 h after BZLF1 induction with the 230,213 peaks identified 15 h after BZLF1 induction. About half of the up-regulated genes comprised one or rarely more peaks, while only in one third of the promoters of the down-regulated genes a peak could be identified. Together the data indicate that BZLF1 is not consistently found in close neighborhood of the promoter regions of regulated genes suggesting that BZLF1 is not a genuine promoter factor of cellular genes.

Table 9-5 in the attachment provides a list of the 25 strongest up- and down-regulated genes, which have at least one BZLF1 binding site within their promoter in DG75 4816 cells 6 h after doxycycline-induced BZLF1 expression.

In Raji 4816 cells a corresponding analysis was performed, which included an additional comparison (Figure 3-27). Genes regulated 3 (panels A and C) and 6 h (panels B and D) after BZLF1 induction were not only linked to peaks identified 15 h (panels A and B) after BZLF1 induction, but also to peaks identified prior to doxycycline-induced BZLF1 induction (0 h, panels C and D).

When BZLF1-regulated genes after 3 h of induction were linked with peaks identified 15 h after induction (panel A) nearly half of the up- and also nearly half of the down-regulated genes contained at least one peak within their promoter regions. The same pattern was

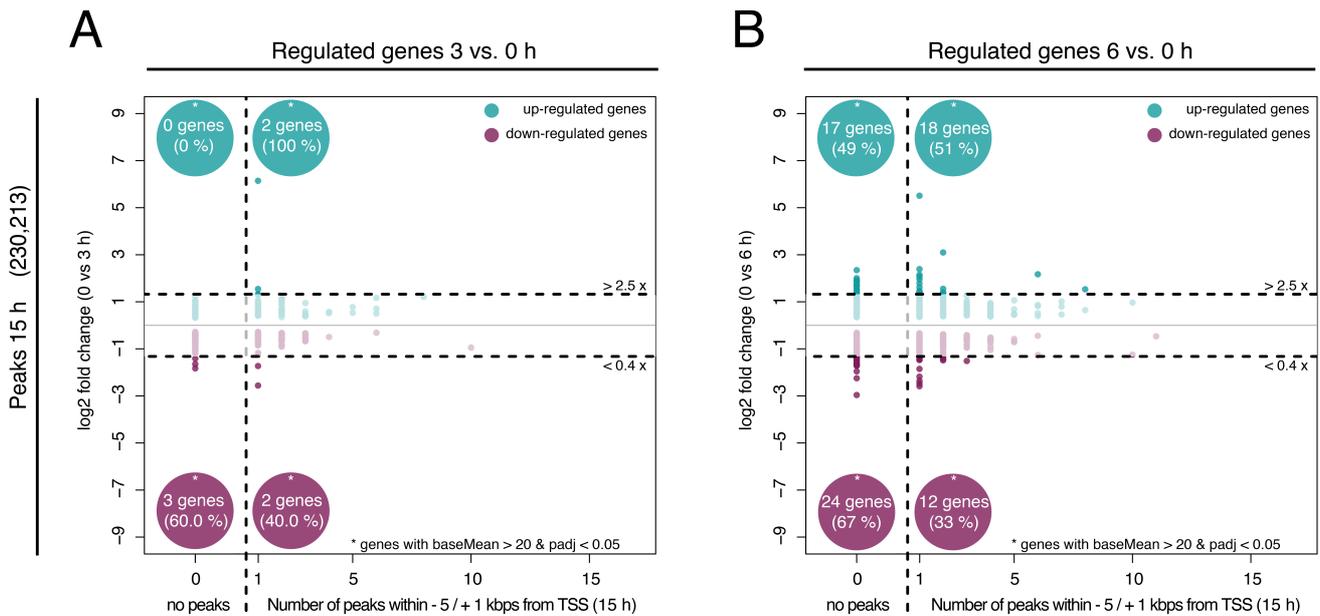


Figure 3-26: Linkage between ChIP-seq peaks of BZLF1 15 h post induction and BZLF1-regulated genes after induction for 3 and 6 h.

The two plots compare BZLF1 peaks identified in DG75 4816 cells induced for 15 h (230,213 peaks in total) with BZLF1-regulated genes 3 h (panel A) and 6 h (panel B) post induction. The comparison considers BZLF1-regulated genes that do not contain an identified BZLF1 site (“no peaks”) or regulated genes that contain at least one or more BZLF1 peaks in close proximity of the genes’ TSSs as indicated (“Number of peaks -5/ +1 kbps of TSS”). Only genes with a base mean of at least 20 RNA-seq reads and adjusted p-value < 0.05 were considered for analysis. Genes indicated with dimmed dots are significantly regulated upon induction, but their factors of regulation do not exceed 2.5 or are smaller than 0.4. These genes locate in a horizontal band together with their x-fold regulation indicated by the y-axis in log2 scale. Cyan and magenta dots indicate up- and down-regulated genes, respectively. The number of genes regulated is provided together with their relative abundance (%). **A:** Only very few genes were regulated in DG75 4816 cells 3 h after induction of BZLF1. **B:** about half of the up-regulated genes contain or do not contain identified BZLF1 peaks within the defined promoter region (-5/ +1 kbps of TSS). The majority of down-regulated genes (n = 24) do not contain an identified BZLF1 peak but a minority (n = 12) does.

identified for down-regulated genes when genes regulated for 6 h were linked to peaks identified 15 h after BZLF1 induction (Panel B). In contrast, in about 60 % of the up-regulated genes a peak could be identified within their promoter region. Counterintuitively, genes with more than one peak within their promoter region revealed a weaker regulation than many genes with only one or no BZLF1 peak (Figure 3-27, panel B). The linkage of genes regulated 3 h after BZLF1 induction with peaks identified prior to doxycycline induced BZLF1 expression (panel C) showed that more than 80 % of the 16 up- and 21 down-regulated genes had no identifiable peak within their promoter regions. Only three up- and three down-regulated genes with one peak within their promoter regions could be identified.

The linkage was also investigated at BZLF1-regulated genes 6 h after induction and BZLF1-related peaks identified prior to doxycycline-induced BZLF1 expression (0 h, panel D). About 90 % of the more than 7,000 down-regulated genes had no peak, but about one fourth of the

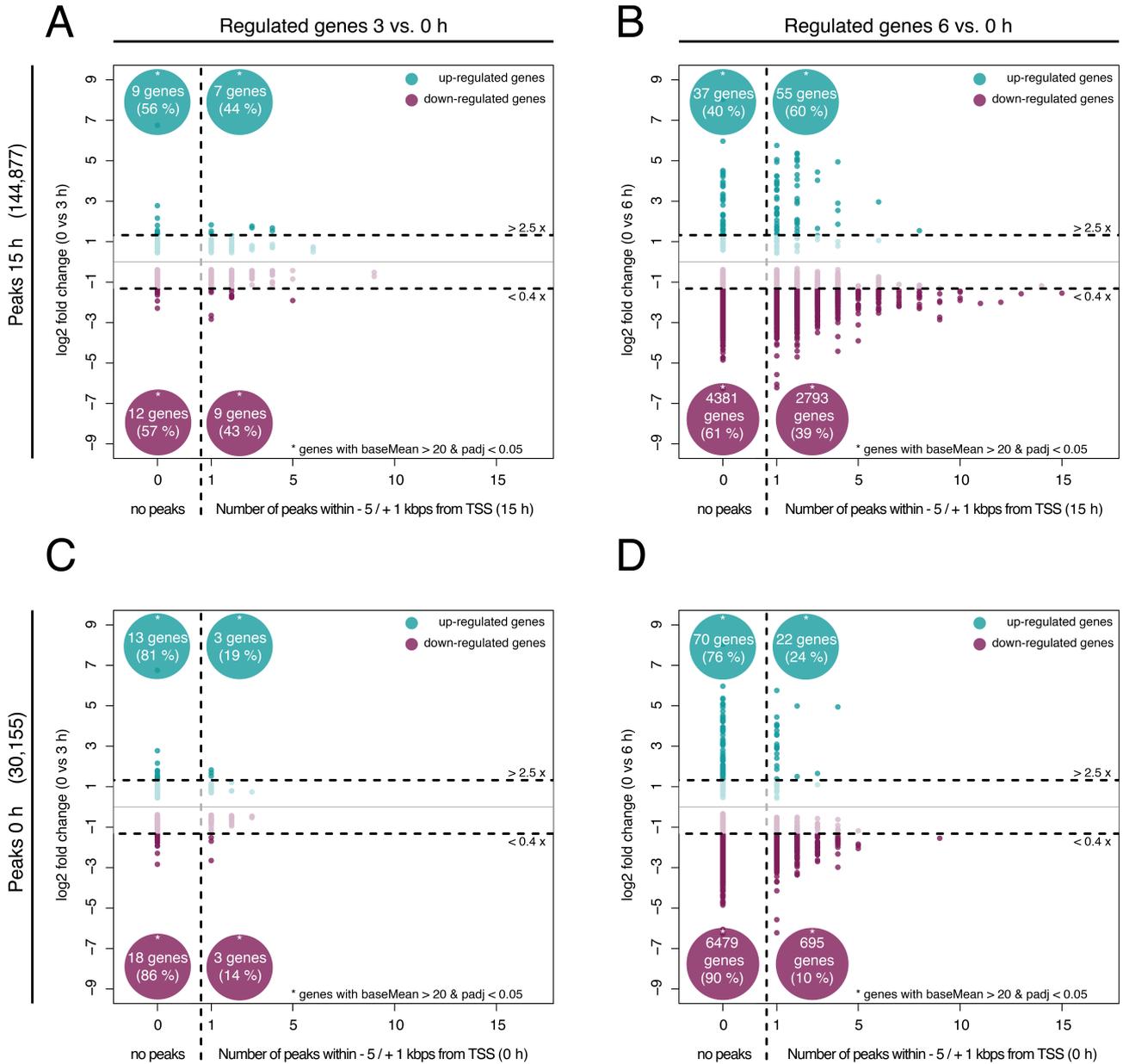


Figure 3-27: Linkage between ChIP-seq peaks of BZLF1 prior to and 15 h post induction and BZLF1-regulated genes induced for 3 and 6 h.

The four plots link BZLF1 peaks in Raji 4816 cells with up- or down-regulated genes 3 h (panels A and C) or 6 h (panels B and D) post induction. In panels A and B 144,877 peaks, which were present at 15 h post induction, were analyzed. The 30,155 peaks (panels C and D) identified in Raji 4816 cell in the absence of doxycycline (0 h) were linked to genes, which were up- or down-regulated 3 or 6 h post induction of BZLF1. The composition of the plots is identical to the plots shown in Figure 3-26. In panels A and B about half of the up-regulated or down-regulated genes do or do not contain a BZLF1 peak within the defined promoter region (-5/+1 kbps of TSS) identified 15 h after BZLF1 induction. The plots indicate that many cellular genes, which do not have an identified BZLF1 binding site within their defined promoter regions, were regulated after BZLF1 induction.

92 up-regulated genes comprised a peak within their promoter regions. The majority of these genes had a single peak and only few genes had more than one peak within their promoter region.

Table 9-6 in the attachment provides a list of the 25 strongest up- and down-regulated genes, which have at least one BZLF1 binding site within their promoter in Raji 4816 cells 6 h after doxycycline-induced BZLF1 expression.

3.7 Chromatin interactions of selected genes prior to and after BZLF1 expression

3.7.1 The experimental setup of the Capture-C approach

BZLF1 binding sites within the promoter region of many BZLF1-regulated genes did not reveal if and how BZLF1 might regulate them. In contrast to viral genes, where BZLF1 acts as a promoter factor, BZLF1 might control enhancers of cellular genes and regulate its targets as an enhancer factor. To test this hypothesis, Capture-C experiments were planned and conducted with chromatin from Raji 4816 cells with a focus on a selection of cellular genes supposedly regulated by BZLF1. On the basis of the RNA-seq experiments strongly up-regulated and down-regulated as well as non-regulated genes were chosen. Additionally, genes in close proximity to strongly regulated genes were chosen to be able to test them for co-regulation.

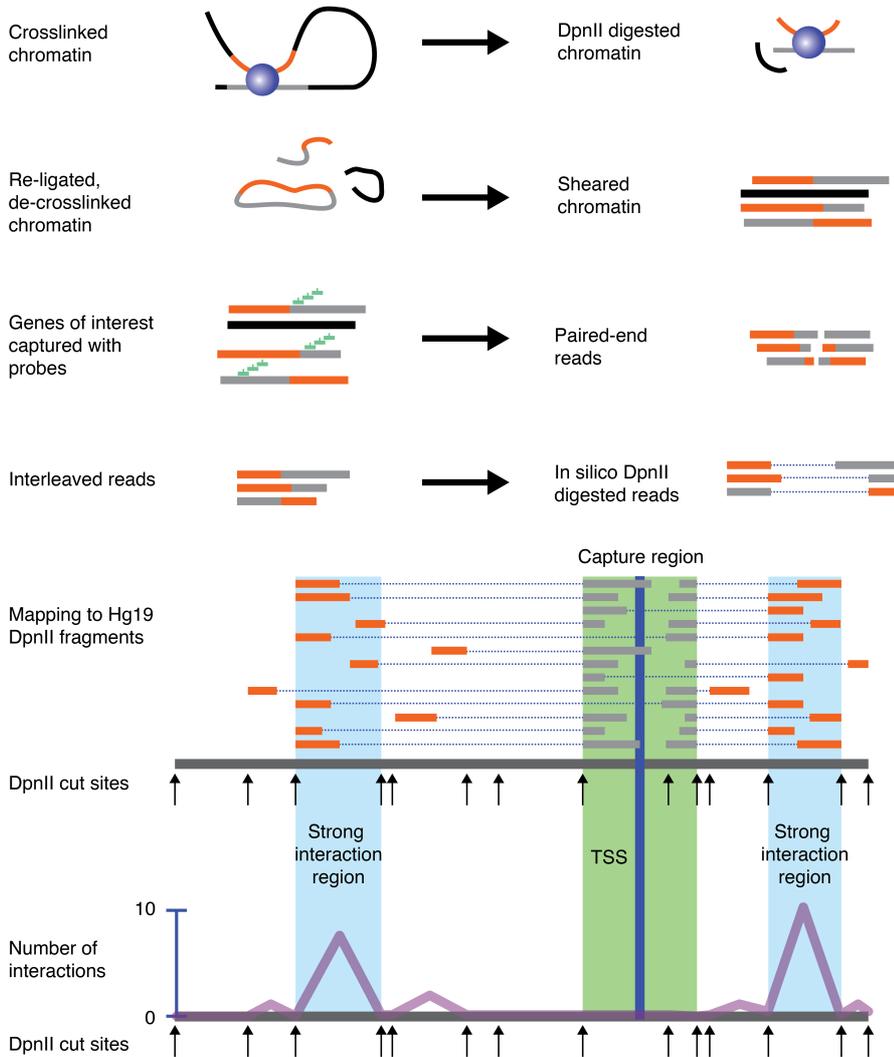
Figure 3-28 summarizes the workflow of the Capture-C experiments. The cells were cross-linked to fixate interacting chromatin in non-induced cells and cells 6 and 15 h after BZLF1 induction. In brief, the DNA was digested with the restriction enzyme DpnII and re-ligated under conditions of low DNA concentration to promote the ligation of chromatin-bound and complexed DNA fragments, only (performed by Paulina Mrozek-Gorska). After removal of the proteins, the DNA fragments were fragmented by sonication and used for library preparation. A set of commercial, custom-made, and overlapping RNA-probes of 120 nucleotides was used to enrich DNA fragments within regions about +/- 5 kbps around the TSSs of the genes of interest. After sequencing, paired-end reads were joined and a DpnII restriction digestion was performed *in silico* before the resulting single fragments were mapped on Hg19. A program (CCAnalyser2) provided by [James Davies](#) (University of Oxford) was used to identify interactions between the capture area and more distant DNA sequences, which had been in close proximity during cross-linkage. To achieve this, Hg19 was digested *in silico* with the restriction enzyme DpnII and the reads were mapped onto the genomic

human DNA sequence. Two categories, reads mapped to the capture region and reads mapped to DpnII fragments located beyond the capture region, were built to assess the number of interactions between both categories. The capture region is depicted as a green column in Figure 3-28 and dashed lines connect reads mapped to DpnII fragments of the

capture area with reads mapped to interacting distant DpnII fragments.

Figure 3-28: Capture-C: from *in vivo* chromatin interactions to *in silico* chromatin interactions.

Raji 4816 cells were crosslinked prior to and after 6 and 15 h of doxycycline-induced BZLF1 expression. The cells with crosslinked chromatin were harvested, accessible DNA was digested with the restriction enzyme DpnII, and ligated at very low DNA concentration. The proteinaceous part of the crosslinked chromatin was removed and the ligated, often circular DNAs were sheared with a Covaris system to fragments of about 200 bps. Probes especially designed to hybridize with genes of interest identified in the previously described RNA-seq experiments were used to pull down the regions around the TSSs (Capture region) of the chosen genes. Together with these DNAs, previously linked interacting DNA fragments were pulled down as well. Paired-end sequencing (100 bps) identified



the sequences of both classes of DNA fragments representing the capture region and interacting DNA targets. If the paired-end reads information yielded overlapping sequences, they were interleaved to one fragment and digested *in silico* by the restriction enzyme DpnII. The single fragments were then mapped to Hg19. In a last step the CCAlyser2 command line program identified read pairs, which contain both the capture region (green) and a distal DpnII fragment on the same chromosome. A dashed line links *in silico* digested read pairs. Their numbers were added up for every DpnII fragment and visualized on the y-axis. A line indicates the frequency of interactions a DpnII fragment enters with the investigated chromatin region. Two strong interacting regions are indicated as examples.

In Figure 3-29 Capture-C data from each time point are visualized by three differently colored lines (①: 0 h, white; ②: 6 h, purple; ③, 15 h: light blue) indicating the numbers of

interactions per DpnII fragment. The colors are partially transparent, which results in mixed colors if the lines overlap. The made-up mock Figure 3-29 exemplifies the setup of the lines and provides the symbols and their legends for BZLF1 binding sites and binding sites of chromatin organizers, histone marks as well as information about the directionality of the regulated gene 6 h after BZLF1 induction (⑥).

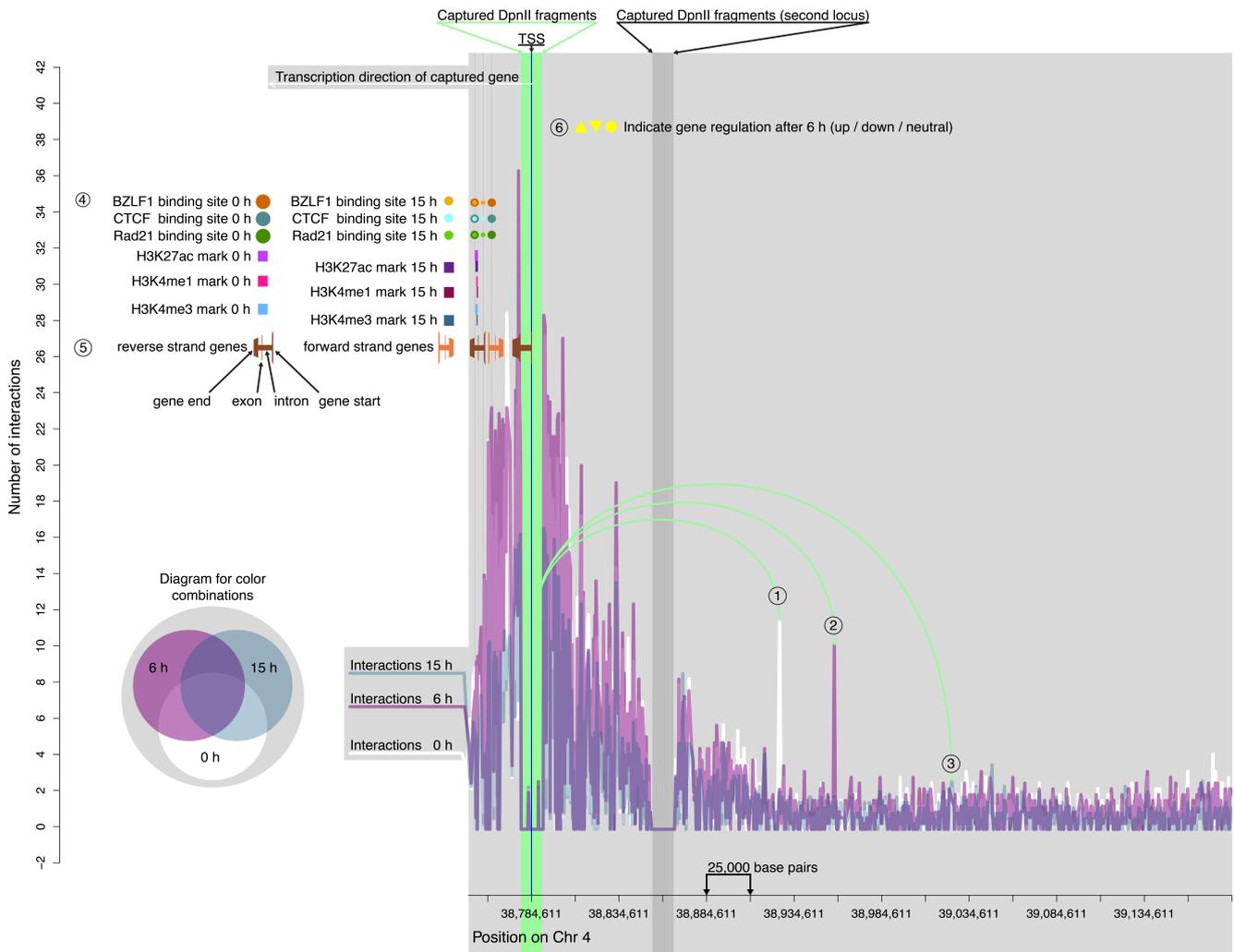


Figure 3-29: Capture-C – an introduction.

The visualization shows a made-up mock example of Capture-C data and their presentation to explain the features, annotations, and legends of the subsequent figures. The shown example does not contain real data. The x-axis shows the nucleotide coordinates on a given chromosome, which is indicated. Each figure shows ± 400 kbps from the TSS and the distance between the ticks is 25 kbps. The y-axis shows the sum of individual interactions between the captured promoter region (green) and distant locations. Three different colors (white, purple, light blue) indicate interactions at different time points (non-induced, 6 h and 15 h) after BZLF1 induction. The diagram with combinations of the three colors indicates the colors that emerge from overlapping data. Capture-C experiments were performed with cellular DNA after mild formaldehyde-mediated cross-linking, cleavage with the restriction enzyme DpnII, and ligation at very low DNA concentrations. Re-ligated DNA fragments were captured by DNA-RNA hybridization with a custom-made set of RNA oligonucleotides linked to biotin. They cover up to 33 DpnII anchor fragments (green column) of selected genes, approximately ± 5 kbps of the TSS (blue). The enriched DNA samples were analyzed by paired-end sequencing and analyzed for interactions, which take place between two DpnII fragments, the captured promoter region (green) and a distant but linked DpnII fragment. The frequency of distal DpnII fragments identified to interact with DpnII fragments of the capture area are represented by single visible peaks or vales per DpnII fragment. For example, the peak of

the DpnII fragment number ① was frequently found to be ligated with one or more of the DpnII fragments of the captured promoter region in non-induced cells indicating a strong interaction at 0 h. DpnII fragment number ② shows a strong interaction with the DpnII anchor fragments of the promoter at 6 h and number ③ a rather weak interaction 15 h post BZLF1 induction. The dark grey bar represents the capture area of a second gene, which is not part of this analysis. The white arrow, which originates from the TSS, indicates the direction of transcription of the captured gene. The localization of BZLF1 binding sites, histone marks and binding sites of the cellular chromatin organizers CTCF and Rad21 are shown and explained in the legend indicated by ④. BZLF1 binding sites were identified with MACS2 after next generation ChIP-sequencing. A large, dark orange dot indicates a BZLF1 peak in non-induced cells, a small bright orange dot stands for a peak 15 h after induction. Peaks identified at both time points show an overlap of both dots. Similar annotations were chosen to visualize CTCF (turquoise) and Rad21 (green). Histone modifications, which differ between non-induced cells and cells induced for 15 h were also identified by MACS2. Histone modifications of H3K27ac are indicated by a bright purple square for 0 h and a dark purple square for 15 h. The H3K4me1 and H3K4me3 modifications are depicted in bright (0 h) and dark pink (15 h) and bright (0 h) and dark blue (15 h) squares, respectively. Histone modifications can cover larger regions, squares then turn into rectangular areas of variable length. ⑤ explains the visualization of genes. Forward strand genes are depicted in bright brown, height tapering from left to right indicates the directionality of the gene. Reverse strand genes are colored in dark brown. A horizontal bar indicates the length of the gene and vertical bars represent its exons. ⑥ shows the regulation of the gene of interest. It can either be up-regulated (upwards directed triangle), down-regulated (downwards directed triangle) or neutral (dot). Anne Wöllmer performed the ChIP-seq wet lab work for the CTCF, Rad21, H3K27ac, H3K4me1 and H3K4me3 samples.

Binding sites of BZLF1, CTCF, and Rad21 are shown prior to and 15 h post doxycycline-mediated induction of BZLF1 (④, the wet lab work for the histone marks, CTCF, and Rad21 was performed by Anne Woellmer). A thin grey line connects each BZLF1 binding site with its corresponding DpnII fragment to simplify the visual association. Additionally, the structure of other genes (exons and introns) and their directionality are depicted in the plot (⑤).

The capture region covers up to 33 DpnII anchor fragments depending on the gene (green column) and covers approximately +/- 5 kbps of the TSS (thin blue line). In several genes the capture region contains gaps, which were introduced to exclude repetitive regions even though these are suspected to play a role in chromatin architecture (Tang, 2011). In such cases the capture region was extended beyond the +/- 5 kbps limits.

3.7.2 Down-regulated genes

In Raji 4816 cells more than 7,000 genes were found to be down-regulated 6 h after doxycycline-induced BZLF1 expression in RNA-seq experiments. One of the most strongly down-regulated genes was the tumor suppressor *BTG2* (0.02 %). Its locus is shown in Figure 3-30 together with DpnII fragments interacting with its capture area. Additionally, the locations of histone marks, chromatin organizers, and BZLF1 binding sites are depicted prior to and after BZLF1 induction.

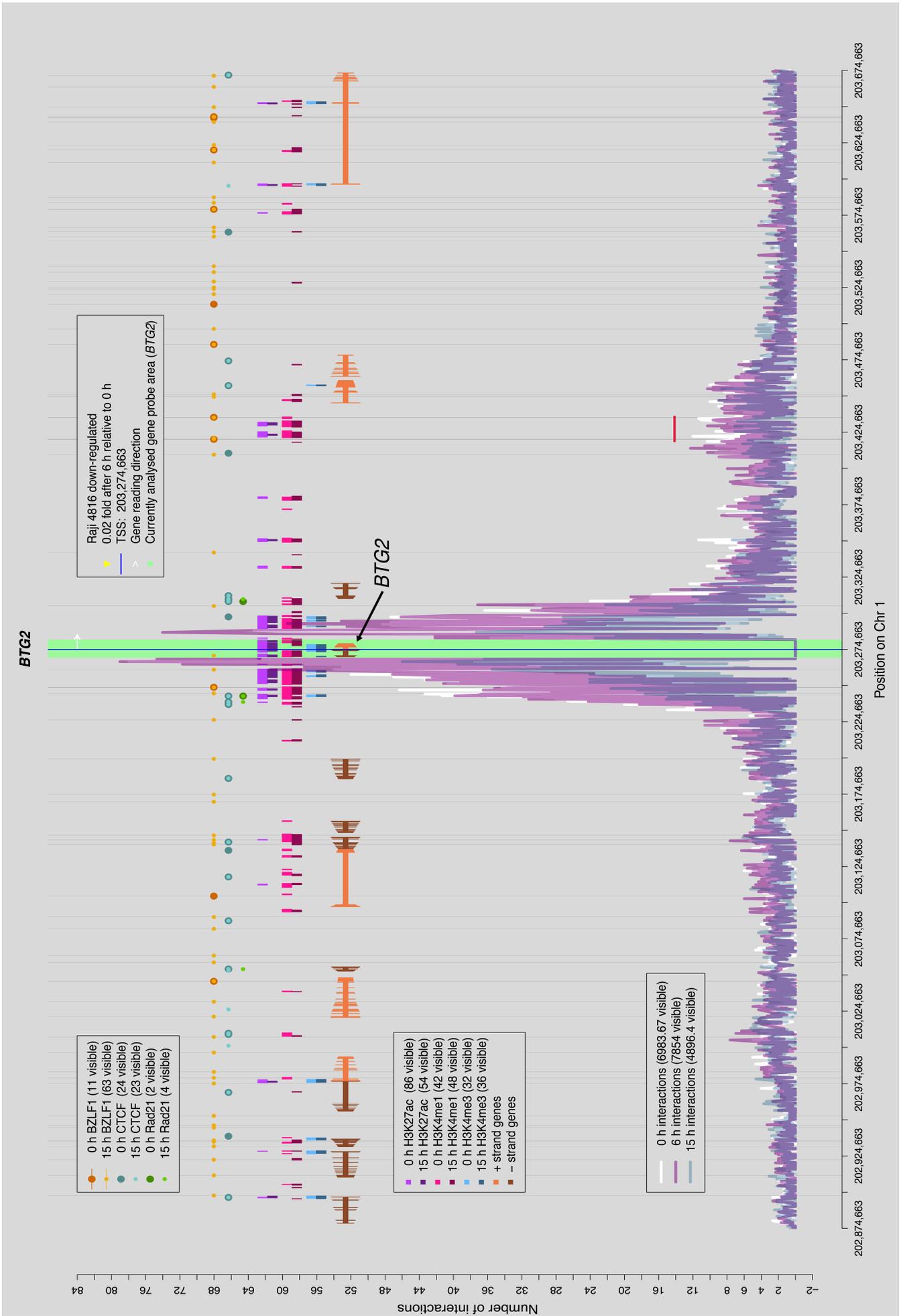


Figure 3-30: Gene regulation, Capture-C and RNA-seq data of the down-regulated *BTG2* gene.

The figure combines results from Raji 4816 cells with a focus on Capture-C, BZLF1 binding sites, histone marks, the chromatin organizers CTCF and Rad21 as well as the regulation of the captured gene, which is down-regulated to 2 % 6 h post BZLF1 induction. The composition of Capture-C plots, their features and annotations are explained in detail in Figure 3-29. The graph legend in the upper right provides information about the regulation of the depicted gene, its TSS, and the direction of transcription of the captured gene. The graph legend in the lower left summarizes the total interactions identified at different time points within this snapshot. The upper left graph legend summarizes the information about the BZLF1 binding sites and the chromatin organizers, while the graph legend below summarizes information about the histone marks. Many interactions with the captured DpnII fragments could be identified within an area +/- 40 kbps around the TSS. This area is flanked by several CTCF and Rad21 sites and strongly covered with the histone marks H3K27ac, H3K4me1, and H3K4me3, which are associated with active chromatin. It appears that the CTCF/ Rad21 sites define a domain with frequent interactions isolated from the neighborhood. An exception is the domain at about nucleotide position #203,424,000 (indicated by the red line), where interactions existed in the non-induced state, which were reduced upon induction of BZLF1. BZLF1 sites were close-by and may influence the loss of interactions. Overall 3,000 interactions were lost comparing the total of interactions of 6 h with 15 h.

The area +/- 40 kbps around *BTG2*'s TSS shows many interactions; their numbers increased 6 h after BZLF1 induction but decreased strongly 15 h after induction. In total about 3,000 interactions were lost between 6 and 15 h after BZLF1 induction. The severe loss of interactions 15 h after BZLF1 induction was characteristic for all analyzed loci irrespective of whether they were up- or down-regulated.

The area with the most frequent interactions very close to the *BTG2* gene body is covered with histone marks associated with active chromatin and flanked by CTCF as well as Rad21 proteins, which organize the chromatin structure. Beyond this clearly addressable region the number of interactions is rather low with the exception of the region at about position #203,424,000 (indicated by the red, horizontal line) downstream of *BTG2*. Within this region BZLF1 bound prior to and after BZLF1 induction but interactions were reduced after BZLF1 induction.

Taken together the interactome map of the *BTG2* gene is characterized by interactions within a close-by domain that is flanked by CTCF and cohesin (Rad21). A region about 150 kbps downstream of the gene body shows a good linkage and probably acts as an enhancer element. This long-range interaction as well as the close-by interactions were severely weakened upon induction of BZLF1 suggesting that its expression induces the relaxation of the chromatin architecture of this locus.

Loss includes the long-distance interactions but also regions up- and downstream of the gene itself. Several BZLF1 sites exist within the *BTG2* locus flanked by CTCF binding sites, but they do neither explain the loss of interactions nor the inhibition of the transcription upon induced expression of BZLF1.

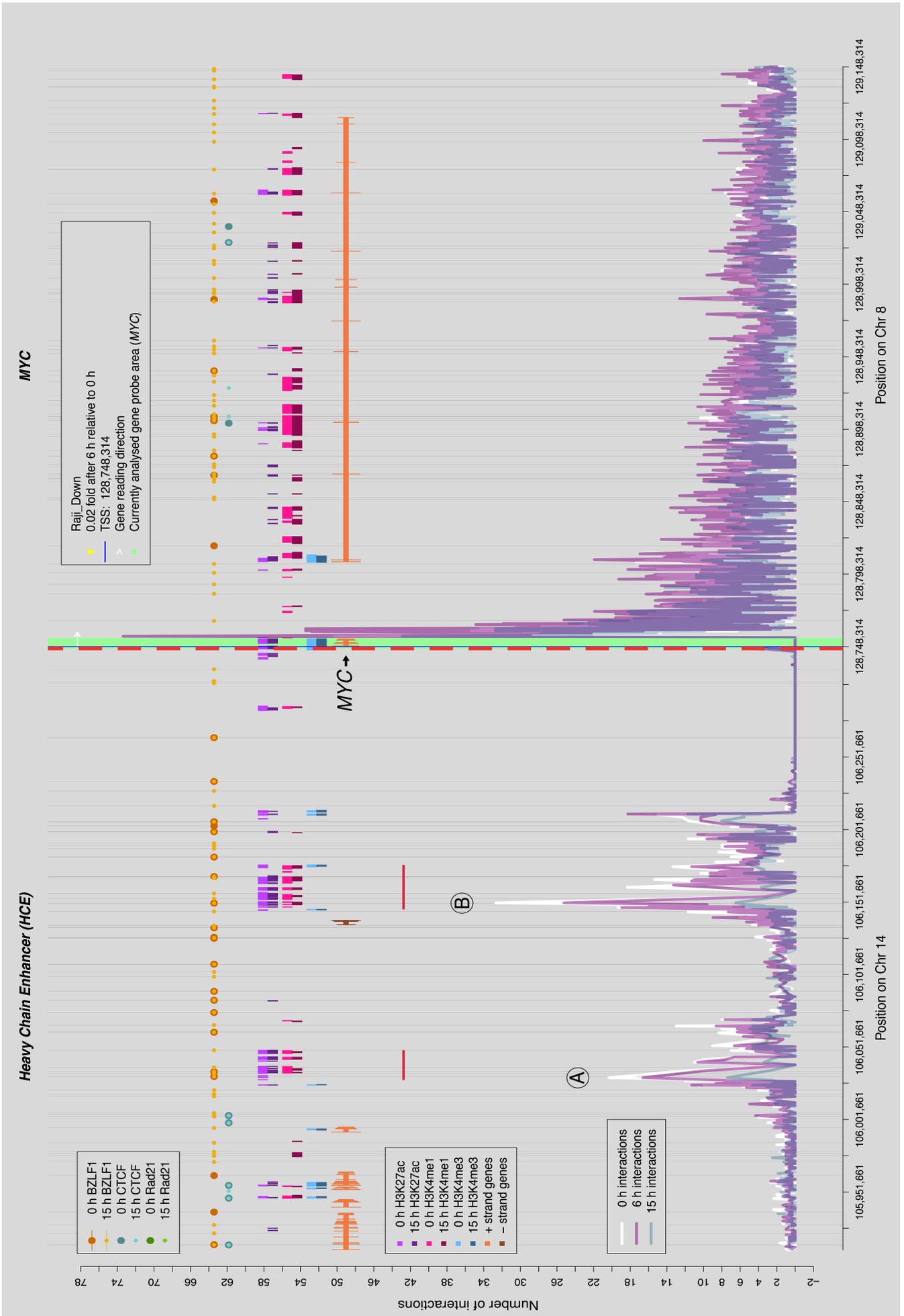


Figure 3-31: Capture-C analysis and chromatin features of the translocated *MYC* gene in Raji 4816 cells.

The figure summarizes results from Raji 4816 cells including Capture-C, ChIP-seq, and RNA-seq data of the *MYC* locus in Raji 4816 cells, which was transcriptionally repressed upon BZLF1 expression. The Raji cell line harbors an 8;14 translocation (Hamlyn and Rabbitts, 1983), which positions the immunoglobulin heavy chain enhancer (*HCE*) of chromosome 14 (upstream of base 106,326,507) upstream of the *MYC* gene on chromosome 8 (downstream of base 128,746,922). The data were mapped on Hg19, which does not include this translocation. Therefore the *MYC* locus was connected with the 8;14 translocation as shown. The vertical dashed red line indicates the junction of chromosomes 8 and 14 (Dyson and Rabbitts, 1985). Features, annotations, and legends of the Capture-C plots are explained in detail in Figure 3-29 and Figure 3-30. Since this plot combines two different loci, the numbers of interactions and chromatin marks cannot be shown. The two regions (A and B) within the *HCE* with the most frequent interactions with the capture region were strongly covered with H3K27ac and H3K4me1 histone marks, while the body of the *MYC* gene contained H3K27ac and H3K4me3 marks. No BZLF1 binding sites could be identified within the capture area, but multiple BZLF1 binding sites were prevalent within the heavy chain enhancer region. Two of them directly bound within the two DpnII fragments A and B, which show the strongest interactions with the capture area. Each of the binding sites contained three adjacent CpG ZRE motifs.

The cellular transcription factor *MYC* was another strongly down-regulated cellular gene 6 h after BZLF1 induction. The *MYC* protein plays very important roles in cell cycle progression, cell growth, and cellular transformation. Prior to BZLF1 induction *MYC* was strongly expressed due to an 8;14 translocation in Raji cells, which positions the immunoglobulin *HCE* upstream in proximity of the *MYC* promoter. Since the data were mapped on Hg19, which does not comprise the 8;14 chromosomal translocation, the two loci had to be manually connected for visualization. As a consequence the graph legends of Figure 3-31 do not provide the numbers of interactions, histone marks, and binding sites.

The *MYC* gene was down-regulated to 2 % 6 h after induced BZLF1 expression. Two DpnII fragments within the *HCE* indicated by A and B showed a particularly high number of interactions with the capture area prior to BZLF1 induction, but about 20 % of these interactions were lost six hours post induction. After 15 h of induced BZLF1 expression about 80 % of the interactions found prior to BZLF1 induction were lost. Downstream of the *MYC* gene the numbers of interactions mainly increased 6 h after BZLF1 induction but decreased again 15 h after BZLF1 induction, but not as strongly as at the *HCE* locus. While the enhancer sites were mainly covered with H3K27ac and H3K4me1 histone marks, the gene body of *MYC* primarily contained H3K27ac and H3K4me3 marks. No BZLF1 binding site could be found within the capture region prior to or post BZLF1 induction. In contrast, in both DpnII fragments with the highest number of interactions within the *HCE* (A and B) a BZLF1 peak was located prior to and post BZLF1 induction. Both DpnII fragments, A and B, harbor the exact same DNA sequence with three CpG ZRE-like sites indicated in bold, capital letters:

TGAGCGTgagcttggcagcgggtgggagaatg**TGTGCGT**cagtgtg**TGAGCGT**ga. The last nucleotide residue was always a T whereas the major motif identified during ChIP-seq ends

with an A (Figure 3-5 and Figure 3-8). These binding sites might explain the loss of interactions, when BZLF1 is induced and expressed after induction.

3.7.3 Up-regulated genes

Six hours after doxycycline-induced BZLF1 expression only 92 up-regulated genes were identified. One of the strongest up-regulated (18.94 fold) genes was the collagen type II alpha 1 (*COL2A1*), whose locus is shown in Figure 3-32.

The numbers of interactions within the depicted *COL2A1* region increased by 14 % already 6 h post induction. No histone marks were identified within the gene body or in close proximity of the TSS, but two BZLF1 peaks were identified 15 h after BZLF1 induction in the capture region and many more downstream of the *COL2A1* gene. By visual inspection the induced expression does not correlate with the BZLF1 binding sites but may correlate with the increased number of interactions within the gene body.

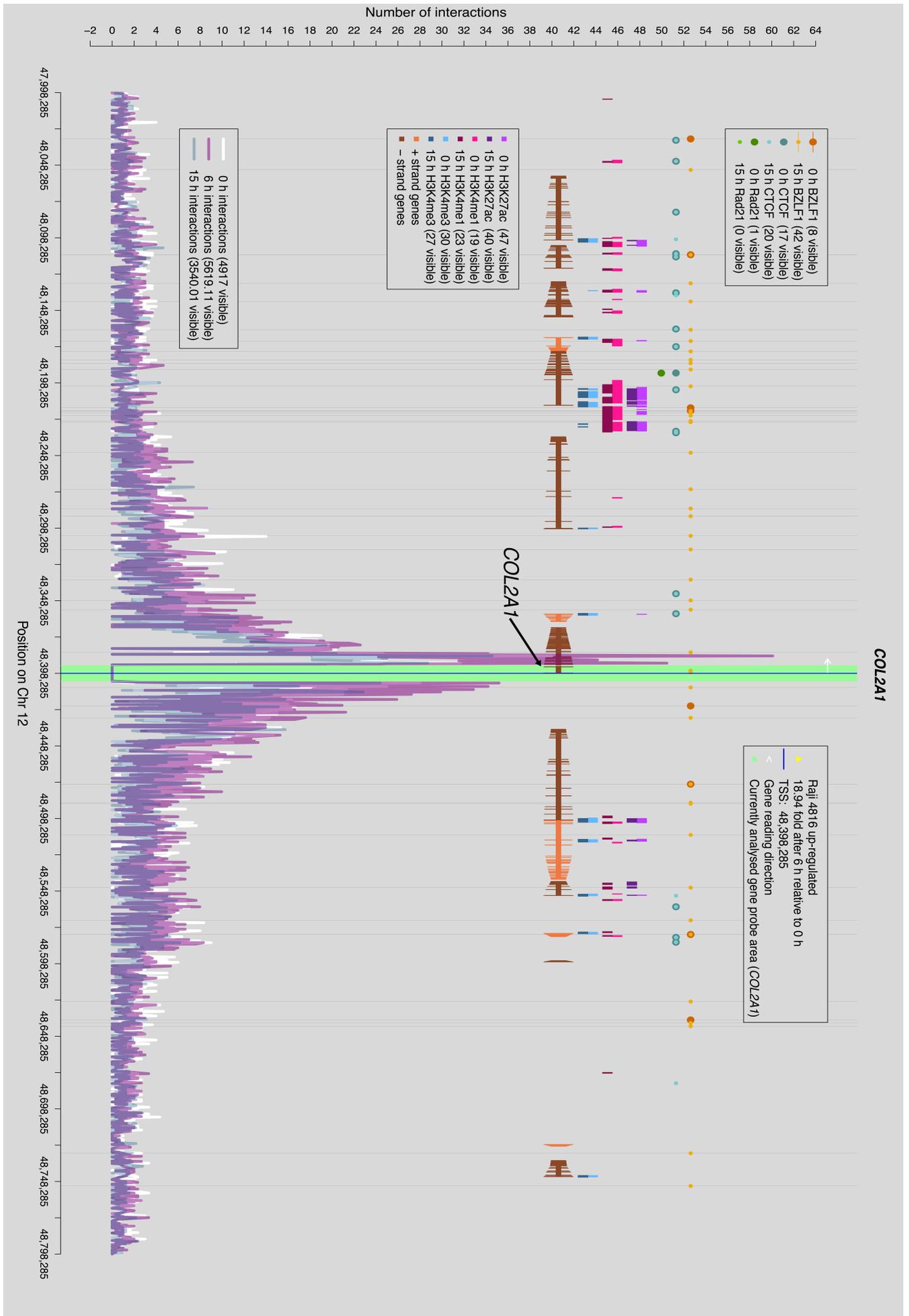
A second gene locus, which was strongly up-regulated (about 14 fold) is the non-coding RNA (ncRNA) *LOC100128288*, which is shown in Figure 3-33.

In total the number of interactions increased 6 h post BZLF1 induction, but many DpnII sites lost some of their interactions within that time frame. The interactions closely upstream of *LOC100128288* were increased 6 h after BZLF1 induction and strongly decreased after 15 h.

The *KRBA2* gene (indicated by arrow and name) is located upstream of the *LOC100128288* gene and was up-regulated about 6 fold. The *KRBA2* site was excluded from the analysis of the ncRNA *LOC100128288* to avoid a bias.

Figure 3-32: The *COL2A1* locus codes for collagen type II alpha 1 and is up-regulated upon BZLF1 expression. (next page)

The figure shows Capture-C and RNA-seq data in combination with additional chromatin attributes of the *COL2A1* locus in Raji 4816 cells prior to and after induced expression of BZLF1. Annotations, features and their legends are explained in Figure 3-29 and Figure 3-30 in detail. The *COL2A1* gene was strongly up-regulated (19 fold) 6 h post BZLF1 induction. The sum of all interactions (5619) 6 h after BZLF1 induction with the capture region was slightly increased when compared to 0 h (4917). Especially noticeable is that certain interactions located within the gene body downstream of the TSS are drastically increased 6 h after BZLF1 induction. No particular chromatin marks were identified within the *COL2A1* gene region. Within the capture region, BZLF1 was present at two sites in close proximity, but bound many more sites downstream of the gene body 15 h post induction. Upon visual inspection there was no apparent correlation between BZLF1's transcriptional activation, binding, and extended interactions 6 h post induction.



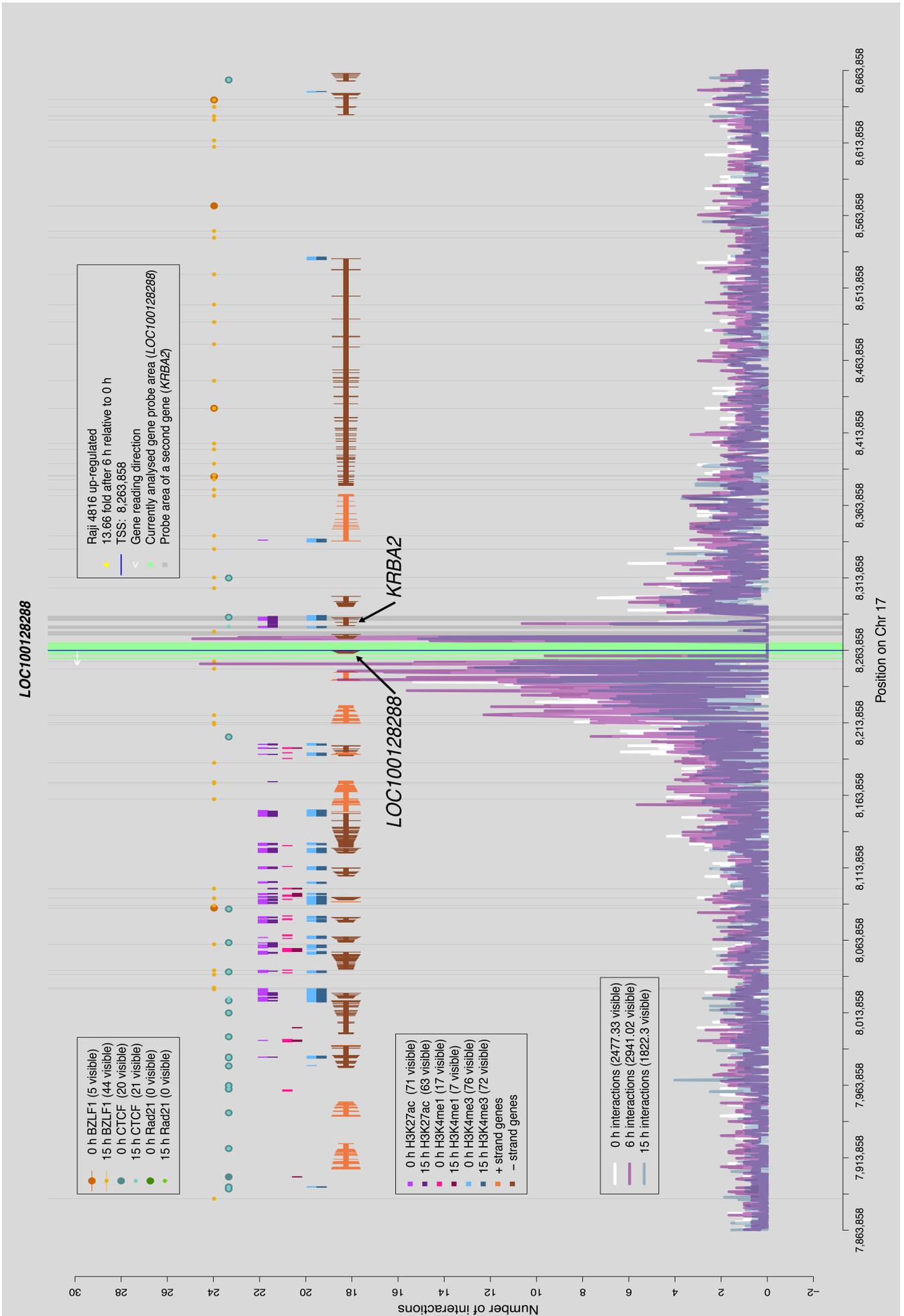


Figure 3-33: The *LOC100128288* locus is a non-coding RNA with unknown functions and was up-regulated upon BZLF1 expression.

The figure shows Capture-C and RNA-seq data in combination with chromatin data of the *LOC100128288* locus in Raji 4816 cells. Annotations, features, and legends are explained in Figure 3-29 and Figure 3-30 in detail. Six hours post BZLF1 induction the ncRNA *LOC100128288* was up-regulated about 14-fold. Six hours after BZLF1 induction the identified interactions increased in general. Especially for the upstream region of the ncRNA, an overall decrease of interactions was apparent 15 h after BZLF1 induction. An adjacent gene, *KRBA2* (light grey areas), which was 6.32 fold up-regulated upon induced expression of BZLF1, was also captured in the Capture-C analysis and was therefore excluded from the analysis of *LOC100128288*. The regions with the strongest interactions downstream of *LOC100128288* did not comprise any identified histone marks. No BZLF1 binding site was identified within the capture region and none of the adjacent BZLF1 binding sites seems to be directly responsible for the up-regulation of *LOC100128288*.

Neither in *LOC100128288*'s gene body nor in the capture region a BZLF1 binding site was identified prior to or after BZLF1 induction. Since none of the DpnII fragments, which interact strongly with the capture region depicts a BZLF1 binding site no connection between BZLF1 expression, gain and loss of interactions, and gene regulation is obvious.

3.7.4 Non-regulated genes

Genes, whose gene expression did not change by more than 10 % were defined as non-regulated. One of these genes is the histone deacetylase 9 (*HDAC9*), which is only slightly regulated (0.91 fold) and can be expressed from several alternative promoters. With the help of RNA-sequencing data the actually used TSS was identified at position #18,535,368 in Raji cells and is indicated in Figure 3-34. Six hours after BZLF1 induction the capture region had about 1,500 interactions in total but certain DpnII fragments lost their interactions in that time window while others gained interactions. As seen in all analyzed genes the interactions 15 h after induction strongly decreased but were not completely lost. No BZLF1 binding sites were identified within these DpnII fragments.

The histone marks H3K27ac and H3K4me1 were prevalent in the region from the TSS to about 75 kbps downstream. Upon BZLF1 induction the coverage with the H3K27ac mark was strongly reduced. The H3K4me3 mark covered a region from the TSS to about 40 kbps downstream.

Several CTCF sites (red circle) upstream of the TSS and one CTCF site (red circle) downstream seemed to flank and encompass a region with a frequent number of interactions with the capture region, forming a discrete chromatin domain.

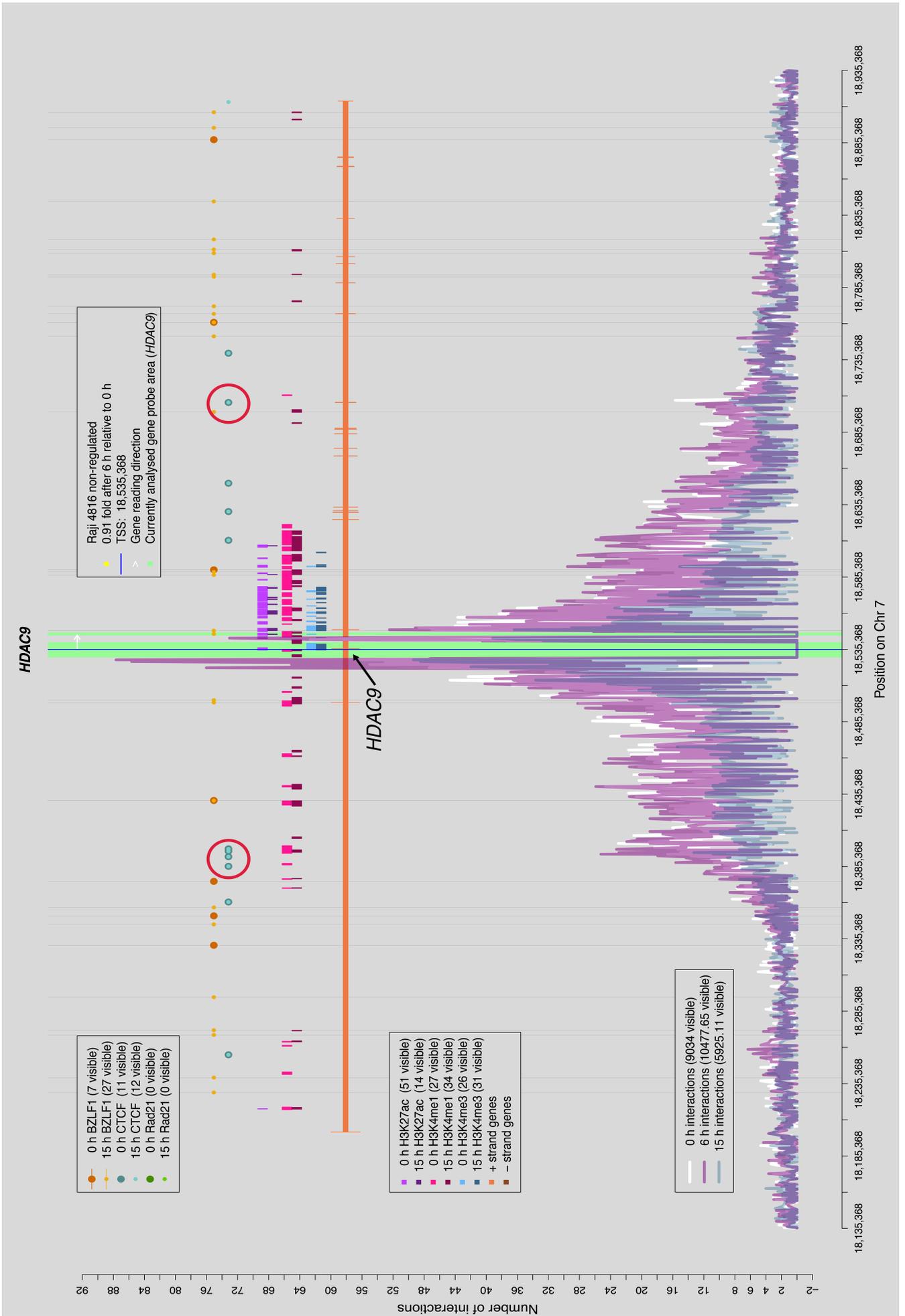


Figure 3-34: The non-regulated HDAC9 gene, its chromatin components and architecture according to ChIP-seq and Capture-C data.

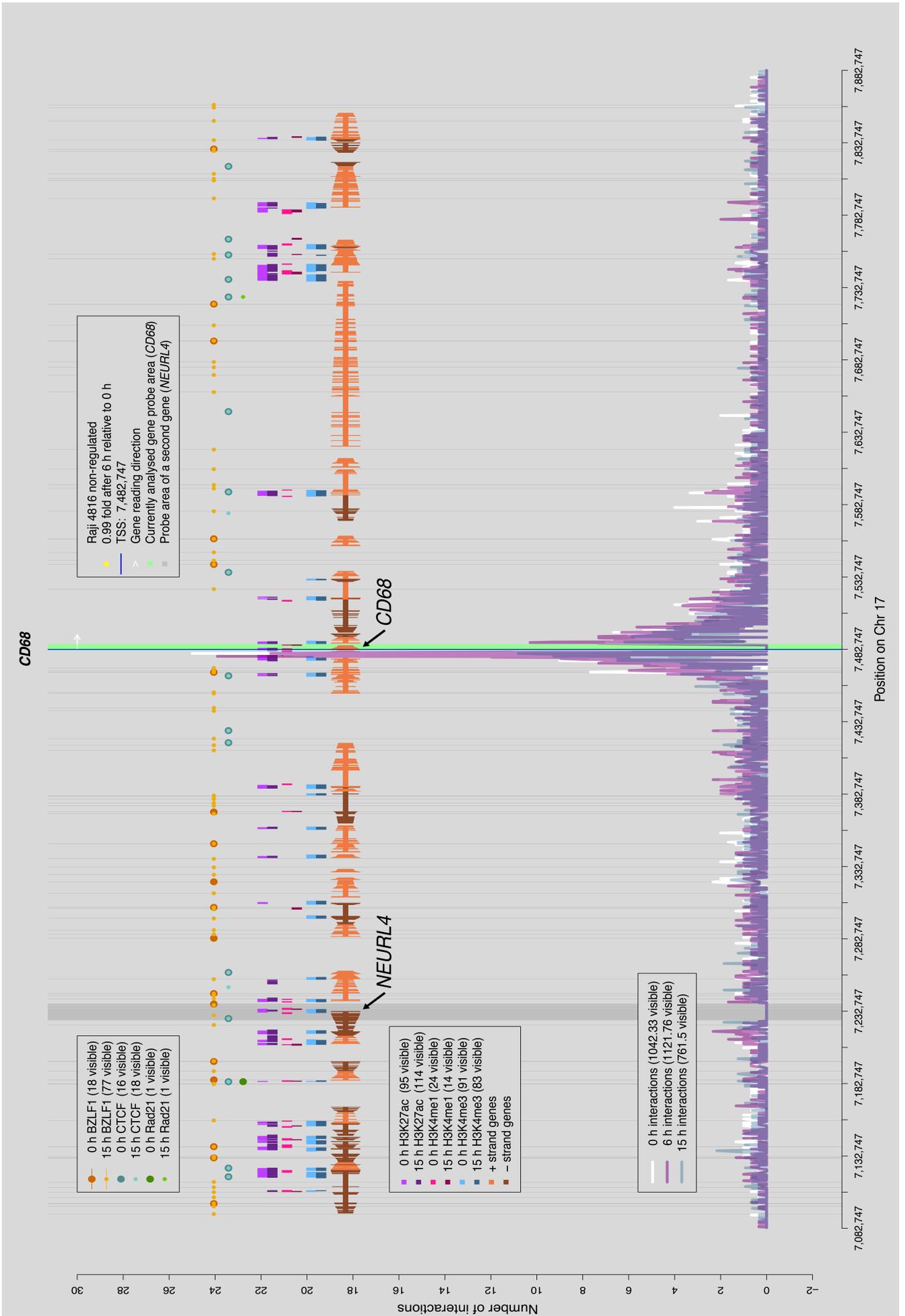
The figure shows the combined results from Raji 4816 cells including Capture-C, ChIP-seq and RNA-seq data for the *HDAC9* gene in Raji 4816 cells. The composition of the plot was explained in greater detail in Figure 3-29, the legends in Figure 3-30. The *HDAC9* variants can be driven from several TSSs. RNA-sequencing data revealed that *HDAC9* expression in Raji cells is driven from the TSS at position #18,535,368 and does not change 6 h post BZLF1 induction (0.91 fold). The total number of interactions within this snapshot increased from 0 to 6 h post BZLF1 induction by about 1,500 reads, but the total number of interaction 15 h post BZLF1 induction strongly decreased to about two third of the interactions found prior to BZLF1 induction. A section of up to 75 kbps downstream of the TSS was highly covered with the histone marks H3K27ac and H3K4me1, while H3K4me3 covered an area about 40 kbps. Upon BZLF1 induction the coverage with H3K27ac was strongly reduced. The area with frequent interactions seems to be flanked by several CTCF sites upstream of *HDAC9* and one particular downstream of *HDAC9* (indicated by red circles). The BZLF1 binding sites do not seem to have any specific role in changing the number of interactions with the capture area or the histone marks.

Another non-regulated gene (0.99 fold) was the *CD68* transmembrane glycoprotein, which is usually expressed in monocytes. Figure 3-35 summarizes the collected information for this locus.

Compared with other loci only few interactions were identified and they were located in DpnII fragments close to the TSS. Certain DpnII fragments showed increased, others decreased numbers of interactions. The DpnII fragments with the highest initial numbers of interactions experienced some loss 6 h after BZLF1 induction but 40 % of the interactions were lost 15 h after BZLF1 induction. BZLF1 binding sites cannot explain this loss of interactions, nor did they have consequences for the expression of this gene.

Figure 3-35: The non-regulated CD68 gene, its chromatin components and architecture according to ChIP-seq and Capture-C data. (next page)

The figure shows the combined results from Raji 4816 cells including Capture-C, ChIP-seq and RNA-seq data for the *CD68* gene in Raji 4816 cells. The composition of the plot was explained in greater detail in Figure 3-29 and Figure 3-30. The transcript level of *CD68* did not change 6 h post BZLF1 induction. Compared with other genes only a very narrow area around the TSS interacted with the capture area. 15 h post BZLF1 induction half of the signal from the most frequent interaction directly upstream of the TSS was lost. No BZLF1 binding site was located within this area.



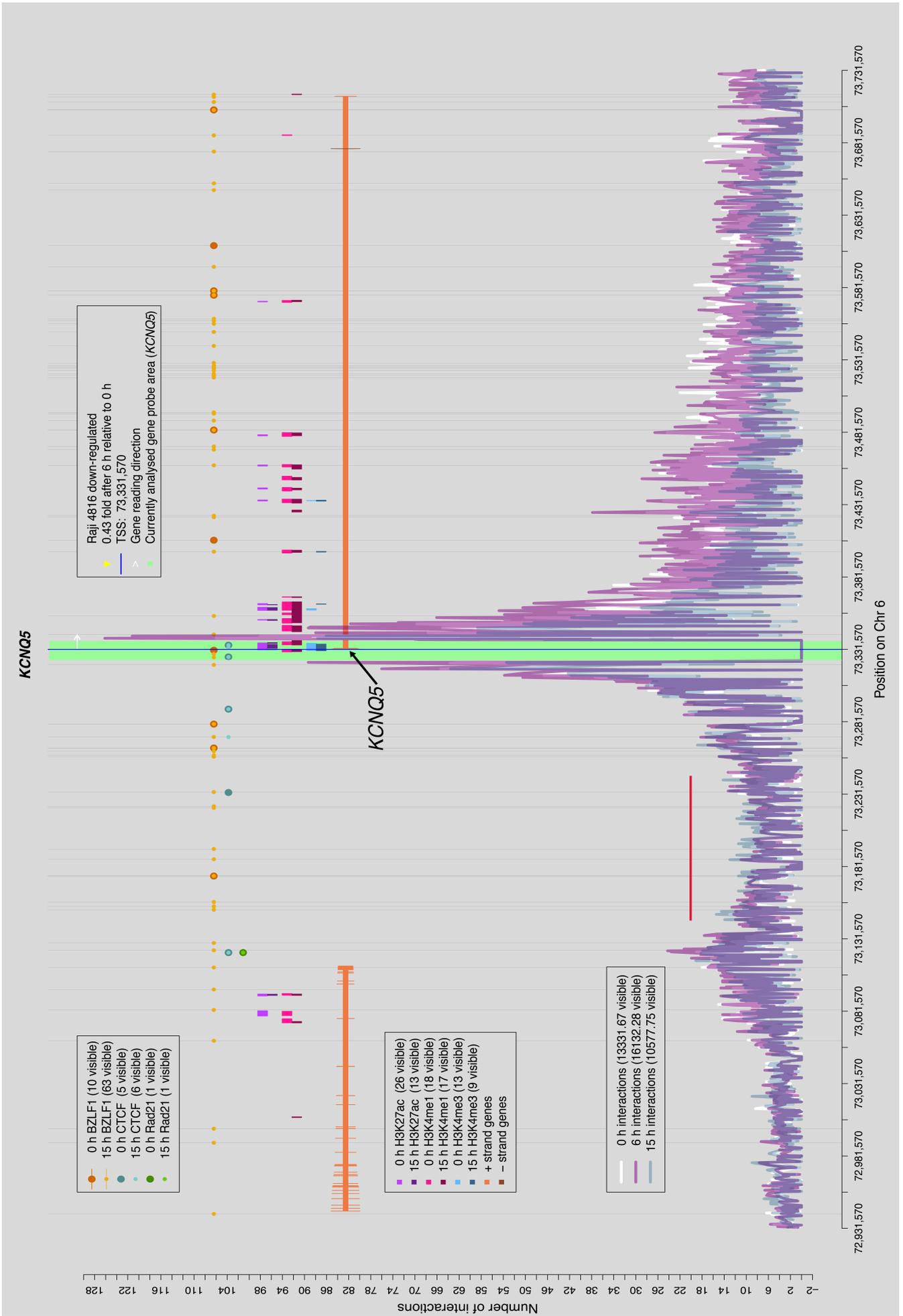
3.7.5 A region with increased numbers of interactions 15 h post BZLF1 induction

Independently of the Capture-C examples introduced so far the total number of interactions of most genes increased and decreased 6 and 15 h after BZLF1 induction, respectively. This is also true for the total number of interactions identified in the *KCNQ5* gene, which codes for a potassium channel. Figure 3-36 shows that another pattern occurs when only the upstream region of the gene is considered.

In the region starting at the TSS and extending about 200 kbps upstream the number of interactions was only slightly increased 6 h after BZLF1 induction. A region of about 100 kbps, which is indicated by a horizontal red line, showed an increased number of interactions 15 h post BZLF1 induction. Compared to all other Capture-C results this observation is an exception among 49 gene loci analyzed in total.

Figure 3-36: Chromatin components and architecture of the *KCNQ5* gene, which encompassed a region with additional chromatin interactions 15 h post BZLF1 induction. (next page)

The figure shows the combined results from Raji 4816 cells including Capture-C, ChIP-seq and RNA-seq data for the *KCNQ5* gene in Raji 4816 cells. The composition of the plot was explained in greater detail in Figure 3-29 and Figure 3-30. All Capture-C examples of up-, down-, and non-regulated genes have in common that interactions with the capture area in general slightly increased 6 h after BZLF1 induction, but were decreased 15 h post BZLF1 induction. This pattern can also be observed within the gene body of *KCNQ5*, but the region 200 kbps upstream of the gene showed a different pattern. Six h post BZLF1 induction, the frequency of interactions stayed almost constant but 15 h post induction several DpnII fragments in an area of about 100 kbps (indicated by a red line) even showed an increase of the number of interactions. This contradicts the assumption that the loss of interactions 15 h post BZLF1 induction is a global effect and suggests that the loss of interactions 15 h post BZLF1 induction is a selective mechanism.



3.7.6 Compartmentalization of interactions

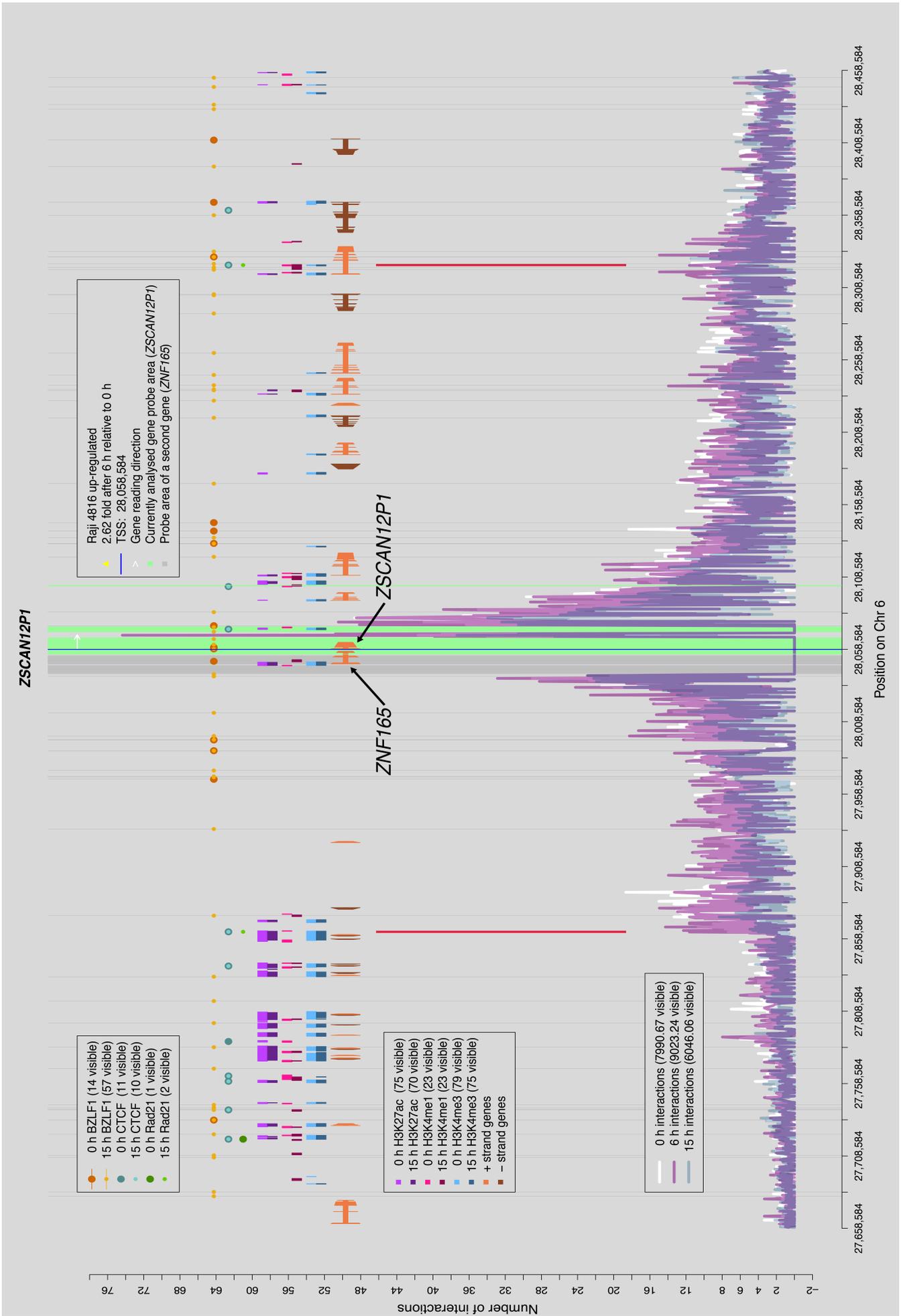
The *BTG2* (Figure 3-30) and *HDAC9* (Figure 3-34) genes already served as examples, in which chromatin organizers like CTCF and Rad21 isolate regions with more interactions from regions with a low number of interactions. The zinc finger protein with a SCAN domain (*ZSCAN12P1*) is an impressive example of a large and isolated area with increased numbers of interactions that were contained in a defined compartment (Figure 3-37). About 200 kbps upstream of the *ZSCAN12P1*'s TSS the number of interactions was strongly reduced with a clear cut (red line). This line co-localizes with a prominent CTCF site, which is accompanied by an additional Rad21 site apparent only after BZLF1 induction. A similar situation was found downstream of *ZSCAN12P1* (red line), even though the borderline was not as sharp as seen upstream. Again, CTCF sites were detected in non-induced and in induced cell chromatin, but the Rad21 mark was detected only 15 h after BZLF1 induction.

It seems as if chromatin organizers can bracket regions with higher numbers of interactions, isolating the regions from flanking chromatin. This observation was confirmed by the chromatin structure found in the zinc finger 165 gene (*ZNF165*) (Figure 3-38), which is located directly upstream of the *ZSCAN12P1* gene and is also highly structured by the chromatin organizers. The compartmentalization pattern of the *ZNF165* gene (Figure 3-38) is extremely similar to the one of *ZSCAN12P1*, supporting the power of the Capture-C technique.

Both genes are located in the same compartment, but the *ZSCAN12P1* gene was up-regulated about 2.6 fold while the *ZNF165* gene was slightly and insignificantly down-regulated (about 0.8 fold).

Figure 3-37: The *ZSCAN12P1* gene is part of a chromatin compartment identified with the help of the Capture-C technique. (next page)

The figure shows the combined results from Raji 4816 cells including Capture-C, ChIP-seq and RNA-seq data for the *ZSCAN12P1* gene in Raji 4816 cells. The composition of the plot was explained in greater detail in Figure 3-29 and Figure 3-30. The Capture-C data from the *ZSCAN12P1* gene is an example of the compartmentalization of interactions. In the upstream section of the *ZSCAN12P1* gene a very sharp cutoff of interactions was apparent and indicated by the left red line. Directly at this position a CTCF protein could be detected prior to and post BZLF1 induction, while a Rad21 was only detected after 15 h of BZLF1 induction. A decrease of interactions per DpnII fragment, even if not as sudden and sharp, could also be seen downstream of the gene (indicated by the right red line). Additionally, the upstream pattern of CTCF and Rad21 proteins was detected as well. It seems as if CTCF and Rad21 flank and define a compartment of DpnII fragments with high frequency interactions.



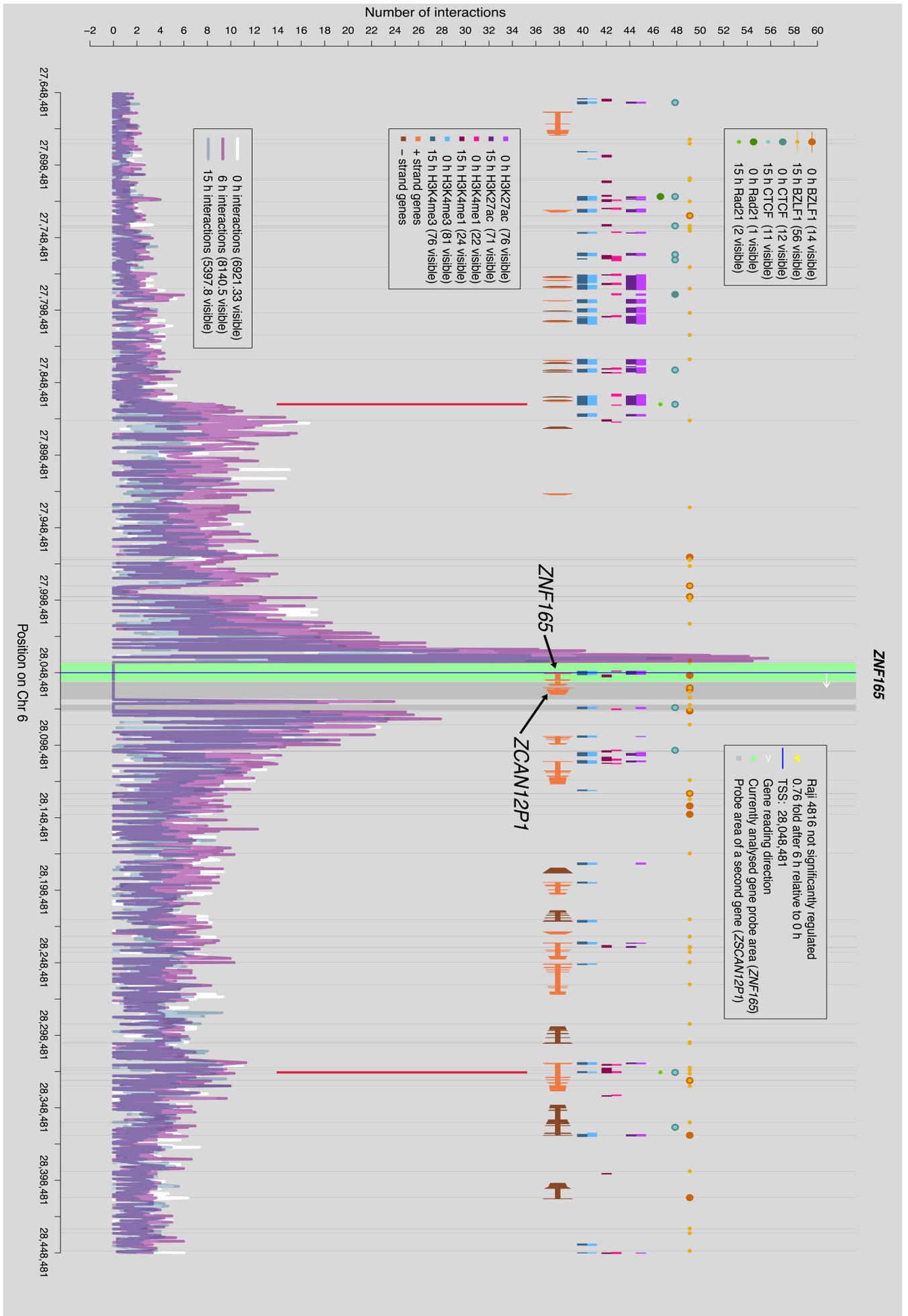


Figure 3-38: The *ZNF165* gene is part of a chromatin compartment identified with the help of the Capture-C technique.

The figure shows the combined results from Raji 4816 cells including Capture-C, ChIP-seq and RNA-seq data for the *ZNF165* gene in Raji 4816 cells. The composition of the plot was explained in greater detail in Figure 3-29 and Figure 3-30. The capture area of the *ZNF165* gene is located directly upstream of the *ZSCAN12P1* gene visualized in Figure 3-37. While *ZSCAN12P1* was 2.62 fold up-regulated, *ZNF165* was 0.76 fold down-regulated, which is not significant. DpnII fragments interacting with the capture area of the *ZNF165* gene showed the same pattern as the *ZSCAN12P1* gene. This supports the idea that the CTCF and Rad21 proteins bracket a compartment, in which interactions occur more frequently compared with DpnII fragments beyond chromatin organizer borders.

3.7.7 The pattern of chromatin interactions in adjacent gene loci

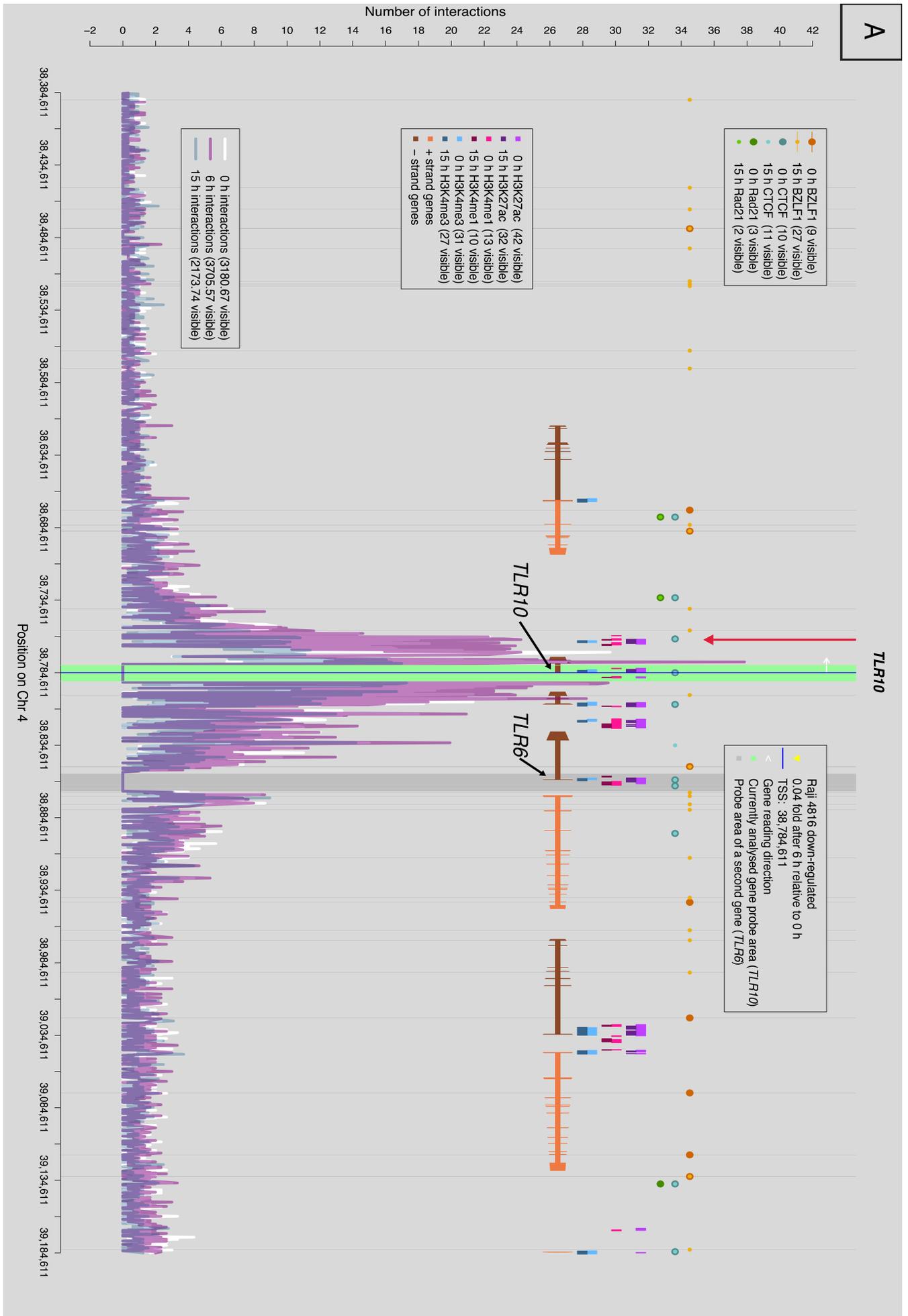
Two genes *ZSCAN12P1* (Figure 3-37) and *ZNF165* (Figure 3-38), which are located in close proximity in a common chromatin compartment, depicted a very similar pattern of chromatin interactions. Since these two genes also seemed to act similarly, two additional examples from the 49 captured genes were selected. (Table 9-7 in the attachment provides the full list of all captured genes)

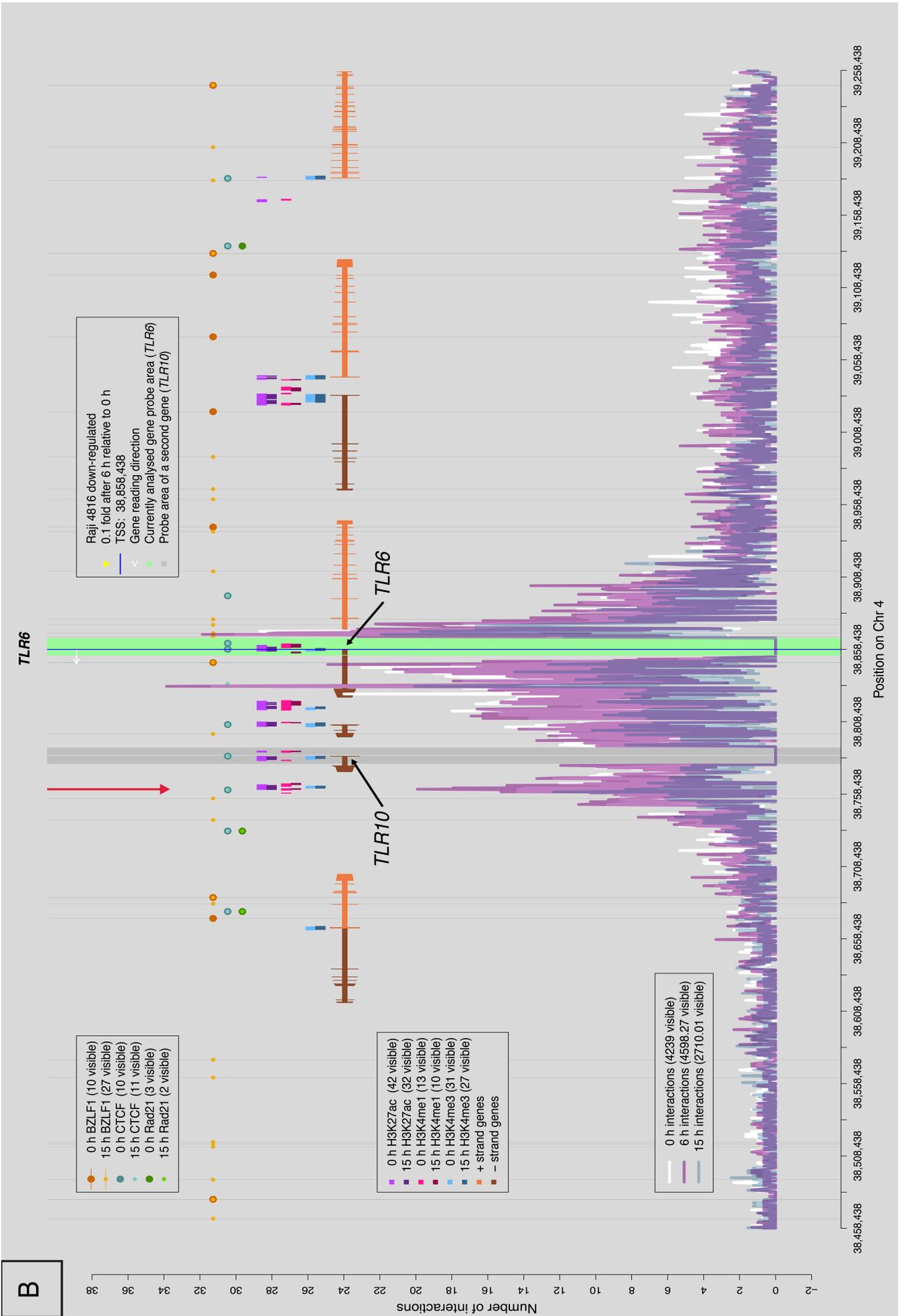
Panel A and B of Figure 3-39 show the toll-like receptor 10 (*TLR10*) and 6 (*TLR6*) gene loci, respectively, which are located in close proximity. Both *TLR10* and *TLR6* were very strongly down-regulated to 4 and 10 %, respectively, and both showed frequent interactions in close proximity of their TSSs.

Both loci shared several DpnII fragments (red arrow), which interacted strongly with the capture regions of both genes. Within these DpnII fragments a CTCF binding site was located and the enhancer marks H3K27ac, H3K4me1, and H3K4me3 prevail prior to and post induction of BZLF1. Neither BZLF1 binding sites were present in the capture regions of both genes nor in the shared DpnII fragments.

Figure 3-39: The adjacent loci of the toll-like receptors 10 (TLR 10) and 6 (TLR 6) show similar interaction pattern with potential enhancer elements. (next page)

The figure shows the combined results from Raji 4816 cells including Capture-C, ChIP-seq and RNA-seq data in Raji 4816 cells. The composition of the plot was explained in greater detail in Figure 3-29 and Figure 3-30. **A** and **B**: Both the *TLR10* (panel A) and the *TLR6* gene (panel B) were strongly down-regulated to 4 and 10 % of their expression levels prior to BZLF1 expression. Both genes have in common that they depicted a strong interaction (indicated by the red arrow) with a cluster of DpnII fragments with enhancer-typical histone marks such as H3K27ac, H3K4me1, and H3K4me3. Additionally, a CTCF binding site was in close neighborhood. BZLF1 sites were absent in the potential enhancer region and in the capture area.





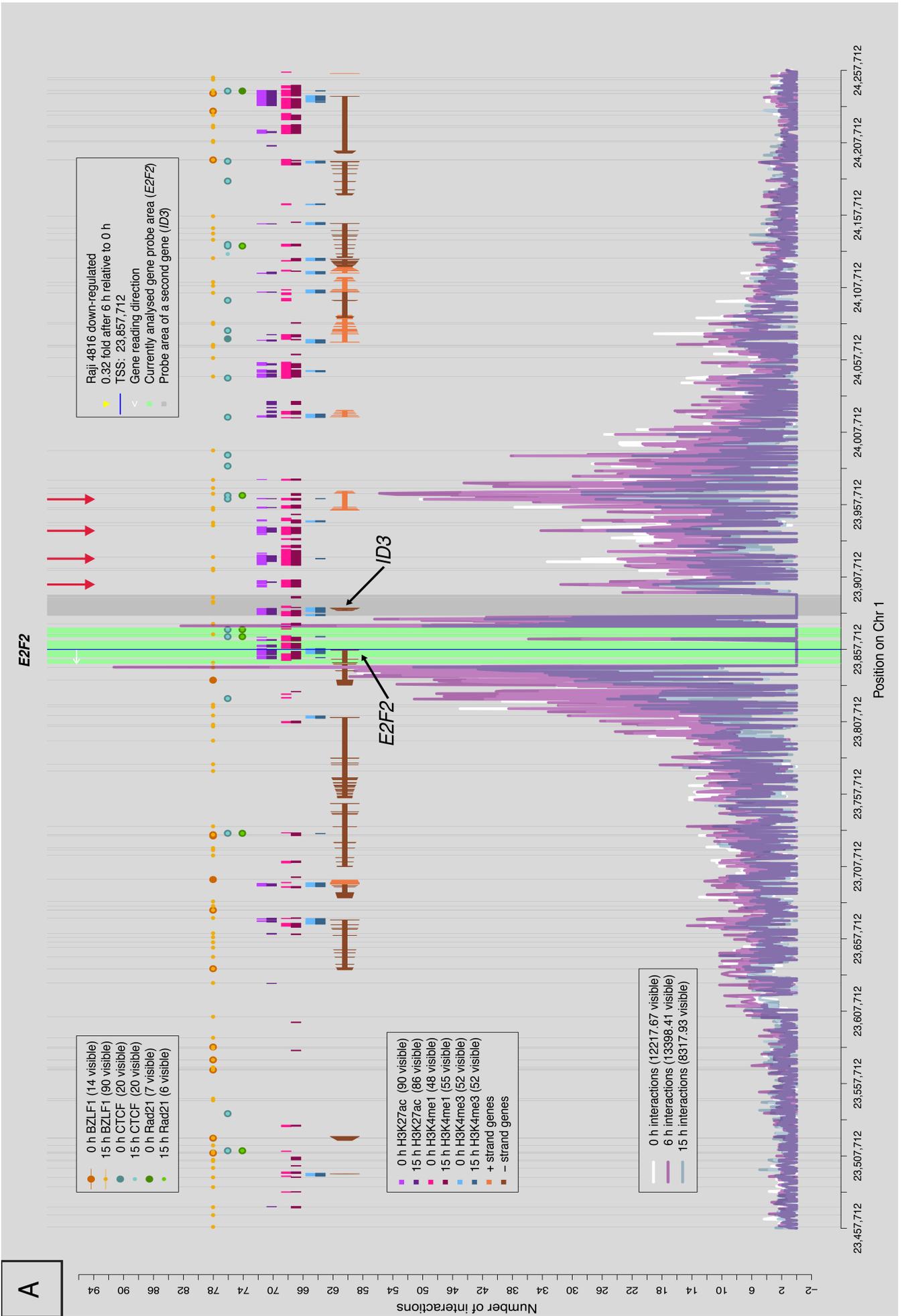
Another pair of genes located in close proximity is the *E2F* transcription factor 2 (*E2F2*) and the inhibitor of DNA binding 3 (*ID3*) shown in panel A and B of Figure 3-40, respectively. The *E2F2* gene was down-regulated to about one third and the *ID3* gene to about 10 % of initial transcript levels prior to the addition of doxycycline.

Downstream of the *E2F2* gene many DpnII fragments with multiple interactions were located. Upstream of the *E2F2* gene four distinct domains with high numbers of interactions showed up (four red arrows), which were characterized by the histone marks H3K27ac and H3K4me1, which are associated with active enhancers. The most distal of the four indicated domains harbored CTCF and Rad21 sites prior to and after induction of BZLF1 and showed more interactions compared with the remaining three sites.

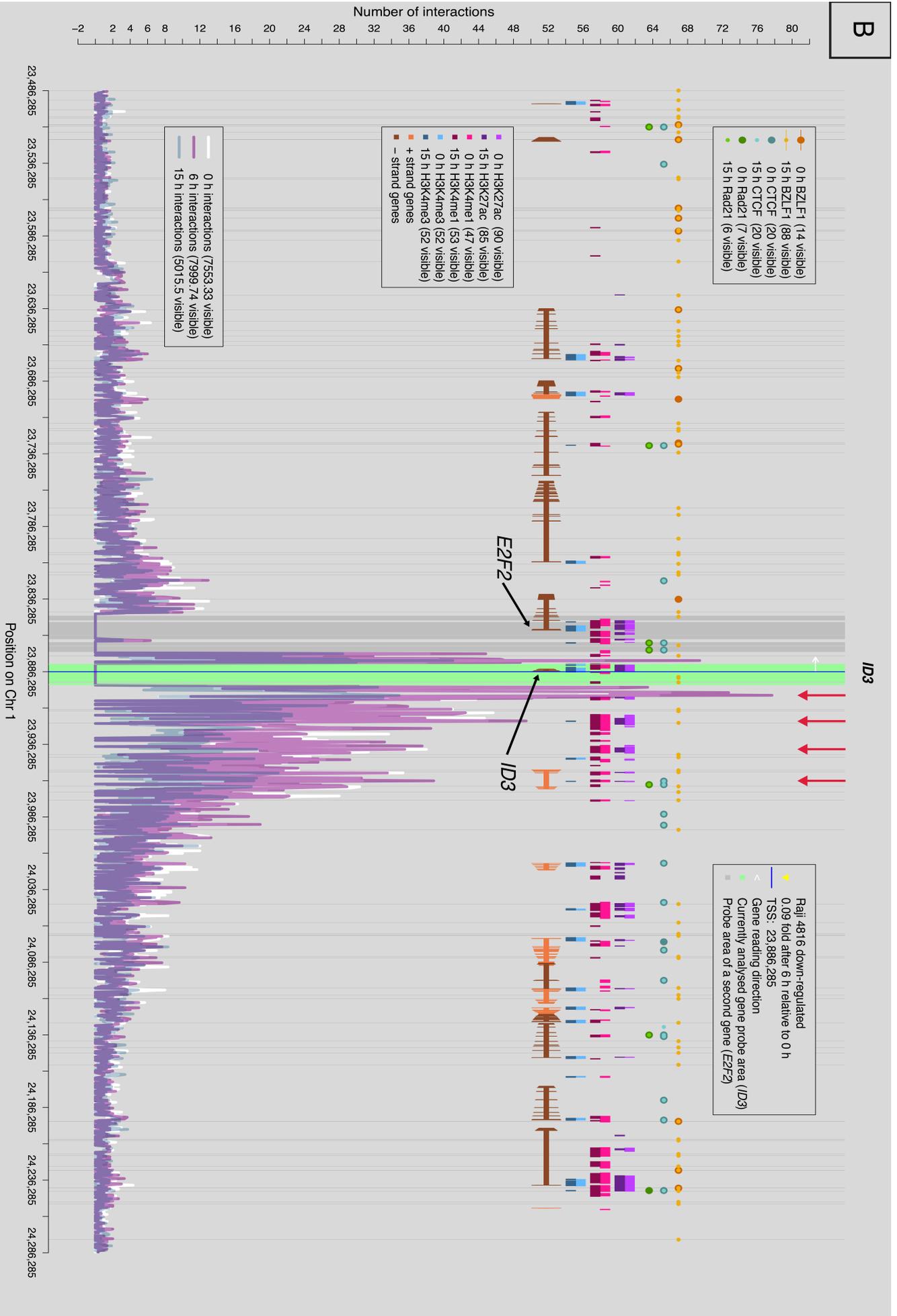
The *ID3* gene (Figure 3-40, panel B) interacts with the same domains (red arrows), which had been identified with the capture region of the *E2F2* gene, but at slight different levels. The domain with the most interactions was the one directly upstream of *ID3*; more distant domains showed fewer interactions.

Figure 3-40: The adjacent *E2F2* and *ID3* genes share four separate regions with enhancer marks. (next page)

The figure shows the combined results from Raji 4816 cells including Capture-C, ChIP-seq and RNA-seq data in Raji 4816 cells. The composition of the plot was explained in greater detail in Figure 3-29 and Figure 3-30. Panel A shows the *E2F2* gene and panel B shows the adjacent *ID3* gene. Both genes were down-regulated, *E2F2* by a factor of 0.32 and *ID3* by a factor of 0.09. **A:** While the downstream region interacted frequently with the *E2F2* gene, the upstream area revealed four particularly strong interaction areas indicated by four red arrows. The most distal area depicted the strongest interaction with *E2F2*. **B:** The *ID3* gene depicted exactly the same interacting regions (red arrows) like the *E2F2* gene but the interactions with the indicated most proximal enhancer region were more frequent.



B



3.7.8 Genes with long-range interactions

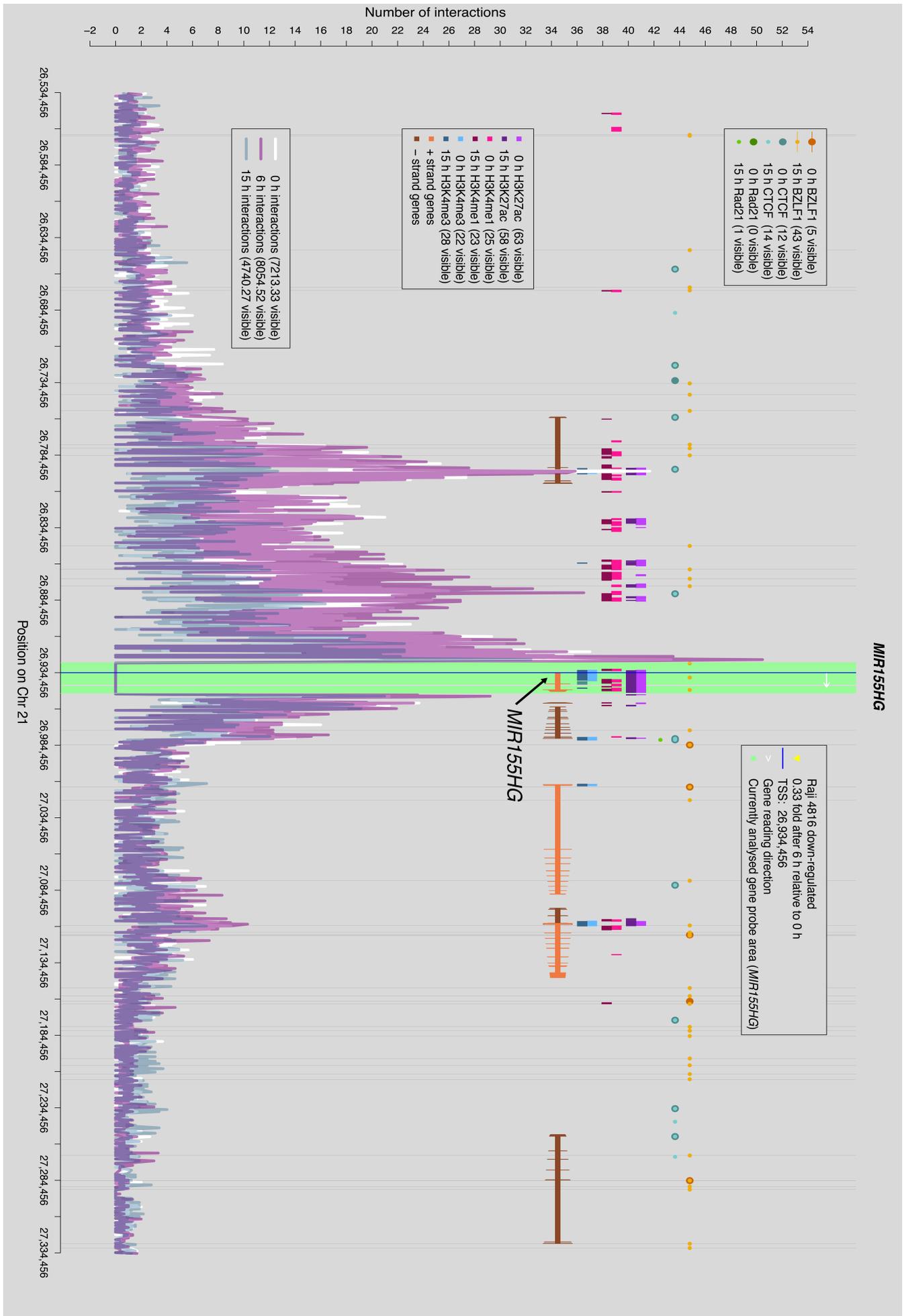
Long-range interactions were already seen in the *HDAC9* (Figure 3-34) and *E2F2* (Figure 3-40) loci and were mostly associated with chromatin organizers. Two especially impressive patterns of long-range interactions were found at the gene locus of the non-coding RNA *MIR155* host gene (*MIR155HG*) and the C-X-C motif chemokine receptor 4 (*CXCR4*). Figure 3-41 shows the *MIR155HG* gene locus, whose expression level was down-regulated to 33 % 6 h after induced BZLF1 expression. Two separate long-range interaction peaks are depicted about 50 kbps and 150 kbps upstream of the *MIR155HG* gene. They comprised domains of several kbps. Both domains were associated with CTCF sites and the histone marks H3K27ac and H3K4me1 prior to and after BZLF1 expression. Upon BZLF1 expression the distal peak interacted less frequently with the capture region, the proximal peak gained several interactions. Both peaks massively lost interactions 15 h post BZLF1 induction, which was especially evident for the distal peak. Three BZLF1 binding sites were identified within the capture region and also three BZLF1 binding sites each were identified closely upstream of the long-range interaction domains.

The second outstanding example of long-range interactions is the *CXCR4* locus (0.04 fold), which was massively down-regulated and is depicted in Figure 3-42. Four domains of long-range interactions with a maximal distance of 270 kbps were detected, which are tagged by red lines. The two distal domains were found associated with CTCF proteins. In addition, the distal domain upstream of *CXCR4* depicted a Rad21 binding site and carried the histone marks H3K27ac and H3K4me1. The two proximal domains were not associated with CTCF sites, but both showed strong H3K27ac and H3K4me1 marks, but few H3K4me3 marks.

The Capture-C data suggest that long-range interactions can be either associated with chromatin organizers or histone marks typical of enhancers.

Figure 3-41: Long-range interaction of potential enhancer elements with the capture region of the *MIR155HG* gene. (next page)

The figure shows the combined results from Raji 4816 cells including Capture-C, ChIP-seq and RNA-seq data in Raji 4816 cells. The composition of the plot was explained in greater detail in Figure 3-29 and Figure 3-30. The *MIR155HG* gene was down-regulated by the factor 0.33 and serves as an example of strong long-range interactions. Two interactions are clearly visible in distances of about 50 kbps and 150 kbps from the TSS, which harbored a CTCF site prior to and post BZLF1 induction. The histone marks H3K27ac and H3K4me1 were both present around the peaks of the interacting DpnII fragments. In this example, the capture region interacted with distant areas of several 10 kbps. The peak of the distal long-range interacting region lost some interactions 6 h post BZLF1 induction, while the proximal peak gained interactions. 15 h post BZLF1 induction both long-range interactions were widely lost.



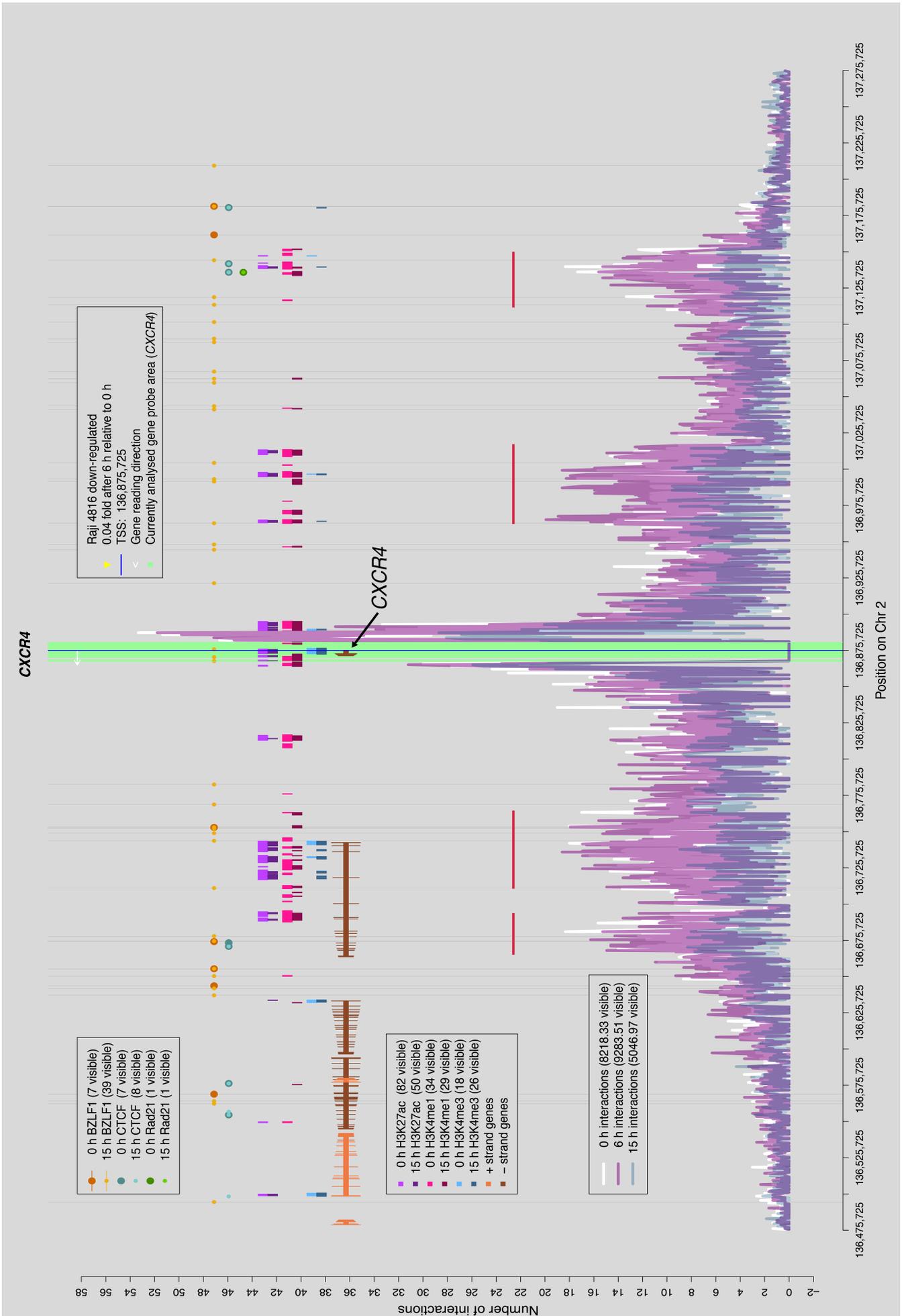


Figure 3-42: Long-range interactions with the capture region of the *CXCR4* gene were either defined by histone marks or by additional chromatin organizers.

The figure shows the combined results from Raji 4816 cells including Capture-C, CHIP-seq and RNA-seq data in Raji 4816 cells. The composition of the plot was explained in greater detail in Figure 3-29 and Figure 3-30. The transcript of the *CXCR4* gene was down-regulated to 4 % compared with its level prior to BZLF1 induction. Four peaks of long-range interactions are depicted (red lines) with a maximal distance of about 270 kbps from the TSS. The two distal long-range interacting regions were associated with two CTCF proteins each and one Rad21 site, while no CTCF molecules were detected for the two proximal interacting regions. Strong H3K27ac and H3Kme1 histone marks indicative of active enhancers were associated, while the H3K4me3 mark was less strongly associated.

4 Discussion

In EBV's life cycle the viral BZLF1 protein mediates the switch from the latent to the lytic phase (Countryman and Miller, 1985; Takada et al., 1986). BZLF1 binds the lytic promoters of viral genes in EBV's genome and activates the expression of these genes (Bergbauer et al., 2010). Additionally, our lab found that the three-dimensional chromatin structure of the EBV episome is disrupted locus specifically upon expression of BZLF1 (unpublished data). The homology in the DNA binding domain of the BZLF1 protein with the cellular transcription factor family AP-1 and the similarity of their cognate DNA binding motifs suggested that BZLF1 might bind cellular DNA and regulate cellular genes (Lieberman et al., 1990; Farrell et al., 1989).

This thesis focuses on the role of BZLF1 in the cellular genome. Two different Burkitt lymphoma B-cell lines were used: DG75 4816 cells, which express no viral proteins except BZLF1 upon doxycycline induction, and Raji 4816 cells, which allow studying BZLF1 in a cell latently infected with EBV. In Raji cells BZLF1 expression switches the cells from EBV's latent to its lytic phase.

I investigated four different aspects to analyze BZLF1's role in the cellular genome. (i) The binding sites and binding motifs of BZLF1 were identified in both cell lines with ChIP-sequencing followed by computational analysis. (ii) Genes differentially regulated upon induced expression of BZLF1 were identified with transcriptome analysis after RNA-sequencing. (iii) The number of genes bound and regulated by BZLF1 was identified by a computational combination of genes regulated after induction of BZLF1 and genes with identified BZLF1 binding site in their promoter sequences. These results revealed that the majority of regulated genes did not contain BZLF1 binding sites in their promoters and recent work in the laboratory indicated that interactions within EBV's chromatin become dissociated upon BZLF1 expression. (iv) Therefore the chromatin architecture of selected cellular genes was analyzed in Capture-C experiments with an emphasis on enhancer elements.

4.1 BZLF1 expression prior to doxycycline induction

In previous years the lab has explored the function of BZLF1 with respect to the regulation of viral genes and their methylation-dependent transactivation. To identify the binding sites of BZLF1 and analyze the changes in the expression levels of viral lytic genes or in the chromatin structure a conditionally inducible BZLF1 was established. To achieve this a pRTR vector was used, which is an improved version of the pRTS vector (Bornkamm et al., 2005). The pRTR vector 4816 (Woellmer et al., 2012) expresses BZLF1, the green fluorescent protein (GFP), and the NGF-receptor upon doxycycline induction.

The human Burkitt-lymphoma cell line Raji is EBV-positive and strictly latently infected in vitro in cell culture. Upon induced expression of BZLF1 from the 4816 plasmid (Figure 3-12 and Figure 3-17), the cells rapidly enter EBV's lytic cycle. Figure 3-1 shows that prior to doxycycline addition the BZLF1 transcript is barely detectable. BZLF1-regulated viral genes such as BMRF1 and BNLF2a are very sensitive to minimal levels of BZLF1 but barely expressed in the absence of doxycycline. Their transcript levels only rose upon BZLF1 induction for 3 h. These data strongly suggested that prior to doxycycline induction cells harboring the 4816 inducible expression plasmid expressed no or only very little and functionally irrelevant levels of BZLF1.

The accuracy and high resolution of NGS, which was used for several of my experiments, revealed that BZLF1 was already expressed prior to doxycycline induction but at levels insufficient to induce EBV's lytic cycle. The Western blot in Figure 3-2, panel A revealed that the BZLF1 protein likely originated from the 4816 plasmid and was expressed prior to induction and independently of the FCS used. The quantification of the Western blot data (Figure 3-2, panel B) revealed that there were about 100 fold more BZLF1 dimers per cell after doxycycline-induced expression of BZLF1. The calculated number of 8.8×10^7 BZLF1 dimers per cell 15 h after induction seems very high when compared with other transcription factor levels (about 10^5 MYC proteins/ cell, Biggin, 2011). The quantification of the BZLF1 dimers is based on the concentration of a bacterially expressed BZLF1 molecule, whose concentration was determined with an SDS page and a BSA standard curve. Another attempt to measure the concentration of the bacterially expressed BZLF1 molecule like a Bradford assay could help to confirm the result.

In the experiment analyzing the expression of BZLF1 responsive viral genes (Figure 3-1) lytic viral genes were not expressed prior to doxycycline-induced BZLF1 expression. After NGS the differential expression of many latent and lytic viral genes was analyzed again 6 h after doxycycline-induced expression of BZLF1 since NGS data are much more sensitive. It was obvious that the read coverage (reads per base) increased dramatically from 0 to 3 h on the EBV genome (Figure 3-3) and increased substantially and further in the following 3 h. The comparison of reads detected on the EBV genome at different time points prior to and after induction of BZLF1 showed that the level of BZLF1 prior to doxycycline induction is not capable to induce the lytic cycle, because the transcription of viral lytic genes was very low and only became substantial after doxycycline induction. An explanation why minor BZLF1 amounts did not activate lytic viral genes prior to addition of doxycycline might be that BZLF1 was reported to activate gene transcription self-synergistically and only once a critical number of BZLF1 molecules is bound at a BZLF1-regulated promoter (Carey et al., 1992). The mechanism of self-synergistical promoter activation of transcription factors was also described for GAL4 in yeast (Carey et al., 1990) and for NF- κ B in mouse cells (Pettersson and Schaffner, 1990).

Clearly, the non-induced expression of BZLF1 was unexpected and went unnoticed for years in other experiments ongoing in the group. Only the more sensitive NGS technique detected the expression of BZLF1 prior to induction, and this led to an interesting finding in cellular DNA, when BZLF1 is expressed at lower levels.

4.2 BZLF1 binding sites in the cellular genome

In 2009 an attempt to enumerate all transcription factors in the human genome came up with about 1,900 factors (Vaquerizas et al., 2009). The binding of transcription factors to their cognate motifs is regulated by several criteria such as chromatin accessibility, nucleosome occupancy, and steric hindrance by other factors, for example. Certain DNA binding motifs contain CpG di-nucleotides. Their status of cytosine methylation is of particular interest because it might interfere with factor binding providing an additional level of regulation.

It is commonly accepted that CpG methylation reduces or even inhibits the binding of transcription factors. Only a few examples do not follow this rule: the DNA binding motif of

Sp1 is GC-rich but its cytosine methylation does neither impair Sp1's binding nor its transcriptional activities (Höller et al., 1988; Harrington et al., 1988). In certain cases the binding site of the CCAAT-enhancer-binding protein alpha (C/EBP α) has to be methylated to activate a tissue-specific set of promoters (Rishi et al., 2010). In recent years additional transcription factors that lack the known methyl-CpG binding domain (MBD) have been identified, which bind methylated DNA motifs (Zhu et al., 2016). The zinc finger protein 57 binds a specific methylated DNA motif in imprinting control regions and is suggested to protect these regions from de-methylation during early embryogenesis (Quenneville et al., 2011). In a systematic study 47 transcription factors, including KLF4, were identified, which bind to methylated DNA. Their tasks in the cell still need to be elucidated (Hu et al., 2013).

A transcription factor with a very dedicated function that binds a methylated motif is EBV's transcriptional activator BZLF1. It regulates the switch from the latent to the lytic cycle in EBV's life cycle (Countryman and Miller, 1985; Takada et al., 1986). BZLF1 was described to preferentially bind methylated ZREs in the viral genome *in vivo* and *in vitro* (Bhende et al., 2004; Bergbauer et al., 2010; Kalla et al., 2010; Woellmer et al., 2012).

It is interesting to note that in my experiment the expression of BZLF1 at lower levels prior to doxycycline addition revealed that BZLF1 exclusively binds to CpG ZREs in DG75 4816 and Raji 4816 cells (Figure 3-5 and Figure 3-8). A control experiment with the parental DG75 cell line excluded an unspecific pull-down of DNA sequences with CpG ZREs (Figure 3-10) indicating the specificity of my finding. The comparison of K_d values of several methylated CpG ZREs (K_d between ~ 13 nM and 20 nM) and different nonCpG ZREs (K_d between ~ 38 nM and 92 nM) (Bergbauer et al., 2010) explains why BZLF1 exclusively binds CpG ZREs when expressed at lower and probably limiting levels. The K_d values of non-methylated CpG ZREs are far higher (K_d between ~ 113 nM and 484 nM) than the ones identified for nonCpG ZREs (K_d between ~ 38 nM and 92 nM, Bergbauer et al., 2010), indicating that the identified CpG ZREs in cellular chromatin of DG75 and Raji cells were indeed methylated. This view is supported by my finding that the BZLF1 binding motifs identified in the thousands of BZLF1 binding sites identified in the cellular genome are nearly identical to the binding motifs identified in the genome of the prototypic EBV strain B95.8 (Bergbauer et al., 2010) and in the EBV infected Akata cell line (Ramasubramanian et al., 2015).

Upon induced expression of BZLF1, which reaches very high levels, far more binding sites were identified than prior to induction (Figure 3-5 and Figure 3-8). This finding is probably linked with the fact that BZLF1 is not longer limiting but binds motifs with a K_d higher than that of methylated CpG ZREs (Bergbauer et al., 2010) at high concentrations 15 h after BZLF1's induced expression (Figure 3-2). A concentration-dependent difference in BZLF1 motif binding was described previously for the promoter and upstream region of the BZLF1 gene locus (Lieberman and Berk, 1990) supporting the assumption that the regulation of BZLF1 target genes depends on local concentrations of BZLF1. The high number of approximately 90,000 BZLF1 binding sites common in both cell lines 15 h after BZLF1 induction (Figure 3-11) speaks for a prevalent but high specificity of BZLF1 binding.

When high doses of BZLF1 proteins were expressed, in each cell line far more than 140,000 binding sites were identified, which is about seven fold higher than the number of annotated genes (20,000) in the human genome (Clamp et al., 2007; Ezkurdia et al., 2014). The short binding motif of BZLF1 in B-cells (Figure 3-5 and Figure 3-8) might explain the large number of binding sites, of which only few are expected to play a critical role in gene regulation. It is known from AP-1 binding site analysis that not all genes comprising an AP-1 binding site are regulated (Zhou et al., 2005) but it is still unclear how exactly transcription factors select their binding sites and how they subsequently influence gene expression (Slattery et al., 2014). The work with BZLF1 seems to be in line with this statement.

Upon induction about 600-fold more dimeric BZLF1 proteins were present in Raji and DG75 cells (Figure 3-2) than BZLF binding sites were detected (Figure 3-5 and Figure 3-8). By statistical distribution, the BZLF1 binding motif identified in Raji cells 15 h after doxycycline addition (Figure 3-8) should exist in human DNA about 4.3 million times. Even though there are enough BZLF1 dimers per cell, they do not bind all these sites. The fact that several criteria i.e. DNA sequence, the chromatin states and histone marks influence the choice of the binding (Ernst and Kellis, 2013) might explain why not all of the statistically possible binding sites were bound and why certain BZLF1 peaks identified prior to induction were lost 15 h after induction (Figure 3-6 and Figure 3-9). Fifteen hours after BZLF1 expression chromatin interactions decreased massively (Figure 3-30 to Figure 3-42), which may change the accessibility of some BZLF1 binding motifs and block BZLF1's admission to previously recognized and bound sites.

From an evolutionary point of view the binding of BZLF1 dimers to cellular motifs may use the cellular genome as a sink to reduce the likelihood of BZLF1 to bind too early to the viral genome and only after a certain threshold of BZLF1 molecules is reached, a sufficient number of viral sites are occupied and the lytic phase may start. This scenario would explain the immediate expression of many viral genes, which are switched on and reach high expression levels in a short time after induction. In this model a gradual increase of BZLF1 would result in a dichromatic switch that flips. This mechanism, which is purely hypothetical in the moment, would also prevent an unwanted, premature activation of EBV's lytic phase that might make the cells vulnerable to an EBV-specific immune response of the host organism.

4.3 Differential gene expression upon induction of BZLF1

More than 140,000 BZLF1 binding sites were identified in both cell lines 15 h after doxycycline-induced BZLF1 expression, but how BZLF1-controlled genes are regulated remains elusive.

In mammalian cells the RNA polymerase II accomplishes the transcription of mRNAs in four steps: the initiation complex is formed; the polymerase II escapes the promoter; the DNA is transcribed to RNA; and transcription is terminated. In eukaryotic cells the resulting mRNAs are commonly polyadenylated and capped at the 3' and 5' end, respectively.

The relative abundance of mRNAs was determined in two different Burkitt lymphoma B-cell lines prior to and after the expression of the viral transactivator BZLF1 with the help of next generation sequencing to identify BZLF1-regulated cellular genes. In both cell lines, DG75 4816 and Raji 4816, only few regulated genes were detected 3 h after doxycycline-induced expression of BZLF1 (Figure 3-13 and Figure 3-20), but 6 h after BZLF1 induction several tens of up- and down-regulated genes were identified in DG75 4816 (Figure 3-14, panel B). Only the introduction and normalization with the ERCC spike-in control RNAs revealed the actual regulation of cellular genes in Raji 4816 cells. Few genes were up-regulated and more than 7,000 genes were down-regulated 6 h post induction (Figure 3-21, panel C). An independent quantification method like qPCR could be helpful to confirm the down-regulation of selected genes found in my transcriptome analysis.

The two groups of the 1,000 most strongly down-regulated genes and the 3,300 most stable genes were used for a KEGG pathway analysis. The most stable genes were found in pathways of important metabolic pathways, while the strongest down-regulated genes were found in pathways associated with immune cells and host immune response.

FRA-1, a member of the AP-1 transcription factor family, was reported to negatively regulate IgG production (Groetsch et al., 2012) and to suppress genes regulating inflammation and immune responses (Rajasekaran et al., 2013). Like the FOS protein, the FRA-1 protein has a protein sequence homology with the BZLF1 protein (Farrell et al., 1989), interacts with the JUN protein to bind AP-1 binding sequences, but lacks the transactivation domain present in FOS and BZLF1 (Cohen and Curran, 1988; Cohen et al., 1989). BZLF1 might not only have adapted the transactivation abilities of the JUN/FOS complex, but depending on the situation, the repressive effects associated with the FRA-1 protein.

Since BZLF1 is known to be a transcriptional activator in the viral genome, the identification of several thousand down-regulated cellular genes was unexpected. An adverse effect of the inducing agent doxycycline that might cause this down-regulation could be excluded with the aid of a control experiment (Figure 3-15 and Figure 3-22). A BZLF1 mutant, which lacks the transactivation domain but still binds to DNA with a K_d of about 16 nM (Schaeffner, 2015), was used to exclude that BZLF1 down-regulates genes by competing with cellular DNA binding transcriptional activators for the DNA binding site like described for proteins of the AP-1 family (Sato et al., 1992). Upon expression of the truncated BZLF1 protein no genes except the co-induced NGF-receptor were identified to be up- or down-regulated (Figure 3-16 and Figure 3-23) indicating that this mutant did not induce the activation or suppression of cellular genes. A similar finding was already described for other members of the AP-1 protein family lacking their transactivation domain (Wisdom and Verma, 1993). The results indicate that BZLF1 does not down-regulate cellular transcripts by competing for DNA binding sites with other transcriptional factors.

The deletion of the transactivation domain might prevent BZLF1 to get in contact with other proteins to stabilize an enhanceosome. BZLF1 is also a regulation factor of the viral DNA replication complex (Baumann et al., 1999) where the activation domain of BZLF1 gets in contact with the viral helicase (BBLF4), the primase subcomplex (BSLF1 and BBLF2/3)

(Gao et al., 1998), the DNA polymerase accessory factor (BMRF1) (Zhang et al., 1996), and the polymerase subunit BALF5 (Kiehl and Dorsky, 1991; Baumann et al., 1999).

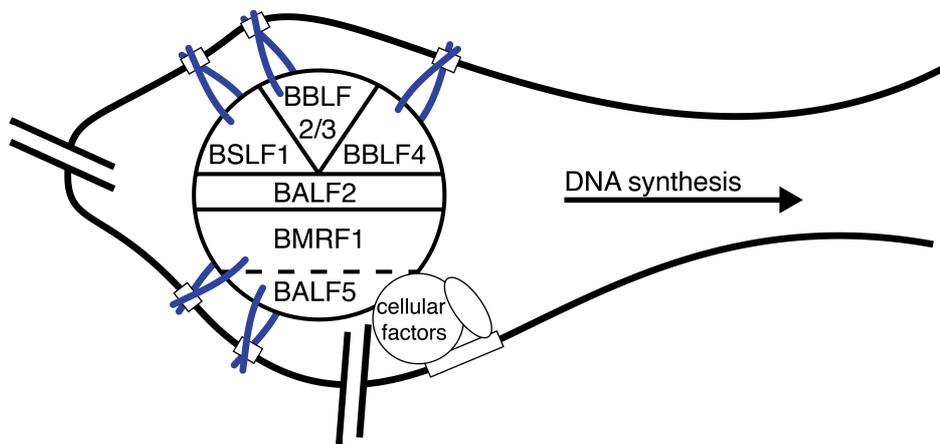


Figure 4-1: The BZLF1 transactivation domain interacts with other viral proteins to stabilize the viral initiation complex.

Suggested and modified from (Baumann et al., 1999).

A similar multifactorial model of transcriptional regulation was proposed with the central cellular CBP in mammalian cells, which interacts with the pre-initiation complex and is synergistically stabilized by several transcription factors (Merika et al., 1998). The transactivation domains of several cellular transcription factors interact with CBP (Dai et al., 1996; Abdel-Wahab et al., 2009; Garcia-Rodriguez and Rao, 1998) as BZLF does. BZLF1 might use its transactivation domain to interact with CBP (Adamson and Kenney, 1999) to reduce the CBP-mediated transcription activity like the adenoviral E1A protein does by competing with proteins of the AP-1 family (Arany et al., 1995; Bannister and Kouzarides, 1995). This way BZLF1 could actively participate in the reduction of cellular mRNA levels by competing with cellular transcription factors for CBP as suggested by Adamson and Kenney earlier (1999).

It remains unclear if the massive down-regulation of cellular mRNAs is directly caused by the expression of BZLF1 or originates from other BZLF1 induced viral genes. The most promising candidate gene to down-regulate the transcriptome of the cell is the viral protein BGLF5, which is responsible for the host shutoff mechanism of EBV. BGLF5 degrades cellular mRNAs actively by its RNase activity (Buisson et al., 2009; Rowe et al., 2007; Horst et al., 2012). After 3 h of doxycycline-induced BZLF1 expression BGLF5 was up-regulated about 30 fold (data not shown) and about 160 fold after 6 h of BZLF1 expression (Figure 3-3). The induced expression of BGLF5 could well explain the down-regulation of the cellular

transcripts observed in Figure 3-20 (panel C) and Figure 3-21 (panel C). To determine the role of BZLF1 in reducing the amount of cellular transcripts the DG75 4816 cell line, which does not express other viral proteins, would need to be re-tested including the ERCC spike-in control RNAs for normalization to exclude a similar effect as observed for Raji 4816 cells. Additionally, a BGLF5 knockout mutant could be used to avoid the problem that BZLF1 might indirectly cause phenotypes by inducing the expression of other viral proteins. Until the experiment with DG75 cells have not been conducted with ERCC spike-in controls it is impossible to distinguish cellular genes, which are directly regulated by BZLF1 from cellular genes, which are indirectly regulated by viral mechanisms like the virus host shutoff.

In contrast to the results found in the RNA-seq section of my thesis a recently published study with BZLF1 identified about 2,000 mainly up-regulated genes 24 h after BZLF1 induction in EBV positive cell lines (Ramasubramanian et al., 2015). This result strongly contradicts the results presented in this thesis and is probably attributed to the missing external control RNAs, which these authors did not consider in their experiments. Additionally, the induction period of 24 h seems inappropriate to identified genes directly regulated by BZLF1 because secondary effects downstream of BZLF1-regulated genes are likely to occur. An EBV-free cell line was not considered for the published NGS analysis (Ramasubramanian et al., 2015), which would be necessary for a distinct assertion.

4.4 Regulated cellular genes with BZFL1 binding sites

Binding sites of BZLF1 and cellular genes regulated 3 and 6 h after the induction of BZLF1 were identified in cells harboring or lacking EBV. BZLF1 is known to be a promoter factor in the viral genome (Bergbauer et al., 2010) suggesting that regulated cellular genes might also contain BZLF1 binding sites within their promoter regions. The promoter region was defined as the region -5000/ +1000 bps from the TSS since BZLF1 binding sites are known to be located in close proximity to viral TSSs.

In DG75 4816 cells, which are not infected with EBV only a fraction of the regulated genes contained a peak within their promoter regions (Figure 3-26). This is unexpected at 3 h post induction because no other viral factors act in these cells and secondary effects are unlikely given the short induction period. Six hours after BZLF1 induction 42 % (Figure 3-26, panel B) of the regulated genes did contain a BZLF1 binding site within their defined

promoter region. Assuming an even distribution of all 230,213 identified BZLF1 peaks in the human genome of 3.2 giga bases a sequence of 6,000 bps in length (-5,000/ +1,000 bps from TSS) harbors a single BZLF1 binding site with a likelihood of 43 % very close to the fraction of genes (42 %), which do comprise a BZLF1 peak within their promoters (Figure 3-26, panel B). Up-regulated (51 %) genes were more likely to contain a BZLF1 motif within their promoters compared with down-regulated genes (33 %). Genes, which do not encompass a BZLF1 binding site within their promoter regions, need to be regulated by a different mechanism or secondary effects. It is difficult to make a final conclusion because the transcriptome analysis of DG75 cells was conducted without ERCC spike-in control RNAs. It cannot be excluded that a shift of the transcriptome similar to the one in the Raji 4816 cell line (Figure 3-21, panel C) might change the interpretation of this part of my results dramatically. Thus it is mandatory to repeat the DG75 4816 experiments with external ERCC spike-in controls to address this uncertainty.

In Raji 4816 cells many more BZLF1 regulated genes were identified than in DG75 4816 cells. Three hours after BZLF1 induction fewer than 50 % of the regulated genes could be associated with BZLF1 binding sites in their promoters (Figure 3-27, panel A). Since BZLF1 is the first gene expressed in the lytic cycle it seems unlikely at first glance that BZLF1 controls genes, which do not comprise a BZLF1 binding site 3 h after induction.

The fact that about 12 % of all genes regulated 6 h after BZLF1 expression were already bound prior to induced expression of BZLF1 but not regulated (Figure 3-27, panel D), supports the assumption that the expression of BZLF1 at lower levels does not influence the expression of cellular genes.

Six hours after BZLF1 induction more up-regulated genes could be associated with a peak compared with down-regulated genes (Figure 3-27, panel B), a finding similar to DG75 4816 cells after 6 h of BZLF1 induction (Figure 3-26, panel B).

It is also noticeable that more BZLF1 peaks were found to be associated with less strongly regulated genes (Figure 3-27, panel B and D). In case of an even distribution the chance for one of the 144,877 peaks identified in Raji 4816 cells to be found within the 6,000 bps long TSS sequences is about 27 %. Both, up- and down-regulated genes exceed this value. This make it likely that BZLF1 induced regulation of target genes involves the promoter sequence especially for up-regulated genes. In contrary more than 50 % of the regulated genes did not contain a BZLF1 binding site suggesting that other viral or cellular

factors or another mechanism might play decisive roles. It is a consistent finding that more frequently up-regulated genes comprised a BZLF1 binding sites compared with down-regulated genes. This finding might indicate that BZLF1 acts as a promoter factor in case of up-regulated genes but uses other mechanisms to down-regulate genes. Such a mechanism could be an alteration in the chromatin architecture that might interfere or disrupt enhancer-promoter interactions.

4.5 Chromatin interactions of strongly regulated cellular genes

The correlation of BZLF-regulated genes and BZLF1 binding sites within their promoters revealed that BZLF1 does not act only as a promoter factor but regulates many genes without BZLF1 binding sites by alternative mechanisms. Regulating the interaction between promoters and their enhancers could be one of them. The modulation of enhancer looping by EBV's proteins EBNA 3A, 3B, and 3C was already observed and is described in McClellan et al., 2013. BZLF1 is known to bind enhancer elements in the EBV genome (Lieberman et al., 1990) and certain chromatin interactions within EBV DNA undergo modifications upon expression of BZLF1 (unpublished data). The Capture-C experiments were performed to learn if the expression of BZLF1 results in the establishment or disruption of three-dimensional chromatin structures, which might explain the regulation of certain BZLF1-controlled cellular genes.

The human genome is about 20,000 times larger than EBV DNA indicating that the organization of human DNA must be far more complex. The complete cellular DNA, which is organized in chromosomes, is compacted in the nucleus of the cell. The chromosomes comprise territories (Cremer and Cremer, 2001; Cremer, 2010), which are made up of compartments of about 5 megabases (Mb) with open (A) and closed (B) chromatin (Lieberman-Aiden et al., 2009; Rivera and Ren, 2013). The compartments consist of several topologically associated domains (TADs) of about 1 Mb, which are frequently flanked by CTCF proteins (Dixon et al., 2012; Nora et al., 2012). The TADs are further compartmentalized to sub-TADs with the help of the chromatin organizers CTCF and cohesin (Phillips-Cremins et al., 2013). Most of the interactions between cis-regulatory elements occur within these compartments (Heinz et al., 2010; Rada-Iglesias et al., 2012;

Zhu et al., 2013). In my work a combination of capturing chromosome conformation (3C) (Dekker et al., 2002) and a DNA-RNA hybridization pull-down approach, termed Capture-C technique (Hughes et al., 2014), was used to identify far distant interactions with strongly regulated genes.

All analyzed genes showed a similar kinetics of chromatin interactions: Six hours after induction of BZLF1 the total numbers of chromatin interactions per locus increased but were strongly reduced 15 h after doxycycline-induced BZLF1 expression. This pattern included up-, down- or non-regulated genes (Figure 3-30 to Figure 3-42) suggesting that the notified chromatin interactions do not correlate with the directionality of gene expression. Why the numbers of interactions increased 6 h after induction of BZLF1 remains unclear, but there is a possible explanation for the massive loss of chromatin interactions 15 h after doxycycline-induced BZLF1 expression. Upon induction of the lytic cycle EBV expresses two proteins, which are capable of introducing double strand breaks to DNA. One is the BGLF5 gene product, which was already described discussing the viral mechanism of host cell shutoff (chapter 4.3). BGLF5 is capable of cutting mRNAs (Buisson et al., 2009) as well as DNA (Baylis et al., 1989). BGLF5 was shown to induce DNA damage in the cellular genome (Wu et al., 2010) and is involved in the maturation of the viral DNA (Feederle et al., 2009). In case BGLF5's DNase activity would be involved in the reduction of chromatin interaction in the cellular genome it appears odd that this reduction is not already visible 6 h after induction of BZLF1 in the Capture-C experiments. An explanation might be that BGLF5's DNase activity might be blocked as long as it cuts mRNAs. The second viral gene product is BALF3, which uses its nuclease activity to support efficient virion production (Chiu et al., 2014b). Additionally, BALF3 mediates genome instability and induces DNA strand breaks (Chiu et al., 2014a). In my RNA-seq experiments BALF3 was not regulated after 3 h but moderately (13 fold) up-regulated after 6 h of BZFL1 induction. Assuming that BALF3 is expressed at much higher levels 15 h after induction it might be critically involved in modifying the cellular chromatin architecture. It needs to be addressed if these indirect mechanisms are involved in the massive alterations of cellular chromatin in cells that undergo the viral lytic cycle.

BZLF1 was reported to prevent the accumulation of cellular DNA damage proteins at double strand breaks in cellular DNA (Yang et al., 2015), but DNA damage response proteins get recruited to replication compartments (Hau et al., 2015; Kudoh et al., 2005), where EBV DNA is amplified during the lytic phase (Takagi et al., 1991; Daikoku et al.,

2005). It thus appears that BZLF1 directly or indirectly modifies central functions of the host cell to support EBV's lytic phase. Redirecting the cellular DNA repair machinery might be responsible for the loss of chromatin interactions 15 h post induction (Figure 3-30 to Figure 3-42) because damaged cellular DNA might not be repaired during EBV's lytic phase and becomes disorganized.

The loss of chromosomal interactions could serve two purposes: (i) remaining cellular transcription is further reduced, which was observed after the disruption of the chromatin architecture in a different BZLF1-independent context (Phillips-Cremins et al., 2013). (ii) the re-organization of cellular chromatin might free space for EBV's replication compartments as has been reported for herpes simplex virus 1, which displaces host chromosomes during the expansion of its replication compartments (Monier et al., 2000).

Nevertheless, my experiments did not prove a direct link between BZLF1 binding sites and the gain or loss of chromatin interactions. The only exception was the strongly down-regulated *MYC* gene (Figure 3-31). BZLF1 bound several DpnII fragments within the HCE (Figure 3-31), which loops to the promoter of the *MYC* gene. The two DpnII fragments with the most frequent interactions with the *MYC* gene are indicated by A and B in Figure 3-31. Both, position A and B were bound by BZLF1 prior to and after induced expression of BZLF1 and contain exactly the same DNA sequence comprising three CpG ZREs (Figure 4-2, "BZLF1 binding motifs"). If the binding of BZLF1 is responsible for the reduction of interactions with the *MYC* promoter still needs to be examined. The mutation of all six CpG ZREs with the CRISPR/Cas9 technology would be a possibility, even though the repetitive character of the sequence might be an obstacle.

Another question is why the HCE enhancer was not disrupted at low concentrations of BZLF1 already. The answer probably is that BZLF1 does not bind the site long enough to cause an effect although BZLF1 binding is clearly detected in ChIP-seq experiments prior to (0 h) doxycycline-induced BZLF1 induction (Figure 4-2). At higher BZLF1 concentrations (15 h) at least one of the three BZLF1 binding motifs (Figure 4-2, "BZLF1 binding motifs") might be bound anytime and enable BZLF1 to cause the reduction of interactions. This hypothesis is supported by a closer look at the sequencing-depth-normalized BZLF1 ChIPs ("α-BZLF1") and MACS2-called peaks ("MACS2 peak") in the IGV browser (Figure 4-2). Prior to and post induction of BZLF1 the binding sites were detected by the peak caller, but the signal 15 h after induction is much stronger, indicating that BZLF1 bound more frequently.

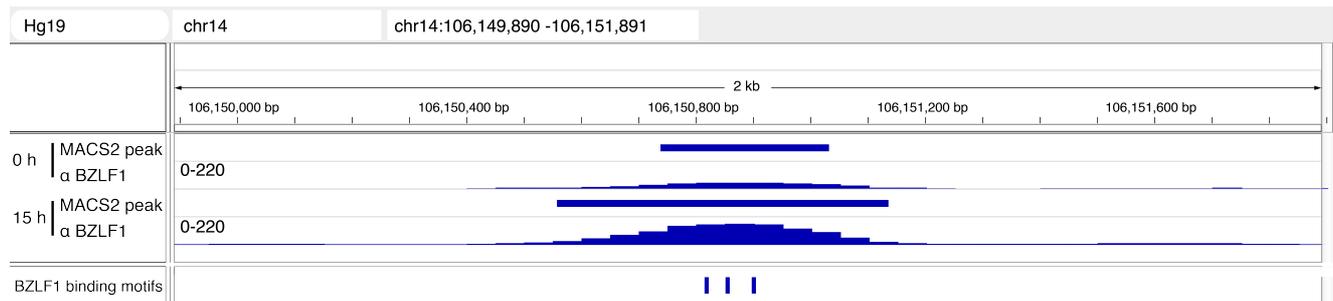


Figure 4-2: BZLF1 binding site in the heavy chain enhancer (HCE) (Figure 3-31, indicated with B) upstream of the *MYC* locus.

The figure shows a snapshot in the IGV browser of chromosome 14 on Hg19 at position #106,149,890-#106,151,891, which reflects the Capture-C peak B in Figure 3-31. Five tracks are shown. The first two tracks show the peak identified by MACS2 and the α -BZLF1 ChIP prior to doxycycline-induced BZLF1 induction (0 h). The y-axis shows the track height from 0 to 220 reads. Tracks three and four show the same information after 15 h of BZLF1 induction. The fifth track shows three CpG ZRE motifs within the peak. Prior to induction the sequencing-depth-normalized BZLF1 peak is much smaller than 15 h post doxycycline-induced BZLF1 expression. The peak strength reflects the binding frequency and the residence time of BZLF1 at this locus.

Another strongly down-regulated gene was the tumor suppressor gene *BTG2* (Mao et al., 2015). The gene was flanked by several CTCF and cohesin proteins, which might probably form a loop. The frequency of interactions within this loop was much higher than beyond. This observation was also found for several other analyzed genes (Figure 3-30, Figure 3-34, Figure 3-37, Figure 3-38, Figure 3-40, Figure 3-42) and might reflect the previously mentioned sub-TADs. The chromatin is probably organized within a loop to increase the chances that certain parts of the loop get in close proximity by the constant movement of the chromatin (Chubb et al., 2002) to form chromatin interactions under appropriate conditions.

The analysis of my Capture-C results revealed several special patterns of chromatin interactions. Certain genes depicted frequent interactions with chromatin in a distance of several hundred thousand base pairs (Figure 3-30, Figure 3-41, Figure 3-42). Depending on the existence or lack of the H3K4me3 mark, which is associated with promoters (Bernstein et al., 2005; Pokholok et al., 2005; Heintzman et al., 2007), and H3K4me1 and H3K27ac marks, which are associated with active enhancer sites (Creighton et al., 2010; Rada-Iglesias et al., 2011; Heintzman et al., 2007), these interactions can be assigned to TSSs or enhancer structures, which form to support and enhance the expression of genes. The formation of a loop brings an enhancer in close proximity to a promoter and increases the number of chromatin interactions in Capture-C experiments. Figure 4-3, panel A shows a miniaturized version of the *BTG2* locus (Figure 3-30) with two marked regions (1 and 2) showing enhancer related chromatin marks, chromatin organizers and BZLF1 binding sites.

In Figure 4-3, panel B the indicated regions 1 and 2 of panel A are magnified and translated to a hypothetical model how chromatin organization might look like at this loci. Region 1 forms a loop, which harbors several BZLF1 binding sites and is stabilized by three CTCF pairs and cohesin (Rad21). Region 2 might also form a loop to contact the *BTG2* gene with its region harboring enhancer marks. This would explain the increased numbers of interactions with the captured region at position 2 (dashed) in Figure 4-3, panel A.

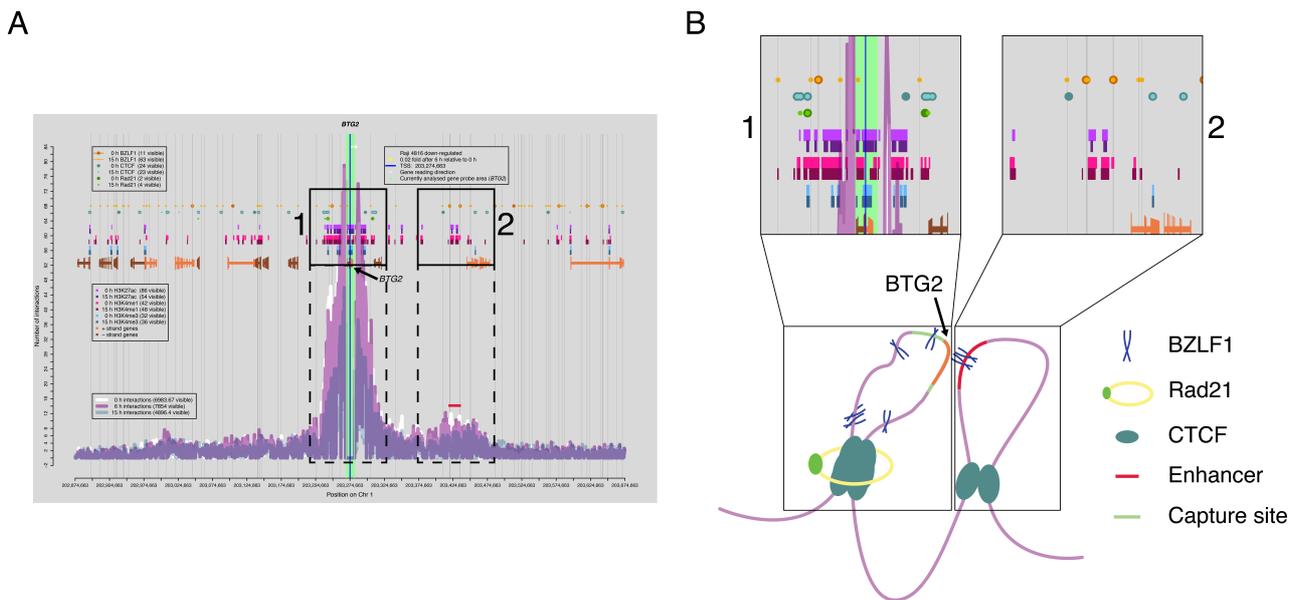


Figure 4-3: A model of the possible chromatin structure at the *BTG2* locus.

A: The figure shows a miniaturized version of the *BTG2* locus (Figure 3-30) and indicates regions of interest within the two squares 1 and 2. **B:** A model of the possible chromatin structure was built on the bases of the data from panel A, including BZLF1 binding sites, chromatin organizers, enhancer and capture sites. Region 1 might form a loop within the three CTCF sites, which are stabilized by cohesin (Rad21). Several BZLF1 sites might get in close proximity to the *BTG2* promoter. Region 2 might form another loop within the two depicted CTCF sites, which harbors chromatin marks associated with enhancers. The model shows how this enhancer might bring additional BZLF1 molecules in close proximity to the *BGT2* gene.

Higher frequencies of chromatin interactions may also result from combining several TSSs into transcription factories (Iborra et al., 1996; Osborne et al., 2004), where many cellular genes get transcribed. This mechanism was suggested in a study describing chromatin interactions by ChIA-PET for the first time (Fullwood et al., 2009). This model would fit to the data found for the *MIR155HG* locus. Figure 4-4, panel A shows a miniaturized version of the *MIR155HG* locus (Figure 3-41) and indicates five regions (1-5) with chromatin organizers, and histone marks, which are associated with enhancers. In Figure 4-4, panel B the five regions are indicated combining the Capture-C and ChIP-seq data resulting in a hypothetical model of chromatin organization. In panel B, regions 1-3

each show three BZLF1 binding sites, which were also transferred to the model. In Figure 4-4, panel A and B, both regions 1 and 2 depict CTCF binding sites. Even though the capture region is not located between these two CTCF sites, many interactions are visible between these two sites. Several loops, which are in close contact with each other (Figure 4-4, panel B) might explain this situation. In this model the most interactions of these loops with the capture region are identified at sites where CTCF molecules fix the interactions. The less frequent interactions with the capture region would probably occur within the loops by the constant movement of the chromatin. This corresponds to the pattern visible in Figure 4-4, panel A. Regions 4 and 5 are also in contact with the captured *MIR155HG* locus and were included in the model (Figure 4-4, panel B).

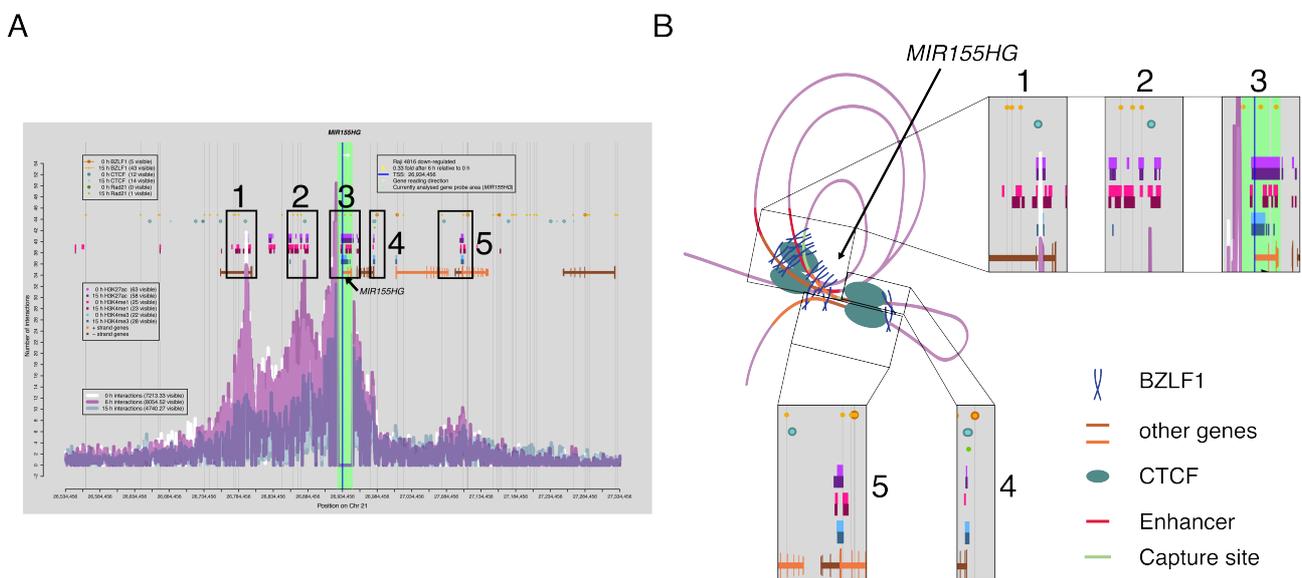


Figure 4-4: A model of the possible chromatin structure at the *MIR155HG* locus.

A: The figure shows a miniaturized version of the *MIR155HG* locus (Figure 3-41) and indicates regions of interest within the five squares 1 to 5. **B:** With the data provided in panel A a model of the possible chromatin structure was built, including BZLF1 binding sites, chromatin organizers, gene positions, enhancer and capture sites. The fluctuating pattern between the CTCF sites indicates that several loops intermingle. With the help of the CTCF sites, depicted in regions 1 and 2, the loops might be stabilized. The BZLF1 binding sites depicted in regions 1-3 were also transferred to the model. Regions 4 and 5 are also interacting with the capture region and were therefore also included in the model.

Another peculiarity was the observation that different genes shared interacting chromatin sequences (Figure 3-39 panel A and B, Figure 3-40 panel A and B). This behavior was already observed in previously published large scale experiments (Thurman et al., 2012; Shen et al., 2012).

Another observation might help to reveal the structure of chromatin and its boundaries even further. On the borders of CTCF sites abrupt reductions of the numbers of interactions were observed at several loci (Figure 3-30, Figure 3-37, and Figure 3-38). Figure 4-5 tries to summarize the observations in a model of chromatin formation. Panel A of Figure 4-5 shows a miniaturized version of the *ZSCAN12P1* locus (Figure 3-37). Within the two CTCF sites (Figure 4-5, panel A, 1 and 2 dashed), which include the *ZSCAN12P1* locus, more interactions were found than outside of these CTCF sites. The regions 1 and 2 (Figure 4-5, panel A) indicate the chromatin interactions at these CTCF sites. A third square (3) indicates an additional pair of CTCF sites within the potential loop.

In Figure 4-5, panel B, the information of panel A were transferred to a model, which tries to explain why the chromatin interactions outside of the CTCF embedded loop are lost abruptly in region 1, while the loss is moderate in region 2. My model suggests that the abrupt loss of chromatin interactions preceding and upstream of the loop results from DNA that is orientated away from the loop structure (Figure 4-5, panel B, square 1). Few interactions with the captured regions are the consequences. The moderate loss of interactions (square 2) might occur because DNA downstream of the loop structure may form a U-turn and orientate towards the loop. By this the 3' region could still be in proximity to the capture site. The two additional CTCF sites within the loop may form a smaller loop and help to bring additional BZLF1 binding sites in close proximity to the capture region (Figure 4-5, panel B, square 3).

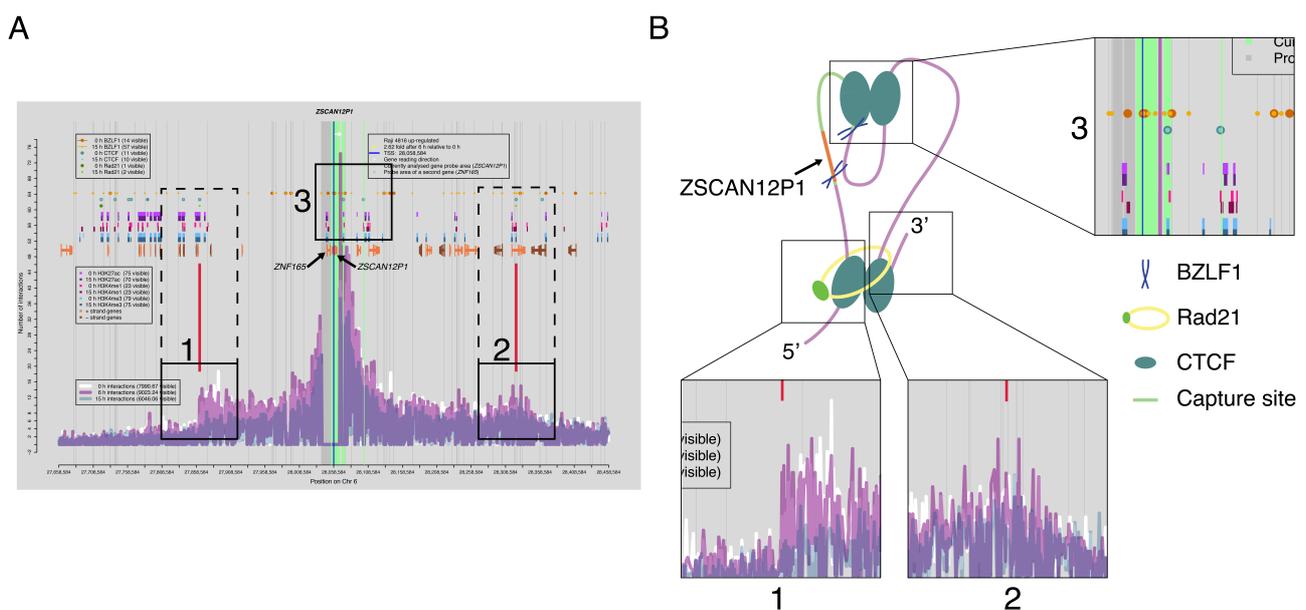


Figure 4-5: A model of the possible chromatin structure at the *ZSCAN12P1* locus.

A: The figure shows a miniaturized version of the *ZSCAN12P1* locus (Figure 3-37) and indicates regions of interest with the three squares 1-3. **B:** BZLF1 binding sites, chromatin organizers, and capture sites were used to build a model of the possible chromatin structure at the *ZSCAN12P1* locus. In this model two CTCF sites link regions 1 and 2. The structure is stabilized by cohesin (Rad21). The CTCF sites in region 3 form another loop within the loop, which might bring additional BZLF1 binding sites in close proximity to the capture region. The loss of chromatin interactions in region 1 is abrupt. This might happen because the orientation of the 5' DNA upstream of the loop orientates away from the loop. The moderate loss of interactions depicted in region 2 might happen because the orientation of the 3' DNA might follow the direction of the loop. Therefore the 5' end cannot get in contact with the capture region any more, while the 3' end is still in close proximity of the loop.

The combination of sites with abrupt and moderate reductions of interactions at sides of chromatin organizers was also found for the *ZNF165* (Figure 3-38) gene, which is very close to the *ZSCAN12P1* (Figure 3-37) gene.

It is clear that these deduced models are highly hypothetical but they might help in data interpretation to resolve the loop orientation within the three-dimensional chromatin architecture.

4.6 Achievements and open questions

This thesis identified two major binding sites of BZLF1 in the cellular genome. One of which, the CpG ZRE sequence, is bound preferentially by BZLF1, but the function of CpG ZREs in the cellular genome remained vague. Upon induced expression of BZLF1, global transcription in Raji cells is reduced but it is currently uncertain if this effect is directly attributable to BZLF1 or an indirect phenomenon.

To address this question RNA-seq experiments in DG75 cells have to be repeated with exogenously added ERCC spike-in control RNAs to allow unambiguous data normalization between data sets. DG75 cells are EBV-negative and this model will reveal whether BZLF1 reduces global transcription directly.

In Raji cells my Capture-C results revealed that binding of BZLF1 and the frequencies of chromatin interactions do not correlate, but the expression of BZLF1 results in a global dissociation of cellular chromatin and its organization.

Further experiments will be needed to address the functions of BZLF1 in regulating transcription and chromatin organization. For example Pol II and CBP ChIP-seq experiments could reveal if the expression of BZLF1 inhibits the recruitment of the transcriptional machinery and ChIP-seq of the H3K36me3 mark could identify the epigenetic status of actively transcribed genes (Li et al., 2007).

A time series experiment is required to analyze if the loss of chromatin interactions is associated with double strand DNA breaks and when this happens. Phosphorylation of serine 139 at histone H2AX, γ -H2AX, (Rogakou et al., 1998) or TUNEL assays (Gavrieli et al., 1992) would indicate the occurrence of DNA double strand breaks.

The heavy chain enhancer in Raji cells is coupled with the *MYC* gene and its promoter element. Concomitantly six BZLF1 binding sites localize to DpnII fragments in the heavy chain enhancer, which show the most frequent interactions with the *MYC* promoter. The deletion or mutational inactivation of these six binding sites via CRISPR/ Cas9 technique (Horvath and Barrangou, 2010; Cong et al., 2013) would certainly be revealing.

5 Material

5.1 Oligonucleotides

The company Metabion (Simmelweisstraße 3, 82152 Planegg, Germany) synthesized all oligonucleotides used for this PhD.

Table 5-1: **Primer pairs for quantitative PCR analysis for ChIP experiments (qChIP), quantitative reverse transcript PCR analysis for RNA expression levels (qRT-PCR), and NGS preparation (NGSPrep).**

| Locus | 5' Forward primer 3' | 5' Reverse primer 3' | Type |
|---------------------------------|------------------------------|--------------------------|---------|
| BBLF4 | GTGCTACACAGCCGCTCCG | CGATACTCCTGATGGTCCTCTCG | qChIP |
| BDLF1 | GTGTCCGTAATGGATGGGGG | CAGCCAGCGACTTGGAGGG | qChIP |
| BRLF1 | CCGGCTGACATGGATTACTGG | AGGAACCAAATAACCGAGCCTC | qChIP |
| EGR1 | GACCCGGAAATGCCATATAA | GCCCAAATAAGGGTTGTTCC | qChIP |
| GAPDH | CCCCGGTTTCTATAAATTGAGC | GGCTGACTGTCTGAACAGGA | qChIP |
| IL8 | CATCAGTTGCAAATCGTGGA | AGAACTTATGCACCCTCATCTTTT | qChIP |
| BBLF4 | GTCCTCCGTGGCTAAAAGCG | CAAGACCAAAAAGTCCATCTG | qRT-PCR |
| BcLF1 | CCTCCCTGACCGTTCCCAG | GCAGTTTGAGACCGCCACATC | qRT-PCR |
| BMRF1 | TTGAGGTTTTACAGGTCTGGCAT C | GGTGGCGGAGGTGAAGGAG | qRT-PCR |
| BNLF2a | TGCTGACGTCTGGGTCT | TGCTTTGCTAGAGCAGCAGT | qRT-PCR |
| BRLF1 | CCTGTCTTGACGAGACCAT | AAGGCCTCCTAAGCTCCAAG | qRT-PCR |
| BZLF1 | GGTTTCCGTGTGCGTCGTG | AGCCTGCTCCTGAGAATGCTT | qRT-PCR |
| CytC | CAATGCTCCGTTGTTGGCAG | CCTGGTGGGCGTGTGCTAC | qRT-PCR |
| Illumina Nextera Adapters | AATGATACGGCGACCACCGA | CAAGCAGAAGACGGCATACGA | NGSPrep |

5.2 Plasmids

Table 5-2: **Plasmids from the collection of the Research Unit Gene Vectors (Helmholtz München).**

| Plasmid | Description |
|---------|---|
| p4816 | pRTR, oriP expression vector with EBNA1, reverse tetracycline controlled transactivator rtTA2(S)-M2 and Tet repressor-tTS (PLDLS-135-L) as a silencer for |

| | |
|-------|--|
| | full-length BZLF1, puromycin resistance, NGF-R-IRES-GFP reporter genes. |
| p5694 | pRTR, oriP expression vector with EBNA1, reverse tetracycline controlled transactivator rtTA2(S)-M2 and Tet repressor- tTS (PLDLS-135-L) as a silencer for a BZLF1 mutant lacking the transactivation domain (aa 175 to 236), puromycin resistance, NGF-R-IRES-GFP reporter genes. |

5.3 Antibodies

Table 5-3: **The antibodies used in this work are named and described in the table**

| Target | Species | Application | Amount/ Dilution | Distributor |
|----------------------------|---------|--------------------|---|---|
| α -BZLF1 | Mouse | ChIP | 400 μ l/ 1 x 10 ⁸ cells | E. Kremmer (HMGU) (BZ1, IgG1k) |
| α -BZLF1 | Mouse | Western blot | 1:50 | E. Kremmer (HMGU) (BZ1, IgG1k) |
| α -rat NGFR | Mouse | Cell sort | 1:10 | E. Kremmer (HMGU) (HB8737-1, Isotyp IgG1) |
| α -mouse IgG | Goat | Magnetic cell sort | 1:10 | Miltenyi Biotec, (Order No. 130-048-401), Bergisch Gladbach, Germany |
| α -mouse IgG | Donkey | Cell sort | 1:50 | e.Bioscience (4 F(ab') ₂ APC polyclonal, Cat. 17.4010.82), San Diego, USA |
| α -mouse IgG HPR | Goat | Western blot | 1:10,000 | Cell Signaling (#7076S), Danvers, USA |

5.4 Eukaryotic cell lines

The following cell lines were either taken from the stock of the Research Unit Gene Vectors (Helmholtz München) or were established (*) during the thesis work.

Table 5-4: **Eukaryotic cell lines**

| Name | Description |
|---------------|---|
| DG75 | Human EBV-negative Burkitt lymphoma cell line (Ben-bassat et al., 1977) |
| DG75 4816 | DG75 cells carrying the full-length BZLF1 (aa 1- 245) under the control of a doxycycline inducible promoter. |
| DG75 5694 (*) | DG75 cells carrying the BZLF1 mutant (aa 175 - 236) lacking the activation domain under the control of a doxycycline inducible promoter. A tandem strep-tag is fused to the BZLF1 gene. |

| | |
|-----------|--|
| Raji | Human EBV-positive Burkitt lymphoma cell line (Pulvertaft, 1964) |
| Raji 4816 | Raji cells carrying the full-length BZLF1 (aa 1- 245) under the control of a doxycycline inducible promoter. |
| Raji 5694 | Raji cells carrying the truncated BZLF1 (aa 175 - 236) lacking the activation domain under the control of a doxycycline inducible promoter. A tandem strep-tag is fused to the BZLF1 gene. |

5.5 Cell culture media and additives for eukaryotic cells

Table 5-5: Cell culture media and additives for eukaryotic cells

| Ingredient | Application | Distributor |
|--|---|--|
| Fetal bovine serum | RPMI 1640 additive (nutritive substance) | BioSell, Feucht / Nürnberg, Germany |
| Fetal bovine serum | RPMI 1640 additive (nutritive substance) | PAA (GE Healthcare Life Science), Easton Turnpike Fairfield, USA |
| Fetal bovine serum (Tet System Approved) | RPMI 1640 additive (nutritive substance) | Takara Bio (former Clontech), Mountain View, USA |
| Penicillin / streptomycin | RPMI 1640 additive (antibiotics) | Life technologies Corporation (gibco), Grand Island, USA |
| RPMI 1640 | Cell culture medium for Raji and DG75 cells | Life technologies Corporation (gibco), Grand Island, USA |
| Sodium pyruvate 100 mM | RPMI 1640 additive (antioxidant reagent) | Life technologies Corporation (gibco), Grand Island, USA |

5.6 Chemicals and enzymes

Table 5-6: Chemicals and enzymes

| Chemical / enzyme | Distributor |
|---|-------------------------------------|
| Agarose | Invitrogen, Karlsruhe, Germany |
| Amersham ECL Western Blotting Detection Reagent | GE Healthcare, Chicago, USA |
| ATX Ponceau S | Sigma-Aldrich, Steinheim, Germany |
| Ammonium persulfate (APS) | SERVA, Heidelberg, Germany |
| beta-mercaptoethanol | Sigma-Aldrich, Taufkirchen, Germany |
| Bradford | Bio Rad, Munich, Germany |
| Bovine Serum Albumin (BSA) | Sigma-Aldrich, Taufkirchen, Germany |
| Doxycycline | Sigma-Aldrich, Taufkirchen, Germany |
| Ethanol (pure) | Merk, Darmstadt, Germany |
| Ethidium bromide | Roth, Karlsruhe, Germany |

| | |
|-------------------------------------|---|
| Milk powder | AppliChem, Darmstadt, Germany |
| NTB buffer | Machery-Nagel, Düren, Germany |
| Restriction enzymes (DpnII) | NEB, Ipswich, USA |
| Rotiphorese Gel30 (acrylamide) | Roth, Karlsruhe, Germany |
| Proteinase inhibitor cocktail (PIC) | Roche Diagnostics GmbH, Unterhaching, Germany |
| Proteinase K | Roth, Karlsruhe, Germany |
| Puromycin | Invitrogen, Grand Island, USA |
| SDS (10 %, ultra pure) | SERVA, Heidelberg, Germany |
| Sepharose protein G beads | GE Healthcare, Chicago, USA |
| T4 DNA Ligase | Affymetrix, Maumee, USA |
| TEMED | Roth, Karlsruhe, Germany |
| Triton-X 100 | Sigma, Steinheim, Germany |
| Trizol | Thermo Fisher Scientific, Waltham, USA |

5.7 Buffers and solutions

Table 5-7: **Buffers and solutions**

| Application | Buffer | Composition |
|-------------|---------------------------------------|---|
| Beads sort | MACS-buffer | PBS, 0.5 % BSA, 2 mM EDTA |
| Buffer | PBS (pH 7.3) | 10 mM Na ₂ HPO ₄ , 1.8 mM KH ₂ PO ₄ , 140 mM NaCl, 2.7 mM KCl |
| Capture-C | 3C lysis buffer (Hagège et al., 2007) | 10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 0.2 % NP-40, in ddH ₂ O (5 ml) add one 10x proteinase inhibitor cocktail (Roche) pill |
| | Fixation mix | 1 % Formaldehyde, 10 % FBS, in PBS |
| | Ligation buffer | 62.5 mM Tris-HCl (pH 7.5), 12.5 mM MgCl ₂ , 1.25 mM ATP, 12.5 mM DTT, in ddH ₂ O |
| ChIP | Hypotonic buffer | 10 mM KCl, 340 mM Sucrose, 1.5 mM MgCl ₂ , 10 mM Hepes (pH 7.9), 1 x proteinase inhibitor cocktail (Roche) |
| | RIPA buffer | 1 % NP-40, 0.05 % SDS, 0.1 % Sodium Deoxycholate, 150 mM NaCl, 1 mM EDTA (pH 8.0), 50 mM Tris-HCl (pH 7.5), 1 x proteinase inhibitor cocktail (Roche) |
| | Low Salt buffer | 20 mM Tris-HCl (pH 8.0), 2 mM EDTA (pH 8.0), 1 % Triton X-100, 150 mM NaCl, 0.1 % SDS |
| | High Salt buffer | 20 mM Tris-HCl (pH 8.0), 2 mM EDTA (pH 8.0), 1 % Triton X-100, 500 mM NaCl, 0.1 % SDS |
| | LiCl buffer | 10 mM Tris-HCl (pH 8.0), |

| | | |
|---------------------|--------------------------|--|
| | | 1 mM EDTA (pH 8.0), 250 mM LiCl, 0.5 % NP-40, 0.5 % Sodium Deoxycholate |
| | TE buffer | 10 mM Tris-HCl (pH 7.9), 1 mM EDTA (pH 8.0) |
| | Elution buffer | 25 mM Tris-HCl (pH 7.5), 10 mM EDTA (pH 8.0), 0.5 % SDS |
| CutSmart buffer | Capture-C | NEB, Ipswich, USA |
| DNA buffer | Tris-HCl | 10 mM Tris-HCl (pH 8.0) |
| | Tris-EDTA | 100 mM Tris (pH 7.2), 0.5 mM EDTA |
| Gel electrophoresis | TBE | 89 mM Tris, 89 mM boric acid, 2 mM EDTA |
| | Loading buffer (6 x) | 10 mM Tris-HCl (pH 7.5), 60 % glycerol, 60 mM EDTA, Orange G |
| Western blot | Running buffer (10 x) | 30.2 g Tris, 144 g Glycine, 100 ml SDS (10 %) filled up to a final volume of 1 l with ddH ₂ O |
| | Blotting buffer | 3.02 g Tris, 14.4 g Glycine, 200 ml methanol filled up to a final volume of 1 l with ddH ₂ O |
| | Wash buffer | 0.1 % Tween-20 in PBS |
| | Blocking solution | 5 g milk powder in 100 ml PBS-T |
| | Laemmli buffer 4 x | 2.5 ml 2 M Tris-HCl (pH 6.8), 1 g SDS, 5 ml 50 % glycerol, 100 μ l 1 M DTT, 10 μ l beta-mercaptoethanol, 0.2 % bromophenol blue, filled up to a final volume of 10 ml with ddH ₂ O |

5.8 Commercial kits

Table 5-8: Commercial kits

| Kit | Application | Distributor |
|---|---------------------------|--|
| Agencourt AMPure XP | PCR clean-up | Beckman Coulter, Munich, Germany |
| Agilent DNA 1000 Kit | DNA quality control | Agilent, Santa Clara, USA |
| Agilent RNA 6000 Nano Kit | RNA quality control | Agilent, Santa Clara, USA |
| Direct-zol™ RNA MiniPrep | RNA clean-up | Zymo Research, Irvine, USA |
| dsDNase | Double strand degradation | Thermo Fisher Scientific, Waltham, USA |
| Dynabeads M-280 Streptavidin | Capture-C | Thermo Fisher Scientific, Waltham, USA |
| ECL Select Western Blotting Detection Reagent | Western blot development | GE-Healthcare, Munich, Germany |
| Encore Complete RNA-Seq Library Systems | NGS library preparation | NuGEN, San Carlos, USA |
| ERCC RNA spike-In control | RNA expression control | Ambion, Thermo Fisher |

| | | |
|---|--|---|
| mixes (4456740) | | Scientific, Waltham, USA |
| GeneRuler 100 bp Plus DNA | DNA analysis | Thermo Fisher Scientific, Waltham, USA |
| Kapa HiFi PCR Kits | DNA amplification | Kapa Biosystems, Wilmington, USA |
| LightCycler 480 SYBR Green I Master Mix | DNA amplification analysis | Roche Diagnostics GmbH, Unterhaching, Germany |
| NEBNext Ultra II DNA Library Prep Kit for Illumina (24 rxn) | NGS library preparation | NEB, Ipswich, USA |
| NEBNext Multiplex Oligos for Illumina (Index Primers Set 1) | NGS library preparation | NEB, Ipswich, USA |
| NEBNext Multiplex Oligos for Illumina (Index Primers Set 2) | NGS library preparation | NEB, Ipswich, USA |
| NucleoSpin Gel and PCR Clean-up | DNA clean-up | Machery-Nagel, Düren, Germany |
| PageRuler Plus Prestained Protein Ladder | Western blot | Thermo Fisher Scientific, Waltham, USA |
| Qubit dsDNA HS Assay Kit | Nucleic acid concentration | Thermo Fisher Scientific, Waltham, USA |
| SureSelect XT2 Custom 1-499 kb | NGS library preparation: SureSelect XT2 | Agilent, Santa Clara, USA |
| SureSelect XT2 Library Prep Kit ILM | | Agilent, Santa Clara, USA |
| SureSelect XT2 Pre-Capture Box 1 | | Agilent, Santa Clara, USA |
| SureSelect XT2 Pre-Capture ILM Module Box 2 | | Agilent, Santa Clara, USA |

5.9 Software

Table 5-9: **Software**

| Software name | Field of application | Distributor / Source |
|----------------------|------------------------------------|--|
| 2100 Expert Software | DNA / RNA quality control | Agilent, Santa Clara, USA |
| BD FACSDiva Software | Fluorescent activated cell sorting | BD Biosciences, San Jose, USA |
| Bedtools 2.25.0 | NGS data processing | (Quinlan and Hall, 2010) http://bedtools.readthedocs.io/en/latest/ |
| Bowtie1 1.1.0 | NGS data processing | (Langmead et al., 2009), Johns Hopkins University, Baltimore, USA http://bowtie-bio.sourceforge.net/index.shtml |

| | | |
|----------------------------|--|--|
| Bowtie2 2.2.6 | NGS data processing | (Langmead and Salzberg, 2012), Johns Hopkins University, Baltimore, USA, http://bowtie-bio.sourceforge.net/bowtie2/index.shtml |
| CCanalyser2.pl | Capture-C interaction finder | James Davies, Oxford, UK https://github.com/telenius/captureC |
| DESeq2 1.12.3 | R package | (Love et al., 2014), http://bioconductor.org/packages/release/bioc/html/DESeq2.html |
| dpnII2E.pl | <i>In silico</i> restriction digestion | James Davies, Oxford, UK https://github.com/telenius/captureC |
| dpnGenome3.1.pl | <i>In silico</i> restriction digestion genome wide | James Davies, Oxford, UK https://github.com/telenius/captureC |
| extrafont 0.17 | R package | Winston Chang (winston@stdout.org), https://CRAN.R-project.org/package=extrafont |
| FastQC 0.11.2 | NGS data control | Babraham Bioinformatics, Babraham, UK, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ |
| Firefox 49.0.1 | Internet browser | The Mozilla Foundation, Mountain View, USA, https://www.mozilla.org |
| Flash 1.2.11 | NGS data processing | (Magoč and Salzberg, 2011), Johns Hopkins University, Baltimore, USA, https://ccb.jhu.edu/software/FLASH/ |
| FlowJo 10.0.7r2 | FACS data visualization | FlowJo, Ashland, USA, http://www.flowjo.com/ |
| Galaxy 15.10 | NGS data processing with user interface | (Afgan et al., 2016), Center for Comparative Genomics and Bioinformatics, State College, USA, https://galaxyproject.org/ |
| GCC compiler 4.2.1 | Compiler | The GNU project, https://gcc.gnu.org/releases.html |
| Homebrew 0.9.5 | Installation helper | Max Howell, http://brew.sh/ |
| IGV 2.3.72 | NGS data visualization | Broad Institute, Cambridge, USA, (Robinson et al., 2011; Thorvaldsdóttir et al., 2013) http://software.broadinstitute.org/software/igv/ |
| Igvttools 2.3.67 | NGS data visualization | http://software.broadinstitute.org/software/igv/download |
| Illumina Demultiplex 1.0.0 | NGS data processing | Galaxy tool (Alex Graf, Gene center, Großhadern, Germany) |
| Illustrator CS5 | Graphic construction | Adobe System Incorporated, San Jose, USA |
| ImageJ 1.50i | Image editing | Wayne Rasband, National Institutes of Health, |

| | | |
|---------------------------------------|--|--|
| | | USA, https://imagej.nih.gov/ij/ |
| LightCycler 480 Software 1.5.1.62 SP3 | Light cycler software | Roche Diagnostics GmbH, Unterhaching, Germany |
| MACS2 2.1.0 | Peak calling | (Zhang et al., 2008), https://github.com/taoliu/MACS |
| MacVector 12.7.5 | Nucleic acid visualization | MacVector, Apex, USA |
| MEME suit 4.10.1 | NGS data processing | (Bailey et al., 2015) http://meme-suite.org/ |
| Mendeley 1.16.3 | Literature citation | Mendeley Inc., New York, USA https://www.mendeley.com/ |
| Microsoft Office 2011 | Text processing | Microsoft Corporation One Microsoft, Way Redmond, USA |
| OS X 10.9.5 | Operating system on Mac | Apple Inc., Cupertino, USA |
| Photoshop CS5 | Image editing | Adobe System Incorporated, San Jose, USA |
| Perl 5.022002 | Programming language | Perl, https://www.perl.org/ |
| Pip 1.5.6 | Installation helper | The Python Packaging Authority, https://www.pypa.io/en/latest/ |
| plotrix 3.6-2 | R package | Jim Lemon (drjimlemon@gmail.com), https://cran.r-project.org/web/packages/plotrix/index.html |
| plyr 1.8.4 | R package | (Wickham, 2011), https://cran.r-project.org/web/packages/plyr/index.html |
| Python 2.7.8 | Programming language | Python, https://www.python.org/downloads/ |
| R 3.3.0 | Programming language used for data compilation and visualization | The R Foundation, Vienna, Austria https://www.r-project.org/ |
| RColorBrewer 1.1-2 | R package | Erich Neuwirth (2014), (erich.neuwirth@univie.ac.at), https://CRAN.R-project.org/package=RColorBrewer |
| reshape2 1.4.1 | R package | (Wickham, 2007), https://cran.r-project.org/web/packages/reshape/index.html |
| RStudio 0.99.902 | Programming language used for data compilation and visualization | RStudio Inc., Boston, USA https://www.rstudio.com/ |
| RUVSeq 1.6.2 | R package | (Risso et al., 2014), http://bioconductor.org/packages/release/bioc/html/RUVSeq.html |
| Samtools 1.0 | NGS data processing | (Li et al., 2009b) |

| | | |
|--------------------------|------------------------------------|--|
| | | http://samtools.sourceforge.net/ |
| Sonolab 7.2 | Covaris Software | Covaris, Woburn, USA |
| splitstackshape 1.4.2 | R package | Ananda Mahto (ananda@mahto.info), https://CRAN.R-project.org/package=splitstackshape |
| Terminal 2.4 | Command line interpretation | Apple Inc., Cupertino, USA |
| TextEdit 1.9 | Text processing | Apple Inc., Cupertino, USA |
| TextMate 2.0-beta.12.4 | Text processing | MacroMates Ltd., Cyprus, http://macromates.com/download |
| Trim galore 0.4.0 | NGS data processing | Babraham Bioinformatics, Babraham, UK, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ |
| VennDiagram 1.6.17 | R package | Hanbo Chen, (Paul.Boutros@oicr.on.ca), https://CRAN.R-project.org/package=VennDiagram |
| XCode 6.2 | Integrated development environment | Apple Inc., Cupertino, USA, https://itunes.apple.com/us/app/xcode/id497799835?mt=12 |
| XML 3.98-1.4 | R package | Duncan Temple Lang (Duncan@r-project.org), https://CRAN.R-project.org/package=XML |

5.10 Special devices and consumables

Table 5-10: **Special devices and consumables**

| Device | Distributor |
|--|--|
| Bioanalyzer 2100 | Agilent, Santa Clara, USA |
| BioRuptor Standard | Diagenode, Seraing, Belgium |
| Covaris M-series | Covaris, Woburn, USA |
| Electroporation device | Bio-Rad, Munich, Germany |
| FACS Canto, FACS Calibur | BD Biosciences, San Jose, USA |
| Film | Agfa HealthCare NV, Mortsel, Belgium |
| Illumina HiSeq 1500 | Illumina, San Diego, USA |
| Illumina NextSeq 500 | Illumina, San Diego, USA |
| Light Cycler 480 Real Time PCR System | Roche Diagnostics GmbH, Unterhaching, Germany |
| MacPro, 3.5 Ghz 6-Core Intel Xeon E5, 32 GB 1866 MHz DDR3 | Apple Inc., Cupertino, USA |
| MACS LS columns and separator | Miltenyi Biotec GmbH, Bergisch Gladbach Germany |

| | |
|------------------------------------|--|
| Mini-PROTEAN Tetra Cell | Bio-Rad, Munich, Germany |
| Nanodrop | Thermo Fisher Scientific, Waltham, USA |
| Nitrocellulose, 0.45 μm | Bio-Rad, Munich, Germany |
| Qubit | Thermo Fisher Scientific, Waltham, USA |

6 Methods

6.1 Eukaryotic cell culture

6.1.1 Cell culture conditions

The handling of cell culture was performed in lamina hoods with sterile glass and plastic pipettes. The cells were kept in sterile cell culture flasks.

All used cell lines were stored in incubators (5 % CO₂, 37 °C, and 95 % air humidity) and cultured in RPMI-1640 medium with the following additions: 1 mM sodium pyruvate; 100 µg/ ml streptomycin and 100 units/ ml penicillin; 10 % fetal bovine serum (FBS). The concentration prior to induction was set to 5 x 10⁵ cells/ ml. All inducible cell lines were induced with doxycycline of a final concentration of 100 ng/ ml.

With an exception of the parental Raji and DG75 cell lines, 1000 and 400 ng/ ml puromycin was added to all Raji and DG75 cell lines, respectively.

6.1.2 Thawing eukaryotic cells

To thaw eukaryotic cells 50 ml of medium were prelayered in a falcon tube. The cryotubes with cells were thawed in a water bath (37 °C). The cells were transferred to the prelayered medium and spun down (300 g, RT, 7 min). Cells were taken up in 5 ml of fresh medium without antibiotics and stored in an incubator over night. The next day required antibiotics were added.

6.1.3 Freezing eukaryotic cells

To freeze eukaryotic cells, 5 x 10⁶ cells per vial were spun down (300 g, 4 °C, 7 min) and taken up in a special freezing medium (80 % medium (described in 6.1.1, page 122), 10 % FCS, 10 % DMSO). The cells were transferred to precooled cryo-tubes, put in cotton

buffered cryo-box and stored at $-80\text{ }^{\circ}\text{C}$ for one week. Afterwards the tubes were transferred to liquid nitrogen storage.

6.1.4 Transfection of eukaryotic cells

For the establishment of a stable cell line, electroporation was used to stably transfect cells with a plasmid of choice. After 5×10^6 cells were spun down (300 g, RT, 10 min) and taken up in 250 μl phosphate buffered saline (PBS), 15 μg plasmid DNA were added to the cells. The mixture was transferred to precooled 4 mm electroporation cuvettes and incubated for 15 min. The following settings were used for electroporations with the gene pulser II electroporator (Biorad): 0.230 kV; 0.975 μF . After pulsing the cells were transferred to 400 μl precooled FCS and incubated in T25 flask with pre-warmed medium over night.

6.1.5 Establishment of stable cell lines

After the electroporation (chapter 6.1.4, page 123) cells were selected for stable plasmid integration by the addition of cell specific amounts of puromycin (chapter 6.1.1, page 122). Therefore, cells grown over night were diluted five times 1:5 and each dilution was distributed on a 96-well plate. The plates were stored in an incubator and the medium was exchanged when necessary for four weeks. The cells were tested for the expression of the fluorescent marker (GFP) by flow cytometry (chapter 6.2, page 123).

6.2 Flow cytometry

To test cell populations viability or the expression of a fluorescent marker the FACS Canto (Becton Dickinson) was used. The cells were either sorted in cell culture medium or PBS.

The output was analyzed with the FlowJo software (version 10.0.7r2).

6.3 ChIP-sequencing (NGS)

6.3.1 Cell preparation

For chromatin immunoprecipitations 1×10^8 cells were used for each induced and non-induced sample. Both samples were kept under puromycin selection (see chapter 6.1.1, page 122 for cell line specific concentrations) and 100 ng/ ml doxycycline was added to induce cells for 15 h.

6.3.2 Chromatin immunoprecipitation

The cells were spun down (300 g, 4 °C, 7 min), washed with ice cold PBS, and spun down again as before. The cells were swollen in a mixture of 90 % hypotonic buffer (chapter 5.7, page 115) and 10 % of phosphatase inhibitor cocktail (Roche) for five minutes. Triton-X was added to the cells to a final concentration of 0.1 % and subsequently the cells were vortexed for five seconds. The isolated nuclei were split up to three tubes, centrifuged (500 g, 4 °C, 5 min), and lysed in 300 μ l of a mixture of 90 % RIPA buffer (chapter 5.7, page 115) and 10 % of PIC. The isolated chromatin was sheared with the BioRuptor (Diagenode) four times on ice (5 min, 30 sec on/ off, high). The sheared chromatin was centrifuged (16,000 g, 4 °C, 15 min) and the supernatant was transferred to a new tube.

To control for appropriate shearing 20 μ l of the supernatant were digested (50 °C, 1 h, 500 rpm) with proteinase K (20 μ l of 20 mg/ ml) in 200 μ l (H₂O). The digested samples were cleaned-up with a NucleoSpin kit (Machery-Nagel) and loaded on a 1 % TBE agarose-gel comprising (chapter 5.7, page 115) 10 μ l Ethidium bromide (1 h, 135 V).

The remaining material of each sample was split up in 45 μ l, which were kept as 10 % input and in two tubes containing each 450 μ l sheared chromatin, 750 μ l RIPA buffer including 10 % PIC, and 200 μ l BZ1 antibody. After the chromatin-antibody-complex was incubated over night (4 °C), 100 μ l of 50 % slurry protein G beads (Table 5-3, page 113) were added to the samples and incubated for another 2 h.

6.3.3 ChIP pull-down control

To finally isolate the precipitated DNA, the chromatin-antibody-beads-complexes were washed with four buffers, whose compositions are shown in chapter 5.7, page 115. One ml of a low salt buffer (ChIP 1) was added to each sample, incubated in a shaker (4 °C, 5 min, 1000 rpm), and the samples were spun down (300 g, RT, 30 seconds). The same procedure was used for the high salt, lithium chloride (LiCl) and tris-EDTA (TE) buffer. To get rid of sepharose beads and chromatin residues 190 μ l elution buffer and 10 μ l proteinase K (20 mg/ ml) were added to each of the chromatin-antibody-beads-complexes after the supernatant had been taken off. After the digestion (50 °C, 1000 rpm, 2 h) the samples were spun down (300 g, RT, 1 min) and the supernatants were purified with the help of NucleoSpin columns (Machery-Nagel). NTB buffer was used instead of the standard NTI buffer to avoid interference with the SDS in the elution buffer.

The input DNA, as well as both ChIP samples of each time point were eluted in 30 μ l of the elution buffer to gain the same concentration. For the evaluation of a successful pull-down of DNA, qPCR was used, whose setup is explained in chapter 6.7 (page 139).

For ChIP experiments the samples were compared against a dilution series of the input (10, 1, and 0.1 %) and a H₂O control.

While the pull-down of the ERG1 and IL8 loci were tested in DG75 4816 cells, in Raji 4816 cells the pull-down of the cellular GAPDH and the viral EBV loci BDLF1, BRLF1, and BBLF4 were tested. The primers for the named genes are listed in table Table 5-1 (page 112).

6.3.4 Library preparation and Illumina sequencing

The library preparation for the DNA precipitated from Raji 4816 cells as well as the Illumina sequencing (paired-end, 150 bp) was performed by the company [Vertis Biotechnologie AG](#) (Lise-Meitner-Straße 30, 85354 Freising, Germany).

The library preparation for the DNA originating from the DG75 4816 cells was prepared with the NEBNext Ultra II DNA Library Prep Kit and the NEBNext Multiplex Oligos

for Illumina kits (chapter 5.8, page 116) in the lab of [Helmut Blum](#) by myself with the help of [Sylvia Mallock](#) and advise from [Stefan Krebs](#).

The concentration of the precipitated DNA was measured with the Qubit system (ThermoFisher scientific) and the double strand DNA high sensitivity kit (Table 5-8, page 116). For every sample 8 ng of DNA were used as input and filled up to 50 μ l with Tris-HCl (10 mM, pH 8.0). With the following exceptions the samples were prepared like described in the protocol version 1.1 (11/2015):

- the NEBNext adaptor was diluted 1:10 in Tris-HCl (10 mM, pH 8.0) in step 1.2 due to the small amount of input DNA.
- as required for ChIP-sequencing, a size selection was performed (follow 1.3 A) with AMPure beads.
- Since the lot number of the kit was higher than 0071412 in step 1.4 option 1.4 A had to be followed.
- At point 1.4 A.1, certain index primer and universal primer were used.
- At point 1.4 A.3 11, cycles were used to amplify the libraries.

After the library preparation, 200 ng of each sample were used to run an additional 1-step PCR to enable partial single-stranded DNA fragments to fill up.

Table 6-1: Composition of the master mix used for 1-step PCR to fill up partial single-stranded DNA fragments

| Ingredients | Amount / Concentration |
|----------------------|------------------------|
| H ₂ O | Fill up to 25 μ l |
| Nextera F primer | 10 μ M |
| Nextera R primer | 10 μ M |
| 5 x KAPA HiFi buffer | 1 x |
| dNTP-Mix | 10 mM |
| KAPA polymerase | 0.5 [U/ μ l] |

5 μ l of the master mix (Table 6-1, page 126) were added to 20 μ l of library DNA and the PCR was run with the program shown in Table 6-2 (page 127).

Table 6-2: PCR program of the 1-step PCR used to fill up partial single-stranded DNA fragments

| Step | Temperature [°C] | Time |
|------|------------------|---------|
| 1 | 95 | pause |
| 2 | 95 | 3.5 min |
| 3 | 98 | 20 sec |
| 4 | 60 | 30 sec |
| 5 | 72 | 1 min |
| 6 | 72 | 5 min |
| 7 | 10 | pause |

After the final elution step, 1 μ l of Tris (100 mM, pH 7.2)- EDTA (0.5 mM) was added to the samples, which were then frozen at -20 °C. Sylvia Mallok used the BioAnalyzer to control the quality of the samples before Stefan Krebs prepared samples for sequencing with on a Illumina HiSeq 1500 machine (single-end, 50 bp).

6.3.5 Bioinformatic analysis

The data received from the sequencer for DG75 4816 and Raji 4816 cells were analyzed the same way, if not stated differently.

Since several samples were sequenced on the same of the eight lanes of a NGS chip, the data had to be sorted by their unique index prior to further analysis (demultiplexing).

DG75 4816 samples were demultiplexed and trimmed in one step on the Galaxy sever of the [Blum laboratory](#) with the Illumina Demultiplex tool. More details about the following software tools can be found in chapter 5.9 (page 117). The Raji 4816 data were already provided demultiplexed by the sequencing service of [Vertis](#) and had to be trimmed (Trim galore) only. The commands shown in the methods are exemplary and need to be adapted for each case to work:

```
trim_galore Input1.fastq Input2.fastq -q 20 --paired --phred33
```

Subsequently, the fastq reads were mapped on Hg19 and the Raji genome with the GenBank accession number KF717093* with the help of the command line program Bowtie2. *The Raji genome used for mapping has an additional G base between the

positions #40,224 and #40,225. The DG75 single-end reads were mapped with the following Bowtie2 settings:

```
bowtie_opts="-p 8 -k 1 --no-unal"  
bowtie_index="-x /hg19_0.9_Raji.ebwt"  
bowtie2 $bowtie_opts $bowtie_index -U Input.fastqsanger > Output.SAM
```

Raji 4816 paired-end reads were mapped with the following Botwie2 settings:

```
bowtie_opts="-p 4 -k 1 -I 0 -X 700"  
bowtie_index= /hg19_0.9_Raji.ebwt"  
bowtie2 $bowtie_opts $bowtie_index -1 InputF.fq -2 InputR.fq >  
Output.SAM
```

The mapped reads were converted from SAM to BAM files, sorted by their position on the genome, and transformed to BED files.

SAMs to BAMs: samtools

```
samtools view -bS Input.SAM > Output.BAM
```

Sort BAMs: samtools

```
samtools sort Input.BAM Output_sorted.BAM
```

Sorted BAMs to BEDs: bedtools

```
bedtools bamtobed -i Sorted_Input.bam > Output.bed
```

The peak calling was done with MACS2 using the following command:

```
macs2 callpeak -t Input.bed -c 10%_Input.bed -n Output.out -g 2.7e9  
--call-summits -f BED
```

The output of the MACS2 peak caller is written in bed format. To go on with the most frequent peaks two experiments per cell line were merged. Only peaks, which were found with the help of bedtools in both experiments, were used for further processing:

```
bedtools intersect -a Input1.bed -b Input2.bed > Intersection.bed
```

Prior to identification of binding motifs the DNA sequences had to be extracted position specifically from Hg19 and the Raji genome with the help of bedtools:

```
bedtools getfasta -fi /Hg19_Raji.fasta -bed Input.bed -fo Output.fa
```

The programming language R was used to make the lines in the bed files unique by deleting all duplicated lines. The script can be found on the compact disk (CD) under:

```
1_Code/0_ChIP-seq/0_R_code/Unqie_BEDs.R.
```

The unique fasta files were then further processed with the ChIP-MEME suit to identify binding motifs of BZLF1:

```
meme-chip -o Output.meme -index-name meme-chip.html -meme-maxsize
10000000 -meme-mod zoops -meme-minw 6 -meme-maxw 12 -meme-nmotifs 4
-dreme-m 10 -centrimo-score 5.0 -centrimo-ethresh 10.0 -ccut 0
Input.fa
```

The sub-algorithm of ChIP-MEME, which identified motifs in large data set is DREME. The results of DREME were used for further analysis, since it takes into account all of the input sequences instead of a subset like the MEME algorithm.

The Venn diagrams were made with R and the scripts can be found on the CD: 1_Code/0_ChIP-seq/0_R_code/VennDiagramms.R

6.4 RNA-sequencing (NGS)

6.4.1 Cell preparation

All DG75 parental / 4816 / 5694 and Raji parental / 4816 / 5694 cell lines were treated the same way with minor exceptions. Prior to doxycycline-induced expression of BZLF1, 2×10^7 cells were cultivated together with 6×10^7 cells (3 h) and 4×10^7 cells (6 h). Puromycin was added cell line specifically as described in chapter 6.1.1 (page 122).

6.4.2 Magnetic beads cell sorting

From here on all steps were performed on ice, with precooled buffers, and centrifuges cooled to 4 °C.

The cells were cooled down on ice, spun down (300 g, 5 min, 4 °C), washed with 15 ml MACS-buffer (chapter 5.7, page 115), and spun down again (300 g, 5 min, 4 °C). The cell were then incubated with the α -rat NGFR antibody (chapter 5.3, page 113) for 15 minutes (three times 5 min, with resuspension).

Table 6-3: Composition of the primary antibody incubation mixture

| Time point [h] | Number of cells | MACS-buffer [μ l] | Antibody [μ l] |
|----------------|-----------------|------------------------|---------------------|
| 0 | 2×10^7 | 360 | 40 |
| 3 | 6×10^7 | 1080 | 120 |
| 6 | 4×10^7 | 720 | 80 |

The cells were washed two times with 15 ml MACS-buffer and spun down (300 g, 5 min, 4 °C) each time. The cells were then incubated with an α -mouse IgG antibody (chapter 5.3, page 113) coupled to magnetic beads for 15 minutes (three times 5 min, with resuspension)

Table 6-4: Composition of the secondary antibody incubation mixture

| Time point [h] | Number of cells | MACS-buffer [μ l] | Antibody [μ l] |
|----------------|-----------------|------------------------|---------------------|
| 0 | 2×10^7 | 180 | 20 |
| 3 | 6×10^7 | 540 | 60 |
| 6 | 4×10^7 | 360 | 40 |

The cells were washed two times with 15 ml MACS-buffer and spun down (300 g, 5 min, 4 °C) each time. Subsequently, the cells were taken up in 1 ml of MACS-buffer and transferred on MACS-buffer equilibrated MACS-columns sticking in a magnet. The columns were washed three times with 3 ml of MACS-buffer. The flow-through was collected as well as the cells, which were eluted from the columns (outside of the magnet) with 6 ml MACS-buffer and a plunger. The cells of each sample were counted, 7.5×10^5 cells were spun down (300 g, 5 min, 4 °C) and resolved in 1 ml of trizol intensively. The solution was frozen in liquid nitrogen and stored @ -80 °C

Since GFP couldn't be detected after the sorting procedure in cells expressing the AD-tBZLF1 mutant anymore, the tNGF-receptor of sorted cells was stained with a secondary, fluorophore-linked α -mouse IgG antibody (chapter 5.3, page 113). Therefore 100 μ l of sorted cells were spun down (300 g, 5 min, 4 °C) and stained for 15 minutes (three times 5 min, with resuspension) in 25 μ l of MACS-buffer and 0.5 μ l of the APC-coupled α -

mouse IgG antibody (Table 5-3, page 113). The remaining cells were analyzed by FACS cytometry (chapter 6.2, page 123) to control the sort efficiency.

6.4.3 Library preparation and sequencing

The library preparation for RNA-sequencing of the DG75 parental / 4816 / 5694 and Raji parental / 5694 cell lines were made by Andrea Klanner from the group of [Helmut Blum](#) in the Gene Center. I prepared the libraries from the Raji 4816 cell line with spike-in controls with support of [Sylvia Mallock](#). The preparation process is the same as for the other cell lines with the exception that ERCCs control RNAs were added to the samples.

6.4.3.1 Addition of ERCC spike-in RNAs:

Each sample (400 μ l) was transferred in a new tube and 10 μ l of a 1:100 dilution of ERCC spike-in RNAs (chapter 5.8, page 116) was added. The solution was vortexed to distribute the spike-in RNAs.

6.4.3.2 RNA clean-up

The samples' RNA was extracted using the Direct-zol™ RNA MiniPrep kit (chapter 5.8, page 116) as described in the manual. The RNA concentration and quality was measured with the Qubit and Nanodrop systems, respectively. The quality of the samples were additionally checked with the BioAnalyzer (Agilent) and diluted to a concentration of 75 ng/ μ l. To get rid of any DNA residues, all samples were incubated with dsDNase (5 min, 37 °C). The dsDNase was heat inactivated afterwards (5 min, 55 °C).

Table 6-5: Composition of the dsDNase reaction

| Ingredient | Amount [μ l] |
|----------------------|-------------------|
| H2O | 2 |
| 10 x buffer | 0.5 |
| dsDNase | 0.5 |
| RNA [75 ng/ μ l] | 2 |

6.4.3.3 Library preparation

For library preparation the Encore Complete RNA-Seq Library Systems kit (chapter 5.8, page 116) was used. The kit uses not-so-random hexamer primer to reduce the amount to ribosomal RNAs. If not stated differently the preparation was performed according to the kit's manual (M01244 v5.1).

- Section B:
 - 62.5 μl instead of 65 μl of master mix was pipetted to every sample due to little amount of liquids.
- Section C:
 - For cDNA fragmentation the snap cap micro tubes were used in the Covaris M-series (Covaris) with the following settings:
(Peak incident power: 50 watts; Duty factor: 20 %; Cycles/ Burst: 200 counts; Duration: 115 seconds).

After library preparation the sample concentrations were measured with the Qubit system using the high sensitivity double strand kit. All samples were diluted with H₂O to a concentration of 20 ng/ μl . To each sample (20 μl) a master mix (30 μl) with the KAPA polymerase (Table 6-6, page 132) was added and a 1-step PCR (Table 6-7, page 133) was run to complement partial single-stranded cDNA fragments.

Table 6-6: Composition of the master mix used for 1-step PCR to fill up partial single-stranded DNA fragments

| Ingredients | Amount / Concentration |
|----------------------|-----------------------------|
| H ₂ O | Fill up to 30 μl |
| Nextera F primer | 10 μM |
| Nextera R primer | 10 μM |
| 5 x KAPA HiFi buffer | 1 x |
| dNTP-Mix | 10 mM |
| KAPA polymerase | 1 [U/ μl] |

Table 6-7: PCR program of the 1-step PCR used to fill up partial single-stranded DNA fragments

| Step | Temperature [°C] | Time |
|------|------------------|---------|
| 1 | 95 | pause |
| 2 | 95 | 3.5 min |
| 3 | 98 | 20 sec |
| 4 | 60 | 30 sec |
| 5 | 72 | 1 min |
| 6 | 72 | 5 min |
| 7 | 10 | pause |

Finally, the cDNA was cleaned up in a 2-step reaction with Agencourt RNAClean XP beads (AMPure beads). To 50 μ l of the amplified sample 20 μ l of AMPure beads were added and mixed by vortexing (10, vortexing every 2.5 min). A magnet collected the magnetic beads (5 min) and the supernatant was transferred to a new tube. Fresh AMPure beads (30 μ l) were added to the supernatant and mixed by vortexing (10, vortexing every 2.5 min). The beads were spun down in a minifuge shortly and a magnet collected the beads (5 min) before the supernatant was discarded. EtOH (180 μ l, 70 %) was added to the beads and incubated for 30 second on the magnet before the supernatant was removed (repeated 3 x). The samples were air-dried (3 min) and the cDNA was resolved from the beads with 12 μ l of H₂O (5 min) before the magnet collected the beads (3 min) again. 11 μ l of the supernatant were transferred to a new tube and 1.2 μ l of a 100 mM Tris – 0.5 mM EDTA solution was added.

Sylvia Mallok analyzed the quality of the samples with the BioAnalyzer and Stefan Krebs prepared the sample for sequencing on a HiSeq 1500 machine (100 bp, single-end).

6.4.4 Bioinformatic analysis

The samples of all cell lines were demultiplexed on the Galaxy server of the [AG Blum](#) (Gene Center) with the Illumina Demultiplex tool. The samples were split up according to their unique indexes, which were ligated to the samples during library preparation. The demultiplexed samples were mapped by Tophat2 to Hg19 and the EBV-Raji genome with the GenBank accession number KF717093* (*The EBV-Raji genome used for mapping had

an additional G base between the positions #40,224 and #40,225). The program HTSeq-count (1.0.0) was used on Galaxy to assign the mapped read counts to annotated genes.

6.4.4.1 Human genome:

The resulting text files were downloaded and processed further on the local machine. The R package DESeq2 was used to calculate the differential expression of genes. Within the same R script the data were visualized. The script with about 500 lines of code is stored on the CD under:

```
1_Code/1_RNA-seq/0_R_code/DESeq2_with_ERCCs.R.
```

This file is an example and differs slightly from cell line to cell line. The ERCC normalization used in this script was not performed for cell lines, which were not spiked with ERCC RNAs.

6.4.4.2 EBV-Raji genome:

To analyze how strong EBV genes were regulated, the differential gene expression of chosen genes was analyzed. A separate R script was used for this.

The script is stored on the CD:

```
1_Code/1_RNA-seq/0_R_code/DESeq2_with_ERCCs_for_EBV_Genes.R
```

6.4.4.3 KEGG pathway analysis:

The non-regulated and strongly regulated genes of the Raji 4816 cell line comprising ERCC RNAs were further analyzed with the online tool [WebGestalt](#) for KEGG pathway enrichment with standard settings and gene_symbol as ID type. The results were downloaded and visualized with the programming language R. The script can be found on the CD:

```
1_Code/1_RNA-seq/0_R_code/KEGG.R.
```

During data processing distinct diseases were excluded from visualization.

6.5 Linkage of ChIP-seq and RNA-seq results

To link the ChIP-seq data with the RNA-seq data the BZLF1 binding sites located in close proximity to TSS had to be identified. Therefore, the RefSeq genes of Hg19 were downloaded from the [UCSC genome browser](#). The output was parsed (dissected) with R to extract the TSSs for further analyses. More details about the following software tools can be found in chapter 5.9 (page 117). The R script for data parsing is stored on the CD:

```
1_Code/2_Comparison_ChIP-RNA-seq/0_R_code/TSS_parsing.R.
```

The command line tool “bedtools” was used to find genes with BZLF1 binding site within 5,000 bps upstream and 1,000 bps downstream of their TSS:

```
bedtools window -l 5000 -r 1000 -sw -a TSS.bed -b Input.bed >
Output.bed
```

An R script with about 300 lines of code was used to count for the number of peaks found in the defined region around the TSS, link these to the regulated genes and finally visualize the results. The code can be found on the CD:

```
1_Code/2_Comparison_ChIP-RNA-seq/0_R_code/Linking_ChIP-RNA-seq_data.R.
```

6.6 Capture-C (NGS)

6.6.1 Cell preparation

Raji 4816 cells (1×10^7) were either not induced or induced for 6 and 15 h with doxycycline (100 ng/ml).

6.6.2 Chromatin Conformation Capture (3C)

Subsequently, 1×10^7 cells were spun down (600 g, RT, 7 min) and washed with PBS. The cells were resuspended in 500 μ l PBS with 10 % FBS and filtered through a 70 μ m cell strainer to obtain a single-cell suspension. A 1 % fixation mixture (chapter 5.7, page 115) was added to the cell suspension (9.5 ml, 30 min, RT, rotating). The cells were put on ice and 2.5 ml 1 M Glycine was added before the cells were spun down (500 g, 4 °C, 10 min).

The cells were washed with ice cold PBS and resuspended and homogenized with a 20G syringe using 15 strokes in 1 ml 3C lysis buffer including proteinase inhibitor (Roche) mix (chapter 5.7, page 115) to isolate the nuclei. The lysate was incubated for 5 min before it was homogenized again by another 15 strokes with the syringe. The cells were spun down (600 g, 4 °C, 7 min), washed with 500 μ l 1.2x CutSmart (NEB), and then resuspended in 500 μ l 1.2x CutSmart. Five μ l of 10 % SDS (biomedical grade) were added to the samples for a final concentration of 0.1 %, mixed and incubated (65 °C, 40 min, 1,200 rpm). 10 % Triton X-100 (200 μ l) were added for a final concentration of 4 % and samples were incubated (37 °C, 1 h, 1,200 rpm). For digestion 80 U DpnII enzyme per tube were added and samples were split up to six tubes for digestion (37 °C, over night, 1,200 rpm).

The next day another 400 U of the DpnII enzyme were added and incubated again (37 °C, 4 h, 1,200 rpm). The remaining lysate was incubated with 99 μ l of pure 10 % SDS to a final concentration of 0.1 % to stop the digestion (65 °C, 30 min, 1,200 rpm). The samples of the same time point were joined in a 15 ml falcon tube and diluted with 6.4 ml 1.25x 3C ligation buffer (chapter 5.7, page 115). Additionally, 720 μ l Triton X-100 were added prior to incubation (37 °C (water bath), 1 h, shaking every 10 min). 100 U T4 DNA Ligase (Affymetrix) were added to the mixtures and incubated (16 °C, 4.5 h + RT, 45 min) before 300 μ g Proteinase K (10 mg/ ml) were added to all samples to de-crosslink them (65 °C (water bath), over night).

The next day 300 μ g of RNase A (10 mg/ ml) were added to the samples and incubated (37 °C, 1 h) to get rid of the RNA. The DNA was cleaned-up by Phenol-Chloroform (2/3 Phenol-Chloroform and 1/3 Buthanol) extraction, precipitated with EtOH (2.5 V EtOH (100 %), 0.1 V 3M NaAcetate) and washed with 70 % EtOH. The pellet was resuspended in TE-buffer (4 °C, overnight) and stored (-20 °C).

6.6.3 Library preparation, sample capture, and sequencing

The concentration of the samples was measured with the Qubit system (Thermo Fisher Scientific) and 1.5 μ g DNA were sheared with the Covaris M-series (Covaris, Peak incident power: 50 watts; Duty factor: 20 %; Cycles/ Burst: 200 counts; Duration: 200 seconds).

To clean up the samples 214.2 μ l AMPure beads were added to 119 μ l of sheared DNA and the mixture was mixed by vortexing (2.5 min, two times). The beads were collected with a magnet and the supernatant was discarded. EtOH (200 μ l, 70 %) was added to the beads and incubated for 30 second on the magnet before the supernatant was removed (repeated 3 x). The beads were air dried (5 min) and incubated with 54 μ l H₂O (5 min, RT) before the magnet collected the beads again and 52.5 μ l of the supernatant with the DNA were transferred to a new tube. The concentration was controlled again with the Qubit system and the samples were frozen overnight (-20 °C).

The library preparation and sample capture was done as described in the SureSelectXT2 Target Enrichment System for Illumina Paired-End Multiplexed Sequencing (Version: E1, June 2015). The probes were captured with Dynabeads (Thermo Fisher Scientific), which were not provided with the kit.

Finally, a 1-step PCR was performed to allow partial single-stranded fragments to complete. To 39 μ l isolated DNA 11 μ l of the 1-step PCR master mix (Table 6-8, page 137) was added and run with the program shown in Table 6-9 (page 137).

Table 6-8: Composition of the master mix used for 1-step PCR to fill up partial single-stranded DNA fragments

| Ingredients | Amount / Concentration |
|----------------------|------------------------|
| H ₂ O | Fill up to 11 μ l |
| Nextera F primer | 10 μ M |
| Nextera R primer | 10 μ M |
| 5 x KAPA HiFi buffer | 1 x |
| dNTP-Mix | 10 mM |
| KAPA polymerase | 1 [U/ μ l] |

Table 6-9: PCR program of the 1-step PCR used to fill up partial single-stranded DNA fragments

| Step | Temperature [°C] | Time |
|------|------------------|---------|
| 1 | 95 | pause |
| 2 | 95 | 3.5 min |
| 3 | 98 | 20 sec |
| 4 | 60 | 30 sec |

| | | |
|---|----|-------|
| 5 | 72 | 1 min |
| 6 | 72 | 5 min |
| 7 | 15 | pause |

The samples (50 μ) were purified with AMPure beads (60 μ l) as described before in this chapter and 1.5 μ l Tris (10 mM) – EDTA (0.05 mM) was added to 14.5 μ l H₂O of cleaned-up DNA to stabilize the DNA. After Sylvia Mallok checked the quality of the samples with the Bioanalyzer, Stefan Krebs prepared the samples for sequencing on a HiSeq 1500 machine (paired-end, 100 bps).

6.6.4 Bioinformatic analysis

The files from the sequencer were demultiplexed with the Illumina Demultiplex tool on the Galaxy server of the <http://www.blum.genzentrum.lmu.de/AG Blum>.

More details about the following software tools can be found in chapter 5.9 (page 117). The paired-end reads were combined to one read with the flash tool if overlapping. The following commands are exemplary and need to be modified case specifically:

```
flash --interleaved-output P1.fastq P2.fastq -M 85
```

The combined and not combined fragments were joined again with the command line using cat command:

```
cat Combined.fastq notCombined.fastq > output.fastq
```

The perl script dpnII2E (kindly provided by [James Davies](#), Oxford, UK) was used to cut the output *in silico* at the DpnII restriction enzyme recognition site.

To map the digested fragments to Hg19 the Bowtie algorithm was used:

```
bowtie -p 6 -m 2 --best --strata --sam --chunkmb 256  
path_to/hg19_0.9_ChrRaji.ebwt input.fastq > output.sam
```

The perl script dpngenome3_1.pl (kindly provided by [James Davies](#), Oxford, UK) was used to digest the Hg19 genome *in silico* at restriction sites of the DpnII enzyme:

```
dpngenome3_1.pl Hg19.fasta
```

The program CCanalyser2 (kindly provided by [James Davies](#), Oxford, UK) identified interactions between fragments, which were mapped to different loci, but were linked together prior to *in silico* DpnII digestion.

```
CCanalyser2.pl -f path_to/input.sam
-o path_to/Captured_DpnII_fragments.txt
-r path_to/Hg19_DpnII_digested.txt -s output.sam --genome Hg19
```

The output provides different file formats, of which the .gff files were used to visualize the data with the help of an R script with about 2,500 lines of code. The fragments interacting with the capture DpnII fragments were summed up and visualized as lines around the TSS of the analyzed gene. This script also took RNA-seq and ChIP-seq results into account, integrated these for the visualization, and can be found on the CD:

```
1_Code/3_Capture-C/0_R_code/Capture-C_visualization.R
```

6.7 Quantitative real time PCR

The LightCycler 480 (Roche) was used to quantify the DNA extracted during ChIP by quantitative real time PCR. During the synthesis of double strand DNA the SYBR-Green I dye is incorporated into the DNA and gets fluorescent, which can be used for quantification after each PCR cycle (Higuchi et al., 1993).

The percent input of DNA was analyzed by relative quantification, which allows the calculation of the amount of DNA in the unknown sample relative to the input DNA with the help of a regression curve. The LightCycler 480 software analyzed the data automatically with the second derivative maximum method.

Table 6-10: Composition of quantitative real-time PCR master mixes

| Ingredients | Amount |
|---------------------------|--------------------------------------|
| H ₂ O | 3.5 μ l |
| Primer | 0.5 μ l (5 pm forward / reverse) |
| SYBR Green I Master (2 x) | 5 μ l |
| DNA (template) | 1 μ l |

Table 6-11: Program used for quantitative real-time PCRs with the LightCycler 480

| Program | Temperature (°C) | Hold (sec) | Acquisition mode | RampRate (°C/ sec) | Cycles | Analysis mode |
|----------------|------------------|------------|------------------|--------------------|--------|----------------|
| Pre-incubation | 95 | 600 | None | 4.4 | 1 | None |
| | 95 | 1 | None | 4.4 | | |
| | 62 | 10 | None | 2.2 | | |
| Amplification | 72 | 10 | None | 4.4 | 45 | Quantification |
| | 75 | 3 | Single | 4.4 | | |
| | 97 | 1 | None | 4.4 | | |
| Melting curve | 67 | 10 | None | 2.2 | 1 | Melting curve |
| | 97 | - | Continuous | 0.11 | | |
| Cool-down | 37 | 15 | None | 2.2 | 1 | None |

6.8 Protein analysis

6.8.1 Western blot

To quantify the amount of BZLF1 dimers per cell a immunodetection using Western blot was performed.

6.8.1.1 Cell preparation

2.5×10^6 cells of the parental DG75 4816 and the parental Raji 4816 cell lines were either not induced or induced with doxycycline for 15 h cell line specifically (chapter 6.1.1, page 122).

6.8.1.2 Cell lysate preparation

The cells (2.5×10^6) were spun down (300 g, 4 °C, 7 min), washed with precooled PBS and spun down (300 g, 4 °C, 7 min) again. The cells were taken up in 250 μ l RIPA buffer with PIC (chapter 5.7, page 115) to a final concentration of 10,000 cells/ μ l). The cell lysates were sheared in the BioRuptor (Diagenode) four times on ice (5 min, 30 sec on/ off, high) and subsequently spun down (15 min, 4 °C, 16,000 g). The supernatant was

transferred to a new tube and the protein concentration was measured using Bradford reagent (1:5, 1000 μ l + 1 μ l sample) in a BioPhotometer plus (Eppendorf). To every sample 83.3 μ l 4x laemmli buffer was added, heat up (95 °C, 5 min) and frozen (-20 °C) afterwards.

6.8.1.3 SDS-gel

A 5 % stacking gel (Table 6-12, page 141) was prepared to condense the samples before they were separated by the separation gel (Table 6-13, page 141).

Table 6-12: Composition of a stacking gel (5 ml)

| Ingredients | Amount [ml] |
|--------------------|-------------|
| H ₂ O | 3.6 |
| Polyacrylamid 30 % | 0.67 |
| 2 M Tris (pH 6.8) | 0.75 |
| 10 % SDS | 0.04 |
| 10 % APS | 0.04 |
| TEMED | 0.005 |

Table 6-13: Composition of a 15 % separation gel (10 ml)

| Ingredients | Amount [ml] |
|--------------------|-------------|
| H ₂ O | 3.1 |
| Polyacrylamid 30 % | 5 |
| 2 M Tris (pH 8.9) | 1.66 |
| 10 % SDS | 0.1 |
| 0,5 M EDTA | 0.0667 |
| 10 % APS | 0.07 |
| TEMED | 0.001 |

The samples were loaded in a SDS gel in a Mini-PROTEAN Tetra Cell system (BioRad) and the PageRuler™ Prestained Protein Ladder was used as a protein size marker. The gel was run with running buffer (chapter 5.7, page 115) at 90 V until the samples passed into the separation gel. Subsequently, the voltage was changed to 110 V and the gel run until the bromophenol blue left the gel.

6.8.1.4 Blot

Afterwards the gel was prepared for blotting with the following setup:

Table 6-14: Structure of the blotting setup

| |
|---|
| - |
| Black plastic |
| Sponge |
| Whatman paper (7 x 10 cm) |
| SDS gel |
| Membrane (Nitrocellulose, 0.45 μ m) |
| Whatman paper (7 x 10 cm) |
| Sponge |
| White plastic |
| + |
| Ice block |

The blot was run in blotting buffer (chapter 5.7, page 115) with 300 mA for 1.2 hours. A Ponceau S staining (2 min) helped to control an appropriate protein transfer to the nitrocellulose membrane.

6.8.1.5 Protein staining

The membrane was blocked with milk (5 %) in PBS-T (30 min, RT, shaking) and then incubated in 5 ml milk (5 %) PBS-T and a 1:50 dilution of the α -BZLF1 antibody (chapter 5.3, page 113) BZ1 (4 °C, over night, rotating). The membrane was washed the next day with PBS-T (10 min, 2 x 5 min, 10 min) and then incubated with the 1:10,000 diluted α -mouse horseradish peroxidase (HRP) antibody (chapter 5.3, page 113) in 5 ml milk (5 %) PBS-T (45 min, RT, rotating).

The membrane was incubated for 5 min in 4 ml ECL (GE healthcare) and developed for different time intervals on a X-ray screen film (Agfa HealthCare).

6.8.2 Western blot protein quantification

For protein quantification the signal intensity of the BZLF1 protein was measured with ImageJ. The results were used to calculate the number of BZLF1 dimers per cells with the help of a self-written R script, which can be found on the CD:

```
1_Code/4_Western_blot/0_R_code/WesternBlot_quantification.R
```

7 Literature

- Abdel-Wahab, O., A. Mullally, C. Hedvat, G. Garcia-Manero, J. Patel, M. Wadleigh, S. Malinge, J. Yao, O. Kilpivaara, R. Bhat, K. Huberman, S. Thomas, I. Dolgalev, A. Heguy, E. Paietta, M.M. Le Beau, M. Beran, M.S. Tallman, B.L. Ebert, H.M. Kantarjian, R.M. Stone, D.G. Gilliland, J.D. Crispino, and R.L. Levine. 2009. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood*. 114:144–7. doi:10.1182/blood-2009-03-210039.
- Adams, A. 1987. Replication of Latent Epstein-Barr Virus Genomes in Raji Cells. *J. Virol.* 61:1743–1746.
- Adamson, A.L., and S. Kenney. 1999. The Epstein-Barr Virus BZLF1 Protein Interacts Physically and Functionally with the Histone Acetylase CREB-Binding Protein. *J. Virol.* 73:6551–6558.
- Afgan, E., D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. Čech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, B. Grüning, A. Guerler, J. Hillman-Jackson, G. Von Kuster, E. Rasche, N. Soranzo, N. Turaga, J. Taylor, A. Nekrutenko, and J. Goecks. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* 44:gkw343. doi:10.1093/nar/gkw343.
- Albanese, M., T. Tagawa, M. Bouvet, L. Maliqi, D. Lutter, J. Hoser, M. Hastreiter, M. Hayes, B. Sugden, L. Martin, A. Moosmann, and W. Hammerschmidt. 2016. Epstein–Barr virus microRNAs reduce immune surveillance by virus-specific CD8⁺ T cells. *Proc. Natl. Acad. Sci.* 201605884. doi:10.1073/pnas.1605884113.
- Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. 2002. *Molecular Biology of the Cell*. 4th edition. Garland Science.
- Allis, C.D., and T. Jenuwein. 2016. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* 17:1–14. doi:10.1038/nrg.2016.59.
- Ameyar, M., M. Wisniewska, and J.B. Weitzman. 2003. A role for AP-1 in apoptosis: The case for and against. *Biochimie.* 85:747–752. doi:10.1016/j.biochi.2003.09.006.
- Angel, P., E. a Allegretto, S.T. Okino, K. Hattori, W.J. Boyle, T. Hunter, and M. Karin. 1988. Oncogene jun encodes a sequence-specific trans-activator similar to AP-1. *Nature.* 332:166–71. doi:10.1038/332166a0.
- Anton, T., S. Bultmann, H. Leonhardt, and Y. Markaki. 2014. Visualization of specific DNA

- sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus*. 5:163–72. doi:10.4161/nucl.28488.
- Arany, Z., D. Newsome, E. Oldread, D.M. Livingston, and R. Eckner. 1995. A family of transcriptional adaptor proteins targeted by the E1A oncoprotein. *Nature*. 374:81–84. doi:10.1038/374081a0.
- Armour, C.D., J.C. Castle, R. Chen, T. Babak, P. Loerch, S. Jackson, J.K. Shah, J. Dey, C. a Rohl, J.M. Johnson, and C.K. Raymond. 2009. Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. *Nat. Methods*. 6:647–649. doi:10.1038/nmeth.1360.
- Avery, M., O. Macleod, and C. McCarty. 1944. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation By a Desoxyribonucleic Acid Fraction Isolated From Pneumococcus Type Iii. *J. Exp. Med.* 79:137–158. doi:10.1084/jem.79.2.137.
- Bailey, T.L., J. Johnson, C.E. Grant, and W.S. Noble. 2015. The MEME Suite. *Nucleic Acids Res.* 43:W39–W49. doi:10.1093/nar/gkv416.
- Banerji, J., S. Rusconi, and W. Schaffner. 1981. Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell*. 27:299–308. doi:10.1016/0092-8674(81)90413-X.
- Bannister, A.J., and T. Kouzarides. 1995. CBP-induced stimulation of c-Fos activity is abrogated by E1A. *EMBO J.* 14:4758–62.
- Bannister, A.J., and T. Kouzarides. 2011. Regulation of chromatin by histone modifications. *Cell Res.* 21:381–395. doi:10.1038/cr.2011.22.
- Bannister, A.J., R. Schneider, F.A. Myers, A.W. Thorne, C. Crane-Robinson, and T. Kouzarides. 2005. Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J. Biol. Chem.* 280:17732–17736. doi:10.1074/jbc.M500796200.
- Bannister, A.J., P. Zegerman, J.F. Partridge, E. a Miska, J.O. Thomas, R.C. Allshire, and T. Kouzarides. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*. 410:120–124. doi:10.1038/35065138.
- Baron, S., and R. Whitley. 1996. Medical Microbiology. 4th edition.
- Barski, A., S. Cuddapah, K. Cui, T.Y. Roh, D.E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao. 2007. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*. 129:823–837. doi:10.1016/j.cell.2007.05.009.

- Baumann, M., R. Feederle, E. Kremmer, and W. Hammerschmidt. 1999. Cellular transcription factors recruit viral replication proteins to activate the Epstein-Barr virus origin of lytic DNA replication, oriLyt. *EMBO J.* 18:6095–6105. doi:10.1093/emboj/18.21.6095.
- Baumann, M., H. Mischak, S. Dammeier, W. Kolch, O. Gires, D. Pich, R. Zeidler, H.J. Delecluse, and W. Hammerschmidt. 1998. Activation of the Epstein-Barr virus transcription factor BZLF1 by 12-O-tetradecanoylphorbol-13-acetate-induced phosphorylation. *J. Virol.* 72:8105–8114.
- Baylis, S.A., E. Littler, C. Hospital, and W. Road. 1989. volume 17 Number 19 1989 Nucleic Acids Research The characterization of the EBV alkaline deoxyribonuclease cloned and expressed in E.coli. 17:7609–7622.
- Ben-bassat, H., N. Goldblum, S. Mitrani, G. Tamar, J.M. Yoffey, M.M. Cohen, Z. Bentwich, B. Ramot, E. Klein, and G. Klein. 1977. Establishment in continuous culture of a new type of Lymphocyte from a “Burkitt-Like” Malignant Lymphoma. *Int. J. Cancer.* 33:27–33.
- Benes, V., J. Blake, and K. Doyle. 2011. Ribo-Zero Gold Kit: improved RNA-seq results after removal of cytoplasmic and mitochondrial ribosomal RNA. *Nat. Methods.* 8:iii–iv. doi:10.1038/nmeth.f.352.
- Bentley, D.R., N.A. Gormley, S. Balasubramanian, P.A. Baybayan, V.A. Benoit, H.P. Swerdlow, G.P. Smith, J.A. Bridgham, A.A. Brown, J. Milton, A.A. Bundu, C.G. Brown, P.A. Granieri, K.P. Hall, T.A. Huw Jones, D.J. Evers, C.L. Barnes, M.A. Laurent, J.A. Loch, H.R. Bignell, J.M. Boutell, M.A. Osborne, J. Bryant, M.A. Smith, R.J. Carter, R. Keira Cheetham, A.J. Cox, D.J. Ellis, M.R. Flatbush, S.J. Humphray, L.J. Irving, M.S. Karbelashvili, S.M. Kirk, H. Li, X. Liu, K.S. Maisinger, L.J. Murray, B. Obradovic, T. Ost, M.L. Parkinson, M.R. Pratt, I.M.J. Rasolonjatovo, M.T. Reed, R. Rigatti, C. Rodighiero, M.T. Ross, A. Sabot, S. V Sankar, A. Scally, G.P. Schroth, M.E. Smith, V.P. Smith, A. Spiridou, P.E. Torrance, S.S. Tzonev, E.H. Vermaas, K. Walter, X. Wu, L. Zhang, M.D. Alam, C. Anastasi, I.C. Aniebo, D.M.D. Bailey, I.R. Bancarz, S. Banerjee, S.G. Barbour, K.F. Benson, C. Bevis, P.J. Black, A. Boodhun, J.S. Brennan, R.C. Brown, D.H. Buermann, J.C. Burrows, N.P. Carter, N. Castillo, M. Chiara E Catenazzi, S. Chang, R. Neil Cooley, N.R. Crake, O.O. Dada, K.D. Diakoumakos, B. Dominguez-Fernandez, D.J. Earnshaw, U.C. Egbujor, D.W. Elmore, S.S. Etchin, M.R. Ewan, M. Fedurco, L.J. Fraser, K. V Fuentes Fajardo, W. Scott Furey, D. George, K.J. Gietzen, C.P. Goddard,

- G.S. Golda, D.E. Green, D.L. Gustafson, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 456:53–9. doi:10.1038/nature07517.
- Bergbauer, M., M. Kalla, A. Schmeinck, C. Göbel, U. Rothbauer, S. Eck, A. Benet-Pagés, T.M. Strom, and W. Hammerschmidt. 2010. CpG-methylation regulates a class of Epstein-Barr virus promoters. *PLoS Pathog*. 6. doi:10.1371/journal.ppat.1001114.
- Bernstein, B.E., M. Kamal, K. Lindblad-Toh, S. Bekiranov, D.K. Bailey, D.J. Huebert, S. McMahon, E.K. Karlsson, E.J. Kulbokas, T.R. Gingeras, S.L. Schreiber, and E.S. Lander. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*. 120:169–181. doi:10.1016/j.cell.2005.01.001.
- Bernstein, E., and S.B. Hake. 2006. The nucleosome: a little variation goes a long way. *Biochem. Cell Biol*. 84:505–507. doi:10.1139/o06-085.
- Bhende, P.M., W.T. Seaman, H.-J. Delecluse, and S.C. Kenney. 2004. The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. *Nat. Genet*. 36:1099–1104. doi:10.1038/ng1424.
- Biggin, M., M. Bodescot, M. Perricaudet, and P. Farrell. 1987. Epstein-Barr virus gene expression in P3HR1-superinfected Raji cells. *J. Virol*. 61:3120–32.
- Biggin, M.D. 2011. Animal Transcription Networks as Highly Connected, Quantitative Continua. *Dev. Cell*. 21:611–626. doi:10.1016/j.devcel.2011.09.008.
- Bird, A.P. 1980. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res*. 8:1499–504.
- Birney, E., J.A. Stamatoyannopoulos, A. Dutta, R. Guigo, T.R. Gingeras, E.H. Margulies, Z. Weng, M. Snyder, E.T. Dermitzakis, R.E. Thurman, M.S. Kuehn, C.M. Taylor, S. Neph, C.M. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J.A. Greenbaum, R.M. Andrews, P. Flicek, P.J. Boyle, H. Cao, N.P. Carter, G.K. Clelland, S. Davis, N. Day, P. Dhami, S.C. Dillon, M.O. Dorschner, H. Fiegler, P.G. Giresi, J. Goldy, M. Hawrylycz, A. Haydock, R. Humbert, K.D. James, B.E. Johnson, E.M. Johnson, T.T. Frum, E.R. Rosenzweig, N. Karnani, K. Lee, G.C. Lefebvre, P.A. Navas, F. Neri, S.C. Parker, P.J. Sabo, R. Sandstrom, A. Shafer, D. Vetrie, M. Weaver, S. Wilcox, M. Yu, F.S. Collins, J. Dekker, J.D. Lieb, T.D. Tullius, G.E. Crawford, S. Sunyaev, W.S. Noble, I. Dunham, F. Denoeud, A. Reymond, P. Kapranov, J. Rozowsky, D. Zheng, R. Castelo, A. Frankish, J. Harrow, S. Ghosh, A. Sandelin, I.L. Hofacker, R. Baertsch, D. Keefe, S. Dike, J. Cheng, H.A.

- Hirsch, E.A. Sekinger, J. Lagarde, J.F. Abril, A. Shahab, C. Flamm, C. Fried, J. Hackermuller, J. Hertel, M. Lindemeyer, K. Missal, A. Tanzer, S. Washietl, J. Korbel, O. Emanuelsson, J.S. Pedersen, N. Holroyd, R. Taylor, D. Swarbreck, N. Matthews, M.C. Dickson, D.J. Thomas, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 447:799–816. doi:10.1038/nature05874.
- Bohmann, D., and R. Tjian. 1989. Biochemical analysis of transcriptional activation by Jun: Differential activity of c- and v-Jun. *Cell*. 59:709–717. doi:10.1016/0092-8674(89)90017-2.
- Bornkamm, G.W., C. Berens, C. Kuklik-Roos, J.M. Bechet, G. Laux, J. Bachl, M. Korndoerfer, M. Schlee, M. Hölzel, A. Malamoussi, R.D. Chapman, F. Nimmerjahn, J. Mautner, W. Hillen, H. Bujard, and J. Feuillard. 2005. Stringent doxycycline-dependent control of gene activities using an episomal one-vector system. *Nucleic Acids Res*. 33:1–11. doi:10.1093/nar/gni137.
- Boyer, L.A., K. Plath, J. Zeitlinger, T. Brambrink, L.A. Medeiros, T.I. Lee, S.S. Levine, M. Wernig, A. Tajonar, M.K. Ray, G.W. Bell, A.P. Otte, M. Vidal, D.K. Gifford, R.A. Young, and R. Jaenisch. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*. 441:349–353. doi:10.1038/nature04733.
- Boyes, J., and A. Bird. 1992. Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J*. 11:327–333. doi:10.1101/gad.1396206.GENES.
- Brinkman, A.B., H. Gu, S.J.J. Bartels, Y. Zhang, F. Matarese, F. Simmer, H. Marks, C. Bock, A. Gnirke, A. Meissner, and H.G. Stunnenberg. 2012. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res*. 22:1128–1138. doi:10.1101/gr.133728.111.
- Bristol, J. a, A.R. Robinson, E. a Barlow, and S.C. Kenney. 2010. The Epstein-Barr virus BZLF1 protein inhibits tumor necrosis factor receptor 1 expression through effects on cellular C/EBP proteins. *J. Virol*. 84:12362–12374. doi:10.1128/JVI.00712-10.
- Brown, P.O., and D. Botstein. 1999. Exploring the new world of the genome with DNA microarrays. *Nat. Genet*. 21:33–37. doi:10.1038/4462.
- Buisson, M., T. Géoui, D. Flot, N. Tarbouriech, M.E. Ressing, E.J. Wiertz, and W.P. Burmeister. 2009. A Bridge Crosses the Active-Site Canyon of the Epstein-Barr Virus

- Nuclease with DNase and RNase Activities. *J. Mol. Biol.* 391:717–728. doi:10.1016/j.jmb.2009.06.034.
- Burkitt, D. 1958. A Sarcoma Involving the Jaws in African Children. *Cancer J. Clin.* 22:349–355. doi:10.3322/canjclin.22.6.349.
- Bustin, S.A. 2000. Absolute quantification of mrna using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol.* 25:169–193. doi:JME00927 [pii].
- Bustin, S.A., V. Benes, T. Nolan, and M.W. Pfaffl. 2005. Quantitative real-time RT-PCR - A perspective. *J. Mol. Endocrinol.* 34:597–601. doi:10.1677/jme.1.01755.
- Cai, S., C.C. Lee, and T. Kohwi-Shigematsu. 2006. SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat. Genet.* 38:1278–88. doi:10.1038/ng1913.
- Carey, M., J. Kolman, D.A. Katz, L. Gradoville, L. Barberis, and G. Miller. 1992. Transcriptional synergy by the Epstein-Barr virus transactivator ZEBRA. *J. Virol.* 66:4803–13.
- Carey, M., Y.S. Lin, M.R. Green, and M. Ptashne. 1990. A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. *Nature.* 345:361–364. doi:10.1038/345361a0.
- Cayrol, C., and E.K. Flemington. 1995. Identification of cellular target genes of the Epstein-Barr virus transactivator Zta: activation of transforming growth factor beta igh3 (TGF-beta igh3) and TGF-beta 1. *J. Virol.* 69:4206–12.
- Chahwan, R., S.N. Wontakal, and S. Roa. 2010. Crosstalk between genetic and epigenetic information through cytosine deamination. *Trends Genet.* 26:443–448. doi:10.1016/j.tig.2010.07.005.
- Chang, Y., H. Lee, Y. Chen, J. Lu, S. Wu, C. Chen, K. Takada, and C. Tsai. 2006. Induction of the early growth response 1 gene by Epstein-Barr virus lytic transactivator Zta. *J. Virol.* 80:7748–55. doi:10.1128/JVI.02608-05.
- Chang, Y.N., D.L. Dong, G.S. Hayward, and S.D. Hayward. 1990. The Epstein-Barr virus Zta transactivator: a member of the bZIP family with unique DNA-binding specificity and a dimerization domain that lacks the characteristic heptad leucine zipper motif. *J. Virol.* 64:3358–69.
- Chen, T., Y. Ueda, J.E. Dodge, Z. Wang, and E. Li. 2003. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b.

Mol. Cell. Biol. 23:5594–605. doi:10.1128/MCB.23.16.5594.

- Chen, X., H. Xu, P. Yuan, F. Fang, M. Huss, V.B. Vega, E. Wong, Y.L. Orlov, W. Zhang, J. Jiang, Y.H. Loh, H.C. Yeo, Z.X. Yeo, V. Narang, K.R. Govindarajan, B. Leong, A. Shahab, Y. Ruan, G. Bourque, W.K. Sung, N.D. Clarke, C.L. Wei, and H.H. Ng. 2008. Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell*. 133:1106–1117. doi:10.1016/j.cell.2008.04.043.
- Cheng, X., and R.M. Blumenthal. 2008. Mammalian DNA methyltransferases: a structural perspective. *Structure*. 16:341–50. doi:10.1016/j.str.2008.01.004.
- Chevallier-Greco, A., E. Manet, P. Chavrier, C. Mosnier, J. Daillie, and A. Sergeant. 1986. Both Epstein-Barr virus (EBV)-encoded trans-acting factors, EB1 and EB2, are required to activate transcription from an EBV early promoter. *EMBO J.* 5:3243–9.
- Chi, T., and M. Carey. 1993. The ZEBRA activation domain: modular organization and mechanism of action. *Mol. Cell. Biol.* 13:7045–7055.
- Chiu, S.-H., C.-C. Wu, C.-Y. Fang, S.-L. Yu, H.-Y. Hsu, Y.-H. Chow, and J.-Y. Chen. 2014a. Epstein-Barr virus BALF3 mediates genomic instability and progressive malignancy in nasopharyngeal carcinoma. *Oncotarget*. 5:8583–8601. doi:10.18632/oncotarget.2323.
- Chiu, S.-H., M.-C. Wu, C.-C. Wu, Y.-C. Chen, S.-F. Lin, J.T. Hsu, C.-S. Yang, C.-H. Tsai, K. Takada, M.-R. Chen, and J.-Y. Chen. 2014b. Epstein-Barr virus BALF3 has nuclease activity and mediates mature virion production during the lytic cycle. *J. Virol.* 88:4962–75. doi:10.1128/JVI.00063-14.
- Chubb, J.R., S. Boyle, P. Perry, and W.A. Bickmore. 2002. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.* 12:439–445. doi:10.1016/S0960-9822(02)00695-4.
- Clamp, M., B. Fry, M. Kamal, X. Xie, J. Cuff, M.F. Lin, M. Kellis, K. Lindblad-Toh, and E.S. Lander. 2007. Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* 104:19428–33. doi:10.1073/pnas.0709013104.
- Cohen, D.R., and T. Curran. 1988. Fra-1: a Serum-Inducible, Cellular Immediate-Early Gene That Encodes a Fos-Related Antigen. *Mol. Cell. Biol.* 8:2063–2069. doi:10.1128/MCB.8.5.2063.Updated.
- Cohen, D.R., P.C. Ferreira, R. Gantz, B.R. Franza, and T. Curran. 1989. The product of a fos-related gene, fra-1, binds cooperatively to the AP-1 site with Jun: transcription factor AP-1 is comprised of multiple protein complexes. *Genes Dev.* 3:173–184.

doi:10.1101/gad.3.2.173.

- Cokus, S.J., S. Feng, X. Zhang, Z. Chen, B. Merriman, C.D. Haudenschild, S. Pradhan, S.F. Nelson, M. Pellegrini, and S.E. Jacobsen. 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature*. 452:215–219. doi:10.1038/nature06745.
- Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, P.D. Hsu, X. Wu, W. Jiang, and L.A. Marraffini. 2013. Multiplex Genome Engineering Using CRISPR/VCas Systems. *Science (80-.)*. 339:819–823. doi:10.1126/science.1231143.Multiplex.
- Countryman, J., and G. Miller. 1985. Activation of expression of latent Epstein-Barr herpesvirus after gene transfer with a small cloned subfragment of heterogeneous viral DNA. *Proc. Natl. Acad. Sci. U. S. A.* 82:4085–4089. doi:10.1073/pnas.82.12.4085.
- Cremer, T., and C. Cremer. 2001. Chromosome Territories, Nuclear Architecture and Gene Regulation in Mammalian Cells. *Nat. Rev. Genet.* 2:292–301. doi:10.1038/35066075.
- Cremer, T.C. and M. 2010. Chromosome Territories. 147:13–24. doi:10.1101/cshperspect.a003889.
- Creyghton, M.P., A.W. Cheng, G.G. Welstead, T. Kooistra, B.W. Carey, E.J. Steine, J. Hanna, M. a Lodato, G.M. Frampton, P. a Sharp, L. a Boyer, R. a Young, and R. Jaenisch. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* 107:21931–21936. doi:10.1073/pnas.1016071107.
- Cross, S., J. Charlton, X. Nan, and A. Bird. 1994. Purification of CpG islands using a methylated DNA binding column. *Nature*. 6:236–344. doi:10.1038/ng0394-236.
- Dai, P., H. Akimaru, Y. Tanaka, D.X. Hou, T. Yasukawa, C. Kanei-Ishii, T. Takahashi, and S. Ishii. 1996. CBP as a transcriptional coactivator of c-Myb. *Genes Dev.* 10:528–540. doi:10.1101/gad.10.5.528.
- Daikoku, T., A. Kudoh, M. Fujita, Y. Sugaya, H. Isomura, N. Shirata, and T. Tsurumi. 2005. Architecture of replication compartments formed during Epstein-Barr virus lytic replication. *J. Virol.* 79:3409–3418. doi:10.1128/JVI.79.6.3409-3418.2005.
- Davison, A.J. 2007. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis.
- Deaton, A., and A. Bird. 2011. CpG islands and the regulation of transcription. *Genes Dev.* 25:1010–1022. doi:10.1101/gad.2037511.1010.
- Decaussin, G., V. Leclerc, and T. Ooka. 1995. The lytic cycle of Epstein-Barr virus in the

nonproducer Raji line can be rescued by the expression of a 135-kilodalton protein encoded by the BALF2 open reading frame. *J. Virol.* 69:7309–7314.

Dekker, J., K. Rippe, M. Dekker, and N. Kleckner. 2002. Capturing chromosome conformation. *Science.* 295:1306–1311. doi:10.1126/science.1067799 [doi].

DeMare, L.E., J. Leng, J. Cotney, S.K. Reilly, J. Yin, R. Sarro, and J.P. Noonan. 2013. The genomic landscape of cohesin-associated chromatin interactions. *Genome Res.* 23:1224–34. doi:10.1101/gr.156570.113.

Deppmann, C.D., R.S. Alvania, and E.J. Taparowsky. 2006. Cross-species annotation of basic leucine zipper factor interactions: Insight into the evolution of closed interaction networks. *Mol. Biol. Evol.* 23:1480–1492. doi:10.1093/molbev/msl022.

Dickerson, S.J., Y. Xing, A.R. Robinson, W.T. Seaman, H. Gruffat, and S.C. Kenney. 2009. Methylation-dependent binding of the Epstein-Barr virus BZLF1 protein to viral promoters. *PLoS Pathog.* 5. doi:10.1371/journal.ppat.1000356.

Dixon, J.R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J.S. Liu, and B. Ren. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 485:376–380. doi:10.1038/nature11082.

Dostie, J., T.A. Richmond, R.A. Arnaout, R.R. Selzer, W.L. Lee, T.A. Honan, E.D. Rubio, A. Krumm, J. Lamb, C. Nusbaum, R.D. Green, and J. Dekker. 2006. Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16:1299–1309. doi:10.1101/gr.5571506.

Dreyfus, D.H., M. Nagasawa, J.C. Pratt, C.A. Kelleher, and E.W. Gelfand. 1999. Inactivation of NF-kappaB by EBV BZLF-1-encoded ZEBRA protein in human T cells. *J Immunol.* 163:6261–6268.

Dyson, P.J., and P.J. Farrell. 1985. Chromatin Structure of Epstein-Barr Virus. *J. Gen. Virol.* 66:1931–1940. doi:10.1099/0022-1317-66-9-1931.

Dyson, P.J., and T.H. Rabbitts. 1985. Chromatin structure around the c-myc gene in Burkitt lymphomas with upstream and downstream translocation points. *Proc Natl Acad Sci U S A.* 82:1984–1988.

Eferl, R., and E.F. Wagner. 2003. AP-1: A double-edged sword in tumorigenesis. *Nat. Rev. Cancer.* 3:859–868. doi:10.1038/nrc1209.

Ehrlich, M., M.A. Gama-Sosa, L.H. Huang, R.M. Midgett, K.C. Kuo, R.A. Mccune, and C.

- Gehrke. 1982. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Res.* 10:2709–2721. doi:10.1093/nar/10.8.2709.
- El-Guindy, A., L. Heston, H.-J. Delecluse, and G. Miller. 2007. Phosphoacceptor site S173 in the regulatory domain of Epstein-Barr Virus ZEBRA protein is required for lytic DNA replication but not for activation of viral early genes. *J. Virol.* 81:3303–3316. doi:10.1128/JVI.02445-06.
- El-Guindy, A.S., Y.P. So, J. Countryman, and G. Miller. 2006. Identification of constitutive phosphorylation sites on the Epstein-Barr virus ZEBRA protein. *J. Biol. Chem.* 281:3085–3095. doi:10.1074/jbc.M506076200.
- Ellenberger, T. 1994. Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains. *Curr. Opin. Struct. Biol.* 4:12–21. doi:10.1016/S0959-440X(94)90054-X.
- Epstein, M., B. Achong, and Y. Barr. 1964. Virus Particles in Cultures Lymphoblasts Form Burkitt'S Lymphoma. *Lancet.* 1964.
- Epstein, M., and Y. Barr. 1964. Cultivation in vitro of human lymphoblasts from burkitt's malignant lymphoma. *Lancet.* 252–253.
- Ernst, J., and M. Kellis. 2013. Interplay between chromatin state, regulator binding, and regulatory motifs in six human cell types. *Genome Res.* 23:1142–1154. doi:10.1101/gr.144840.112.
- Ezkurdia, I., D. Juan, J.M. Rodriguez, A. Frankish, M. Diekhans, J. Harrow, J. Vazquez, A. Valencia, and M.L. Tress. 2014. Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Hum. Mol. Genet.* 23:5866–5878. doi:10.1093/hmg/ddu309.
- Farrell, P.J., D.T. Rowe, C.M. Rooney, and T. Kouzarides. 1989. Epstein-Barr virus BZLF1 trans-activator specifically binds to a consensus AP-1 site and is related to c-fos. *EMBO J.* 8:127–32.
- Feederle, R., H. Bannert, H. Lips, N. Muller-Lantzsch, and H.J. Delecluse. 2009. The Epstein-Barr Virus Alkaline Exonuclease BGLF5 Serves Pleiotropic Functions in Virus Replication. *J. Virol.* 83:4952–4962. doi:10.1128/jvi.00170-09.
- Fixman, E.D., G.S. Hayward, and S.D. Hayward. 1992. trans-acting requirements for replication of Epstein-Barr virus ori-Lyt. *J. Virol.* 66:5030–5039.
- Flemington, E., and S.H. Speck. 1990a. Evidence for coiled-coil dimer formation by an

- Epstein-Barr virus transactivator that lacks a heptad repeat of leucine residues. *Proc. Natl. Acad. Sci. U. S. A.* 87:9459–9463.
- Flemington, E., and S.H. Speck. 1990b. Epstein-Barr Virus BZLF1 trans Activator Induces the Promoter of a Cellular Cognate Gene , c-fos. 64:4549–4552.
- Flemington, E.K., a M. Borrás, J.P. Lytle, and S.H. Speck. 1992. Characterization of the Epstein-Barr virus BZLF1 protein transactivation domain. *J. Virol.* 66:922–929.
- Franklin, R., and R. Gosling. 1953. Molecular Configuration in Sodium Thymonucleate. 740–741. doi:10.1038/171740a0 ISSN.
- Frommer, M., L.E. McDonald, D.S. Millar, C.M. Collis, F. Watt, G.W. Grigg, P.L. Molloy, and C.L. Paul. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U. S. A.* 89:1827–31. doi:10.1073/pnas.89.5.1827.
- Fullwood, M.J., M.H. Liu, Y.F. Pan, J. Liu, H. Xu, Y. Bin Mohamed, Y.L. Orlov, S. Velkov, A. Ho, P.H. Mei, E.G.Y. Chew, P.Y.H. Huang, W.-J. Welboren, Y. Han, H.S. Ooi, P.N. Ariyaratne, V.B. Vega, Y. Luo, P.Y. Tan, P.Y. Choy, K.D.S.A. Wansa, B. Zhao, K.S. Lim, S.C. Leow, J.S. Yow, R. Joseph, H. Li, K. V Desai, J.S. Thomsen, Y.K. Lee, R.K.M. Karuturi, T. Herve, G. Bourque, H.G. Stunnenberg, X. Ruan, V. Cacheux-Rataboul, W.-K. Sung, E.T. Liu, C.-L. Wei, E. Cheung, and Y. Ruan. 2009. An oestrogen-receptor- α -bound human chromatin interactome. *Nature.* 461:58–64. doi:10.1038/nature08497.
- Gall, J.G., and M. Lou Pardue. 1969. Formation and Detection of Rna-Dna Hybrid Molecules in Cytological Preparations*. *Proc Natl Acad Sci U S A.* 63:378–383. doi:10.1073/pnas.63.2.378.
- Gao, Z., A. Krithivas, J.E. Finan, O.J. Semmes, S. Zhou, Y. Wang, and S.D. Hayward. 1998. The Epstein-Barr virus lytic transactivator Zta interacts with the helicase-primase replication proteins. *J. Virol.* 72:8559–8567.
- Garcia-Rodriguez, C., and A. Rao. 1998. Nuclear Factor of Activated T Cells (NFAT) - dependent. 187:2031–2036.
- Gavrieli, Y., Y. Sherman, and S.A. Ben-Sasson. 1992. Identification of Programmed Cell Death In Situ via Specific Labeling of Neuclear DNA Fragmentation. *J. Cell Biol.* 119:493–501.
- Ghisletti, S., I. Barozzi, F. Mietton, S. Polletti, F. De Santa, E. Venturini, L. Gregory, L. Lonie, A. Chew, C.L. Wei, J. Ragoussis, and G. Natoli. 2010. Identification and

Characterization of Enhancers Controlling the Inflammatory Gene Expression Program in Macrophages. *Immunity*. 32:317–328. doi:10.1016/j.immuni.2010.02.008.

Griffith, F. 1928. The Significance of Pneumococcal Types. *Journ. Hyg.* 27:113–159. doi:10.1017/S0022172400031879.

Groetsch, B., C. Lang, K. Neubert, S. Brachs, C. Zech, A. Derer, J. Luther, D. Voehringer, E. Wagner, G. Schett, D. Mielenz, and J.-P. David. 2012. The AP-1 transcription factor Fra-1 is a negative regulator of B2B cell proliferation and of immunoglobulin class switch. *J. Immunol.*

Gruenbaum, Y., H. Cedar, and a Razin. 1982. Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature*. 295:620–622. doi:10.1038/295620a0.

Gustems, M., A. Woellmer, U. Rothbauer, S.H. Eck, T. Wieland, D. Lutter, and W. Hammerschmidt. 2014. C-Jun/c-Fos heterodimers regulate cellular genes via a newly identified class of methylated DNA sequence motifs. *Nucleic Acids Res.* 42:3059–3072. doi:10.1093/nar/gkt1323.

Hagège, H., P. Klous, C. Braem, E. Splinter, J. Dekker, G. Cathala, W. de Laat, and T. Forné. 2007. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* 2:1722–1733. doi:10.1038/nprot.2007.243.

Halazonetis, T.D., K. Georgopoulos, M.E. Greenberg, and P. Leder. 1988. c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities. *Cell*. 55:917–924. doi:10.1016/0092-8674(88)90147-X.

Hamlyn, P., and T. Rabbitts. 1983. Translocation joins c-myc and immunoglobulin γ 1 genes in a Burkitt lymphoma revealing a third exon in the c-myc oncogene. *Nature*.

Harrington, M.A., P.A. Jones, M. Imagawat, and M. Karint. 1988. Cytosine methylation does not affect binding of transcription factor Spl (DNA methylation/metallothionein gene expression/“GC box”). *Biochemistry*. 85:2066–2070. doi:10.1073/pnas.85.7.2066.

Hashimoto, H., Y. Liu, A.K. Upadhyay, Y. Chang, S.B. Howerton, P.M. Vertino, X. Zhang, and X. Cheng. 2012. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* 40:4841–4849. doi:10.1093/nar/gks155.

Hau, P.M., W. Deng, L. Jia, J. Yang, T. Tsurumi, A.K.S. Chiang, M.S.-Y. Huen, and S.W. Tsao. 2015. Role of ATM in the Formation of the Replication Compartment during Lytic Replication of Epstein-Barr Virus in Nasopharyngeal Epithelial Cells. *J. Virol.* 89:652–

668. doi:10.1128/JVI.01437-14.

zur Hausen, H., F.J. O'Neill, U.K. Freese, and E. Hecker. 1978. Persisting oncogenic herpesvirus induced by the tumour promoter TPA. *Nature*. 272:373–375. doi:10.1038/272373a0.

zur Hausen, H., H. Schulte-Holtenhausen, G. Klein, W. Henle, G. Henle, P. Clifford, and L. Santesson. 1970. Epstein–Barr Virus in Burkitt's Lymphoma and Nasopharyngeal Carcinoma: EBV DNA in Biopsies of Burkitt Tumours and Anaplastic Carcinomas of the Nasopharynx. *Nature*. 228:1056–1058. doi:10.1038/2281056a0.

Hayatsu, H., Y. Wataya, K. Kai, and S. Iida. 1970. Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry*. 9:2858–2865. doi:10.1021/ja00706a062.

He, Y.-F., B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He, and G.-L. Xu. 2011a. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*. 333:1303–7. doi:10.1126/science.1210944.

He, Y.-F., B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C.-X. Song, K. Zhang, C. He, and G.-L. Xu. 2011b. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*. 333:1303–7. doi:10.1126/science.1210944.

Heintzman, N.D., G.C. Hon, R.D. Hawkins, P. Kheradpour, A. Stark, L.F. Harp, Z. Ye, L.K. Lee, R.K. Stuart, C.W. Ching, K. a Ching, J.E. Antosiewicz-Bourget, H. Liu, X. Zhang, R.D. Green, V. V Lobanenkov, R. Stewart, J. a Thomson, G.E. Crawford, M. Kellis, and B. Ren. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. 459:108–112. doi:10.1038/nature07829.

Heintzman, N.D., R.K. Stuart, G. Hon, Y. Fu, C.W. Ching, R.D. Hawkins, L.O. Barrera, S. Van Calcar, C. Qu, K.A. Ching, W. Wang, Z. Weng, R.D. Green, G.E. Crawford, and B. Ren. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39:311–8. doi:10.1038/ng1966.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y.C. Lin, P. Laslo, J.X. Cheng, C. Murre, H. Singh, and C.K. Glass. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell*. 38:576–589. doi:10.1016/j.molcel.2010.05.004.

Henle, G., W. Henle, P. Clif-, V. Diehl, G.W. Kafuko, B.G.K. A, K. Richard, O. Morrow, and

- G.M.R. Munube. 1969. Antibodies to Epstein-Barr Virus in Burkitt ' s. *J. Natl. Cancer Inst.* 1147–1157.
- Henle, G., W. Henle, and V. Diehl. 1967. Relation of Burkitt's tumor-associated herpes-type virus to infectious mononucleosis. *Proc. Natl. Acad. Sci. U. S. A.* 59:94–101.
- Hicks, M.R., S.S. Al-mehairi, and A.J. Sinclair. 2003. The Zipper Region of Epstein-Barr Virus bZIP Transcription Factor Zta Is Necessary but Not Sufficient To Direct DNA Binding The Zipper Region of Epstein-Barr Virus bZIP Transcription Factor Zta Is Necessary but Not Sufficient To Direct DNA Binding. *J. Virol.* 77:8173–8177. doi:10.1128/JVI.77.14.8173.
- Hicks, M.R., S. Balesaria, C. Medina-Palazon, M.J. Pandya, D.N. Woolfson, and A.J. Sinclair. 2001. Biophysical analysis of natural variants of the multimerization region of Epstein-Barr virus lytic-switch protein BZLF1. *J. Virol.* 75:5381–5384. doi:10.1128/JVI.75.11.5381-5384.2001.
- Higuchi, R., C. Fockler, and G. Dollinger. 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Nat. Biotechnol.* 11:1026–1030. doi:10.1038/nbt0993-1026.
- Hinnebusch, A.G., and K. Natarajan. 2002. Gcn4p , a Master Regulator of Gene Expression , Is Controlled at Multiple Levels by Diverse Signals of Starvation and Stress Gcn4p , a Master Regulator of Gene Expression , Is Controlled at Multiple Levels by Diverse Signals of Starvation and Stress. 1:22–32. doi:10.1128/EC.1.1.22.
- Holland, B., J. Wong, M. Li, and S. Rasheed. 2013. Identification of Human MicroRNA-Like Sequences Embedded within the Protein-Encoding Genes of the Human Immunodeficiency Virus. *PLoS One.* 8. doi:10.1371/journal.pone.0058586.
- Höller, M., G. Westin, J. Jiricny, and W. Schaffner. 1988. Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes Dev.* 2:1127–1135. doi:10.1101/gad.2.9.1127.
- Horike, S., S. Cai, M. Miyano, J.-F. Cheng, and T. Kohwi-Shigematsu. 2005. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat. Genet.* 37:31–40. doi:10.1038/ng1491.
- Horst, D., W.P. Burmeister, I.G.J. Boer, D. van Leeuwen, M. Buisson, A.E. Gorbalenya, E.J.H.J. Wiertz, and M.E. Rensing. 2012. The “Bridge” in the Epstein-Barr virus alkaline exonuclease protein BGLF5 contributes to shutoff activity during productive infection. *J.*

Viol. 86:9175–87. doi:10.1128/JVI.00309-12.

Horvath, P., and R. Barrangou. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science*. 327:167–170. doi:10.1126/science.1179555.

Hsieh, C. 1994. Dependence of transcriptional repression on CpG methylation Dependence of Transcriptional Repression on CpG Methylation Density. 14:5487–5494.

Hsu, M., S.-Y. Wu, S.-S. Chang, I.-J. Su, C.-H. Tsai, S.-J. Lai, A.-L. Shiau, K. Takada, and Y. Chang. 2008. Epstein-Barr virus lytic transactivator Zta enhances chemotactic activity through induction of interleukin-8 in nasopharyngeal carcinoma cells. *J. Virol.* 82:3679–88. doi:10.1128/JVI.02301-07.

Hu, S., J. Wan, Y. Su, Q. Song, Y. Zeng, H.N. Nguyen, J. Shin, E. Cox, H.S. Rho, C. Woodard, S. Xia, S. Liu, H. Lyu, G.L. Ming, H. Wade, H. Song, J. Qian, and H. Zhu. 2013. DNA methylation presents distinct binding sites for human transcription factors. *Elife*. 2013:1–16. doi:10.7554/eLife.00726.

Huang, T.H., D.E. Laux, B.C. Hamlin, T. Hw-ming, E. Laux, C. Hamlin, and B. Lubahn. 1997. Identification of DNA Methylation Markers for Human Breast Carcinomas Using the Methylation-sensitive Restriction Fingerprinting Technique Advances in Brief Identification of DNA Methylation Markers for Human Breast Carcinomas Using the Methylation-sensit. 1030–1034.

Hughes, J.R., N. Roberts, S. McGowan, D. Hay, E. Giannoulatou, M. Lynch, M. De Gobbi, S. Taylor, R. Gibbons, and D.R. Higgs. 2014. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* 46:205–12. doi:10.1038/ng.2871.

Iborra, F.J., a Pombo, D. a Jackson, and P.R. Cook. 1996. Active RNA polymerases are localized within discrete transcription "factories" in human nuclei. *J. Cell Sci.* 109 (Pt 6:1427–1436.

Ito, S., L. Shen, Q. Dai, S.C. Wu, L.B. Collins, J. a Swenberg, C. He, and Y. Zhang. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*. 333:1300–3. doi:10.1126/science.1210597.

Jin, C., C. Zang, G. Wei, K. Cui, W. Peng, K. Zhao, and G. Felsenfeld. 2009. H3.3/H2A.Z double variant-containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. *Nat. Genet.* 41:941–5. doi:10.1038/ng.409.

Jochum, S., A. Moosmann, S. Lang, W. Hammerschmidt, and R. Zeidler. 2012. The EBV

- immuno-evasins vIL-10 and BNLF2a protect newly infected B cells from immune recognition and elimination. *PLoS Pathog.* 8. doi:10.1371/journal.ppat.1002704.
- Johannsen, E., M. Luftig, M.R. Chase, S. Weicksel, E. Cahir-McFarland, D. Illanes, D. Sarracino, and E. Kieff. 2004. Proteins of purified Epstein-Barr virus. *Proc. Natl. Acad. Sci. U. S. A.* 101:16286–91. doi:10.1073/pnas.0407320101.
- Jones, P.A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13:484–492. doi:10.1038/nrg3230.
- Jones, R.J., W.T. Seaman, W.H. Feng, E. Barlow, S. Dickerson, H.J. Delecluse, and S.C. Kenney. 2007. Roles of lytic viral infection and IL-6 in early versus late passage lymphoblastoid cell lines and EBV-associated lymphoproliferative disease. *Int. J. Cancer.* 121:1274–1281. doi:10.1002/ijc.22839.
- Kalla, M., A. Schmeinck, M. Bergbauer, D. Pich, and W. Hammerschmidt. 2010. AP-1 homolog BZLF1 of Epstein-Barr virus has two essential functions dependent on the epigenetic state of the viral genome. *Pnas.* 107:850–855. doi:10.1073/pnas.0911948107.
- Kaneko-Ishino, T., and F. Ishino. 2010. Retrotransposon silencing by DNA methylation contributed to the evolution of placentation and genomic imprinting in mammals. *Dev. Growth Differ.* 52:533–543. doi:10.1111/j.1440-169X.2010.01194.x.
- Kauder, S.E., A. Bosque, A. Lindqvist, V. Planelles, and E. Verdin. 2009. Epigenetic regulation of HIV-1 latency by cytosine methylation. *PLoS Pathog.* 5. doi:10.1371/journal.ppat.1000495.
- Kenney, S., J. Kamine, E. Holley-Guthrie, J.C. Lin, E.C. Mar, and J. Pagano. 1989. The Epstein-Barr virus (EBV) BZLF1 immediate-early gene product differentially affects latent versus productive EBV promoters. *J. Virol.* 63:1729–1736.
- Keshet, I., Y. Schlesinger, S. Farkash, E. Rand, M. Hecht, E. Segal, E. Pikarski, R. a Young, A. Niveleau, H. Cedar, and I. Simon. 2006. Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat. Genet.* 38:149–153. doi:10.1038/ng1719.
- Kiehl, A., and D.I. Dorsky. 1991. Cooperation of EBV DNA Polymerase and EA-D (BMRF1) in Vitro and colocalization in nuclei of infected cells. 340:330–340.
- Kim, T.-K., M. Hemberg, J.M. Gray, A.M. Costa, D.M. Bear, J. Wu, D.A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P.F. Worley, G. Kreiman, and M.E. Greenberg. 2010. Widespread transcription at

- neuronal activity-regulated enhancers. *Nature*. 465:182–7. doi:10.1038/nature09033.
- Kintner, C., and B. Sugden. 1981. Conservation and progressive methylation of Epstein-Barr viral DNA sequences in transformed cells. *J. Virol.* 38:305–16.
- Kirchmaier, a L., and B. Sugden. 1995. Plasmid maintenance of derivatives of oriP of Epstein-Barr virus. *J. Virol.* 69:1280–1283.
- Knipe, D.M. 2015. Nuclear sensing of viral DNA, epigenetic regulation of herpes simplex virus infection, and innate immunity. *Virology*. 479–480:153–9. doi:10.1016/j.virol.2015.02.009.
- König, P., and T.J. Richmond. 1993. The X-ray Structure of the GCN4-bZIP Bound to ATF/CREB Site DNA Shows the Complex Depends on DNA Flexibility. *J. Mol. Biol.* 233:139–154. doi:10.1006/jmbi.1993.1490.
- Kornberg, R. 1974. Chromatin structure: a repeating unit of histones and DNA. *Science* (80-). 184:3–6.
- Kornberg, R. 1977. Structure of chromatin. *Annu. Rev. Biochem.* 46:931–54. doi:10.1146/annurev.bi.46.070177.004435.
- Kouzarides, T. 2007. Chromatin Modifications and Their Function. *Cell*. 128:693–705. doi:10.1016/j.cell.2007.02.005.
- Kouzarides, T., G. Packham, A. Cook, and P.J. Farrell. 1991. The BZLF1 protein of EBV has a coiled coil dimerisation domain without a heptad leucine repeat but with homology to the C/EBP leucine zipper. *Oncogene*. 6:195–204.
- Kudoh, A., M. Fujita, L. Zhang, N. Shirata, T. Daikoku, Y. Sugaya, H. Isomura, Y. Nishiyama, and T. Tsurumi. 2005. Epstein-Barr virus lytic replication elicits ATM checkpoint signal transduction while providing an S-phase-like cellular environment. *J. Biol. Chem.* 280:8156–8163. doi:10.1074/jbc.M411405200.
- Laity, J.H., B.M. Lee, and P.E. Wright. 2001. Zinc finger proteins: New insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* 11:39–46. doi:10.1016/S0959-440X(00)00167-6.
- Lan, Y.Y., T.H. Yeh, W.H. Lin, S.Y. Wu, H.C. Lai, F.H. Chang, K. Takada, and Y. Chang. 2013. Epstein-Barr Virus Zta Upregulates Matrix Metalloproteinases 3 and 9 That Synergistically Promote Cell Invasion In Vitro. *PLoS One*. 8:1–10. doi:10.1371/journal.pone.0056121.
- Landschulz, W.H., P.F. Johnson, and S.L. McKnight. 1988. The leucine zipper: a

- hypothetical structure common to a new class of DNA binding proteins. *Science*. 240:1759–64.
- Langmead, B., and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9:357–359. doi:10.1038/nmeth.1923.
- Langmead, B., C. Trapnell, M. Pop, and S. Salzberg. 2009. 2C- Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10:R25. doi:10.1186/gb-2009-10-3-r25.
- Lei, H., S.P. Oh, M. Okano, R. Jüttermann, K. a Goss, R. Jaenisch, and E. Li. 1996. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development*. 122:3195–3205.
- Leonhardt, H., a W. Page, H.U. Weier, and T.H. Bestor. 1992. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*. 71:865–73.
- Levin, J.Z., M. Yassour, X. Adiconis, C. Nusbaun, D.A. Thompson, N. Friedman, A. Gnirke, and A. Regev. 2010. Comprehensive comparative analysis of strand specific RNA sequencing methods. *Nat. Methods*. 7:709–715. doi:10.1038/nmeth.1491.Comprehensive.
- Li, B., M. Carey, and J.L. Workman. 2007. The Role of Chromatin during Transcription. *Cell*. 128:707–719. doi:10.1016/j.cell.2007.01.015.
- Li, D., L. Qian, C. Chen, M. Shi, M. Yu, M. Hu, L. Song, B. Shen, and N. Guo. 2009a. Down-regulation of MHC class II expression through inhibition of CIITA transcription by lytic transactivator Zta during Epstein-Barr virus reactivation. *J. Immunol*. 182:1799–1809. doi:10.4049/jimmunol.0802686.
- Li, G., X. Ruan, R.K. Auerbach, K.S. Sandhu, M. Zheng, P. Wang, H.M. Poh, Y. Goh, J. Lim, J. Zhang, H.S. Sim, S.Q. Peh, F.H. Mulawadi, C.T. Ong, Y.L. Orlov, S. Hong, Z. Zhang, S. Landt, D. Raha, G. Euskirchen, C.L. Wei, W. Ge, H. Wang, C. Davis, K.I. Fisher-Aylor, A. Mortazavi, M. Gerstein, T. Gingeras, B. Wold, Y. Sun, M.J. Fullwood, E. Cheung, E. Liu, W.K. Sung, M. Snyder, and Y. Ruan. 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*. 148:84–98. doi:10.1016/j.cell.2011.12.014.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009b. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*.

25:2078–2079. doi:10.1093/bioinformatics/btp352.

Lieberman-Aiden, E., N.L. Van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B.R. Lajoie, P.J. Sabo, M.O. Dorschner, R. Sandstrom, B. Bernstein, M.A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, and L.A. Mirny. 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the human genome. *33292:289–294*.

Lieberman, P.M., and A.J. Berk. 1990. In vitro transcriptional activation, dimerization, and DNA-binding specificity of the Epstein-Barr virus Zta protein. *J. Virol.* 64:2560–2568.

Lieberman, P.M., and A.J. Berk. 1991. The Zta trans-activator protein stabilizes TFIID association with promoter DNA by direct protein-protein interaction. *Genes Dev.* 5:2441–2454. doi:10.1101/gad.5.12b.2441.

Lieberman, P.M., J.M. Hardwick, J. Sample, G.S. Hayward, and S.D. Hayward. 1990. The zta transactivator involved in induction of lytic cycle gene expression in Epstein-Barr virus-infected lymphocytes binds to both AP-1 and ZRE sites in target promoter and enhancer regions. *J. Virol.* 64:1143–55.

Lister, R., R.C. O'Malley, J. Tonti-Filippini, B.D. Gregory, C.C. Berry, A.H. Millar, and J.R. Ecker. 2008. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell.* 133:523–536. doi:10.1016/j.cell.2008.03.029.

Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi:10.1186/s13059-014-0550-8.

Lovén, J., D.A. Orlando, A.A. Sigova, C.Y. Lin, P.B. Rahl, C.B. Burge, D.L. Levens, T.I. Lee, and R.A. Young. 2012. Revisiting global gene expression analysis. *Cell.* 151:476–482. doi:10.1016/j.cell.2012.10.012.

Lu, J., S.Y. Chen, H.H. Chua, Y.S. Liu, Y.T. Huang, Y. Chang, J.Y. Chen, T.S. Sheen, and C.H. Tsai. 2000. Upregulation of tyrosine kinase TKT by the Epstein-Barr virus transactivator Zta. *J. Virol.* 74:7391–7399. doi:10.1128/JVI.74.16.7391-7399.2000.

Luger, K., A.W. Mäder, R.K. Richmond, D.F. Sargent, and T.J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *7*.

Magoč, T., and S.L. Salzberg. 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* 27:2957–2963. doi:10.1093/bioinformatics/btr507.

- Mahot, S., A. Sergeant, E. Drouet, and H. Gruffat. 2003. A novel function for the Epstein-Barr virus transcription factor EB1/Zta: Induction of transcription of the hIL-10 gene. *J. Gen. Virol.* 84:965–974. doi:10.1099/vir.0.18845-0.
- Maiti, A., and A.C. Drohat. 2011. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *J. Biol. Chem.* 286:35334–35338. doi:10.1074/jbc.C111.284620.
- Mäntylä, E., K. Salokas, M. Oittinen, V. Aho, P. Mäntysaari, L. Palmujoki, O. Kalliolinna, T.O. Ihalainen, E.A. Niskanen, J. Timonen, K. Viiri, and M. Vihinen-Ranta. 2016. Promoter targeted histone acetylation of chromatinized parvoviral genome is essential for infection progress. *J. Virol.* 90:JVI.03160-15. doi:10.1128/JVI.03160-15.
- Mao, B., Z. Zhang, and G. Wang. 2015. BTG2: A rising star of tumor suppressors (Review). *Int. J. Oncol.* 46:459–464. doi:10.3892/ijo.2014.2765.
- Marazzi, I., J.S.Y. Ho, J. Kim, B. Manicassamy, S. Dewell, R. a. Albrecht, C.W. Seibert, U. Schaefer, K.L. Jeffrey, R.K. Prinjha, K. Lee, A. García-Sastre, R.G. Roeder, and A. Tarakhovsky. 2012. Suppression of the antiviral response by an influenza histone mimic. *Nature.* 483:428–433. doi:10.1038/nature10892.
- Marschall, M., U. Leser, R. Seibl, and H. Wolf. 1989. Identification of proteins encoded by Epstein-Barr virus trans-activator genes. *J. Virol.* 63:938–942.
- Mausser, A., S. Saito, E. Appella, C.W. Anderson, W.T. Seaman, and S. Kenney. 2002. The Epstein-Barr Virus Immediate-Early Protein BZLF1 Regulates p53 Function through Multiple Mechanisms. *J. Virol.* 76:12503–12512. doi:10.1128/JVI.76.24.12503-12512.2002.
- Maze, I., K.-M. Noh, A.A. Soshnev, and C.D. Allis. 2014. Every amino acid matters: essential contributions of histone variants to mammalian development and disease. *Nat. Rev. Genet.* 15:259–71. doi:10.1038/nrg3673.
- McClellan, M.J., C.D. Wood, O. Ojeniyi, T.J. Cooper, A. Kanhere, A. Arvey, H.M. Webb, R.D. Palermo, M.L. Harth-Hertle, B. Kempkes, R.G. Jenner, and M.J. West. 2013. Modulation of Enhancer Looping and Differential Gene Targeting by Epstein-Barr Virus Transcription Factors Directs Cellular Reprogramming. *PLoS Pathog.* 9. doi:10.1371/journal.ppat.1003636.
- McDonald, C.M., C. Petosa, and P.J. Farrell. 2009. Interaction of Epstein-Barr virus BZLF1 C-terminal tail structure and core zipper is required for DNA replication but not for

promoter transactivation. *J. Virol.* 83:3397–3401. doi:10.1128/JVI.02500-08.

Merika, M., A. Williams, G. Chen, T. Collins, and D. Thanos. 1998. Recruitment of CBP/p300 by the IFN β enhanceosome is required for synergistic activation of transcription. *Mol. Cell.* 1:277–287. doi:10.1016/S1097-2765(00)80028-3.

Mikkelsen, T.S., M. Ku, D.B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T.-K. Kim, R.P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E.S. Lander, and B.E. Bernstein. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature.* 448:553–560. doi:10.1038/nature06008.

Mishiro, T., K. Ishihara, S. Hino, S. Tsutsumi, H. Aburatani, K. Shirahige, Y. Kinoshita, and M. Nakao. 2009. Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J.* 28:1234–1245. doi:10.1038/.

Monier, K., J.C. Armas, S. Etteldorf, P. Ghazal, and K.F. Sullivan. 2000. Annexation of the interchromosomal space during viral infection. *Nat. Cell Biol.* 2:661–665. doi:10.1038/35023615.

Mora, A., G.K. Sandve, O.S. Gabrielsen, and R. Eskeland. 2015. In the loop: promoter-enhancer interactions and bioinformatics. *Brief. Bioinform.* 1–16. doi:10.1093/bib/bbv097.

Morris, K. V, and J.S. Mattick. 2014. The rise of regulatory RNA. *Nat. Rev. Genet.* 15:423–37. doi:10.1038/nrg3722.

Morrison, T.B., J.J. Weis, and C.T. Wittwer. 1998. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques.* 24:954–8, 960, 962.

Morrison, T.E., A. Mauser, A. Klingelutz, and S.C. Kenney. 2004. Epstein-Barr virus immediate-early protein BZLF1 inhibits tumor necrosis factor alpha-induced signaling and apoptosis by downregulating tumor necrosis factor receptor 1. *J Virol.* 78:544–549. doi:10.1128/JVI.78.1.544.

Morrison, T.E., A. Mauser, A. Wong, J.P.Y. Ting, and S.C. Kenney. 2001. Inhibition of IFN γ signaling by an Epstein-Barr virus immediate-early protein. *Immunity.* 15:787–799. doi:10.1016/S1074-7613(01)00226-6.

Nakabeppu, Y., and D. Nathans. 1989. The basic region of Fos mediates specific DNA binding. *EMBO J.* 8:3833–41.

- Natarajan, K., M.R. Meyer, M. Belinda, D. Slade, C. Roberts, G. Alan, M.J. Marton, B.M. Jackson, and A.G. Hinnebusch. 2001. Transcriptional Profiling Shows that Gcn4p Is a Master Regulator of Gene Expression during Amino Acid Starvation in Yeast. *Mol. Cell. Biol.* 21:4347–4368. doi:10.1128/MCB.21.13.4347.
- Nitzsche, A., C. Paulus, and M. Nevels. 2008. Temporal dynamics of cytomegalovirus chromatin assembly in productively infected human cells. *J. Virol.* 82:11167–80. doi:10.1128/JVI.01218-08.
- Nora, E.P., B.R. Lajoie, E.G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N.L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, and E. Heard. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature.* 485:381–385. doi:10.1038/nature11049.
- O’Neil, K.T., R.H. Hoess, and W.F. DeGrado. 1990. Design of DNA-binding peptides based on the leucine zipper motif. *Science.* 249:774–8. doi:10.1126/science.2389143.
- O’Neil, K.T., J.D. Shuman, C. Ampe, and W.F. DeGrado. 1991. DNA-induced increase in the alpha-helical content of C/EBP and GCN4. *Biochemistry.* 30:9030–4.
- Ohmori, H., J. Tomizawa, and A. Maxam. 1978. Detection of 5-methylcytosine in DNA sequences. *Nucleic Acids Res.* 5:1779–1785.
- Okano, M., S. Xie, and E. Li. 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* 19:219–20.
- Ørom, U.A., T. Derrien, M. Beringer, K. Gumireddy, A. Gardini, G. Bussotti, F. Lai, M. Zytnicki, C. Notredame, Q. Huang, R. Guigo, and R. Shiekhattar. 2010. Long noncoding RNAs with enhancer-like function in human cells. *Cell.* 143:46–58. doi:10.1016/j.cell.2010.09.001.
- Osborne, C.S., L. Chakalova, K.E. Brown, D. Carter, A. Horton, E. Debrand, B. Goyenechea, J. a Mitchell, S. Lopes, W. Reik, and P. Fraser. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* 36:1065–1071. doi:10.1038/ng1423.
- Parkhomchuk, D., T. Borodina, V. Amstislavskiy, M. Banaru, L. Hallen, S. Krobitsch, H. Lehrach, and A. Soldatov. 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* 37. doi:10.1093/nar/gkp596.

- Pennisi, E. 2012. For Junk DNA. *Science* (80-). 337:1159–1161. doi:doi:10.1126/science.337.6099.1159.
- Petosa, C., P. Morand, F. Baudin, M. Moulin, J.B. Artero, and C.W. Müller. 2006. Structural basis of lytic cycle activation by the Epstein-Barr virus ZEBRA protein. *Mol. Cell.* 21:565–572. doi:10.1016/j.molcel.2006.01.006.
- Pettersson, M., and W. Schaffner. 1990. Synergistic activation of transcription by multiple binding sites for NF-kappa B even in absence of co-operative factor binding to DNA. *J. Mol. Biol.* 214:373–380. doi:10.1016/0022-2836(90)90187-Q.
- Phillips-Cremins, J.E., M.E.G. Sauria, A. Sanyal, T.I. Gerasimova, B.R. Lajoie, J.S.K. Bell, C.T. Ong, T.A. Hookway, C. Guo, Y. Sun, M.J. Bland, W. Wagstaff, S. Dalton, T.C. McDevitt, R. Sen, J. Dekker, J. Taylor, and V.G. Corces. 2013. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell.* 153:1281–1295. doi:10.1016/j.cell.2013.04.053.
- Pinkel, D., T. Straume, and J.W. Gray. 1986. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. U. S. A.* 83:2934–8. doi:10.1073/pnas.83.9.2934.
- Pokholok, D.K., C.T. Harbison, S. Levine, M. Cole, N.M. Hannett, I.L. Tong, G.W. Bell, K. Walker, P.A. Rolfe, E. Herbolsheimer, J. Zeitlinger, F. Liewitter, D.K. Gifford, and R.A. Young. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell.* 122:517–527. doi:10.1016/j.cell.2005.06.026.
- Pulvertaft, V. 1964. Cytology of Burkitt's tumour (African lymphoma). *Lancet.* 283:238–240. doi:10.1016/S0140-6736(64)92345-1.
- Quenneville, S., G. Verde, A. Corsinotti, A. Kapopoulou, J. Jakobsson, S. Offner, I. Baglivo, P. V. Pedone, G. Grimaldi, A. Riccio, and D. Trono. 2011. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell.* 44:361–372. doi:10.1016/j.molcel.2011.08.032.
- Quinlan, A.R., and I.M. Hall. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics.* 26:841–842. doi:10.1093/bioinformatics/btq033.
- Rada-Iglesias, A., R. Bajpai, S. Prescott, S.A. Brugmann, T. Swigut, and J. Wysocka. 2012. Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell.* 11:633–648. doi:10.1016/j.stem.2012.07.006.

- Rada-Iglesias, A., R. Bajpai, T. Swigut, S. a Brugmann, R. a Flynn, and J. Wysocka. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. 470:279–83. doi:10.1038/nature09692.
- Rajasekaran, S., N.M. Reddy, W. Zhang, and S.P. Reddy. 2013. Expression profiling of genes regulated by Fra-1/AP-1 transcription factor during bleomycin-induced pulmonary fibrosis. *BMC Genomics*. 14:381. doi:10.1186/1471-2164-14-381.
- Ramasubramanian, S., K. Osborn, R. Al-Mohammad, I.B. Naranjo Perez-Fernandez, J. Zuo, N. Balan, A. Godfrey, H. Patel, G. Peters, M. Rowe, R.G. Jenner, and A.J. Sinclair. 2015. Epstein-Barr virus transcription factor Zta acts through distal regulatory elements to directly control cellular gene expression. *Nucleic Acids Res*. 43:3563–3577. doi:10.1093/nar/gkv212.
- Reinke, A.W., G. Grigoryan, and A.E. Keating. 2010. Identification of bZIP interaction partners of viral proteins HBZ, MEQ, BZLF1, and K-bZIP using coiled-coil arrays. *Biochemistry*. 49:1985–1997. doi:10.1021/bi902065k.
- Ren, L., Y. Wang, M. Shi, X. Wang, Z. Yang, and Z. Zhao. 2012. CTCF mediates the cell-type specific spatial organization of the *kcnq5* locus and the local gene regulation. *PLoS One*. 7:1–9. doi:10.1371/journal.pone.0031416.
- Rennekamp, A.J., and P.M. Lieberman. 2011. Initiation of Epstein-Barr virus lytic replication requires transcription and the formation of a stable RNA-DNA hybrid molecule at OriLyt. *J. Virol*. 85:2837–50. doi:10.1128/JVI.02175-10.
- Rhee, H.S., and B.F. Pugh. 2011. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*. 147:1408–1419. doi:10.1016/j.cell.2011.11.013.
- Rinn, J.L., M. Kertesz, J.K. Wang, S.L. Squazzo, X. Xu, S.A. Brugmann, L.H. Goodnough, J.A. Helms, P.J. Farnham, E. Segal, and H.Y. Chang. 2007. Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell*. 129:1311–1323. doi:10.1016/j.cell.2007.05.022.
- Rishi, V., P. Bhattacharya, R. Chatterjee, J. Rozenberg, J. Zhao, K. Glass, P. Fitzgerald, and C. Vinson. 2010. CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. *Proc. Natl. Acad. Sci. U. S. A*. 107:20311–6. doi:10.1073/pnas.1008688107.
- Risso, D., J. Ngai, T.P. Speed, and S. Dudoit. 2014. Normalization of RNA-seq data using

- factor analysis of control genes or samples. *Nat Biotech.* 32:896–902. doi:10.1038/nbt.2931.
- Rivera, C.M., and B. Ren. 2013. Mapping human epigenomes. *Cell.* 155:39–55. doi:10.1016/j.cell.2013.09.011.
- Robertson, G., M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. Euskirchen, B. Bernier, R. Varhol, A. Delaney, N. Thiessen, O.L. Griffith, A. He, M. Marra, M. Snyder, and S. Jones. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods.* 4:651–657. doi:10.1038/nmeth1068.
- Robinson, J.T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, and J.P. Mesirov. 2011. Integrative genomics viewer. *Nat. Biotechnol.* 29:24–26. doi:10.1038/nbt.1754.
- Rogakou, E.P., D.R. Pilch, A.H. Orr, V.S. Ivanova, and W.M. Bonner. 1998. Double-stranded Brekas Induce Histone H2AX phosphorylation on Serine 139. *J. Biol. Chem.* 273:5858–5868. doi:10.1074/jbc.273.10.5858.
- Roh, T.-Y., S. Cuddapah, K. Cui, and K. Zhao. 2006. The genomic landscape of histone modifications in human T cells. *Pnas.* 103:15782–7. doi:10.1073/pnas.0607617103.
- Rougier, N., D. Bourc, D.M. Gomes, A. Niveleau, and M. Plachot. 1998. Chromosome methylation patterns during mammalian preimplantation? development service Chromosome methylation patterns during mammalian preimplantation development. 2108–2113. doi:10.1101/gad.12.14.2108.
- Rowe, M., B. Glaunsinger, D. van Leeuwen, J. Zuo, D. Sweetman, D. Ganem, J. Middeldorp, E.J.H.J. Wiertz, and M.E. Rensing. 2007. Host shutoff during productive Epstein-Barr virus infection is mediated by BGLF5 and may contribute to immune evasion. *Pnas.* 104:3366–71. doi:10.1073/pnas.0611128104.
- Sanger, F., S. Nicklen, and a R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74:5463–7. doi:10.1073/pnas.74.12.5463.
- Sanyal, A., B.R. Lajoie, G. Jain, and J. Dekker. 2012. The long-range interaction landscape of gene promoters. *Nature.* 489:109–113. doi:10.1038/nature11279.
- Sato, H., H. Takeshita, M. Furukawa, and M. Seiki. 1992. Epstein-Barr virus BZLF1 transactivator is a negative regulator of Jun. *J Virol.* 66:4732–4736.
- Saxonov, S., P. Berg, and D.L. Brutlag. 2006. A genome-wide analysis of CpG dinucleotides

- in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* 103:1412–1417. doi:10.1073/pnas.0510310103.
- Schaeffner, M. 2015. Chromatin remodeling in Epstein-Barr virus after induction of the lytic phase: molecular characterization of the role of BZLF1 and its interactions. Ludwig-Maximilians-Universität. 142 pp.
- Schmeink, A. 2011. Acquisition and loss of chromatin modifications during an Epstein-Barr virus infection. Ludwig-Maximilians-Universität.
- Shandilya, J., and S.G.E. Roberts. 2012. The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1819:391–400. doi:10.1016/j.bbagr.2012.01.010.
- Shaulian, E., and M. Karin. 2001. AP-1 in cell proliferation and survival. *Oncogene.* 20:2390–2400. doi:10.1038/sj.onc.1204383.
- Shaulian, E., and M. Karin. 2002. AP-1 as a regulator of cell life and death. *Nat Cell Biol.* 4:E131-6. doi:10.1038/ncb0502-e131.
- Shaw, J.E., L.F. Levinger, and C.W. Carter. 1979. Nucleosomal structure of Epstein-Barr virus DNA in transformed cell lines. *J. Virol.* 29:657–665.
- Shemer, R., a Y. Hershko, J. Perk, R. Mostoslavsky, B. Tsuberi, H. Cedar, K. Buiting, and a Razin. 2000. The imprinting box of the Prader-Willi/Angelman syndrome domain. *Nat. Genet.* 26:440–3. doi:10.1038/82571.
- Shen, Y., F. Yue, D.F. McCleary, Z. Ye, L. Edsall, S. Kuan, U. Wagner, J. Dixon, L. Lee, V. V. Lobanenko, and B. Ren. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature.* 488:116–120. doi:10.1038/nature11243.
- Shuman, J.D., C.R. Vinson, and S.L. McKnight. 1990. Evidence of changes in protease sensitivity and subunit exchange rate on DNA binding by C/EBP. *Science (80-).* 249:771–774.
- Simonis, M., P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, B. van Steensel, and W. de Laat. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* 38:1348–1354. doi:10.1038/ng1896.
- Slattery, M., T. Zhou, L. Yang, A.C. Dantas Machado, R. Gordân, and R. Rohs. 2014. Absence of a simple code: How transcription factors read the genome. *Trends Biochem. Sci.* 39:381–399. doi:10.1016/j.tibs.2014.07.002.

- Song, L., Z. Zhang, L.L. Gräf, A.P. Boyle, P.G. Giresi, B.K. Lee, N.C. Sheffield, S. Gräf, M. Huss, D. Keefe, Z. Liu, D. London, R.M. McDaniell, Y. Shibata, K.A. Showers, J.M. Simon, T. Vales, T. Wang, D. Winter, Z. Zhang, N.D. Clarke, E. Birney, V.R. Iyer, G.E. Crawford, J.D. Lieb, and T.S. Furey. 2011. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* 21:1757–1767. doi:10.1101/gr.121541.111.
- Statham, A.L., M.D. Robinson, J.Z. Song, M.W. Coolen, C. Stirzaker, and S.J. Clark. 2012. Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res.* 22:1120–1127. doi:10.1101/gr.132076.111.
- Sutherland, J.A., A. Cook, A.J. Bannister, and T. Kouzarides. 1992. Conserved motifs in Fos and Jun define a new class of activation domain. *Genes Dev.* 6:1810–1819. doi:10.1101/gad.6.9.1810.
- Tagawa, T., M. Albanese, M. Bouvet, A. Moosman, J. Mautner, V. Heissmeyer, C. Zielinski, J. Hoser, M. Hastreiter, M. Hayes, B. Sugden, and W. Hammerschmidt. 2016. Epstein-Barr Viral miRNAs inhibit antiviral CD4 + T cell responses targeting IL-12 and antigen presentation. 1–16. doi:10.1084/jem.20160248.
- Tahiliani, M., K. Koh, and Y. Shen. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (80-.).* 324:930–935. doi:10.1126/science.1170116.Conversion.
- Takada, K. 1984. Cross-linking of cell surface immunoglobulins induces Epstein-Barr virus in Burkitt lymphoma lines. *Int. J. Cancer.* 33:27–32. doi:10.1002/ijc.2910330106.
- Takada, K., and Y. Ono. 1989. Synchronous and sequential activation of latently infected Epstein-Barr virus genomes. *J. Virol.* 63:445–449.
- Takada, K., N. Shimizu, S. Sakuma, and Y. Ono. 1986. trans activation of the latent Epstein-Barr virus (EBV) genome after transfection of the EBV DNA fragment. *J. Virol.* 57:1016–22.
- Takagi, S., K. Takada, and T. Sairenji. 1991. Formation of intranuclear replication compartments of Epstein-Barr virus with redistribution of BZLF1 and BMRF1 gene products. *Virology.* 185:309–315. doi:10.1016/0042-6822(91)90778-A.
- Talanian, R. V, C.J. McKnight, and P.S. Kim. 1990. Sequence-specific DNA binding by a short peptide dimer. *Science (80-.).* 249:769–771. doi:10.1126/science.2389142.

- Tan, M., H. Luo, S. Lee, F. Jin, J.S. Yang, E. Montellier, T. Buchou, Z. Cheng, S. Rousseaux, N. Rajagopal, Z. Lu, Z. Ye, Q. Zhu, J. Wysocka, Y. Ye, S. Khochbin, B. Ren, and Y. Zhao. 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*. 146:1016–1028. doi:10.1016/j.cell.2011.08.008.
- Tang, S.J. 2011. Chromatin organization by repetitive elements (CORE): A genomic principle for the higher-order structure of chromosomes. *Genes (Basel)*. 2:502–515. doi:10.3390/genes2030502.
- Taylor, N., E. Flemington, J.L. Kolman, R.P. Baumann, S.H. Speck, and G. Miller. 1991. ZEBRA and a Fos-GCN4 chimeric protein differ in their DNA-binding specificities for sites in the Epstein-Barr virus BZLF1 promoter . ZEBRA and a Fos-GCN4 Chimeric Protein Differ in Their DNA-Binding Specificities for Sites in the Epstein-Barr Virus BZLF1. 65:4033–4041.
- Thorvaldsdóttir, H., J.T. Robinson, and J.P. Mesirov. 2013. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* 14:178–192. doi:10.1093/bib/bbs017.
- Thurman, R.E., E. Rynes, R. Humbert, J. Vierstra, M.T. Maurano, E. Haugen, N.C. Sheffield, A.B. Stergachis, H. Wang, B. Vernet, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T.K. Canfield, M. Diegel, D. Dunn, A.K. Ebersol, T. Frum, E. Giste, A.K. Johnson, E.M. Johnson, T. Kutayavin, B. Lajoie, B.-K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E.D. Nguyen, H. Qu, A.P. Reynolds, V. Roach, A. Safi, M.E. Sanchez, A. Sanyal, A. Shafer, J.M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M.O. Dorschner, R.S. Hansen, P.A. Navas, G. Stamatoyannopoulos, V.R. Iyer, J.D. Lieb, S.R. Sunyaev, J.M. Akey, P.J. Sabo, R. Kaul, T.S. Furey, J. Dekker, G.E. Crawford, and J.A. Stamatoyannopoulos. 2012. The accessible chromatin landscape of the human genome. *Nature*. 489:75–82. doi:10.1038/nature11232.
- Tian, Z., N. Tolić, R. Zhao, R.J. Moore, S.M. Hengel, E.W. Robinson, D.L. Stenoien, S. Wu, R.D. Smith, and L. Paša-Tolić. 2012. Enhanced top-down characterization of histone post-translational modifications. *Genome Biol.* 13:1–9. doi:10.1186/gb-2012-13-10-r86.
- Tremblay, K.D., J.R. Saam, R.S. Ingram, S.M. Tilghman, and M.S. Bartolomei. 1995. A paternal-specific methylation imprint marks the alleles of the mouse H19 gene. *Nat.*

Genet. 9:407–413. doi:10.1038/ng0595-111.

- Tsai, S.C., S.J. Lin, P.W. Chen, W.Y. Luo, T.H. Yeh, H.W. Wang, C.J. Chen, and C.H. Tsai. 2009. EBV Zta protein induces the expression of interleukin-13, promoting the proliferation of EBV-infected B cells and lymphoblastoid cell lines. *Blood.* 114:109–118. doi:10.1182/blood-2008-12-193375.
- Ushijima, T., K. Morimura, Y. Hosoya, H. Okonogi, M. Tatematsu, T. Sugimura, and M. Nagao. 1997. Establishment of methylation-sensitive-representational difference analysis and isolation of hypo- and hypermethylated genomic fragments in mouse liver tumors. *Proc. Natl. Acad. Sci. U. S. A.* 94:2284–9. doi:10.1073/pnas.94.6.2284.
- Valinluck, V., and L.C. Sowers. 2007. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.* 67:946–50. doi:10.1158/0008-5472.CAN-06-3123.
- Vaquerezas, J.M., S.K. Kummerfeld, S. a Teichmann, and N.M. Luscombe. 2009. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet.* 10:252–263. doi:doi: 10.1038/nrg2538.
- Venkatesh, S., and J.L. Workman. 2015. Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* 16:178–189. doi:10.1038/nrm3941.
- Vinson, C.R., P.B. Sigler, and S.L. McKnight. 1989. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science.* 246:911–6. doi:10.1126/science.2683088.
- Viré, E., C. Brenner, R. Deplus, L. Blanchon, M. Fraga, C. Didelot, L. Morey, A. Van Eynde, D. Bernard, J.-M. Vanderwinden, M. Bollen, M. Esteller, L. Di Croce, Y. de Launoit, and F. Fuks. 2006. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature.* 439:871–874. doi:10.1038/nature04431.
- Visel, A., M.J. Blow, Z. Li, T. Zhang, J. a Akiyama, A. Holt, I. Plajzer-Frick, M. Shoukry, C. Wright, F. Chen, V. Afzal, B. Ren, E.M. Rubin, and L. a Pennacchio. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature.* 457:854–8. doi:10.1038/nature07730.
- Volpe, T. 2002. Regulation oh heterochromatic silencing and histone H3 Lysine-9 by RNAi. *Science (80-.).* 297:1833–1837. doi:10.1126/science.1074973.
- Watson, J., and F. Crick. 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature.* 171:737–738. doi:10.1038/171737a0.

- Weber, M., J.J. Davies, D. Wittig, E.J. Oakeley, M. Haase, W.L. Lam, and D. Schübeler. 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* 37:853–62. doi:10.1038/ng1598.
- Wen, W., D. Iwakiri, K. Yamamoto, S. Maruo, T. Kanda, and K. Takada. 2007. Epstein-Barr virus BZLF1 gene, a switch from latency to lytic infection, is expressed as an immediate-early gene after primary infection of B lymphocytes. *J. Virol.* 81:1037–42. doi:10.1128/JVI.01416-06.
- Whyte, W.A., D.A. Orlando, D. Hnisz, B.J. Abraham, C.Y. Lin, M.H. Kagey, P.B. Rahl, T.I. Lee, and R.A. Young. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell.* 153:307–319. doi:10.1016/j.cell.2013.03.035.
- Wickham, H. 2007. Reshaping Data with the `\pkg{reshape}` Package. *J. Stat. Softw.* 21:1–20. doi:10.1016/S0142-1123(99)00007-9.
- Wickham, H. 2011. The Split-Apply-Combine Strategy for Data. *J. Stat. Softw.* 40:1–29. doi:10.1.1.182.5667.
- Wilkins, H., M. Stokes, and A. Wilson. 1953. Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. 738–740. doi:10.1038/171738a0.
- Wisdom, R., and I.M. Verma. 1993. Transformation by Fos proteins requires a C-terminal transactivation domain. *Mol. Cell. Biol.* 13:7429–38. doi:10.1128/MCB.13.12.7429.Updated.
- Woellmer, A., J.M. Arteaga-Salas, and W. Hammerschmidt. 2012. BZLF1 Governs CpG-Methylated Chromatin of Epstein-Barr Virus Reversing Epigenetic Repression. *PLoS Pathog.* 8. doi:10.1371/journal.ppat.1002902.
- Wu, C.-C., M.-T. Liu, Y.-T. Chang, C.-Y. Fang, S.-P. Chou, H.-W. Liao, K.-L. Kuo, S.-L. Hsu, Y.-R. Chen, P.-W. Wang, Y.-L. Chen, H.-Y. Chuang, C.-H. Lee, M. Chen, W.-S. Wayne Chang, and J.-Y. Chen. 2010. Epstein-Barr virus DNase (BGLF5) induces genomic instability in human epithelial cells. *Nucleic Acids Res.* 38:1932–1949. doi:10.1093/nar/gkp1169.
- Wyatt, G. 1952. The nucleic acids of some insect viruses. *J. Gen. Physiol.* 201–205.
- Yang, J., W. Deng, P.M. Hau, J. Liu, V.M.Y. Lau, A.L.M. Cheung, M.S.Y. Huen, and S.W. Tsao. 2015. Epstein-Barr virus BZLF1 protein impairs accumulation of host DNA damage proteins at damage sites in response to DNA damage. *Lab. Invest.* 95:937–

950. doi:10.1038/labinvest.2015.69.

- Yates, J.L., N. Warren, and B. Sugden. 1985. Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells. *Nature*. 313:812–5. doi:10.1038/314731a0.
- Young, M.D., T.A. Willson, M.J. Wakefield, E. Trounson, D.J. Hilton, M.E. Blewitt, A. Oshlack, and I.J. Majewski. 2011. ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Res*. 39:7415–7427. doi:10.1093/nar/gkr416.
- Zalckvar, E., C. Paulus, D. Tillo, A. Asbach-Nitzsche, Y. Lubling, C. Winterling, N. Strieder, K. Mucke, F. Goodrum, E. Segal, and M. Nevels. 2013. Nucleosome maps of the human cytomegalovirus genome reveal a temporal switch in chromatin organization linked to a major IE protein. *Proc. Natl. Acad. Sci. U. S. A.* 110:13126–13131. doi:10.1073/pnas.1305548110.
- Zamenhof, S., G. Brawerman, and E. Chargaff. 1952. On the desoxyribose nucleic acids from several microorganisms. *Biochim. Biophys. Acta*. 9:402–405. doi:10.1016/0006-3002(52)90184-4.
- Zhang, C.X., G. Decaussin, J. Daille, and T. Ooka. 1988. Altered expression of two Epstein-Barr virus early genes localized in BamHI-A in nonproducer Raji cells. *J. Virol.* 62:1862–9.
- Zhang, Q., Y. Hong, D. Dorsky, E. Holley-Guthrie, S. Zalani, N.A. Elshiekh, A. Kiehl, T. Le, and S. Kenney. 1996. Functional and physical interactions between the Epstein-Barr virus (EBV) proteins BZLF1 and BMRF1: Effects on EBV transcription and lytic replication. *J. Virol.* 70:5131–5142.
- Zhang, Y., T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nussbaum, R.M. Myers, M. Brown, W. Li, and X.S. Liu. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* 9:R137. doi:10.1186/gb-2008-9-9-r137.
- Zhao, S., Y. Zhang, W. Gordon, J. Quan, H. Xi, S. Du, D. von Schack, B. Zhang, A. Mortazavi, B. Williams, K. McCue, L. Schaeffer, B. Wold, Z. Wang, M. Gerstein, M. Snyder, K. Mutz, A. Heilkenbrinker, M. Lönne, J. Walter, F. Stahl, J. Malone, B. Oliver, S. Zhao, W.-P. Fung-Leung, A. Bittner, K. Ngo, X. Liu, C. Sanna, W.-H. Li, L. Zhang, D. Parkhomchuk, T. Borodina, V. Amstislavskiy, M. Banaru, L. Hallen, S. Krobitsch, H. Lehrach, A. Soldatov, S. Zhong, J.-G. Joung, Y. Zheng, B. Liu, Y. Shao, J. Xiang, Z. Fei,

- J. Giovannoni, B. Weissenmayer, J. Prendergast, A. Lohan, B. Loftus, J. Levin, M. Yassour, X. Adiconis, C. Nusbaum, D. Thompson, N. Friedman, A. Gnirke, A. Regev, B. Sigurgeirsson, O. Emanuelsson, J. Lundeberg, J. Mills, Y. Kawahara, M. Janitz, W. Zhao, X. He, K. Hoadley, J. Parker, D. Hayes, C. Perou, J. Harrow, A. Frankish, J. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. Aken, D. Barrell, A. Zadissa, S. Searle, A. Dobin, C. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. Gingeras, Y. Liao, G. Smyth, W. Shi, M. Robinson, D. McCarthy, G. Smyth, C. Law, Y. Chen, et al. 2015. Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. *BMC Genomics*. 16:675. doi:10.1186/s12864-015-1876-7.
- Zhao, W., X. He, K.A. Hoadley, J.S. Parker, D.N. Hayes, and C.M. Perou. 2014. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics*. 15:419. doi:10.1186/1471-2164-15-419.
- Zhou, H., T. Zarubin, Z. Ji, Z. Min, W. Zhu, J.S. Downey, S. Lin, and J. Han. 2005. Frequency and distribution of AP-1 sites in the human genome. *DNA Res*. 12:139–150. doi:10.1093/dnares/12.2.139.
- Zhu, H., G. Wang, and J. Qian. 2016. Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet*. 17:551–65. doi:10.1038/nrg.2016.83.
- Zhu, J., M. Adli, J.Y. Zou, G. Verstappen, M. Coyne, X. Zhang, T. Durham, M. Miri, V. Deshpande, P.L. De Jager, D.A. Bennett, J.A. Houmard, D.M. Muoio, T.T. Onder, R. Camahort, C.A. Cowan, A. Meissner, C.B. Epstein, N. Shores, and B.E. Bernstein. 2013. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell*. 152:642–654. doi:10.1016/j.cell.2012.12.033.
- Zuo, J., W.A. Thomas, T.A. Haigh, L. Fitzsimmons, H.M. Long, A.D. Hislop, G.S. Taylor, and M. Rowe. 2011. Epstein-barr virus evades CD4 + T cell responses in lytic cycle through BZLF1-mediated downregulation of CD74 and the cooperation of vBcl-2. *PLoS Pathog*. 7. doi:10.1371/journal.ppat.1002455.

8 Abbreviations

| | |
|-----------------|---|
| -TET | Lacking tetracycline |
| % | Percent |
| °C | Degree Celsius |
| 3C | Chromosome conformation capture |
| 4816 | pRTR 4816 plasmid |
| 4C | Chromosome conformation capture-on-chip |
| 5C | Chromosome confirmation capture carbon copy |
| 5caC | 5-carboxylcytosine |
| 5fC | 5-formylcytosine |
| α | Anti (in context of antibody specificity) |
| A | Adenosine (in context of nucleotides) |
| aa | Amino acids |
| ac | Acetylation |
| AD-tBZLF1 | BZLF1 mutant lacking the transactivation domain |
| AP-1 | Activator protein 1 |
| APS | Ammonium persulfate |
| ATF | Activating transcription factor |
| BAM | Binary Alignment/Map |
| BED | Browser extensible data |
| bp | Base pair |
| BSA | Bovine Serum Albumin |
| bZIP | Basic leucine zipper |
| C | Cytosine (in context of nucleotides) |
| C-terminus | Carboxy-terminus |
| C/EBPα | CCAAT-enhancer-binding protein alpha |
| cAMP | Cyclic adenosine monophosphate |
| CBP | CREB binding protein |
| CBP | CREB-binding protein |
| CD | Compact disk |
| cDNA | Complementary DNA |
| ChIA-PET | Paired-end tag sequencing |
| ChIP | Chromatin immunoprecipitation |
| ChIP | Chromatin immunoprecipitations |
| ChIP-loop | Combination of the 3C approach with ChIP |
| ChIP-seq | Chromatin immunoprecipitation sequencing |
| cm | Centimeter |
| CO ₂ | Carbon dioxide |
| COL2A1 | Collage type II alpha 1 |
| CpG | Cytosine-phosphatidyl-guanine |
| CpG ZRE | ZRE with a CpG di-nucleotide |
| CRE | cAMP response elements |
| CREB | cAMP response element binding protein |
| CTCF | CCCTC-binding factor |
| CXCR4 | C-X-C motif chemokine receptor 4 |

| | |
|------------------|--|
| cytC | Cytochrome C |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxyribonucleic acid |
| DNase | Deoxyribonuclease |
| DNMT | DNA methyltransferase |
| dNTP | Deoxynucleotide triphosphate |
| dsDNase | Double strand DNase |
| DTT | Dithiothreitol |
| E2F2 | E2F transcription factor 2 |
| EBER | EBV-encoded RNA |
| EBNA | EBV-encoded nuclear antigen |
| EBNA1 | Nuclear antigen 1 |
| EBV | Epstein-Barr virus |
| EDTA | Ethylenediaminetetraacetic |
| EGR1 | Growth response 1 gene |
| ERCC | External RNA Controls Consortium |
| et al. | And others |
| EtOH | Ethanol |
| FBS | Fetal bovine serum |
| FISH | Fluorescent in situ hybridization |
| g | Gram (in context of chemicals) |
| g | Earth gravitational force (in context of centrifugation) |
| G | Guanine (in context of nucleotides) |
| GAPDH | Glyceraldehyde 3-phosphate dehydrogenase |
| GFP | Green fluorescent protein |
| h | Hour |
| H ₂ O | Dihydrogen monoxide |
| H3K27ac | Acetylated histone H3 at lysine 27 |
| H3K27me3 | Trimethylated histone H3 at lysine 27 |
| H3K36me3 | Trimethylated histone H3 at lysine 36 |
| H3K4me3 | Trimethylated histone H3 at lysine 4 |
| HCE | Heavy chain enhancer |
| HCL | Hydrogen chloride |
| HDAC9 | Histone deacetylase 9 |
| Hg19 | Human genome 19 |
| HIV | Human immunodeficiency virus |
| HMGU | Helmholtz center Munich |
| HPLC | High performance liquid chromatography |
| ID3 | Inhibitor of DNA binding 3 |
| IgG | Immunoglobulin G |
| IGV | Integrated genome viewer |
| IL8 | Interleukin 8 gene |
| INO80 | Inositol requiring 80 |
| IQR | Interquartile range |
| IRES | Internal ribosomal entry site |
| ISWI | Imitation switch |
| kbp | Kilo base pairs |
| K _d | Equilibrium dissociation constant |
| kDa | Kilo dalton |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| l | Liter |
| LiCl | Lithium chloride |

| | |
|------------|---|
| μF | Microfarad |
| μg | Microgram |
| μl | Microliter |
| μM | Micromolar |
| M | Molar |
| M | Adenine or cytosine (in context of nucleotides) |
| mA | Milliampere |
| MA | Mean average |
| MACS | Magnetic activated cell sorting |
| MACS2 | Model-based Analysis of ChIP-Seq |
| MBD | Methyl-CpG binding domain |
| MEME | Multiple expectation maximization for Motif Elicitation |
| min | Minute |
| MIR155HG | Non-coding RNA MIR155 host gene |
| ml | Milliliter |
| mM | Millimolar |
| mRNA | Messenger RNA |
| N-terminus | Amino-terminus |
| NaCl | Sodium chloride |
| ncRNA | Non-coding |
| NEB | New England Biolabs |
| ng | Nanogram |
| NGF | Truncated nerve growth factor |
| NGF-R | Nerve growth factor receptor |
| NGS | Next generation sequencing |
| nmol | Nanomole |
| no ZRE | Binding site without ZRE motif |
| nonCpG ZRE | ZRE without a CpG di-nucleotide |
| oriLyt | Origin of lytic replication |
| oriP | Origin of plasmid replication |
| PAA | Polyacrylamide |
| Padj | adjusted p-value |
| PBS | Phosphate buffered saline |
| PBS-T | Phosphate buffered saline with tween |
| PCA | Principal component analysis |
| PCR | Polymerase chain reaction |
| peaks | Enriched sequencing reads |
| pH | Potential of hydrogen |
| PIC | Proteinase inhibitor cocktail |
| pm | Picomolar |
| Pol II | RNA polymerase II |
| PTM | Posttranscriptional modifications |
| qPCR | Quantitative PCR |
| RLE | Relative log expression |
| RNA | Ribonucleic acid |
| RNA-seq | Ribonucleic acid sequencing |
| ROI | Regions of interest |
| RPMI | Roswell Park Memorial Institute |
| rRNA | Ribosomal RNA |
| RT | Room temperature |
| RT-PCR | Reverse transcriptase PCR |
| SAM | Sequence Alignment/Map (in terms of sequence alignment) |

| | |
|-----------|--|
| SAM | S-adenosyl methionine (in terms of cytosine methylation) |
| SDS | Sodium dodecyl sulfate |
| SDS | Sodium dodecyl sulfate |
| sec | Second |
| T | Thymine (in context of nucleotides) |
| TBE | Buffer containing Tris base, boric acid and EDTA |
| TBP | TATA box binding protein |
| TBP | TATA-box-binding protein |
| TE | Tris-EDTA |
| TEMED | Tetramethylethylenediamine |
| TET | Ten-eleven-translocation |
| TLR10 | Toll-like receptor 10 |
| TLR6 | Toll-like receptor 6 |
| TPA | 12-O-tetradecanoylphorbol-12-acetate |
| TRE | TPA response elements |
| Tris | 2-amino-2-hydroxymethyl-propane-1,3-diol |
| TSS | Transcriptional start sites |
| U | Units |
| UK | United Kingdom |
| USA | United state of America |
| V | Adenosine, guanine or cytosine (in context of nucleotides) |
| V | Volt |
| W | Adenosine or thymine (in context of nucleotides) |
| w/v | Weight per volume |
| Y | Cytosine or thymine (in context of nucleotides) |
| ZNF165 | Zinc finger 165 |
| ZRE | BZLF1 responsive element |
| ZSCAN12P1 | Zinc finger and SCAN domain containing 12 pseudogene 1 |

9 Appendix

9.1 Raji EBV's gene expression

Table 9-1: Raji EBV's gene expression and gene coordinates

Selected Raji EBV genes are indicated with base mean (baseMean) and log2 fold change (log2FoldChange). The genes are sorted by log2 fold change. Additionally the start ("start") end ("end") positions of the coding regions of the genes are given together with the strand ("strand") and the used genome ("Genome"). Values are visualized in Figure 3-3.

| Gene | baseMean | log2FoldChange | Start | End | Strand | Genome |
|-------------|----------|----------------|--------|--------|--------|-----------------------------|
| BBLF3-BBLF2 | 24650.1 | 8 | 101511 | 103805 | - | 0_9_Consensus_Sequence_Raji |
| BZLF1 | 27030.86 | 7.71 | 86936 | 87067 | - | 0_9_Consensus_Sequence_Raji |
| | | | 87152 | 87256 | - | |
| | | | 87381 | 87881 | - | |
| BBLF4 | 18785.72 | 7.47 | 96556 | 98985 | - | 0_9_Consensus_Sequence_Raji |
| BGLF5 | 62307.49 | 7.35 | 105653 | 107065 | - | 0_9_Consensus_Sequence_Raji |
| BALF5 | 26742.18 | 6.82 | 150252 | 153299 | - | 0_9_Consensus_Sequence_Raji |
| BRLF1 | 6698.53 | 6.45 | 88092 | 89909 | - | 0_9_Consensus_Sequence_Raji |
| BMRF2 | 60118.2 | 6.14 | 68836 | 69909 | + | 0_9_Consensus_Sequence_Raji |
| BMLF1-BSLF1 | 63544.48 | 6.13 | 70466 | 74600 | - | 0_9_Consensus_Sequence_Raji |
| BMRF1 | 48222.18 | 6 | 67617 | 68831 | + | 0_9_Consensus_Sequence_Raji |
| BHLF1 | 71961.99 | 5.14 | 38298 | 40280 | - | 0_9_Consensus_Sequence_Raji |
| BNLF2b | 5647.99 | 5.11 | 160893 | 161189 | - | 0_9_Consensus_Sequence_Raji |
| BNLF2a | 11111.24 | 5.01 | 161203 | 161420 | - | 0_9_Consensus_Sequence_Raji |
| BHRF1 | 39009.62 | 4.85 | 42099 | 42674 | + | 0_9_Consensus_Sequence_Raji |
| BALF3 | 224.8 | 3.72 | 155852 | 157919 | - | 0_9_Consensus_Sequence_Raji |
| BcLF1 | 196.6 | 2.3 | 118042 | 122187 | - | 0_9_Consensus_Sequence_Raji |
| EBER-2 | 244.05 | 0.6 | 6956 | 7128 | + | 0_9_Consensus_Sequence_Raji |
| | | | 162055 | 162860 | - | |
| | | | 162937 | 163023 | - | |
| LMP1 | 252.3 | 0.21 | 163102 | 163369 | - | 0_9_Consensus_Sequence_Raji |
| | | | 6629 | 6795 | + | |
| EBER-1 | 46856.27 | -0.36 | 92676 | 94598 | + | 0_9_Consensus_Sequence_Raji |
| EBNA1 | 1585.98 | -1.23 | 36228 | 37691 | + | 0_9_Consensus_Sequence_Raji |
| EBNA-2 | 429.32 | -1.86 | | | | |

9.2 ERCC RNAs

Table 9-2: ERCC RNAs

The table shows the ERCC spike-in control RNAs (“ERCC.ID”) and their concentration (“Mix.1:attomoles/ μ l”)

| ERCC.ID | Mix.1:attomoles/ul | ERCC.ID | Mix.1:attomoles/ul |
|------------|--------------------|------------|--------------------|
| ERCC-00002 | 15000 | ERCC-00083 | 0.02861023 |
| ERCC-00003 | 937.5 | ERCC-00084 | 29.296875 |
| ERCC-00004 | 7500 | ERCC-00085 | 7.32421875 |
| ERCC-00009 | 937.5 | ERCC-00086 | 0.11444092 |
| ERCC-00012 | 0.11444092 | ERCC-00092 | 234375 |
| ERCC-00013 | 0.91552734 | ERCC-00095 | 117.1875 |
| ERCC-00014 | 3.66210938 | ERCC-00096 | 15000 |
| ERCC-00016 | 0.22888184 | ERCC-00097 | 0.45776367 |
| ERCC-00017 | 0.11444092 | ERCC-00098 | 0.05722046 |
| ERCC-00019 | 29.296875 | ERCC-00099 | 14.6484375 |
| ERCC-00022 | 234375 | ERCC-00104 | 0.22888184 |
| ERCC-00024 | 0.22888184 | ERCC-00108 | 937.5 |
| ERCC-00025 | 58.59375 | ERCC-00109 | 0.91552734 |
| ERCC-00028 | 3.66210938 | ERCC-00111 | 468.75 |
| ERCC-00031 | 1.83105469 | ERCC-00112 | 117.1875 |
| ERCC-00033 | 1.83105469 | ERCC-00113 | 3750 |
| ERCC-00034 | 7.32421875 | ERCC-00116 | 468.75 |
| ERCC-00035 | 117.1875 | ERCC-00117 | 0.05722046 |
| ERCC-00039 | 3.66210938 | ERCC-00120 | 0.91552734 |
| ERCC-00040 | 0.91552734 | ERCC-00123 | 0.22888184 |
| ERCC-00041 | 0.22888184 | ERCC-00126 | 14.6484375 |
| ERCC-00042 | 468.75 | ERCC-00130 | 30000 |
| ERCC-00043 | 468.75 | ERCC-00131 | 117.1875 |
| ERCC-00044 | 117.1875 | ERCC-00134 | 1.83105469 |
| ERCC-00046 | 3750 | ERCC-00136 | 1875 |
| ERCC-00048 | 0.01430512 | ERCC-00137 | 0.91552734 |
| ERCC-00051 | 58.59375 | ERCC-00138 | 0.11444092 |
| ERCC-00053 | 29.296875 | ERCC-00142 | 0.22888184 |
| ERCC-00054 | 14.6484375 | ERCC-00143 | 3.66210938 |
| ERCC-00057 | 0.01430512 | ERCC-00144 | 29.296875 |
| ERCC-00058 | 1.83105469 | ERCC-00145 | 937.5 |
| ERCC-00059 | 14.6484375 | ERCC-00147 | 0.91552734 |
| ERCC-00060 | 234375 | ERCC-00148 | 14.6484375 |
| ERCC-00061 | 0.05722046 | ERCC-00150 | 3.66210938 |
| ERCC-00062 | 58.59375 | ERCC-00154 | 7.32421875 |
| ERCC-00067 | 3.66210938 | ERCC-00156 | 0.45776367 |
| ERCC-00069 | 1.83105469 | ERCC-00157 | 7.32421875 |
| ERCC-00071 | 58.59375 | ERCC-00158 | 0.45776367 |
| ERCC-00073 | 0.91552734 | ERCC-00160 | 7.32421875 |
| ERCC-00074 | 15000 | ERCC-00162 | 58.59375 |
| ERCC-00075 | 0.01430512 | ERCC-00163 | 14.6484375 |
| ERCC-00076 | 234375 | ERCC-00164 | 0.45776367 |
| ERCC-00077 | 3.66210938 | ERCC-00165 | 58.59375 |
| ERCC-00078 | 29.296875 | ERCC-00168 | 0.45776367 |
| ERCC-00079 | 58.59375 | ERCC-00170 | 14.6484375 |
| ERCC-00081 | 0.22888184 | ERCC-00171 | 3750 |

9.3 Boxplots

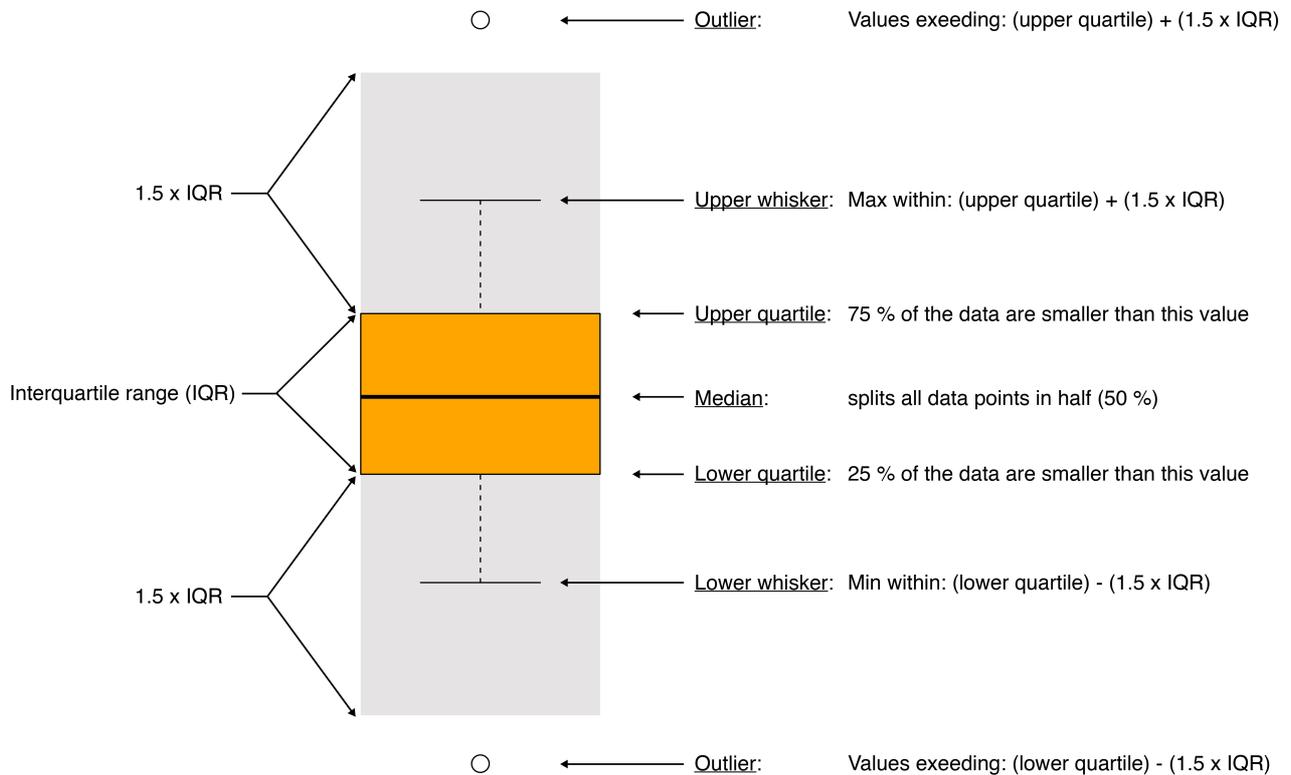


Figure 9-1: The setup of a boxplot

The figure shows an example of a boxplot. Boxplots categorize data to different quartiles, which contain 25 % of the data each. The line within the box indicates the median, which splits all data points in half (50 %). The colored box is described as interquartile range (IQR), contains 50 % of the data, and is comprised by the upper and the lower quartiles. These quartiles indicate that 75 % and 25 % of all data points, respectively, are below this level. The upper whisker shows the largest value within the area (grey) defined by the position of the (upper quartile) + (1.5 x IQR), while the lower whisker shows the smallest value within the area (grey) defined as the (lower quartile) – (1.5 x IQR). All values exceeding the limits of the grey areas are outliers and depicted as circles.

9.4 Top 25 up- and down-regulated genes of DG75 4816 and Raji 4816 cells 6 h after induction of BZLF1

Table 9-3: Top 25 up- and down-regulated genes in DG75 4816 cell 6 h after induction of BZLF1 and DESeq2 normalization.

The two columns show the up-regulated (left) and down-regulated (right) genes 6 h after BZLF1 induction. Each of the two columns depict the gene symbol ("Gene_Symbol"), log2 fold change of gene expression ("log2FoldChange") prior to and 6 h after BZLF1 induction, and the mean of reads counted for each gene ("baseMean").

| Top 25 UP-regulated genes (6 h) | | |
|---------------------------------|----------------|----------|
| Gene_Symbol | log2FoldChange | baseMean |
| ZSCAN12P1 | 3.1 | 163.26 |
| ZNF165 | 2.39 | 228.76 |
| NIPAL1 | 2.35 | 53.29 |
| AMPD3 | 2.17 | 236.96 |
| COL2A1 | 2.15 | 98.13 |
| GPR68 | 2.07 | 69.9 |
| NCF2 | 2 | 62.67 |
| ITK | 1.95 | 82.09 |
| FBXO27 | 1.92 | 225.99 |
| UNC5CL | 1.9 | 99.47 |
| CN5H6.4 | 1.86 | 42.23 |
| RAB3B | 1.84 | 92.32 |
| CCDC103 | 1.84 | 33.37 |
| PPM1E | 1.78 | 68.71 |
| DHRS2 | 1.71 | 132.88 |
| VWA5A | 1.71 | 102.77 |
| FAM70A | 1.68 | 61.42 |
| TXLNB | 1.63 | 139.04 |
| APBB3 | 1.62 | 91.1 |
| TRIM22 | 1.55 | 165.52 |
| DST | 1.53 | 730.34 |
| CPNE3 | 1.52 | 2947.54 |
| HCG22 | 1.51 | 120.43 |
| CENPBD1 | 1.46 | 126.85 |
| RSPH3 | 1.44 | 86.75 |

| Top 25 DOWN-regulated genes (6 h) | | |
|-----------------------------------|----------------|----------|
| Gene_Symbol | log2FoldChange | baseMean |
| MYC | -2.96 | 1344.82 |
| BTG2 | -2.59 | 203.45 |
| TLR10 | -2.48 | 1035.74 |
| ID3 | -2.36 | 196.57 |
| ID1 | -2.25 | 101.12 |
| IL21R | -2.18 | 522.96 |
| BEST3 | -1.96 | 91.85 |
| MARC2 | -1.85 | 104.34 |
| TLR6 | -1.72 | 77.48 |
| C15orf62 | -1.7 | 153.54 |
| CD180 | -1.69 | 94.21 |
| DDIT4 | -1.69 | 136.13 |
| ELL3 | -1.67 | 189.56 |
| PRDM10 | -1.66 | 1128.07 |
| DUSP2 | -1.61 | 38.88 |
| KIAA0182 | -1.61 | 517.94 |
| PPP1R18 | -1.6 | 320.77 |
| ICOSLG | -1.6 | 117.53 |
| CREB5 | -1.52 | 252.15 |
| WNT10A | -1.51 | 31.03 |
| HSPA1B | -1.5 | 424.81 |
| POU2F2 | -1.49 | 829.33 |
| MARC1 | -1.47 | 49.91 |
| LRRC25 | -1.47 | 110.74 |
| ZNF296 | -1.44 | 107.97 |

Table 9-4: Top 25 up- and down-regulated genes in Raji 4816 cells 6 h after induction of BZLF1 and ERCC normalization.

The columns show the up-regulated (left) and down-regulated (right) genes. Each of the two columns depict the gene symbol ("Gene_Symbol"), log2 fold change of gene expression ("log2FoldChange") prior to and 6 h after BZLF1 induction, and the mean of reads counted for each gene ("baseMean").

| Top 25 UP-regulated genes (6 h) | | |
|---------------------------------|----------------|----------|
| Gene_Symbol | log2FoldChange | baseMean |
| DPCR1 | 5.97 | 103.12 |
| L3MBTL1 | 5.76 | 53.23 |
| ASAP2 | 5.37 | 59.58 |
| TNXB | 5.32 | 82.95 |
| GRIN2A | 5.11 | 35.58 |
| NKD1 | 5.06 | 24.6 |
| RNF144A | 5.05 | 24.98 |
| PIWIL2 | 4.99 | 22.7 |
| PTGER3 | 4.94 | 29.71 |
| KIF5A | 4.93 | 28.59 |
| DNAH2 | 4.73 | 36.71 |
| SEPT4 | 4.52 | 77.73 |
| USP17L2 | 4.44 | 25.77 |
| SHROOM4 | 4.44 | 25.05 |
| SYNJ2BP | 4.38 | 42.35 |
| ANKRD26P1 | 4.31 | 28.18 |
| TEX14 | 4.29 | 380.79 |
| SCARB2 | 4.25 | 22.71 |
| COL2A1 | 4.24 | 21.76 |
| GPR98 | 4.22 | 88.15 |
| FAM184B | 4.08 | 22.93 |
| ULK2 | 4.03 | 30 |
| MYO5C | 4 | 84.47 |
| PAK6 | 3.93 | 21.27 |
| DAND5 | 3.86 | 20.39 |

| Top 25 DOWN-regulated genes (6 h) | | |
|-----------------------------------|----------------|----------|
| Gene_Symbol | log2FoldChange | baseMean |
| AICDA | -6.28 | 105.9 |
| KCNJ1 | -6.22 | 40.17 |
| BTG2 | -6.05 | 1275.11 |
| MYC | -5.58 | 7698.26 |
| IFIT2 | -4.86 | 305.85 |
| S1PR2 | -4.76 | 1717.94 |
| DUSP2 | -4.74 | 205.27 |
| CXCR7 | -4.73 | 196.47 |
| CXCR4 | -4.7 | 843.83 |
| WNT10A | -4.65 | 109.52 |
| TLR10 | -4.5 | 1353.61 |
| RHOBTB2 | -4.42 | 776.66 |
| IL21R | -4.41 | 659.49 |
| TLR7 | -4.35 | 49.46 |
| SPIB | -4.33 | 580.85 |
| SNORD93 | -4.18 | 30.51 |
| RNF144B | -4.18 | 1656.18 |
| SOWAHD | -4.16 | 53.85 |
| CXCR5 | -4.15 | 144.16 |
| CD79B | -4.12 | 4053.56 |
| SLC5A3 | -4.12 | 3724.07 |
| DTX1 | -4.12 | 1089.66 |
| CD83 | -4.11 | 3328.1 |
| RHOV | -4.06 | 170.51 |
| MIDN | -4.04 | 520.45 |

9.5 Top 25 up- and down-regulated genes of DG75 4816 and Raji 4816 cells 6 h after induction of BZLF1 with at least one peak within their promoters

Table 9-5: Top 25 up- and down-regulated genes in DG75 4816 cell 6 h after induction of BZLF1 and DESeq2 normalization, which had at least one BZLF1 peak within their promoter.

The columns show the up-regulated (left) and down-regulated (right) genes. Each of the two columns depict the gene symbol ("Gene_Symbol"), log2 fold change of gene expression ("log2FoldChange") prior to and 6 h after BZLF1 induction, and the mean of reads counted for each gene ("baseMean"), which had at least one BZLF1 binding peak ("Peaks") in their promoter.

| Top 25 UP-regulated genes with peak (6 h) | | | |
|---|----------------|----------|-------|
| Gene_Symbol | log2FoldChange | baseMean | Peaks |
| ZSCAN12P1 | 3.1 | 163.26 | 2 |
| ZNF165 | 2.39 | 228.76 | 1 |
| AMPD3 | 2.17 | 236.96 | 6 |
| COL2A1 | 2.15 | 98.13 | 1 |
| GPR68 | 2.07 | 69.9 | 1 |
| ITK | 1.95 | 82.09 | 1 |
| PPM1E | 1.78 | 68.71 | 1 |
| APBB3 | 1.62 | 91.1 | 1 |
| TRIM22 | 1.55 | 165.52 | 2 |
| DST | 1.53 | 730.34 | 8 |
| CENPBD1 | 1.46 | 126.85 | 1 |
| GCNT3 | 1.43 | 47.39 | 1 |
| LOC644669 | 1.41 | 93.02 | 1 |
| HIST1H2AL | 1.4 | 361.47 | 1 |
| ZNF43 | 1.4 | 277.19 | 2 |
| INSC | 1.36 | 20.59 | 1 |
| PIK3CG | 1.36 | 457.73 | 2 |

| Top 25 DOWN-regulated genes with peak (6 h) | | | |
|---|----------------|----------|-------|
| Gene_Symbol | log2FoldChange | baseMean | Peaks |
| BTG2 | -2.59 | 203.45 | 1 |
| TLR10 | -2.48 | 1035.74 | 1 |
| ID3 | -2.36 | 196.57 | 1 |
| IL21R | -2.18 | 522.96 | 1 |
| MARC2 | -1.85 | 104.34 | 1 |
| CREB5 | -1.52 | 252.15 | 3 |
| POU2F2 | -1.49 | 829.33 | 2 |
| MARC1 | -1.47 | 49.91 | 1 |
| LRRC25 | -1.47 | 110.74 | 1 |
| HEG1 | -1.43 | 586.98 | 2 |
| RAB3A | -1.4 | 93.32 | 1 |
| CD48 | -1.35 | 101.44 | 2 |

Table 9-6: Top 25 up- and down-regulated genes in Raji 4816 cell 6 h after induction of BZLF1 and ERCC normalization, which have at least one BZLF1 peak within their promoter.

The columns show the up-regulated (left) and down-regulated (right) genes. Each of the two columns depict the gene symbol ("Gene_Symbol"), log2 fold change of gene expression ("log2FoldChange") prior to and 6 h after BZLF1 induction, and the mean of reads counted for each gene ("baseMean"), which had at least one BZLF1 binding peak ("Peaks") in their promoter.

| Top 25 UP-regulated genes with peak (6 h) | | | |
|---|----------------|----------|-------|
| Gene_Symbol | log2FoldChange | baseMean | Peaks |
| L3MBTL1 | 5.76 | 53.23 | 1 |
| ASAP2 | 5.37 | 59.58 | 2 |
| TNXB | 5.32 | 82.95 | 2 |
| GRIN2A | 5.11 | 35.58 | 2 |
| NKD1 | 5.06 | 24.6 | 1 |
| RNF144A | 5.05 | 24.98 | 2 |
| PIWIL2 | 4.99 | 22.7 | 2 |
| PTGER3 | 4.94 | 29.71 | 4 |
| KIF5A | 4.93 | 28.59 | 1 |
| DNAH2 | 4.73 | 36.71 | 2 |
| USP17L2 | 4.44 | 25.77 | 3 |
| ANKRD26P1 | 4.31 | 28.18 | 2 |
| SCARB2 | 4.25 | 22.71 | 1 |
| COL2A1 | 4.24 | 21.76 | 1 |
| FAM184B | 4.08 | 22.93 | 2 |
| ULK2 | 4.03 | 30.00 | 3 |
| MYO5C | 4.00 | 84.47 | 1 |
| PAK6 | 3.93 | 21.27 | 1 |
| DAND5 | 3.86 | 20.39 | 1 |
| MAP7 | 3.81 | 22.81 | 1 |
| LGR4 | 3.77 | 139.11 | 2 |
| MYO15A | 3.58 | 21.76 | 1 |
| POTEM | 3.55 | 66.86 | 1 |
| HIST1H2AA | 3.55 | 20.49 | 1 |
| LTBP2 | 3.2 | 31.34 | 1 |

| Top 25 DOWN-regulated genes with peak (6 h) | | | |
|---|----------------|----------|-------|
| Gene_Symbol | log2FoldChange | baseMean | Peaks |
| KCNJ1 | -6.22 | 40.17 | 1 |
| BTG2 | -6.05 | 1275.11 | 1 |
| MYC | -5.58 | 7698.26 | 1 |
| CXCR4 | -4.7 | 843.83 | 2 |
| WNT10A | -4.65 | 109.52 | 1 |
| RHOBTB2 | -4.42 | 776.66 | 4 |
| IL21R | -4.41 | 659.49 | 2 |
| CXCR5 | -4.15 | 144.16 | 2 |
| DTX1 | -4.12 | 1089.66 | 1 |
| RHOV | -4.06 | 170.51 | 2 |
| MIDN | -4.04 | 520.45 | 1 |
| CDKN1A | -3.91 | 121.32 | 5 |
| TSC22D3 | -3.84 | 1310.86 | 2 |
| AKAP2 | -3.79 | 41.97 | 1 |
| ADORA2A | -3.78 | 324.42 | 1 |
| KREMEN2 | -3.77 | 81.07 | 1 |
| IL411 | -3.74 | 31.57 | 2 |
| TLR9 | -3.71 | 773 | 1 |
| PVRIG | -3.71 | 125.55 | 1 |
| BCL9L | -3.71 | 205.11 | 2 |
| POU2F2 | -3.7 | 1291.89 | 4 |
| ARL11 | -3.68 | 135.7 | 1 |
| Orai2 | -3.68 | 1285.01 | 1 |
| NFKBIE | -3.68 | 468.05 | 1 |
| SIT1 | -3.58 | 94.41 | 1 |

9.6 List of genes and positions captured during Capture-C analysis

Table 9-7: List of genes captured and positions captured during Capture-C analysis.

The table shows the gene symbol, the capture start and end, the TSS, the start and end of the gene annotation as well as the chromosome the gene is positioned on.

| Gene_Symbol | Capture.start | Capture.end | TSS | Chr | Gene.start | Gene.end |
|--------------|---------------|-------------|-------------|-------|-------------|-------------|
| AMPD3 | 10,471,665 | 10,481,665 | 10,476,665 | chr11 | 10,472,224 | 10,529,126 |
| APBB3 | 139,939,189 | 139,949,189 | 139,944,189 | chr5 | 139,937,853 | 139,944,189 |
| BTG2 | 203,269,663 | 203,279,663 | 203,274,663 | chr1 | 203,274,664 | 203,278,729 |
| CCDC103 | 42,972,079 | 42,982,079 | 42,977,079 | chr17 | 42,977,080 | 42,981,047 |
| CCR7 | 38,716,736 | 38,726,736 | 38,721,736 | chr17 | 38,710,022 | 38,721,736 |
| CD68 | 7,482,247 | 7,483,247 | 7,482,747 | chr17 | 7,482,805 | 7,485,429 |
| CD79B | 62,004,704 | 62,014,704 | 62,009,704 | chr17 | 62,006,098 | 62,009,704 |
| CEBPB | 48,802,119 | 48,812,119 | 48,807,119 | chr20 | 48,807,120 | 48,809,227 |
| COL2A1 | 48,393,285 | 48,403,285 | 48,398,285 | chr12 | 48,366,748 | 48,398,285 |
| CXCR4 | 136,870,725 | 136,880,725 | 136,875,725 | chr2 | 136,871,919 | 136,875,725 |
| CXCR5 | 118,749,474 | 118,759,474 | 118,754,474 | chr11 | 118,754,475 | 118,766,980 |
| CXCR7 | 237,473,379 | 237,483,379 | 237,478,379 | chr2 | 237,478,380 | 237,490,994 |
| E2F2 | 23,852,712 | 23,862,712 | 23,857,712 | chr1 | 23,832,920 | 23,857,712 |
| GPR68 | 91,705,852 | 91,715,852 | 91,710,852 | chr14 | 91,698,876 | 91,710,852 |
| HARS | 140,066,312 | 140,076,312 | 140,071,312 | chr5 | 140,053,490 | 140,071,312 |
| HDAC9 | 18,530,368 | 18,540,368 | 18,535,368 | chr7 | 18,535,885 | 19,036,992 |
| HLA-DQA1 | 32,600,182 | 32,610,182 | 32,605,182 | chr6 | 32,605,183 | 32,611,429 |
| HLA-DRB1 | 32,552,614 | 32,562,614 | 32,557,614 | chr6 | 32,546,547 | 32,557,613 |
| HN1L | 1,723,277 | 1,733,277 | 1,728,277 | chr16 | 1,728,278 | 1,752,073 |
| ID3 | 23,881,285 | 23,891,285 | 23,886,285 | chr1 | 23,884,421 | 23,886,285 |
| IL21R | 27,408,482 | 27,418,482 | 27,413,482 | chr16 | 27,413,483 | 27,460,604 |
| IL7R | 35,851,976 | 35,861,976 | 35,856,976 | chr5 | 35,856,977 | 35,879,705 |
| KANSL2 | 49,071,035 | 49,081,035 | 49,076,035 | chr12 | 49,046,995 | 49,076,008 |
| KCNQ5 | 73,326,570 | 73,336,570 | 73,331,570 | chr6 | 73,331,571 | 73,908,573 |
| KRBA2 | 8,275,029 | 8,285,029 | 8,280,029 | chr17 | 8,271,973 | 8,274,858 |
| Loc100128288 | 8,258,858 | 8,268,858 | 8,263,858 | chr17 | 8,261,731 | 8,263,859 |
| LPIN1 | 11,876,474 | 11,886,474 | 11,881,474 | chr2 | 11,817,705 | 11,967,533 |
| MAPK1 | 22,216,970 | 22,226,970 | 22,221,970 | chr22 | 22,113,947 | 22,221,970 |
| MIR155HG | 26,929,456 | 26,939,456 | 26,934,456 | chr21 | 26,934,457 | 26,947,480 |
| MYC | 128,743,314 | 128,753,314 | 128,748,314 | chr8 | 128,748,315 | 128,753,680 |
| NCF2 | 183,555,056 | 183,565,056 | 183,560,056 | chr1 | 183,524,697 | 183,560,056 |
| NEURL4 | 7,227,639 | 7,237,639 | 7,232,639 | chr17 | 7,218,951 | 7,232,638 |
| NIPBL | 36,871,860 | 36,881,860 | 36,876,860 | chr5 | 36,876,861 | 37,065,921 |
| PKD1 | 2,180,899 | 2,190,899 | 2,185,899 | chr16 | 2,138,711 | 2,185,899 |
| RAB27A | 55,577,013 | 55,587,013 | 55,582,013 | chr15 | 55,495,164 | 55,563,107 |
| RSPH3 | 159,416,198 | 159,426,198 | 159,421,198 | chr6 | 159,398,266 | 159,421,198 |
| SAAL1 | 18,122,638 | 18,132,638 | 18,127,638 | chr11 | 18,101,890 | 18,127,638 |
| SLC43A3 | 57,190,054 | 57,200,054 | 57,195,054 | chr11 | 57,174,427 | 57,195,053 |
| STRN4 | 47,245,251 | 47,255,251 | 47,250,251 | chr19 | 47,222,768 | 47,249,720 |
| SWAP70 | 9,678,627 | 9,688,627 | 9,685,627 | chr11 | 9,685,628 | 9,774,507 |
| TFEC | 115,794,950 | 115,804,950 | 115,799,950 | chr7 | 115,575,202 | 115,670,867 |
| TIFA | 113,202,059 | 113,212,059 | 113,207,059 | chr4 | 113,196,782 | 113,207,059 |

| | | | | | | |
|-----------|-------------|-------------|-------------|-------|-------------|-------------|
| TLR10 | 38,779,611 | 38,789,611 | 38,784,611 | chr4 | 38,773,860 | 38,784,611 |
| TLR6 | 38,853,438 | 38,863,438 | 38,858,438 | chr4 | 38,825,329 | 38,858,438 |
| WHSC1L1 | 38,234,790 | 38,244,790 | 38,239,790 | chr8 | 38,132,561 | 38,239,790 |
| ZBTB37 | 173,832,492 | 173,842,492 | 173,837,492 | chr1 | 173,837,493 | 173,855,774 |
| ZNF143 | 9,477,511 | 9,487,511 | 9,482,511 | chr11 | 9,482,512 | 9,550,071 |
| ZNF165 | 28,043,481 | 28,053,481 | 28,048,481 | chr6 | 28,048,482 | 28,057,340 |
| ZSCAN12P1 | 28,053,584 | 28,063,584 | 28,058,584 | chr6 | 28,058,585 | 28,063,493 |

10 Curriculum vitae

Personal data

Name: Buschle Alexander

Mail-address:

Professional experience

Since 11/ 2012 Doctoral studies, Independent research on the topic:
“The impact of Epstein-Barr virus protein BZLF1 on the cellular transcriptome and genome structure”
AG Eick (Molecular epigenetics), Helmholtz Zentrum München

University education

10/ 2010 – 09/ 2012 Master of Science,
Biology, LMU Munich,
Master thesis (1.3), AG Leonhardt, Depart. Bio. 2, LMU Munich

10/ 2007 – 09/ 2010 Bachelor of Science,
Biology, LMU Munich,
Bachelor thesis (1.0), AG Boshardt, Depart. Bio. 1, LMU Munich

Language skills

German: native language

English: full professional proficiency (ILR scale)

Electronic data processing

- Usage of Unix command line programs for the analysis of next-generation-sequencing data
- R (statistical software used for the analysis and visualization of next-generation-sequencing data)
- MS-Office Word / LibreOffice: advanced
- Adobe Illustrator: advanced

11 Danksagungen

Zuerst einmal danke ich allen, die mich auf meinem Weg zu dieser Arbeit unterstützt haben!

Mein besonderer Dank geht an:

Wolfgang: Vielen Dank für die intensive Betreuung, das Interesse an meinem Projekt und die Mühe bei der Korrektur meiner Arbeit. Ein besonderer Dank für die Zeit, die du mir gegeben hast, um mich mit der Bioinformatik zu beschäftigen, sie zu erlernen und mich mit Kursen darin fortbilden zu dürfen.

Dirk: Vielen lieben Dank für das Interesse an meinem Projekt, deine Vorschläge, Ideen und die spannenden Diskussionen. Danke, dass du die Aufgabe als mein Doktorvater übernommen hast und mich im Thesis Advisory Committee (TAC) beraten hast.

Tobias Straub: Vielen lieben Dank für deine Beratung, deine Tipps und vor allem für deinen Kurs, der mir die Türen zur Welt der Command line tools eröffnet hat. Vielen Dank für deine besonders hilfreichen Hinweise in den TACs.

Heinrich Leonhardt: Vielen Dank, dass du die Aufgabe des Zweitkorrektors übernommen hast. Das freut mich besonders, weil ich dir den größten Teil meiner epigenetischen Ausbildung zu verdanken habe.

Andreas Ladurner: Vielen Dank für deine Beratung im Rahmen des TACs.

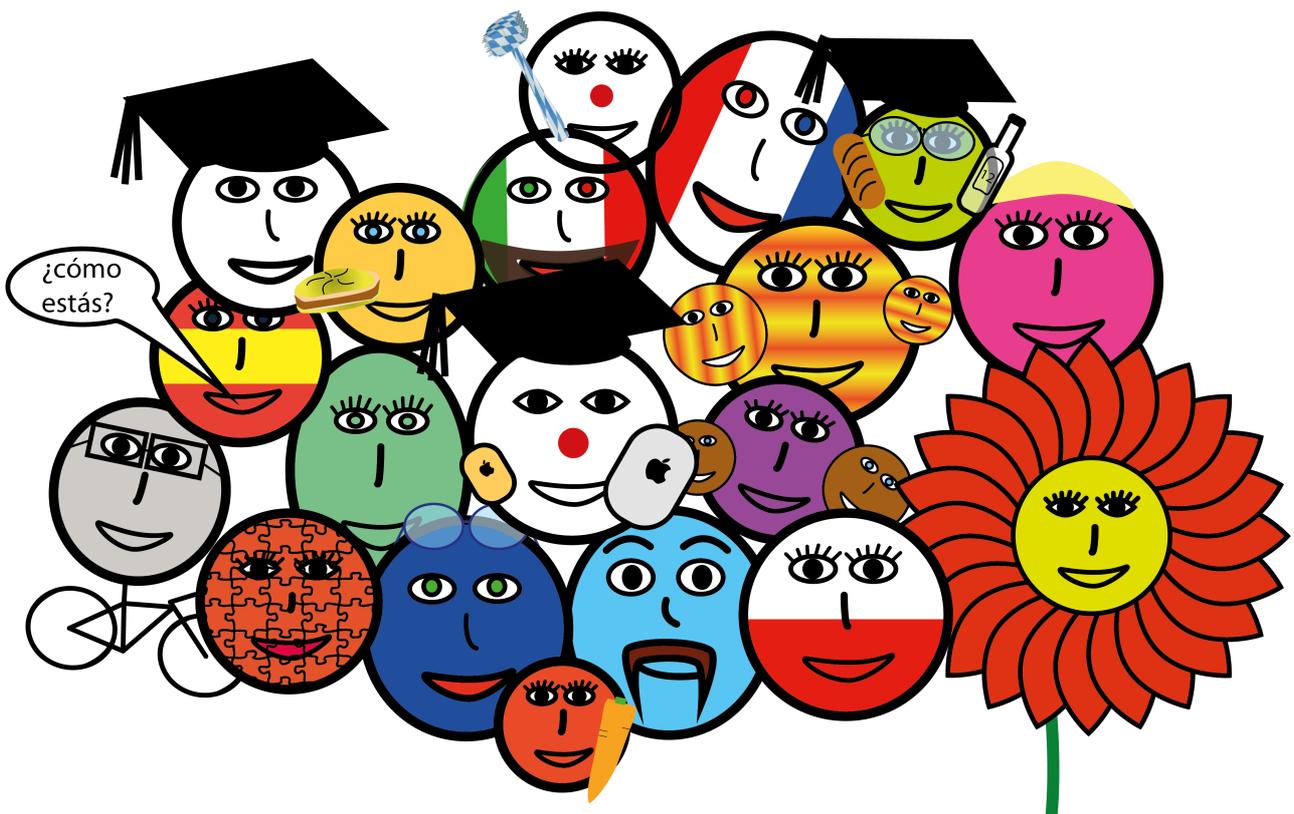
Corinna: Mein herzlichster Dank geht an dich, weil du trotz deiner eigenen Doktorarbeit immer ein offenes Ohr für mich hattest und mich immer unterstützt hast. In guten Phasen hast du dich mit mir gefreut, in schlechten hast du mich aufgefangen. Dass wir es in dieser Zeit auch noch geschafft haben unsere Hochzeit zu organisieren und zu heiraten hat nur so gut funktioniert, weil wir immer zusammenhalten. Ein ganz besonderer Dank für deine Unterstützung in der Zeit als ich geschrieben habe, wir uns teils tagelang nicht gesehen

haben und du fast alle wichtigen Aufgaben übernommen hast, obwohl du selber geschrieben hast. Vielen lieben Dank! Du machst mich sehr glücklich!

Familie: Vielen lieben Dank für euer Interesse an meiner Arbeit und dafür, dass ihr mich bei dieser Arbeit und auch sonst immer unterstützt habt. Vielen Dank für das Essen und die Essenslieferungen. Ich weiß, dass ich immer auf euch zählen kann.

Opa: Auch wenn du das hier nicht mehr lesen kannst möchte ich dir trotzdem danken für dein Interesse an meiner Arbeit, all die besonderen Geschichten, die du mir über dein Leben erzählt hast und dein schelmisches Lachen, wenn du einen deiner Witze gemacht hast.

Hammerschmidts: Ich danke allen Hammerschmidts für die angenehme Arbeitsatmosphäre und die große Hilfsbereitschaft, die ich in den letzten Jahren erfahren durfte. Plopp.



Paula: My special thanks goes to you for the every day discussions, your reliability, the private conversations, and the positive atmosphere you are spreading in the lab.

Manuel: Thanks a lot for your help the last days before handing in. That helped a lot!

Level 1: Vielen Dank an alle aus Stockwerk 1 für besonders lustige und entspannte Mittagspausen und den Willen gemeinsame Unternehmungen zu organisieren.

Lafuga: Vielen Dank an Dr. Blum und sein Lafuga Team, dass ich meine Libraries in eurem Labor erstellen durfte. Mein besonderer Dank geht an Sylvia Mallok für all ihre Hilfe und Stefan Krebs für die Beratung, die Tipps und das Sequenzieren. Außerdem möchte ich Andrea Klanner für die Erstellung einiger Libraries für mehrere Projekte danken. Vielen Dank an Alexander Graf und Ashok Varadharajan für die Einführung in die Handhabung der Galaxy Software sowie weiteren Support.

Thomas Wieland: Ein besonderer Dank geht an Thomas Wieland, der mir die ersten Schritte in der Analyse von ChIP-seq Daten erklärt hat und mich motiviert hat es auch mal selber mit der Bioinformatik zu versuchen. Vielen Dank!

Catherine Regnard: Vielen lieben Dank für die Diskussionen bezüglich der ERCCs und dafür, dass du den Kontakt mit Tamas hergestellt hast.

Tamas Schauer: Vielen Dank, dass du mir dein Script für die Normalisierung mit ERCCs zur Verfügung gestellt hast. Erst dadurch hat sich gezeigt, was für einen massiven Effekt die Expression von BZLF1 verursacht.

Matthias Siebert und Anja Sophie Kiesel: Vielen Dank für die Tipps und Tricks, die ihr mir für die Bioinformatik mitgegeben habt und für die Erklärungen bezüglich der Sequenzlogos.

James Davies: Thanks a lot for the discussion about the different steps of the Capture-C analysis, which helped me to understand how the bioinformatic concept works.

Aicha Jeridi und Vigo Heissmeyer: Vielen Dank für die 50 ml des tetracyclinfreien FCS.

Chris Lindley: Vielen Dank für das Konvertieren eines MacVector files in das .gtf format