A University of Sussex PhD thesis

Available online via Sussex Research Online:

http://sro.sussex.ac.uk/

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details

Exploring interactions between Epstein-Barr virus transcription factor Zta and The Human Genome

By IJIEL BARAK NARANJO PEREZ FERNANDEZ

A Thesis submitted for the degree of Doctor of Philosophy

University Of Sussex

School of Life Sciences

September 2017

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signature:……

Acknowledgements

I want to thank Professor Alison J Sinclair for her guidance, mentoring and above all continuous patience. During the time that I've been part of her lab I've appreciated her wisdom as an educator her foresight as a scientist and tremendous love as a parent. I wish that someday soon rather than later her teachings are reflected in my person and career; hopefully inspiring others like me.

Thanks to Professor Michelle West for her help whenever needed or offered. Her sincere and honest feedback, something that I only learned to appreciate after my personal scientific insight was developed. I wish I had learned this sooner.

Thanks to the people in the virology labs that make this place feel like home, Dr Kay Osborne has always been helpful and kind, Dr David Wood always lend his intelligence and "know how", even when he didn't have to! Dr Michelle Brocard, Dr Andrea Gunnell, Dr Helen Webb and Dr Sharada Ramasubramanyan have been missed greatly since their departure. Dr Lina Chen and her flourishing spirit always makes any day brighter. Thank you to Yaqi Zhou, Anja Godfrey, Sarika Katnis, Hildegonda Veenstra, and Rajaei Al-Mohammad for their friendship and support. We should travel together more often.

I know I take with me great friendships that will span far beyond my stay in the United Kingdom, so I hope it won't be long before Dr Chris Traylen, Rob Simmons, Jake Evans and Dr Mike McClellan stop for a visit wherever I'm living.

I want to thank the Mexican Council of Science and Technology for their economic support in the development of my scientific formation and by extension our nation's progress. This of course is possible thanks to the Mexican taxpayer that still does what is correct, even when averse with the current administration.

I want to thank my family for their support and direct involvement in my personal development. I thank my parents and the rest of the family. In particular I want to thank my sister Thais B. Naranjo P.-F. whose selflessness and generosity brings life to everyone that surrounds her. Special thanks to Anna K, you've made this so much easier I'm forever grateful to you.

UNIVERSITY OF SUSSEX

IJIEL BARAK NARANJO PEREZ FERNANDEZ

PhD BIOCHEMISTRY

"EXPLORING INTERACTIONS BETWEEN EPSTEIN-BARR VIRUS TRANSCRIPTION FACTOR Zta AND THE HUMAN GENOME"

SUMMARY

Epstein-Barr virus is a gamma herpesvirus that is present in human adult's B-lymphocytes infecting 90% of the global population. EBV causes many types of lymphoma and carcinoma. The virus life cycle can be divided in two stages, latency and lytic cycle. Viral gene BZLF1 codes for the viral transcription and replication factor Zta (also known as BZLF1, ZEBRA, EB1, and Z) which is part of the signalling required to switch from latency to the lytic cycle. Zta is part of the bZIP family of proteins, it forms homodimers and can bind to specific sequences termed Zta Response Elements (ZREs). It binds to the EBV lytic origin of replication as well as to specific targeted promoters in the viral genome and regulates its expression. Recent research found and mapped interactions between the key viral transcription factor Zta and the B-cell genome, this showed interactions of Zta proximal (closer than 2Kb) and distal (farther than 2 Kb) to the transcription start site of several genes. In this work, I asked the questions: Can enhancer properties be found in the sequences where Zta binds to? Is Zta distally regulating expression by looping of DNA? This was approached first by identifying potential sequences that could be conferring enhancing activity, then inserting them into vectors and transfecting them into two different cell lines. In this way, through luciferase reporter assays, any enhancing capabilities of the sequences were tested when placed in a proximal and distal manner to promoters known to be regulated by Zta, as well as mutated promoters not regulated by Zta. This resulted in finding discreet enhancer activity in the sequences analysed, with some being specific to the cell type that was used in the experiment. To answer the second question, Chromosome conformation capture (3C) was used to test the possibility of a spatial rearrangement bringing together distal Zta binding regions and promotor regions of selected genes (looping). However, I did not find evidence of looping between Zta binding sites and the neighbouring promoters analysed, in the cell context employed.

Table of contents

Chapter 1. Introduction 1
1.1 Herpesvirales 1
1.2 Epstein-Barr Virus 4
1.2.1 EBV genome structure 4
1.2.2 Viral Entry 6
1.2.3 EBV life cycle 6
1.1.3.1 EBV latency cycle 7
1.1.3.2 EBV associated pathologies and latency 9
1.1.3.2.1 EBV malignancies related to B cell infection 9
1.1.3.2.2 EBV malignancies in atypical cells 12
1.1.3 EBV lytic stage 14
Lytic Reactivation 14
1.1.3 Zta 16
Structure and DNA binding 17
1.3 Computational methods for sequence analysis 19
1.4 Zta interacting with the human genome 19
1.5 Enhancers 19
1.6 Chromatin looping events 23
1.7 Aims of project 24
Chapter 2. Materials and Methods 25
2.1 Materials 25
2.1.1 Vectors 25
2.1.2 Cell lines 25
2.1.3 Antibodies 26
2.1.4 Primers (oligonucleotides) 27
2.1.5 Solutions and buffers 28

2.1.6 Kits and reagents 29
2.2 Methods 31
2.2.1 Molecular cloning by digestion-ligation 31
2.2.2 Bacterial Transformation 32
2.2.3 Plasmid DNA extraction 33
2.2.4 Cell Culture 33
2.2.5 Cell Transfection 34
2.2.6 Protein gels and Western Blots 35
2.2.7 Luciferase reporter assays 36
2.2.7.1 Bicinchoninic Acid Assay (BCA) 36
2.2.7.2 Normalization 37
2.2.8 Polymerase Chain Reaction 37
2.2.9 Agarose Gel Electrophoresis 38
2.2.10 Chromatin conformation capture assay 38
2.2.10.1 Formaldehyde crosslink and cell lysis 38
2.2.10.2 DNA digestion 39
2.2.10.3 DNA ligation 39
2.2.10.4 DNA purification 39
2.2.11 Computational methods 40
2.2.11.1 Regulatory Sequence Analysis Tools 42
2.2.11.1.1 Sequence upstream from genes 43
2.2.11.1.2 Consensus sequence 43
2.2.11.1.3 Matrix scan 43
2.2.11.1.4 Matrix mapping 43
2.2.11.2 BLASTn alignment 43
2.2.11.3 Primer design 43
Chapter 3. Enhancer capability of Zta binding sites 44

3.1 Introduction 44

3.2 Results 45

3.2.1 *Enhancers* method of identification of genomic sequences 45

3.2.2 *Enhancers* method of analysis of sequences and element design 45

3.2.3 *Upregulated genes* method of identification of genomic sequences 50

3.2.4 *Upregulated genes* method of analysis of sequences and element design 50

3.2.4.1 Derivation of matrices of Zta binding sites 51

3.2.4.1.1 Identification of genomic sequence of proposed interactions 51

3.2.4.1.2 Basic Local Alignment Search Tool 51

3.2.4.1.3 Regulatory Sequence Analysis Tools 51

3.2.4.1.4 Alignments and verifications 55

3.2.4.2 Reporter genes with Zta binding sites 55

3.2.4.3 Synthetic generation of elements through GENESTRINGS ...62

3.2.5 Element enhancing effect 62

3.2.5.1 Transfections and Reporter Assays 65

3.2.5.1.1 Is element AHNAK an enhancer in DG75 cells? 70

3.2.5.1.2 Is element AHNAK an enhancer in 293T cells? 70

3.2.5.1.3 Is element FOSB an enhancer in DG75 cells? 71

3.2.5.1.4 Is element FOSB enhancing in 293T cells? 71

3.2.5.1.5 Is 1OEM enhancing in DG75 cells? 72

3.2.5.1.6 1OEM enhances in 293T cells? 73

3.2.5.1.8 Is 4OEM enhancing in DG75 cells? 74

3.2.5.1.9 Is 4OEM enhancing in 293T cells? 74

3.3 Discussion 108

Chapter 4. Potential looping interactions of Zta with promoters 111

4.1 Introduction 111

4.2 Results 114

4.1.1 Identification of potential looping targets between Zta binding sites and promoters 114

4.1.1.1 Selection Criteria 114

4.1.1.1.1 Highest regulated genes 114

4.1.1.2.1 Upregulated gene clusters 115

4.3 Discussion 122

Chapter 5. Chromatin Conformation Capture for the proposed Zta loops ... 124

5.1 Introduction 124

5.2 Results 128

5.2.1 Preparation of the Template 128

5.2.1.1 Lytic Cycle confirmation 128

5.2.1.2 Enzymatic digestion of Chromatin 128

5.2.2 Zta and promoter sequence of SLC6A7 131

5.2.3 Zta and promoter sequence of LHX1 140

5.2.4 Zta and promoter sequence of FZD10 149

5.2.5 Zta and promoter sequence of cluster VWF and CD9 158

5.2.6 Template controls 164

5.3 Discussion 169

Chapter 6. Discussion 171

List of Figures

Figure 1.1	Herpesvirus.	2
Figure 1.2	Schematic of the EBV genome.	5
Figure 1.3	Zta dimerization and binding to DNA.	18
Figure 3.1	Diagram of how the different elements were designed and produced.	46
Figure 3.2	Clusters of peaks proximal to FOSB gene proposed for super-enhancer activity.	47
Figure 3.3	Clusters of peaks proximal to AHANK gene proposed for super-enhancer activity.	48
Figure 3.4	Sequences corresponding to peaks proximal to AHNAK.	49
Figure 3.5	BLASTn alignment of sequences with FOSB finds several matches.	52
Figure 3.6	Three Matrices were generated through RSAT analysis.	53
Figure 3.7	Matrices incidences in FOSB (fragment).	54
Figure 3.8	Mapped matrices occurrences in FOSB.	56
Figure 3.9	Annotation and alignment of sequence upstream of FOSB.	57
Figure 3.10	Alignment of analysis maps for the region upstream of FOSB.	59
Figure 3.11	Alignment of analysis maps for the region upstream of SCIMP.	60
Figure 3.12	Alignment of analysis maps for the region upstream of BCL2A1.	61
Figure 3.13	Compiled Matrices instances into 4OEM and 1OEM.	63
Figure 3.14	Elements designed from the peak clusters proximal to FOSB and AHNAK genes.	64
Figure 3.15	Three vectors were used to test the effect of the presence of the elements.	66
Figure 3.16	Sample sequencing and comparison for element FOSB.	67
Figure 3.17	Influence over BHLF1 activity by AHNAK in DG75 cells.	76
Figure 3.18	Influence over BHLF1 activity by AHNAK when ZREs are mutated in DG75 cells.	78
Figure 3.19	Influence over BHLF1 activity by AHNAK in 293T cells.	80
Figure 3.20	Influence over BHLF1 activity by AHNAK in 293T cells when ZREs are mutated.	82
Figure 3.21	Influence over BHLF1 activity by FOSB in DG75 cells.	84
Figure 3.22	Influence over BHLF1 activity by FOSB in DG75 cells when there are no ZREs available.	86
Figure 3.23	Influence over BHLF1 activity by FOSB in 293T.	88

Figure 3.24	Influence over BHLF1 activity by FOSB when ZREs are mutated in 293T cells.	90
Figure 3.25	Influence over BHLF1 activity of 1OEM in DG75 cells.	92
Figure 3.26	Influence over BHLF1 activity of 1OEM in DG75 cells when ZREs are mutated.	94
Figure 3.27	Influence over CIITA activity of 1OEM in DG75 cells.	96
Figure 3.28	Influence over BHLF1 activity by 1OEM in 293T cells.	98
Figure 3.29	Influence over BHLF1 activity of 1OEM in 293T cells when ZREs are mutated.	100
Figure 3.30	Influence over CIITA activity by 1OEM in 293T cells.	102
Figure 3.31	Influence over CIITA activity by 4OEM in DG75 cells.	104
Figure 3.32	Influence over CIITA activity by 4OEM in 293T cells.	106
Figure 4.1	Genes selected for analysis.	116
Figure 4.2	Genes were grouped into clusters.	117
Figure 4.3	Zta peaks were found within 100 kb from gene clusters.	119
Figure 4.4	Clusters selected.	120
Figure 5.1	Schematic of 3C detection library.	125
Figure 5.2	Expression of Zta only in lytic cell lines and higher expression of EBV viral load in induced cell lines.	129
Figure 5.3	Enzymatic Digestion of Chromatin analysed on 1.2% agarose Gel.	130
Figure 5.4	Maps of potential looping regions at SLC6A7.	132
Figure 5.5	Sequences used to design primers for SLC6A7.	134
Figure 5.6	Sequences used as template for synthetic controls of SLC6A7.	137
Figure 5.7	Amplification of PCR products from synthetic positive controls for loops with promoter of SLC6A7.	138
Figure 5.8	Temperature gradients used to optimise melting temperatures for loops with SLC6A7.	139
Figure 5.9	Chromosome conformation capture analysis of the SLC6A7 promotor region.	141
Figure 5.10	Maps of potential looping regions at LHX1.	142
Figure 5.11	Sequences used as template for synthetic controls of LHX1.	144
Figure 5.12	Amplification of PCR products from synthetic positive controls for loops with promoter of LHX1.	145
Figure 5.13	Non specific amplification of a false positive.	146
Figure 5.14	Chromosome conformation capture analysis of the LHX1 promotor region.	148
Figure 5.15	Map of potential loop regions at FZD10.	150
Figure 5.16	Sequences used as template for synthetic controls for FZD10.	152
Figure 5.17	Amplification of PCR products from synthetic positive controls for loops with promoter of FZD10.	153

Figure 5.18	Temperature gradients used to optimise melting temperatures for loops with FZD10.	155
Figure 5.19	Chromosome conformation capture analysis of the FZD10 promotor region.	157
Figure 5.20	Maps of potential looping regions at cluster VWF-CD9.	159
Figure 5.21	Sequences used as template for synthetic controls for cluster VWF and CD9.	162
Figure 5.22	Amplification of PCR products from synthetic positive controls for loops with cluster VWF and CD9.	166
Figure 5.23	Chromosome conformation capture analysis of cluster VWF-CD9.	167
Figure 5.24	Amplification of DNA in 3C templates.	168

List of tables

Table 1.1	Summary of EBV-associated malignancies and viral latency.	15
Table 2.1	List of Vectors used.	25
Table 2.2	List of Cell lines used.	25
Table 2.3	List of Antibodies used.	26
Table 2.4	List of primers used.	27
Table 2.5	List of solutions and buffers used.	28
Table 2.6	List of kits and reagents used.	29
Table 3.1	Table of matrices incidences and weight scores.	62
Table 3.2	Summary table of transfections.	68
Table 4.1	Eight gene clusters with Zta peaks less than 100 kb away from TSS.	118
Table 4.2	Fragment sizes for the clusters.	121
Table 5.1	Cell lines used to generate 3C library.	127

List of Abbreviations

3C	Chromosome conformation capture
4C	Circularized chromosome conformation capture
5C	Chromosome conformation capture carbon copy
aa	amino acids
AP1	Activator protein 1
ATP	Adenosine triphosphate
BACs	Bacterial Artificial Chromosomes
BARTs	BamHI A rightward transcripts
BCR	B cell receptor
BED	Browser Extensible Data (file extension)
BL	Burkitt’s lymphoma
BLASTn	Basic Local Alignment Search Tool
bp	Base pair
bZIP	Basic leucine zipper
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation coupled with sequencing
cHL	Classical Hodgkin lymphoma
CIITA	Class II major histocompatibility complex transactivator
CMV	Cytomegalovirus
Cp	Latency promoter within BamHI C digestion fragment
CpG	Cytosine-phosphate-Guanine
CR2	Complement receptor type 2
DMEM	Dulbecco/Vogt Modified Eagle's Minimal Essential Medium
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
DPBS	Dulbecco’s Phosphate-Buffered Saline
DTT	Dithiothreitol
EBER	Epstein-Barr virus-encoded ribonucleic acid
EBNA	Epstein Barr Nuclear Antigen
EBNA-LP	Epstein Barr Nuclear Antigen Leader Protein
EBV	Epstein-Barr virus
ECL	Enhanced chemiluminescence
EDTA	Ethylene diamine tetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
EICE	Ets/ISRE-consensus element
FCS	Foetal Calf Serum
GFP	Green Fluorescent Protein
Gp	Glycoprotein
H3K27ac	Histone H3 acetylated at Lysine 27
H3K27me3	Histone H3 trimethylated at Lysine 27
H3K9ac	Histone H3 acetylated at Lysine 9
H3K9me3	Histone H3 trimethylated at Lysine 9
HCMV	Human Cytomegaolvirus
HEK	Human embrionic kidney (cells)

HHV-1	Human Herpes Virus 1, or Herpes Simplex Virus 1 (HSV1)
HHV-2	Human Herpes Virus 2, or Herpes Simplex Virus 2 (HSV2)
HHV-3	Human Herpes Virus 3, or Varicella Zoster Virus (VZV)
HHV-4	Human Herpes Virus 4, or Epstein Barr Virus (EBV)
HHV-5	Human Herpes Virus 5, or Human Cytomeglovirus (HCMV)
HHV-8	Human Herpes Virus 8, or Kaposi's Sarcoma-associated Herpesvirus (KSHV)
Hi-C	high throughput chromosome conformation capture
HL	Hodgkin's Lymphoma
HSV-1	Herpes simplex virus-1
HSV-2	Herpes simplex virus-2
ICTV	International Committee on Taxonomy of Viruses
IgG	Immunoglobulin G
IM	Infectious Mononucleosis
kbp	Kilobase pair
KDa	Kilo Dalton
KLF4	Kruppel like factor 4
KSHV	Kaposi's Sarcoma-associated Herpesvirus
LB	Luria broth
LCL	Lymphoblastoid Cell Lines
LMP	Latent Membrane Protein
MHC	Major Histocompatibility Complex
miRNA	Micro RNA
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	Messenger RNA
NaCl	Sodium Chloride
NCBI	National Center for Biotechnology Information
NFκB	Nuclear factor kappa enhancer of activated B cell
NK	Natural killer
NPC	Nasopharyngeal Carcinoma
oriLyt	Origin of lytic replication
oriP	Origin of plasmid replication
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PKC	Protein Kinase C
PSG	Penicillin, Streptomycin, L-Glutamine
PTLD	Post Transplantation Lymphoproliferative Disease
Qp	Latency promoter within BamHI Q digestion fragment
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
Rp	Promoter of *BRLF1*
Rpm	Rotation per minute
RPMI	Roswell Park Memorial Institute medium
RSAT	Regulatory Sequences Analysis Tools
SDS	Sodium dodecyl sulfate
TBE	Tris-Borate-EDTA

TPA	12-0-tetradecanoyl phorbol-13-acetate
TR	Terminal repeats
TSS	Transcription start site
UCSC	University of California SantaCruz
v/v	Volume / volume
VZV	Varicella-zoster virus
w/v	Weight / volume
WB	Western blot
Wp	Latency promoter within BamH1 W digestion fragment
XBP1	X-box binding protein 1
Zp	Promoter of *BZLF1*
ZRE	Zta Response Element

Chapter 1. Introduction

1.1 Herpesvirales

Herpesviruses have a long historical presence in humanity; early records of diseases kept by ancient practitioners of medicine in Classical Greece described particular cutaneous lesions typical of infections now known to be caused by herpes simplex viruses (HSV) as well as varicella-zoster virus (VZV). The family name of the viruses “Herpes” comes from the Greek description of the lesions appearing to crawl over the skin. Studies on herpesvirus have evolved with the times, and their transmission as well as their occurrence have been followed by scientists through centuries; in the 19th century, experiments showed that HSV and VZV could be transmitted from fluid recovered from HSV and VZV lesions, and in the first half of the 20th century research showed great advancements in the understanding of the latency and reactivation of herpesvirus and how the infection persists for life (Arvin and Abendroth 2007).

In the early 50’s the human cytomegalovirus (HCMV) as well as VZV were the first herpesvirus isolated in cultured cells, and in the following decade Epstein-Barr virus (EBV) was shown to be associated with mononucleosis after being discovered in cells derived from a tumour from a patient with Burkitt’s Lymphoma (BL). As time progressed and technology progress with it, more human herpesviruses were discovered and studied, showing biological diversity and evolutionary similarities. Herpesviruses share many common genes (see Figure 1.1 A) suggesting a common ancestor 150 million years ago from which several herpesviruses diverged (Mocarski 2007).

Members of the *Herpesviridae* family share common traits, such as a large size (130-250 kb) linear double stranded DNA genome. Structurally, the virions are spherical, with a diameter of approximately 200nm and consist of 3 main structures: Envelope, tegument, and nucleocapsid. The lipid envelope of the *Herpesviridae* viruses is embedded with viral membrane glycoproteins and cellular proteins of different function (Davison 2007). The tegument, a protein compartment, is surrounded by the envelope and in turn the tegument surrounds the nucleocapsid. The nucleocapsid consists of capsomeres which form an icosahedron joined via triplexes, the densely compacted linear double stranded

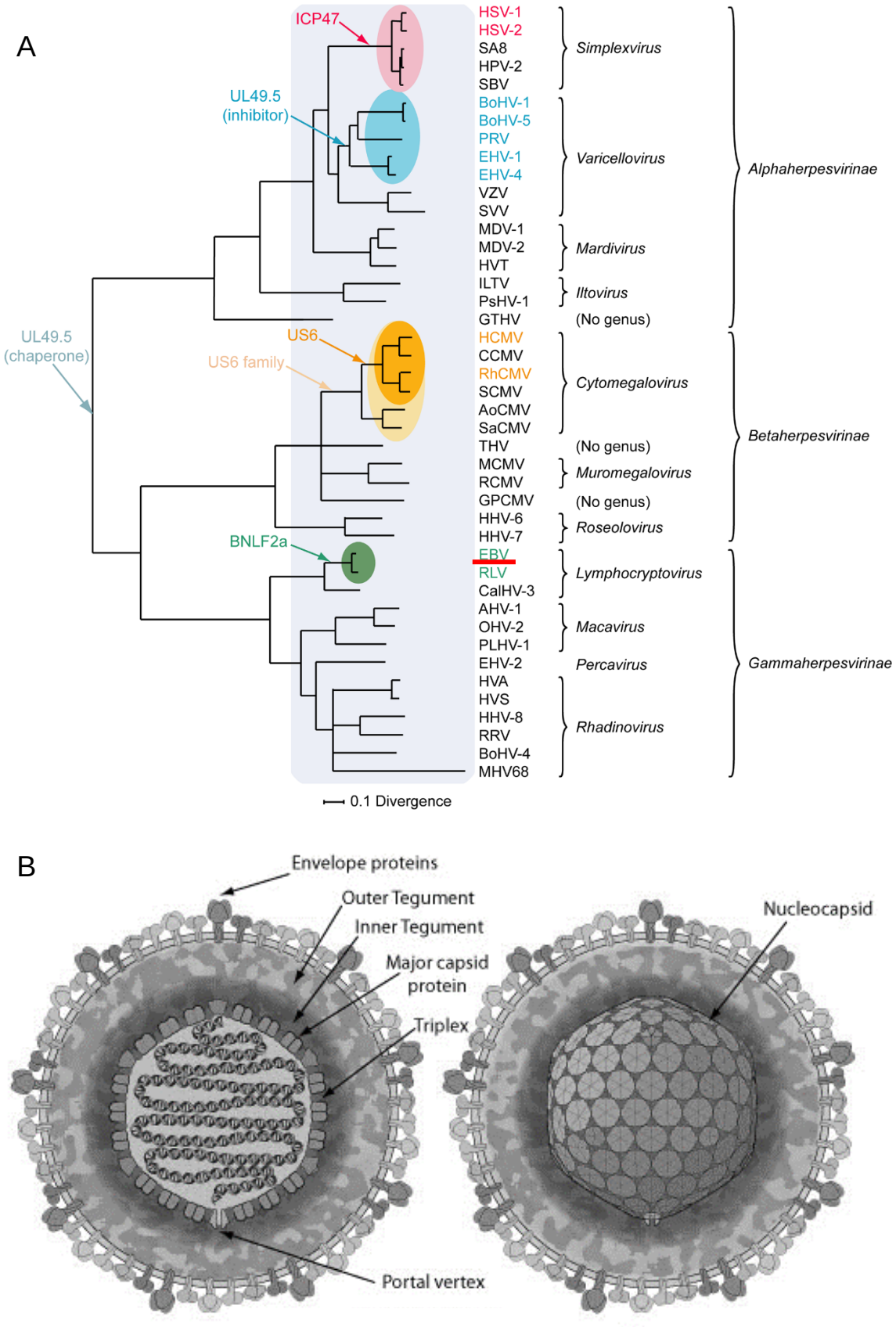


Figure 1.1 Herpesvirus. (**A**) Composite phylogenetic tree for herpesviruses showcasing genes that are shared in different species. Underlined in red EBV. Modified from (Verweij et al. 2015). (**B**) Herpesvirus virion structure. Taken from (Swiss Institute of Bioinformatics 2017).

DNA is packed in its core (See Figure 1.1 B) (F. Y. Liu and Zhou 2007).

Initially, Herpesviruses were classified as such based on the shared biological properties, such as host range or broad pathogenic and epidemiological features of the disease; however, the type of cell infected was later used to subdivide the *Herpesviridae* family into three subfamilies: *Alphaherpesvirinae, Betaherpesvirinae* and *Gammaherpesvirinae*. Later, with the development of nucleic acid sequencing, the taxonomy of Herpesvirus also advanced; now they are classified into an order called *Herpesvirales*, which consists of the *Herpesviridae* family, subdivided into 3 subfamilies *Alpha- Beta-* and *Gamma-herpesvirinae* which encompass 17 genera and 90 species (Davison 2010).

Some studies have shown that Herpesviruses have evolved interesting mechanisms to modulate or evade the host reaction to the infection, and therefore establishing a very prolonged or even lifelong latent infection (Tellam et al. 2012). Severe symptoms of infection by Herpesviruses are usually shown only in immunosuppressed or very young individuals, the usual infection routes can range from aerosol spread to mucosal contact (Davison 2007). Although occasional inter-species transfers do occur, when studied in vitro, single species of virus from the Beta- and Gammaherpesvirinae subfamily are limited to infection of specific host cells from a single species, whereas some virus from Alphaherpesvirinae can target more than one species of hosts cells (Barton et al. 2011). For instance the porcine pseudorabies virus (PrV) can infect its natural host cells as well as mouse cells in vitro (Menotti et al. 2000).

Nine herpesviruses are known to infect humans; in the *Alphaherpesvirinae* subfamily: herpes simplex virus type 1 (HSV1, HHV1) and type 2 (HSV2, HHV2). In the *Betaherpesvirinae* subfamily: varicella-zoster virus (VZV, HHV3), human cytomegalovirus (HCMV, HHV5), human herpesviruses 6A, 6B, and 7 (HHV6A, HHV6B, HHV7). And in the *Gammaherpesvirinae* subfamily; Epstein-Barr virus (EBV, HHV4), and Kaposi's sarcoma-associated herpesvirus (KSHV, HHV8) (Verweij et al. 2015).

Found in all vertebrate species, the *Gammaherpesvirinae* subfamily of herpesviruses is divided into 4 genera, *Lymphocryptovirus*, *Macavirus*, *Percavirus* and *Rhadinovirus* (Davison 2010). The *Gammaherpesvirinae* species

are as varied as their multiple various hosts: mice, cows, horses, humans and other primates, these viruses were first distinguished by their tropism towards lymphocytes, inducing lymphoproliferation and cancer (Longnecker and Neipel 2007).

As with all other Herpesvirus, infection of cells by gamma herpesvirus places the previously linear viral DNA inside the host cell nucleus, where it circularizes as an episome that is not integrated into the cell genome. Then, two different situations can occur, latency is established or lytic cycle is followed. In latency, viral expression of genes is limited to a minimum and the production of new viruses is suspended; and in lytic cycle, the viral genome expresses all necessary proteins for the generation of new infectious viral particles (Barton et al. 2011).

1.2 Epstein-Barr Virus

1.2.1 EBV genome structure

Mature EBV virions are 120-180 nm in diameter, the core of the viral particle holds inside the linear DNA, which has a large, nucleosome free, double stranded ~172Kb DNA genome. It is flanked by terminal repeats, which bind and circularize after entry to the nucleus through homologous recombination; where it remains as an episome (Kieff et al. 1982; Kintner and Sugden 1979; Nonoyama and Pagano 1972; Shaw et al. 1979).

When in latency, the viral episome is maintained in the nucleus and associated with histones, where it is replicated through the origin of replication (OriP) (Tempera and Lieberman 2010). When in lytic stage, the viral genome is replicated through the two origins of replication; OriLyt Left and OriLyt Right (Hammerschmidt and Sugden 1988)

Fragments within the genome are named alphabetically based on the initial sequencing of EBV that digested the DNA with BamHI (See fig 1.2) (Baer et al. 1984).

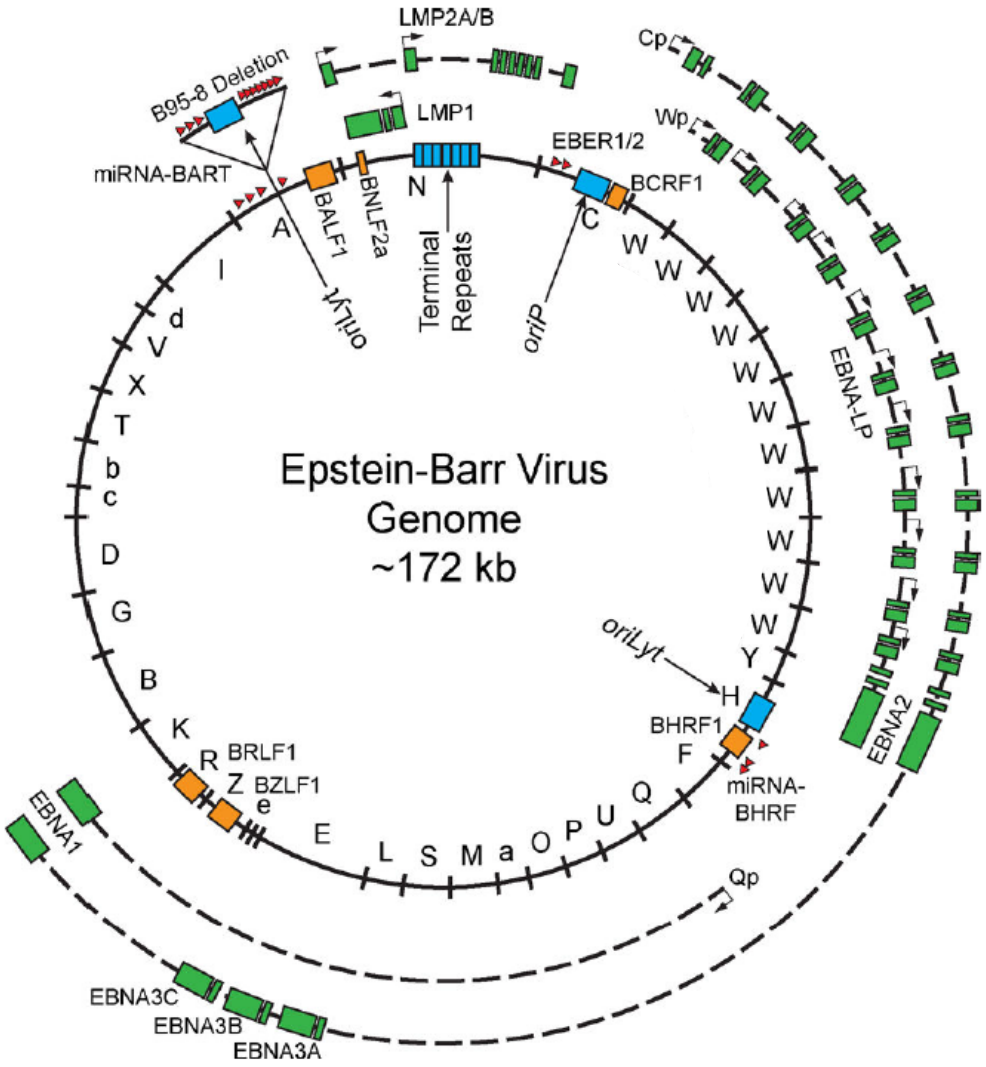


Fig 1.2 Schematic of the EBV genome. Letters correspond to BamHI digestion fragments. Origin of plasmid replication (oriP), the two origins of lytic replication (oriLyt) and the terminal repeats in blue. Lytic genes of pre-latent phase are in orange. Coding exons for latency genes are shown in green. Latent mRNAs can be initiated from different promoters depending on latency type and time after infection: the W promoter (Wp), the C Promoter (Cp), the Q promoter (Qp, only in Latency I), and the LMP promoters are labelled. The non-spliced pre-mRNAs driven from these promoters is shown as a dashed line. Latency 0 miRNAs shown as red triangles. Taken from (Price et al 2014).

1.2.2 Viral Entry

Initial entry of the virus into host cells occurs differently depending on the cell type; in epithelial cells it occurs by direct fusion of the viral envelope with the cell plasma membrane, whereas entry to B cells require endocytosis of the virus, which is followed by a subsequent membrane fusion event (Miller and Hutt-Fletcher 1992; Nemerow and Cooper 1984).

Several viral glycoproteins participate when EBV is entering B cells; gp350/220 allows attachment by binding to the cell's receptor CD21, gp42 binds to the cell's major histocompatibility complex (MHC) class II to initiate both the entry event, and viral proteins gH/gL involved with the fusion of membrane (Fingeroth et al. 1984; Nemerow et al. 1985).

Once the virion enters the B cell and escapes endocytosis, viral proteins contained in the tegument are released into the host cell. Viral tegument protein BNRF1 binds to cellular protein Daxx and disrupts the Daxx-ATRX complex, which is known to supress transcription through histone methylation; this could mean that BNRF1 is therefore preventing the suppression of viral gene expression (Stanfield and Luftig 2017; Tsai et al. 2011; Xue et al. 2003). Recent research has suggested that upon fusion of the membranes, viral RNA transcripts that were contained in the virus are released into the host cell, undergo translation and modify the host cell response to infection (Jochum et al. 2012; Stanfield and Luftig 2017).

1.2.3 EBV life cycle

Depending on the expression pattern of EBV nuclear antigens (EBNAs) as well as latent membrane proteins (LMPs), and small non-polyadenylated RNAs called Epstein-Barr encoded small RNAs (EBERs) and BamHI-A rightward transcripts (BARTs) miRNAs; EBV latency can be subdivided into 4 patterns of expression: Latency 0, I, II and III (Abate et al. 2015; Rowe et al. 1992).

In Latency 0, EBERs and BART miRNA transcripts are expressed, however they are also expressed in all other Latency programs. In Latency I, EBNA1 is expressed, in Latency II LMP1, LMP2A, and LMP2B are expressed in conjunction with what is expressed in Latency I, and finally, in Latency III all the latency genes

are expressed (EBNA1, EBNA2, EBNA3A, 3B, 3C, EBNA-LP, LMP1, LMP2A, LMP2B, BARTs, EBERs and BHRF1 miRNA).

1.1.3.1 EBV latency cycle

Immediately after the viral genome is deposed into the nucleus of the newly infected B cell, a short period where both lytic and latent genes are expressed occurs; this period is called Prelatency or Latency 0. In Latency 0 the expression program allows the infection to be persistent but without expression of proteins that aid cell proliferation. This program is typically present when the infected B cells are part of the memory pool; however it is important to clarify that EBERs and BARTs are also being expressed in all subsequent latency programs (Price and Luftig 2015). BART miRNAs are a product of alternative splicing, which can be grouped according to the introns of BARTs, they are very highly expressed and are believed to serve several different purposes related to latent replication (Edwards et al. 2008; Pfeffer et al. 2004). EBER1 and EBER 2 are small RNAs that also expressed at high levels; they are known to form RNA-protein complexes with cellular proteins and although their function is still debated, they are believed to hold a central role in cell transformation (Glickman et al. 1988; Lerner et al. 1981; Toczyski and Steitz 1991)

In latency I, viral DNA replication is facilitated with EBNA1 by attaching the viral episome to host chromosomes and taking over the host cell's replication machinery. EBNA1 is transcribed from promoter Q (Qp), and it also activates transcription of other EBNAs from viral C promoter (Cp) in Latency III. EBNA1 decreases p53 tumour suppression activity, and activates nuclear factor κB (NF-κB) which migrates to the nucleus and activates lytic gene promoters (see lytic reactivation) (Mogensen and Paludan 2001; Price and Luftig 2014; Reisman and Sugden 1986; Yates et al. 1985).

Regarding proteins expressed in Latency II, LMP1 mimics B cell receptor CD40 in vivo hampering communication with T cells, it also acts as an activated tumour necrosis factor receptor and activates signalling pathways that stimulates oncogenesis (Dawson et al. 2012; Kaye et al. 1993; D. Wang et al. 1985). LMP2A is a B cell receptor mimic that helps the infected cell evade apoptosis (Caldwell et al. 1998). LMP2B is still being studied and its function has not been entirely

elucidated, in a study where expression of LMP2A and LMP2B were knocked down *in vitro*, a critical loss of efficient activation and proliferation of EBV infected B cells was shown when LMP2A was not being expressed. In contrast, LMP2B did not appear to have an effect in those processes (Wasil et al. 2013).

In latency III the EBNA transcripts are driven by viral promoters Cp or Wp. EBNA2 has been shown to affect transformation of B cells in vitro, activates the expression of LMP1 and LMP2, and in recent studies EBNA2 has been shown to drive the formation of new binding sites for cell factors RBP-jκ and EBF1, as well as regulating RUNX1 and RUNX3 through distal super enhancers (Gunnell et al. 2016; Hsieh et al. 1996; Lu et al. 2016; Rabson et al. 1982). EBNA3A and EBNA3C are essential for transformation *in vitro* of B cells, however, this proliferation fails in experiments where these EBNAs are rendered non-functional; adding to this, EBNA3B expression did not alter transformation when absent, however it may function as a tumour suppressor (Tomkinson and Kieff 1992a, 1992b; White et al. 2012). Expression of EBNA-LP with EBNA2 was shown to induce cellular cyclin D2, and it is required for the efficient outgrowth of LCLs; subsequent studies confirmed this coactivation (Harada and Kieff 1997; Sinclair et al. 1994; Young et al. 2016). BHRF1 miRNAs are known to help in cell proliferation as well as aiding in the evasion of apoptosis (Feederle et al. 2011; Seto et al. 2010).

Making the first association between EBV and human malignancies, EBV was first identified in cells derived from a patient in Africa with Burkitt's Lymphoma. The pathology was described by Denis Burkitt in 1958, and the virus was isolated and identified using electron microscopy by tumour virologist Michael Anthony Epstein and Yvonne Barr (Burkitt 1958; Epstein et al. 1964). Later, a link was established between the presence of antibodies for EBV and patients with infectious mononucleosis (IM) (G. Henle and Henle 1966); these antibodies were also found in 90% of tested individuals worldwide (de-The et al. 1975; G. Henle et al. 1969). Then, EBV was used to produce lymphoblastoid cell lines (LCLs), by causing immortalization of B lymphocytes in culture, establishing a new method of studying yet another important link between EBV and cell pathology *in vitro*. These were the first steps into the study of how EBV relates to disease (W. Henle et al. 1967; Nilsson et al. 1971).

1.1.3.2 EBV associated pathologies and latency

The sequence of events in EBV infection follow a general outline. First, an initial infection exposure occurs typically in epithelial cells of the oral mucosa, followed by lytic replication of virus and shedding to neighbouring epithelial cells and, to a lesser extent, neighbouring B cells in the oropharynx. This is followed by proliferation of infected B cells in lymphoid tissues of the pharynx. Then, there is a downregulation of the proliferation program in the infected B cells, and entry into the memory B cell pool, giving way into a lifelong persistence as a truly latent infection. Finally some occasional latently infected cells switch to lytic cycle, starting new foci of virus replication and an intermittent low-level virus shedding (Taylor et al. 2015).

There are two models to explain how EBV enters the memory B cell pool; the germinal centre model describes an initial infection of naive B cells that undergo latency III and proliferate, followed by entry to the germinal centre where expression defaults to latency II, subsequently the infected cells leave the germinal centre and enter the memory pool changing the program to latency 0, and ultimately in the switch from memory cell to plasma cell, virus is replicated and infection spreads; however, in the direct infection model, EBV infects B memory cells which supposes an intermediary latency III before going to latency 0 and true persistence (Thorley-Lawson et al. 2013; Young et al. 2016).

1.1.3.2.1 EBV malignancies related to B cell infection

Burkitt's Lymphoma

As mentioned before, BL was first described as a lymphosarcoma prevalent in children in Africa, and years later EBV was discovered in cultures of BL cells. Children whom tested positive for the presence of EBV exhibited a higher risk for developing tumours (reviewed by (Thorley-Lawson and Allday 2008)). Tumour cells exhibited latent infection and carried the viral genome as an episome, expressing a putative tumour antigen in their nuclei termed EBNA (EBV nuclear antigen). Shortly prior to this, it was discovered that LCLs could be established by infecting primary B cells with EBV. This supported the idea that EBV was the defining factor in BL tumorigenesis, however, a very important discovery was made; the patterns of viral gene expression were restricted in cells from BL

tumours (Latency I) compared to the expression in the LCLs (Latency III) generated with cells coming from B cells of the same patients. This is what allowed to conceptualize latency being subdivided into different states (reviewed by (Rowe et al. 2009; Thorley-Lawson and Allday 2008)).

In parallel, work by (Kohn et al. 1967), found abnormalities in the chromosomes of BL cells, and further research discovered that there was a translocation between chromosome 8 and 14, juxtaposing the MYC gene to one of the immunoglobulin G (IgG) enhancers. This region is highly active and increases the expression of MYC, thus driving the uncontrolled proliferation of BL cells. When considered jointly, the research showed the translocation of the gene MYC to be the common trait in the variants of BL (EBV positive vs EBV negative), which challenged the theory that EBV infection is the only factor leading to lymphoma (reviewed by (Boerma et al. 2009; Thorley-Lawson and Allday 2008)).

Three different clinical variants of BL are used for classification: endemic, sporadic, and immunodeficiency-related. Endemic BL is present mainly in Africa and is associated with endemic occurrence of malaria, and EBV infection; whereas, sporadic BL occurs outside of Africa and is typically free from EBV infection. Immunodeficiency-related BL also occurs outside Africa, but is approximately 60% EBV-positive (Abate et al. 2015).The gene expression program of EBV in BL cells follows latency I; expression of EBNA1 and Epstein-Barr virus-encoded small RNAs (EBERs).

Hodgkin's Lymphoma

Hodgkin's Lymphoma (HL) was first linked to EBV infection when Levine and collaborators found high levels of antibodies specific to EBV antigens in patients with HL; followed by experiments where DNA and RNA from the virus were detected in Hodgkin/Reed Stemberg (HRS) cells (Levine et al. 1971; Weiss et al. 1989; Wu et al. 1990).

HL can be classified into two types based on morphologic, immunophenotypic and clinical differences: classical HL (cHL) and nodular lymphocyte predominant HL (NLPHL). The disease is characterized by the presence of malignant cells called Hodgkin/Reed Stemberg cells, surrounded by non-neoplastic B and T cells (Reviewed (M. R. Chen 2011)). HRS cells are derived from B cells, however they

show a characteristic loss of a functional B cell receptor and are suspected to depend on acquiring anti-apoptotic functions from external sources. There's a high level of Latent membrane protein-1 (LMP1) and 2 (LMP2) expression in EBV-positive HRS cells; these two proteins are used by the virus to evade apoptosis, and drive persistence in B cell memory pool (Vockerodt et al. 2014). The pattern of latency expression in HL is latency II.

PostTransplant Lymphoproliferative Disease

Posttransplant lymphoproliferative disease (PT-LPD) is manifest after medical intervention, where high EBV loads can be detected in saliva, following a solid organ transplantation or hematopoietic stem cell transplantation that was complemented with immunosuppressive therapy. Most cases of early onset of PT-LPD show EBV-transformed LCL-like cells expressing all of the latent cycle antigens, and even some cells entering lytic cycle. This phenomena can be seen in 1 to 11% of transplants peaking 2-3 months after the event depending on the degree of immune suppression. However, patients carry risk of late onset of PT-LPD for life if they continue immunosuppressed. The risk for developing PT-LPD is increased in children that are EBV-negative receiving a transplant that is EBV-positive compared to children that have been previously exposed to the virus; since the viral load is large in the transplanted organ, and the immune system of EBV-negative children can't give a strong enough defence response. There is a correlation between loss of EBV-specific T cells and the development of PT-LPD, therefore new therapies involving receiving preparations of virus-specific cytotoxic T cells are arising and treating PT-LPD (Taylor et al. 2015).The pattern of latency expression in PT-LPD cells is type III, however some tumours display a Latency I/II pattern of viral expression (Rickinson 2014).

EBV-positive diffuse large B cell lymphoma of the elderly

Diffuse large B cell lymphoma (DLBCL) is a common malignant lymphoma, with a particular higher incidence in males of Asian countries, it presents as polymorphic or monomorphic B cell proliferation in patients without any known immunodeficiency or prior lymphoma. If in situ hybridization for EBER confirms the presence of EBV, the prognosis of overall survival is affected negatively. A common trait seen in EBV-positive DLBCL of the elderly patients is an activated

B cell immunophenotype as well as an greatly activated NFκB (Jaffe and Pittaluga 2011).

Although it has been mostly studied in patients older than 50 years old, it can also affect younger patients in a similar way. Some cases show morphological overlap with Hodgkin's Lymphoma. The EBV-positive tumours have a latency I or II pattern of expression, similar to what is observed in PT-LPD (H. J. Kim et al. 2017; T. X. Lu et al. 2015).

Infectious Mononucleosis

Since the symptoms of IM are so easily discernible, IM has been used to study the events occurring in the primary exposure to the virus. IM was first recognized as a disease in 1880 by Nil Filatov, a Russian paediatrician; however, the name "infectious mononucleosis" was coined in 1920 by Sprunt and Evans when noticing large numbers of atypical mononuclear cells in blood of patients exhibiting fever and pharyngitis (Balfour et al. 2015).

As mentioned before, EBV was linked to IM by the Henle group in 1969. The disease has a particular set of symptoms that last 2 to 6 weeks; these are: a sore throat, cervical lymphadenopathy, fever, fatigue and sometimes inflammation of the spleen. Primary EBV infection produce two or more of the symptoms in over 75% of cases (Luzuriaga and Sullivan 2010; Taylor et al. 2015). The IM infection, is characterized by activated Natural Killer (NK) cells, large increment in number of virus-specific CD8 T cells, responsive to latent antigens such as EBNA-2 and EBNA-3 (Latency III), and a smaller increment in number of virus-specific CD4T cells. This expansion of EBV-specific cytotoxic T cells destroy most infected cells. (Balfour et al. 2015; H. J. Kim et al. 2017; Taylor et al. 2015).

1.1.3.2.2 EBV malignancies in atypical cells

NK and T cell lymphoma

Lymphomas commonly develop in the nose and upper parts of the aerodigestive tract, however they can occur in the skin, testis, gastrointestinal tract, and muscle; if untreated, the malignancy eventually leads to death. The defining feature of NK/T cell lymphoma is the presence of episomal EBV DNA making the detection of EBV infection necessary for diagnosis, achieved by in situ hybridization for EBV early RNA (EBER) is used to diagnose NK/T cell lymphoma. The affected

cells can be identified as large granular lymphocytes regardless of the localisation; they can alter the functionality of the liver, spleen and bone marrow, where the lymphoma cells can be found as cellular infiltrate (reviewed by (Tse and Kwong 2017)).

Cases of NK/T cell lymphomas are geographically found in Asian and South American populations, being rare in other places. The lymphoma is locally invasive and it can destroy the facial structure and even manifest hard palate perforation. Treatment usually involves chemotherapy and radiotherapy, but new therapies are targeting the overexpression of the programmed death protein ligand 1 (PDL1) by NK/T cell lymphoma cells, which upregulates expression through the MAPK/NFκB pathway, and inhibits T cell activity aiding to the evasion of immunosurveillance by lymphoma cells (reviewed by (Tse and Kwong 2017)). The expression program of NK/T cell lymphoma is Latency I or II with abundance of EBER followed by BARTs, expression of LMP2A 2B and EBNA1 are comparatively lower.

Nasopharyngeal carcinoma

Nasopharyngeal carcinoma (NPC) has a very unique pattern of geographical distribution, with the vast majority of cases diagnosed in Southeast Asia, and southern China. There seems to be a predisposition to NPC in certain ethnic groups, and in concordance to environmental factors like exposure to carcinogenic substances. NPC can be classified as non-keratinising and keratinising. Non-keratinising NPC is the most common in regions where NPC is endemic, constituting over 95% of the cases, and commonly associated with EBV infection(Chua et al. 2016).

EBNAs are detected in cells coming from tumours but not in normal nasopharyngeal epithelium cells, suggesting that EBV is necessary for the pathogenesis of NPC. EBV latency expression in NPC cells seems to be an intermediary program between latency I and II with expression of LMP1, LMP2 and EBNA1, and BART miRNAs and EBER (Kutok and Wang 2006).

EBV associated Gastric carcinoma

EBV associated Gastric carcinoma (EBVaGC) forms tumours in the proximal stomach, these tumours manifest as multiple lesions not necessarily appearing at the same time. It is morphologically almost identical to EBV-negative gastric carcinoma, therefore, in situ hybridisation against EBER1 is necessary to identify EBVaGC. Another characteristic of EBVaGC is the presence of inflammatory infiltrate in the tumour, primarily lymphocytic, and with an abundance of B cells (Shinozaki-Ushiku et al. 2015).

Gastric epithelial cells lack the EBV receptor CD21 (CR2) which is necessary for an EBV particle to attach to the host cell before endocytosis. This lack of CD21 supports the idea that infection in epithelial cells happens through cell-to-cell contact, especially in inflammatory responses where B lymphocytes and gastric epithelial cells are involved, like gastritis (Shinozaki-Ushiku et al. 2015). EBVaGC shows Latency I or II, expressing EBERs, EBNA1, BARTS and BART miRNA, with some cases expressing LMP2A.

A summary of the latency stage on several EBV-associated malignancies are shown in Table 1.1.

1.1.3 EBV lytic stage

Lytic Reactivation

When lytic cycle is reactivated, EBV gene expression changes for the production of new viral particles and the infection to spread. The lytic gene expression pattern can be sub-classified into 3 subsequent lytic phases; immediate early (IE), early (E), and late (L). The two genes that are transcribed in the IE phase, BRLF1 and BZLF1 are distinguished by the capacity of their produced proteins, Rta and Zta, to reactivate lytic cycle (Chevallier-Greco et al. 1986; J. Countryman and Miller 1985). Zta and Rta transactivate other lytic genes and begin the switch to full lytic production, E genes are activated and viral proteins required for the replication of the EBV genome are expressed, then L genes are transcribed, producing structural proteins as well as capsid agents and membrane proteins, which allow for viral genome encapsidation and ultimately new viral particles; however, this reactivation cannot occur without Zta or Rta (continues in page 16)

Table 1.1. Summary of EBV-associated malignancies and viral latency. Modified from (Taylor et al. 2015).

Tumor	Subtype	EBV proteins expressed	Latency
Burkitt's lymphoma	Endemic	EBNA1	I
Hodgkin's lymphoma	Classical	EBNA1, LMP1, and LMP2A	II
Lymphoproliferative disease	Posttransplant	EBNA1, EBNA2, EBNA3A, EBNA3B, EBNA3C, EBNA-LP, LMP1, and LMP2A	III
Diffuse large B cell lymphoma	Elderly DLBCL	EBNA1, LMP1, and LMP2A	I/II
T/NK cell lymphoma	Extranodal	EBNA1 and LMP2B	I/II
Nasopharyngeal carcinoma	Undifferentiated	EBNA1, LMP1, and LMP2A	I/II
Gastric carcinoma	EBV associated	EBNA1 and LMP2A	I/II

and they are both required for full expression of EBV proteins expressed in lytic cycle (Feederle et al. 2000).

Differentiation of infected memory B cells into plasma cells promotes Lytic cycle reactivation; plasma cell associated transcription factors XBP1 and BLIMP1 activate the promoter of BZLF1 (Zp) and BRLF1 (Rp). This also happens when the B cell receptor (BCR) is stimulated and cellular signalling pathways activate; for instance, with Protein kinase C (PKC) pathway active, transcription factors NF-κB and AP1 (activating protein 1) activate, which in turn bind to Zp and Rp and they in turn activate early genes (Goswami et al. 2012; Laichalk and Thorley-Lawson 2005; Shaffer et al. 2002; Thorley-Lawson 2015).

A similar activation caused by differentiation happens in epithelial cells, with BLIMP1 in conjunction with KLF4 coactivating Zp and Rp (Murata et al. 2013). Lytic reactivation can also be stimulated with the use of external agents: 12-*O*-tetradecanolyphorbol-13-acetate (TPA) increases the concentration of calcium ions, which triggers PKC pathway activating NF-κB and AP1 which as mentioned above activates Zp and Rp; sodium butyrate inhibits histone deacetylase and therefore facilitates transcription and activation; and anti-Ig which mimics BCR stimulation and activates lytic cycle (Gao et al. 2001; H. Li et al. 2016a; Murata 2014; Takada and Ono 1989; Westphal et al. 2000).

Zta and Rta activate several different early genes encoding proteins required for replication, for instance BALF5 for DNA polymerase, BALF2 a single stranded DNA binding protein homologue, BMRF1 the DNA polymerase processivity factor, BSLF1 the primase homologue, BBLF4 an helicase homologue, BBLF2/3 formed from open reading frames of BBLF2 and BBLF3, which are suspected to function as a potential homologue of the third component of the helicase-primase complex (Fixman et al. 1992; Young et al. 2016).

1.1.3 Zta

As mentioned previously, one of the major players in lytic switch activation is gene BZLF1; several different methods of lytic reactivation involve the binding of cellular activation factors binding to its promotor Zp in order to produce Zta. The expression of Zta happens in the first 30 minutes after lytic activation and in turn initiates the signalling needed for expression of E and L genes. (Chevallier-Greco

et al. 1986; J. Countryman and Miller 1985; H. Li et al. 2016a; Sinclair 2003; Takada et al. 1986).

Histone modifications have an interesting role in regulating activation of Zp and Rp. Histones H3 and H4, around Zp and Rp, are acetylated in distinct Lysine residues; H3K27ac, H3K9ac, and H4K8ac help establish an open chromatin configuration and allow access of transcription factors, like Zta which can auto-regulate its own promoter Zp. In contrast, methylation modifications such as H3K27me3, H3K9me2/me3, and H4K20me3 correlate with inactivation of the IE promoters, maintaining latency; however H3K4me3 allows the virus to express Zta (J. K. Countryman et al. 2008; E. Flemington and Speck 1990a; Jenkins et al. 2000; H. Li et al. 2016a).

Structure and DNA binding

Zta is composed of 245 amino acids with a total molecular weight of 35 KDa and 4 distinct regions; an N-terminal transactivation domain, a basic DNA binding region, a dimerization “zipper” region, and a C-terminal region. Basic-leucine zipper (bZIP) proteins are characterized to have a basic DNA contact region and a dimerization “zipper” region which are present in Zta, giving Zta homology to AP1 transcription factors c-fos and c-jun. (Chang et al. 1990; Farrell et al. 1989; Flemington et al. 1992; Marschall et al. 1989; Seibl et al. 1986).

The N-terminal correspond to amino acids 1 to 167 of Zta, it encompasses the transactivation domain (aa 27 to 78), where the pre initiation complexes bind to as well as other factors that form complexes to activate transcription (Chi and Carey 1993; Lieberman et al. 1990)

The DNA binding domain of Zta corresponds to amino acids 175 to 196; it is necessary for the activation of lytic genes. If serine 186 is mutated, binding to Rp is altered, as well as the transcription activation of late genes through binding of Zta response elements (ZREs); in contrast, a substitution of cysteine 189 for serine maintained recognition to Rp but the transcription of late genes is affected. (Baumann et al. 1998; Bhende et al. 2005; P. Wang et al. 2005).

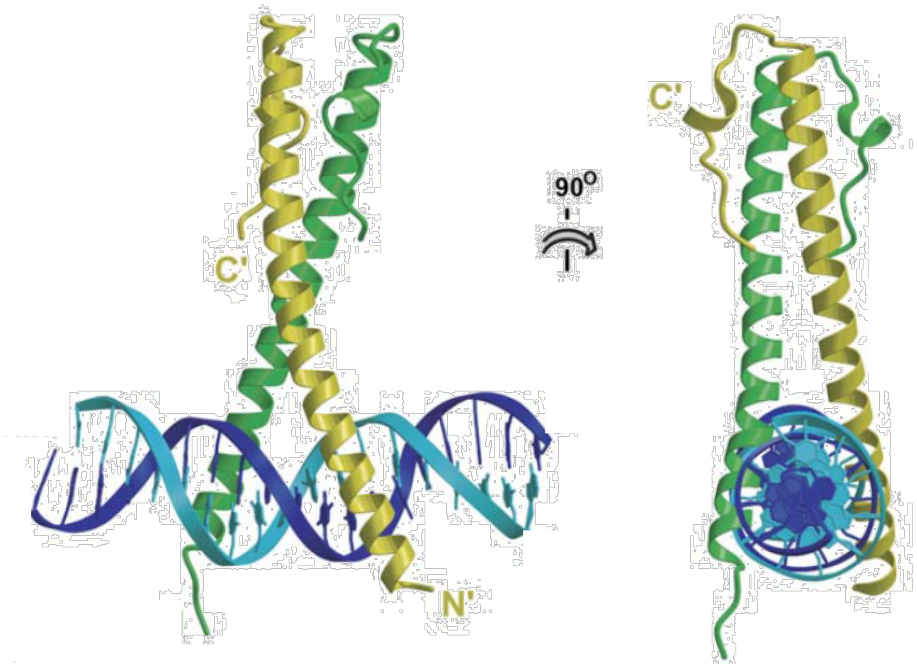


Fig 1.3 Zta dimerization and binding to DNA. Only the DNA binding region of Zta is displayed dimerised in green and in gold, the rest of Zta is not shown. The C terminal coil folding back to the “zipper” region can be seen on top, and the amino N’ terminal in the bottom. Adapted from (Petosa et al 2006).

The leucine “zipper” dimerization domain correspond to amino acids 197 to 221. As stated before, Zta has a distinct non-leucine coiled-coil that is different from the typical leucine “zipper”; the C-terminal region of Zta is contiguous to the “zipper” region, folding the structure back towards the “zipper” region, giving stability to the homodimer (See Image 1.3). (McDonald et al. 2009; Petosa et al. 2006; Sinclair 2006). This C terminal region is necessary for transcriptional activation, and is necessary for the interaction with DNA; however the formation of a homodimer is independent from the interaction with DNA (Hicks et al. 2001; Hicks et al. 2003).

As mentioned before, Zta is a homolog of AP1 and binds to AP1 sites as well as sites that resemble AP1 binding motif, these 7bp sequences are called Zta response elements (ZREs) (Farrell et al. 1989; Urier et al. 1989). There are at least 32 variants of ZREs and can be classified into classes I, II and III; in class I there is no CpG motif, but there are CpG motifs in class II and III. In classes I and II there is no negative change on the binding of Zta when there is no methylation, but in class III the binding of Zta is negated when there is no methylation, and

finally, the effect on the binding of Zta to methylated ZREs is enhanced in class II (Flower et al. 2011; Karlsson et al. 2008).

1.3 Computational methods for sequence analysis

Sequence analysis has been used before to predict and identify variants of Zta response elements. Parting from well characterized Zta-responsive promoters from the EBV genome (Zp, Rp, BMRF1 promoter) an algorithm allowed for the generation of a position frequency matrix, followed by a viral genome-wide analysis and ultimately the identification of 32 variants of ZREs (Flower et al. 2011).

1.4 Zta interacting with the human genome

Besides regulating the expression of cellular genes such as FOS, E2F1, EGR1, IL8, IL10 and IL13, Zta can inhibit the expression of core genes of interferon alpha receptor as well as impair the function of the interferon regulatory factor 7 (IRF7) whose role is to augment the production of type I interferons; Zta can also regulate cellular cytokine expression levels and inhibit the expression of MHC all of these mechanisms to alter or evade the natural defence response from the infected host (Bailey et al. 2009; E. Flemington and Speck 1990b; Heather et al. 2009; Hsu et al. 2008; Kenney 2007; Morrison et al. 2004)

The Zta interactions with the human genome were recently explored using chromatin immunoprecipitation coupled to next-generation DNA sequencing. This allowed to find Zta binding in proximity (≤ 2 kb) to the TSS of 278 regulated genes as well as Zta binding distally (≥ 4 kb) relative to the TSS of regulated genes. Also, the promoter sequence from upregulated genes FOSB and RASA3 showed enhancer activity when placed distally from a Luciferase Reporter gene (Ramasubramanyan et al. 2015).

1.5 Enhancers

Enhancers are defined as cis-acting DNA sequences that can increase the transcription of genes, independently of orientation and distance from target promoter or promoters (Pennacchio et al. 2013). Identifying enhancers is challenging, they are scattered through the whole genome and their location relative to their target gene or genes is variable, upstream, downstream, or within introns. They also don't necessarily act on the closest promoter but can bypass

neighbouring genes to regulate distant genes. Sometimes a single enhancer regulates multiple genes (Mohrs et al. 2001; Pennacchio et al. 2013; Visel et al. 2009)

There is no consensus sequence found across all different enhancers, and even more, the enhancer activity might be restricted to a particular tissue or cell type, time point, or other particular conditions. Chromatin immunoprecipitation coupled to high throughput sequencing (ChIP sequencing) has helped define enhancers, by finding that enhancers have regions that are heavily enriched in closely spaced recognition motifs for sequence specific transcription factors. (Heinz et al. 2015; Koch et al. 2011; Pennacchio et al. 2013).

Most enhancers are active in a cell type-specific manner, predominantly in regions of accessible chromatin, primed by lineage-determining transcription factors. This event is followed by the binding of signal-dependent transcription factors to the primed enhancers; therefore, enabling the enhancer cell type-specific activity. The binding of transcription factors to enhancers can be aided with the binding of co-regulators, whose recruitment and transcription are accompanied by methylation or acetylation of histone tails; epigenetic marks like acetylation of lysine 27 in Histone H3 (H3K27ac) and methylation of lysine 4 (H3K4me1) which have been useful in finding enhancers in mammals (Heinz et al. 2015).

Depending on the activity found in regions corresponding to enhancers, they can be classified as inactive, primed, poised or active. Inactive enhancers are found in tightly compacted chromatin and without any bound transcription factors or histone modifications. Primed enhancers do have sequence-specific transcription factors bound to it, in an open chromatin conformation, but are still to be activated due to for instance a missing co-activator. Poised enhancers can be thought of as primed enhancers that also contain repressive epigenetic chromatin marks such as H3K27me3. Active enhancers are in an open chromatin conformation, with bound sequence-specific transcription factors activating them, with epigenetic markers of active histone (H3K27ac) with recruited RNA polymerase II enhancing the transcription of the target gene (Heinz et al. 2015; Shlyueva et al. 2014).

The first link between virus and an enhanced gene transcription was in 1981 when two different labs found that the expression of reporter genes raised several hundred times over when a particular 72 bp sequence found upstream to the TSS of the simian virus SV40 gene, was present. This initial research served as a take-off point for subsequent investigation that focused on finding key components in DNA sequence that could serve to predict the presence of enhancers. Computational scanning of long DNA sequences and even genomes, have been used to find the motifs to which transcription factors bind after enrichment, however these studies can be limited by the contextual conditions of the study. Still, this approach is currently used and has found similarities between the sequence found in viral enhancers and host enhancers, supporting the notion that evolutionary convergence allowed virus to exploit and mimic the host enhancers as well as transcription factors (Arnone and Davidson 1997; Banerji et al. 1981; Kropp et al. 2014; Moreau et al. 1981; Yanez-Cuna et al. 2012).

Due to the chromatin structure needed for an enhancer to exert its function over a target gene, following research focused on finding an open and accessible conformation that would allow binding of transcription factors. By crosslinking with formaldehyde and then employing DNase I coupled to deep sequencing, researchers have been able to find footprints of transcription factors that protect a stretch of DNA susceptible to DNase excision, in other words, open chromatin in the process of recruiting factors to initiate transcription, an indication of an enhancer sequence. However, the information that can be gathered does not focus exclusively on enhancers, since not all enhancers are in a state of open chromatin at all time, nor all stretches of DNA ready to undergo transcription correspond to enhancers (Giresi et al. 2007; Neph et al. 2012; Shlyueva et al. 2014).

Research that mapped epigenetic marks, found histone modifications on nucleosomes surrounding active and inactive promoters and enhancers. This allowed to find that active promoters presented H3K4me1 and H3K27ac, while active enhancers showed H3K4me1 and H3K27ac, whereas silent regions showed H3K9me3. (Heintzman et al. 2007; Kouzarides 2007; C. L. Liu et al. 2005). Eventually a generalized custom developed of considering epigenetic marks for the prediction of enhancers, however there is no general agreement on

which epigenetic marks signal the presence of an enhancer; this is due to the variation of different marks found on different enhancers and on different data sets used to study such enhancers. It is also important to know that there is no histone modification or combination of modifications that flawlessly mark an enhancer; some active enhancers don't even present epigenetic marks (Arnold et al. 2013; Bonn et al. 2012; Rada-Iglesias et al. 2011; Shlyueva et al. 2014).

Other research has focused on a different method of predicting enhancers, by instead using ChIP-sequencing mapping the binding of transcription factors, co-activators or other chromatin regulators like Pol II, Cohesin, or Mediator (Kagey et al. 2010; G. L. Li et al. 2012). The reason why these can work as an indicator of the presence of an enhancer is that the sequence of the enhancer undergoes transcription of non coding RNA, termed enhancer RNA (eRNA) (De Santa et al. 2010; T. K. Kim et al. 2010).

Although it is disputed whether the eRNAs have a particular function in the enhancing effect or they are just a temporal consequence of the proximity of Pol II to an active enhancer; currently, the detected transcription of eRNA has expanded the previous methods of finding potential enhancers besides locating histone modifications near the putative enhancer regions. For instance, recent research has postulated the possibility of altering the RNA surveillance mechanisms and therefore extending the period of detection of eRNA to reveal functioning enhancers without epigenetic marks (Andersson et al. 2014; W. Li et al. 2016b; Lubas et al. 2015; Shlyueva et al. 2014).

Regions in the genome with a high density of enhancers bound by an array of transcription factors to drive transcription, like Mediator, are called super-enhancers. These regions also present eRNA transcription, histone acetylation and other epigenetic regulation markers, and are typically cell type specific as well. Enhancers within a super-enhancer form 3D interactions to other enhancers within the same super-enhancer region, which is feature of the folded genome in the nucleus. The majority of these super-enhancers are formed with genes that are enriched for the biological processes that define the identities of the cell types (Dowen et al. 2014; Hah et al. 2015; Heinz et al. 2015; Hnisz et al. 2013; Z. J. Liu et al. 2014).

1.6 Chromatin looping events

A major challenge in deciphering cell type-specific enhancer functions is connecting active enhancers to their target genes in vivo, however Chromosome conformation capture (3C) has helped show evidence of the formation of loops between enhancers and distal promoters in vitro (Heinz et al. 2015; Pennacchio et al. 2013).

An interesting instance of expression control through 3D rearrangement of chromatin (looping) happens in the Hox gene cluster, which contribute to the body plan during vertebrate development. The Hox genes are found linearly, in a way that reflects the temporal and sequential expression of the genes, however as development of expression pattern advances, the Hox genes re-arrange spatially, in such a way that a cluster of Hox genes that are actively being expressed forms an “active array”. This “active array” gradually integrates clusters of previously inactive Hox genes, and as they spatially incorporate into the “active array”, they become active themselves. This re-arrangement is accompanied by a change of active histone epigenetic markers (Denker and de Laat 2016; Montavon et al. 2011; Noordermeer et al. 2011).

In the EBV genome, three-dimensional rearrangement has been documented in Latency I when OriP forms a loop with Qp mediated in part by EBNA1, or in Latency III when OriP forms a loop with cohesin to the LMP1 promoter (Arvey et al. 2012; Avolio-Hunter and Frappier 1998; Lieberman 2013; Tempera et al. 2011). Similar events are also seen in the host cell genome, EBV has been known to upregulate MYC, driving cell proliferation and counteracting apoptosis by silencing BCL211 (bim). Interesting recent research showed that this is achieved by EBNA2 binding to enhancers of MYC, which induces a structural reorganisation of the MYC promoter to enhancer loops, which upregulates tumorigenesis. Part of their study also explored how EBNA3A and EBNA3C repress BCL2L11 by inactivating a long range enhancer hub recruiting H3K27 methyltransferase EZH2 and silencing promoter-enhancer looping interactions. As mentioned previously, epigenetic regulation markers like acetylation or methylation of H3K27 was an important marker for the identification of the interactions they set to explore (McClellan et al. 2012; Wood et al. 2016).

1.7 Aims of project

Viral transcription factor Zta is known to regulate cellular expression of genes. A previously performed genome wide ChIP-seq analysis in BL Akata cells showed a great proportion of the binding sites of Zta distal (≥ 4 kb) to promoter regions of cellular genes, and a smaller number of binding sites proximal (≤ 2 kb) to TSS of genes (Ramasubramanyan et al. 2015). The mapped binding sites could have enhancer properties, and perhaps when binding sites are distal, Zta regulation is occurring through chromatin looping. In order to research this the aims are:

- To develop a method to test if Zta is binding to sequences with enhancing properties. Potential sequences that could be conferring enhancing activity will be identified. Components in the sequences that could work as enhancers will be analysed. The sequences will be inserted into vectors and its enhancing capabilities will be tested through Luciferase Assays in EBV negative, epithelial and BL cells.
- To develop a method to discriminate possible Zta interactions that would be good candidates for looping activity. Relating genes highly regulated by Zta a selection criteria will aid to select which interactions to further research.
- To design a 3C assay that would capture the interactions between distal Zta interactions and promoter regions that are candidates to be looping in EBV positive inducible LCL and BL cells.

Chapter 2. Materials and Methods

2.1 Materials

2.1.1 Vectors

Table 2.1. List of Vectors used.

Vector Name	Design	Backbone plasmid	Application
pcDNA3	Invitrogen	pcDNA3	Control for Zta expression
His-Zta	Sinclair Lab (Bailey et al. 2009)	pcDNA3	Zta expression
CIITA (CIITA -214+52)	Sinclair Lab Nicolae Balan (Balan et al. 2016)	pGL3-Control	Zta enhancer element analysis
BHLF1	Sinclair Lab (Ramasubramanyan et al. 2015)	pCpGL-basic	Zta enhancer element analysis
BHLF1-Mut	Sinclair Lab (Ramasubramanyan et al. 2015)	pCpGL-basic	Zta enhancer element analysis

2.1.2 Cell lines

Table 2.2. List of Cell lines used.

Cell line Name	Description	Reference	Application
DG75	EBV- Burkitt's Lymphoma cells	(Ben-Bassat et al. 1976)	Zta enhancer analysis
293T	EBV- embryonic renal cells	(DuBridge et al. 1987)	
LCL#3	LCL, EBV+ Spontaneously lytic 6 to 20%	(Sinclair et al. 1994)	Chromatin conformation capture

GM2188	LCL, EBV+, Tightly latent	(Stiff et al. 2005)	
Akata	BL, EBV+, anti-IgG inducible	(Takada and Ono 1989)	
Akata-Zta	BL, EBV+, BZLF1 expression vector, Doxycycline inducible	(Ramasubramanyan et al. 2015)	

2.1.3 Antibodies

Table 2.3. List of Antibodies used.

Antibody	Description	Supplier	Dilution	Application
Zta antibody (BZ1)	Zta primary mouse monoclonal	Martin Rowe Lab, Birmingham. Described in (Young et al. 1991)	1:500	Western Blot (WB)
β-Actin	Rabbit polyclonal	Sigma (A2066)	1:5000	WB
IRDye 680CW	Anti-mouse secondary fluorescent Ab	LI-COR	1:5000	WB
IRDye 800CW	Anti-rabbit secondary fluorescent Ab	LI-COR	1:5000	WB
IgG	Rabbit polyclonal	Dako (A0423)	1:800	Akata cell line Induction

2.1.4 Primers (oligonucleotides)

Table 2.4. List of primers used.

Primer Name	Sequence 5'-3'	Designed for 3C amplification of fragment containing
1	GTCCCAAACCATCCGTCTTCA	LHX1 promoter
2	CTTGGACAGGCTCCATCCAG	peak near LHX1 1
3	TGGTCTCTCCTCCTCCAATCTT	peak near LHX1 2
4	GACTTCCAAGCCCTTGCTCA	peak near LHX1 3
5	GCAGAAGAGGGCAAGGGATTT	LHX1 negative control in loop
6	CCAGTTTCCAAACAAACCGGTAA	LHX1 negative control out of loop
7	GGAACCCTGGTTCTTTCCCA	SLC6A7 promoter
8	TGGCAGTAAGCCTGCATAACC	peak near SLC6A7 1
9	GTCCACCTGCCCATGTAACC	peak near SLC6A7 2
10	TACCTGGCTTGATGAGCAGAA	SLC6A7 negative control in loop
11	AAAGAGGGAGCAAAGGAATGC	SLC6A7 control out of loop
12	GCTGTCTTTCCCTCAGCTCTTT	FZD10 Promoter
13	CTCTACGTCCCCTGCTTTCTC	peak near FZD10 1
14	ACAGTTGGTCCAAGCAGTAG	peak near FZD10 3

15	AACGCCATCAAGGGAATGGG	peak near FZD10 4
16	TACAACCAGATGACGCCCTG	FZD10 control out of loop
17	AGCCCAACCTTTATTTGGCTG	FZD10 negative control in loop 5’
18	GCAAGCATCCTTTCTTTTCCC	FZD10 negative control in loop 3’
C1	AACATTCCTTGGAAGTACATGAGC	Peak 898 near cluster VWF CD9
C2	AAACTTACAAGAATCCACGCAGT	Peak 899 near cluster VWF CD9
C3	GGAAGGATGATGGACGGAAGATT	cluster VWF CD9 negative control in loop
C4	GACTGCTGATGGAAGTGAGTGG	cluster VWF CD9 negative control out of loop
C5	GGAATTAAGTCTTTTTTTAGAATA	VWF promoter large
C6	TCCTTAGGGCAAAAGAGGGAAAG	CD9 promoter
C7	cTAGAATTTTCATCCTTTGCAC	VWF promoter short

2.1.5 Solutions and buffers

Table 2.5. List of solutions and buffers used.

Name	Composition	Purpose
Tris-acetate-EDTA (TAE)	40mM Tris (pH 7.6), 20mM acetic acid, 1mM EDTA	Agarose gels, electrophoresis.
Transfer Buffer	25 mM Tris-HCl (pH8.3), 192 mM, Glycine, 20% v/v methanol	Western Blot

PBS tween	138 mM NaCl, 2.7mM KCl, 10mM NA_2HPO_4	Western Blot
Ponceau Stain	0.1% w/v Ponceau in 5% v/v acetic acid	Nitrocellulose membrane staining
Lysis buffer	10mM TRIS-HCl pH 7.7; 10mM NaCl; 5mM MgCl2; 0.1mM EGTA; 1X complete protease inhibitor	Chromosome Conformation Capture
Ligation Buffer	660mM TRIS-HCl, pH 7.5; 50mM DTT; 50mM MgCl2; 10mM ATP	Chromosome Conformation Capture

2.1.6 Kits and reagents

Table 2.6 List of kits and reagents used.

Kit/reagent	Supplier	Purpose
CutSmart buffer	NEB	Restriction digest
SalHI	NEB	Restriction digest
HindII	NEB	Restriction digest
BamHI	NEB	Restriction digest
EcoRI	NEB	Restriction digest
PstI	NEB	Restriction digest
rAPid Alkaline phosphatase	Roche	Digestion treatment
PCR extraction kit	Qiagen	DNA products clean up
T4 DNA ligase	Invitrogen	Sub-cloning
5x T4 ligase buffer	Invitrogen	Sub-cloning
Ampicillin	Sigma	Bacterial cell selection
Zeocin	Invitrogen	Bacterial cell selection

Top10 competent cells	Invitrogen	Plasmid DNA transformation
PIR competent cells	Invitrogen	Plasmid DNA transformation
Luria Broth	LifeSci school service department	Bacterial cultures
Miniprep	Qiagen	Small scale plasmid DNA extraction and purification
Midi/Maxiprep	Qiagen	Large scale plasmid DNA extraction and purification
RPMI 1640 medium	GIBCO, Invitrogen	Cell culture
DMEM medium	GIBCO, Invitrogen	Cell culture
DPBS	GIBCO, Invitrogen	Cell culture
Foetal Calf serum (FCS)	GIBCO, Invitrogen	Cell culture supplementation
Penicilin, Streptomicin, L-Glutamine (PSG)	GIBCO, Invitrogen	Cell culture supplementation
Effectene transfection kit	Qiagen	Cell transfection
2x protein sample Laemmli buffer	Sigma	Protein Gel electrophoresis
MOPS SDS buffer	NuPAGE	Protein Gel electrophoresis
NuPAGE Bis-Tris SDS-PAGE 12%	Invitrogen	Protein Gel electrophoresis
Odyssey blocking buffer	LI-COR	WB
SeeBlue Pre-stained protein marker	Invitrogen	WB
Reporter Lysis Buffer	Promega	Luciferase Reporter Assay

Luciferase Reagent	Promega	Luciferase Reporter Assay
Bicinchoninic Acid Assay	Pierce ThermoFisher	Protein measurement
T4 Ligase	NEB	3C
T4 Ligation buffer	NEB	3C
RNase	Sigma	3C
Proteinase K	Sigma	3C
Phenol Chloroform	Sigma	3C
Ethanol	Sigma	3C
Sodium acetate	AnalaR Normapur	3C
Phusion high fidelity polymerase	NEB	PCR
Agarose	Fisher bioreagents	DNA electrophoresis
GelRed	Biotium	DNA electrophoresis
Hyperladder	Bioline	DNA electrophoresis

2.2 Methods

2.2.1 Molecular cloning by digestion-ligation

Designed elements were synthetically produced through Invitrogen's GeneStrings service. The sequence of the elements included restriction sites present in the plasmids aimed to be inserted at. In this way, when ligated after being mixed and digested, the fragments would be introduced into the targeted plasmids. The synthetic fragments were first reconstituted in nuclease-free de-ionized water; then, the concentration was determined with the absorbance measure of UV light using a NanoDrop spectrophotometer (Invitrogen) for the designed elements as well as the employed plasmids. Digestions were prepared in a 30 µL volume, the components for each digestion consisted of 3 µL of 10X restriction enzyme buffer, 1 µL of restriction enzyme (10 Units), 250 ng of synthetic DNA or 1 µg of plasmid DNA, and enough nuclease-free de-ionized

water to reach 30 µL. The reaction was incubated for 1 hour at 37°C. Then, the temperature was increased to 65°C and maintained for 20 minutes in order to deactivate the restriction enzymes. The digested destination vector DNA was treated with 10 Units of rAPid alakine phosphatase (Roche) for 10 min at 37°C to reduce self-ligation; then the phostatase enzyme was inactivated by increasing and maintaining the temperature to 75°C for 2 minutes. The products were purified using a PCR extraction kit from (Qiagen).

A ligation reaction was set with a 1:3 (vector : Insert) ratio, typically 50 ng of vector DNA were used and 1 µL of T4 DNA ligase (Invitrogen) was added. The amount of insert DNA was calculated in the following way:

$$\frac{ng\ of\ vector * Kb\ size\ of\ insert}{Kb\ size\ of\ vector} * \frac{molar\ ratio\ vector}{molar\ ratio\ insert}$$

Enough water was added for a final volume of 15 µL. The ligation samples were incubated in water with ice overnight. This was typically followed by a subsequent bacterial transformation.

2.2.2 Bacterial Transformation

Two strains of *E. coli* were used for the transformation of plasmid DNA containing the designed inserts; TOP10 *E. coli* cells were used for vectors PGL3 and pcDNA3 vectors, while PIR *E. coli* were used for constructs inserted in the pCpGL plasmid. A small concentration of plasmid DNA (<10 ng) was mixed with 50 µL of competent *E. coli* cells. The antibiotic resistance cassettes in the vectors allowed for the use of 100 µg/mL of ampicillin for selection of PGL3 and pcDNA3 vectors, whereas 25 µg/mL of zeocin was employed for selection of colonies with pCpGL vector. Competent cells were mixed with plasmid DNA, left on ice and then heatshocked at 42 °C for 45 seconds, immediately placing them back on ice for 2 minutes. Sterile Luria broth (LB) (1 mL) was added to foster the recovery heat-shocked competent cells, after which the sample was shaken in an orbital shaker at 225 rpm and 37 °C for 1 hour.

After this, 10% of the mixture was spread in an agar plate containing the appropriate antibiotic for colony selection. The agar plates were then cultured at 37 °C for 16 hours, after this, plates with colonies were stored at 4°C. Different colonies from a single agar plate were picked to inoculate different tubes with

5mL of LB culture supplemented with the appropriate antibiotic. These broth samples were kept in agitation at 37°C for 16 hours to increase the yield of transformed *E.coli.* If long term storage of the bacterial stocks was needed, 800 µL of verified starter culture was mixed with 200 µL of sterile glycerol and stored at -80°C.

To check that the intended DNA was present in the cultured bacterial cells, the plasmid DNA was extracted (see Plasmid DNA extraction), digested again with the appropriate restriction enzimes, and then visualised in an agarose gel (see Agarose Gel Electrophoresis). Bands of the appropriate molecular weight were excised from the gel and sequenced using the Eurofins service (MWG Eurofins, UK).

2.2.3 Plasmid DNA extraction

When aiming to obtain small amounts of plasmid DNA, the Mini-prep plasmid extraction kit (Qiagen) was used. Bacterial cells from 3 mL of LB broth starter cultures were pelleted, lysed and centrifuged in a DNA binding silica column, as directed in the protocol from the kit. For higher yields of plasmid DNA, Midi-prep or Maxi-prep kits (Qiagen) were used, 500 mL of LB broth containing the selected antibiotic were used to culture cells, were incubated overnight at 37°C with agitation (225 rpm), the rest of the DNA extraction was executed following the protocol from the kit.

2.2.4 Cell Culture

Cells employed for the different experiments were kept at 37°C in aseptic conditions in a humidified incubator with 5% CO2 gas distribution. Counting of cells was performed using a Neubauer chamber, mixing 10 µL of cells in media with 10 µL of Trypan blue.

Some experiments employed the EBV-negative Burkitt's Lymphoma DG75 cell line (Ben-Bassat et al. 1976), the cells were cultured in suspension, maintained at a concentration between 3×10^5 and 3×10^6 cells/mL, fed with RPMI 1640 medium (GIBCO, Invitrogen) supplemented with 10% (v/v) fetal calf serum and 100 U/mL of penicillin, 100 µg/mL of streptomycin and 2 mM of L-glutamine (GIBCO, Invitrogen). The lymphoblastoid cell line 3, referred as LCL#3, is a B-

cell transformed with B95.8 EBV strain (Sinclair et al. 1994); cell line GM2188 is a Lymphoblastoid EBV positive and tightly latent cell line. These cells were cultured in the same way as described for DG75 cells.

The Akata cell line is an EBV positive, Burkitt's Lymphoma derived cell line; inducible to lytic cycle after treatment with IgG (Takada and Ono 1989). It served as the base for the generation of the Akata-Zta cell line, which was designed to facilitate the study of cells in lytic cycle, since it contains a doxycycline inducible Zta expression vector (Ramasubramanyan et al. 2015). These two cell lines were also cultured with RPMI medium supplemented with fetal bovine serum, penicillin and streptomycin and L-glutamine, as described with the DG75 cell line. The Akata cells were harvested 48 hours after induction with IgG, and Akata-Zta cells were harvested 24 hours after induction with Doxycycline, a kind gift from Anja K. Godfrey.

In some experiments, adherent 293T cells were employed, these are EBV-negative, embryonic kidney, epithelial cells (DuBridge et al. 1987; Pear et al. 1993). The cells were cultured in 75 cm^2 flasks, were fed with DMEM medium supplemented with 10% (v/v) fetal calf serum and 100 U/mL of penicillin, 100 µg/mL of streptomycin and 2 mM of L-glutamine (GIBCO, Invitrogen). Since they are adherent, they were rinsed and lifted with warm D-PBS, and subcultured with a ratio of 1:3 approximately every other day, or after confluency.

When cells were harvested, they were counted and an appropriate volume reflecting the number needed for each experiment, was spun down 5 min at 1300 rpm. The media was discarded, and the cell pellet washed with warm D-PBS. A new centrifugation ensured that the cell pellet was formed and the D-PBS was discarded. Cell pellets were immediately used or flash frozen in liquid nitrogen for later use, kept at -80°C.

2.2.5 Cell Transfection

Transfection of plasmid DNA was performed through electroporation for suspension DG75 cells. The cells were split 1:3 the day before transfection to ensure viability. For each transfection 1×10^7 cells in 250 µL of warm serum free medium were placed in a disposable 4mm polycarbonate cuvette; with 10 µg of total plasmid DNA (5 µg of either pcDNA3 (control) DNA or 5 µg of Zta expression

vector (HisZta-pcDNA3) + 5 µg of luciferase reporter vector). The electroporation was executed using a BIORAD Genepulser II using 250 volts with 975 µF of capacitance.

Transfection of 293T cells was executed using the Effectene transfection kit (Qiagen). The protocol was followed as instructed by the kit, except changing the amount of Effectene reagent to 2.5 µL per each µg of DNA. Each transfection started by seeding 4×10^5 cells the day previous to the transfection. A maximum of 1 µg of plasmid DNA was added to each transfection (0.5 µg of pcDNA or 0.5 µg of Zta expression vector + 0.5 µg of luciferase reporter vector). Cells were harvested 48 hours after transfection, scraped and washed with warm D-PBS.

2.2.6 Protein gels and Western Blots

After mixing with vortex in 250 µL of 2x Laemmli protein buffer, cell pellets were boiled for 5 minutes at 125°C, then they were spun down for 1 min at 13000 rpm. A volume of 10 µL was loaded per well for each sample, and a control sample of 5 µL of SEE blue marker (Invitrogen), into commercially available pre-casted Bis-Tris SDS-PAGE gels to separate proteins and conduct western blots (NuPAGE, Invitrogen). Gel electrophoresis was performed at 200 V for 50 min after covering the inner chamber of the tank with MOPS SDS buffer (NuPAGE).

In order to transfer the proteins, after the electrophoresis, the gels were placed in contact to a nitrocellulose membrane and sandwiched between filter papers and sponges. The sandwich was then placed in a transfer cassette which was inserted in the transfer chamber of the tank, preceded by being filled with transfer buffer. Current was applied to migrate the proteins and transfer them to the nitrocellulose membrane (75V 90min).

To check if the transfer was successful, after disassembling the sandwich, 50 mL of Ponceau S stain was added to the nitrocellulose membrane. After 2 minutes of rocking, the Ponceau stain was removed and washed with PBS-Tween buffer.

After the transfer was completed, the nitrocellulose membrane was blocked for 45 min with 5 mL of blocking buffer (LI-COR) and 5 mL of PBS-Tween buffer with a slow rocking motion. The blocking buffer was discarded and primary antibody was diluted in new blocking buffer, added to the membrane and incubated at 4°C

overnight with a slow rocking motion. The solution of primary antibody was discarded and the membrane was washed with PBS-Tween three times, for five minutes at room temperature, while rocked gently. This was followed with a new incubation of 45 minutes with secondary antibodies diluted in blocking buffer, at room temperature while rocked gently. After 3 subsequent washes with PBS-Tween, the membrane was placed in the LI-COR imaging system and visualized under the 700 and 800 wavelength channels. Images were captured after a 10 min exposure and imported at 600bpi quality.

2.2.7 Luciferase reporter assays

Pellets of 5×10^6 DG75 cells and 2×10^5 293T cells were resuspended in 250 µL of 1x Reporter Lysis Buffer (Promega) by vortexing for 10 seconds, followed by a short 15 second spin down at 13000 rpm. The supernatant was transferred to a new tube and the cellular debris discarded. After this, 10 µL of lysate was loaded per well (manually, in triplicate) of a white 96 well plate; the plate was then placed in a plate reader machine (Promega GloMax multi-detection system) which added 50 µL of luciferase reagent (LAR reagent, Promega) to each well. After a 2 second delay, the plate reader integrated a 10 second reading of the sample and moved on to do the same to the following well. To obtain a value of background noise, 3 wells were loaded with reporter lysis buffer that did not lysate any cells. The obtained values from each plate were automatically saved as a comma separated value (csv) file. The results reported the measured light as relative light units (RLU) for each well. The values of the transfected cells were subtracted the background noise values, and were then normalized with the values from the bicinchoninic acid assay (BCA, see BCA section and Normalization section).

2.2.7.1 Bicinchoninic Acid Assay (BCA)

To account for any possible differences in the concentration of protein across the different samples being compared in the luciferase assays, BCA assays were performed. In each clear plastic 96 well plate, a standard of known increasing concentration of Bovine Serum Albumin (Pierce ThermoFisher) was placed in triplicates 25 µL, and in parallel 25 µL of the lysate from the transfected cells was placed in triplicate in the plate. The following step was adding 200 µL of working

reagent (WR, Pierce ThermoFisher) in the wells, and reading the absorbance in the plate reader (Promega GloMax multi-detection system) at 560 nm using the default BCA program. Three wells which were devoid of BSA or lysate were used to obtain a value of background to be subtracted from the wells with BSA and lysate. The readings were automatically saved as a csv file accessible using Excel. Once the background value was averaged and subtracted to the readings of the BSA and lysate, a standard curve was generated with the average absorbance value from the wells with BSA, a linear regression allowed to find the protein concentration of the wells containing lysate from transfected cells.

2.2.7.2 Normalization and Statistical Analysis

The protein concentration values from the BCA assay were averaged and this value was used to normalize the triplicate RLU values (with the background subtracted) from the luciferase assays. This resulted in three values which were in turn averaged and named Normalized Luciferase Activity. The obtained values were used to run a Student T-test used to attain p value for statistical significance between different conditions. The T-test was performed through Excel T-test function, set as double tailed, with equal variances and an alpha equal to 0.05. To test for equal variances across sample arrays, prior to the T-test, an F-test was performed in Excel with an alpha equal to 0.05.

2.2.8 Polymerase Chain Reaction

The amplification of products for the 3C experiments, was performed based on the protocol recommended by New England Biolabs for the use of their Phusion High-Fidelity DNA Polymerase. Each reaction was typically set for a 25 µL volume, the components were: 5 µL of 5x Phusion HF buffer, 0.5 µL of 10 mM dNTPs, 2.5 µL of 10 µM primer pairs (forward and reverse), 0.25 µL of Phusion DNA Polymerase, DNA template (400 ng for cellular template, 10 fg for genestrings controls), and enough nuclease-free water for 25 µL. Several different concentrations of cellular template were tested, ranging from 10 fg to 800ng, however the results presented in following chapters are from assays performed with 400ng of template. Except where stated otherwise, the thermal cycling parameters were:

Initial Denaturation of 30 seconds at 98°C

30 cycles of

98°C for 10 second

60°C for 30 seconds

72°C for 30 seconds

A final extension of 72°C for 2 minutes

Hold at 4°C

2.2.9 Agarose Gel Electrophoresis

Gels were prepared to 1.2% (w/v) of Agarose and TAE buffer. Each gel was composed of 50 mL of 1x TAE and 600 µg of Agarose, the suspension was boiled to help agarose dissolve. Then, 4 µL of 12500X GelRed DNA stain (Biotium) was added to the cooled solution. The liquid was then left to be solidified in the casting tray fitted with a gel comb to generate wells. Gels were covered inside of an electrophoresis tank, with 1x TAE buffer. Loaded samples were mixed with blue DNA loading dye (Bioline), a marker was also loaded into the wells as a control, 1kb DNA hyperladder (Bioline). The electrophoresis tank was closed, and the electrophoresis was performed at 90 volts for 90 minutes. Gels were imaged and visualized with the use of the LI-COR imaging system (600 channel).

2.2.10 Chromatin conformation capture assay

The cells were centrifuged for 1min at 400g at room temperature, the supernatant was discarded and the cell pellets were re-suspended in 500µL of 10% (v/v) FCS/PBS per $1x10^7$ cells. Then this suspension of cells was filtered through the 70µm cell strainer to make single-cell suspension.

2.2.10.1 Formaldehyde crosslink and cell lysis

Per each cell pellet of $1x10^7$ cells, 9.5 mL of 1% formaldehyde/10% FCS/PBS was added, incubated while tumbling for 15 min at room temperature. The fixation was quenched by adding glycine to 0.125M. The tubes were centrifuged 8 min at 225g at 4°C, and the supernatant removed. Pellets were taken up in 5mL of cold lysis buffer and incubated for 10 min on ice. Then, the cells were centrifuged for 5 min at 400g at 4°C, and the supernatant was removed.

2.2.10.2 DNA digestion

The nuclei were taken up in 0.5mL of restriction enzyme buffer 1.2X (cutsmart), and then SDS added to a 0.3% concentration at 37°C, incubating for 1 hour while shaking at 900 rpm. Then 50µL of 20% Triton X-100 was added. After this 400 U of the selected restriction enzyme was added to the sample, and incubated at 37°C overnight while shaking at 900 rpm.

2.2.10.3 DNA ligation

Once the DNA was digested, 40µL of 20% SDS was added to the sample, it was then incubated for 25 min at 65°C, while shaking at 900 rpm. Then 6.125mL of 1.15 X ligation buffer (New England Biolabs) was added, followed by 375µL of 20% Triton X-100. A 1 hr incubation at 37°C was performed with gentle shaking. After this 100 U of ligase were added and incubated for 4 hrs at 16°C, followed by 30 min at room temperature. Then 300µg of Proteinase K were added, followed by an overnight incubation at 65°C to reverse cross link the sample.

2.2.10.4 DNA purification

After the ligation, 300µg of RNAse was added, and later incubated for 45 min at 37°C. Then 7mL of phenol-chloroform was added, and mixed. After centrifugation for 15min at 2,200g at RT, the supernatant was saved, 7ml of water was added, followed by 1.5mL of 2M sodium acetate pH 5.6 followed by 35mL of ethanol for precipitation. The samples were stored overnight at -20°C, this was then followed by centrifugation for 15 min at 2,200g at 4°C. The supernatant was discarded and the DNA pellet was reconstituted in water.

2.2.11 Computational methods

As word processors were needed to alter or analyse sequence, Word and Notepad (Microsoft Office) were employed; for the generation of figures Microsoft PowerPoint was used, and for calculations and correlations Microsoft Excel was employed. When finding ZREs in sequences, 3 different Macros were generated and programmed to be executed in Word, a different macro for each different class of ZREs:

```
Sub HighlightTargetsZREsCLASS2()

Dim range As range
Dim i As Long
Dim TargetList

TargetList = Array("TGAGCGA", "TCGCTCA", "TGAGCGC", "GCGCTCA")  '
For i = 0 To UBound(TargetList)

Set range = ActiveDocument.range

With range.Find
.Text = TargetList(i)
.Format = True
.MatchCase = False
.MatchWholeWord = False
.MatchWildcards = False
.MatchSoundsLike = False
.MatchAllWordForms = False

Do While .Execute(Forward:=True) = True
range.Underline = wdUnderlineDotted

Loop

End With
Next

End Sub
```

```
Sub HighlightTargetsZREsCLASS1()


Dim range As range

Dim i As Long

Dim TargetList


TargetList = Array("TGAGCCA", "TGGCTCA", "TTAGCAA", "TTGCTAA", "TGTGTAA", "TTACACA", "TGAGCAA", "TTGCTCA", "TGTGTCA", "TGACACA", "TGTGCAA", "TTGCACA", "TGAGTCA", "TGACTCA", "TGACTAA", "TTAGTCA", "TGTGTCT", "AGACACA", "TGGCACA", "TGTGCCA", "TGAGTAA", "TTACTCA", "GTTGCAA", "TTGCAAC") '

For i = 0 To UBound(TargetList)


Set range = ActiveDocument.range


With range.Find

.Text = TargetList(i)

.Format = True

.MatchCase = False

.MatchWholeWord = False

.MatchWildcards = False

.MatchSoundsLike = False

.MatchAllWordForms = False


Do While .Execute(Forward:=True) = True

range.Underline = wdUnderlineDouble


Loop


End With

Next


End Sub
```

```
Sub HighlightTargetsZREsCLASS3()


Dim range As range

Dim i As Long

Dim TargetList


TargetList = Array("GGAGCGA", "TCGCTCC", "TTCGCGA", "TCGCGAA", "TAAGCGA", "TCGCTTA", "TGAGCGG",
"CCGCTCA", "TGTGCGA", "TCGCACA", "TGAGCGT", "ACGCTCA", "CGGGCGA", "TCGCCCG", "AGAGCGA",
"TCGCTCT", "CGAGCGA", "TCGCTCG", "TCAGCGA", "TCGCTGA", "TGCGCGA", "TCGCGCA", "TGGGCGA",
"TCGCCCA", "TTAGCGA", "TCGCTAA", "CGCGCGA", "TCGCGCG", "CGTGCGA", "TCGCACG", "TGTGCGT",
"ACGCACA", "TTTGCGA", "TCGCAAA", "TTGGCGA", "TCGCCAA")  '

For i = 0 To UBound(TargetList)


Set range = ActiveDocument.range


With range.Find

.Text = TargetList(i)

.Format = True

.MatchCase = False

.MatchWholeWord = False

.MatchWildcards = False

.MatchSoundsLike = False

.MatchAllWordForms = False


Do While .Execute(Forward:=True) = True

range.Underline = wdUnderlineDotDashHeavy


Loop


End With

Next


End Sub
```

2.2.11.1 Regulatory Sequence Analysis Tools

Several of the tools that are freely available on the internet were used to generate different analysis, however they were all accessed from the main website as it was in the second half of 2013 (http://rsat.sb-roscoff.fr/).

2.2.11.1.1 Sequence upstream from genes

Sequences from EnsEMBL database were obtained through “RSAT – retrieve EnsEMBL sequence”. The parameters were set to obtain the upstream sequence (-2000 to +2), relative to the CDS.

2.2.11.1.2 Consensus sequence

The sequences were transferred from the previous analysis to the RSAT – oligo analysis tool; the parameters were set to mask non-DNA, to purge sequences, and to use the human genome as background model, no other parameter was changed from the default (van Helden et al. 1998).

2.2.11.1.3 Matrix scan

The obtained sequences were transferred to the RSAT – matrix-scan tool, where the DNA sequences were scanned with the profile matrices generated in the previous step, the analysis was performed with the default settings (Turatsinze et al. 2008).

2.2.11.1.4 Matrix mapping

The alignments obtained from the matrix scan were mapped with the RSAT – feature-map tool, with default settings.

2.2.11.2 BLASTn alignment

To analyse the alignments of fragments across sequences, Standard Nucleotide BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch) was used, checking the box to align two or more sequences, setting “Somewhat similar sequences (blastn)” program selection for a low level of stringency, and with a filter for species-specific repeats for Human, no other setting was changed.

2.2.11.3 Primer design

Primers were designed using NCBI Primer-Blast web tool, parameters were set to obtain primers of at least 20 bp in length and 60° C as optimal melting temperature, no other parameters were changed. The specificity of the primers was checked against human and EBV genomes (NC_007605.1) (www.ncbi.nlm.nih.gov/tools/primerblast).

Chapter 3. Enhancer capability of Zta binding sites

3.1 Introduction

In the Introduction, the process through which EBV inserts its DNA into the cell was reviewed; as well as the way in which the infection remains latent until Zta initiates a cascade of signals that switch the infection into a lytic state of gene expression.

Zta binds to the AP1 like sites, known as Zta response elements (ZREs) (Hardwick et al. 1988; Lieberman et al. 1990). ZREs can be classified as unmethylated and methylated (Bergbauer et al. 2010). Zta can preferentially bind to ZREs with an integral CpG motif when they are methylated (Bhende et al. 2004). It activates its own promoter as well as many other viral genes (Bergbauer et al. 2010; E. Flemington and Speck 1990a; Ramasubramanyan et al. 2012).

Previous experiments performed in the lab mapped Zta binding sites to the human genome through ChIP-sequencing (Ramasubramanyan et al. 2015). The numerous interactions between Zta and the human genome provided a large number of regions to analyse for the exploration of Zta influence over gene expression.

Although Zta has previously been characterized to be acting at promoter-proximal regions in the EBV genome (Adamson and Kenney 1998; Bhende et al. 2004; Packham et al. 1990; Ramasubramanyan et al. 2012), when studied in the human genome, some of the mapped peaks of Zta were found near promotor regions of genes; and interestingly, most other interactions are found in places not contiguous with promoters.

This pointed towards an idea; sequences where Zta binds could have enhancer properties. In this work, I asked the question: Is Zta binding to sequences with enhancing properties? the exploration of this idea was executed, first by identifying potential sequences that could be conferring enhancing activity, then by trimming and editing components in the sequence that could work as enhancers, and lastly by inserting this sequence into vectors and testing its enhancing capabilities through Luciferase Assays.

To select which interactions to examine, previous research performed in the Sinclair Lab analysed the change of mRNA levels in the transcriptome of B cells (Ramasubramanyan et al. 2015). When the mRNA levels were compared between latency and lytic cycle, a list ranking the genes from highest fold change to lowest was generated. This helped prioritize which interactions to analyse.

3.2 Results

Aiming to discover if the sequences to which Zta binds had enhancing properties not analyzed before, an initial step was taken; selecting which genomic sequences were of interest. Two different methods of selection were used: the first to be presented focused on 6 specific sequences that due to their characteristics could be candidates to be considered *enhancers*. The second method of selection focused on a higher number of sequences which were associated with *Upregulated genes* (Figure 3.1 left and right, respectively). In the interest of clarity, first I will present the approach for the *enhancers* method, and then the *Upregulated genes* method.

3.2.1 *Enhancers* method of identification of genomic sequences

Initially, a simple and straightforward process of selection was used in the *enhancers* method. A previous master's project analysis confirmed Zta binding to the promotor region of the FOSB gene using qPCR, as well as a many fold increase in the levels of expression of mRNA when compared to in latency. A similar change in the mRNA levels happens with AHNAK; also, they both showed Zta peaks within 2kb away to their transcription start site (TSS) (Figure 3.2 and 3.3). The Zta peaks close to these genes are not isolated, but surrounded by smaller peaks, making them interesting potential enhancers. This, therefore, presented the opportunity to consider the sequences from prominent Zta peaks proximal to genes FOSB and AHNAK to be analyzed.

3.2.2 *Enhancers* method of analysis of sequences and element design

The sequences selected to be analyzed through the *enhancers* method were retrieved through the use of the UCSC Genome Browser. Sequence under three peaks in proximity (less than 20 kb away) to AHNAK were retrieved, as well as sequence under five peaks close to FOSB. (continues in page 50).

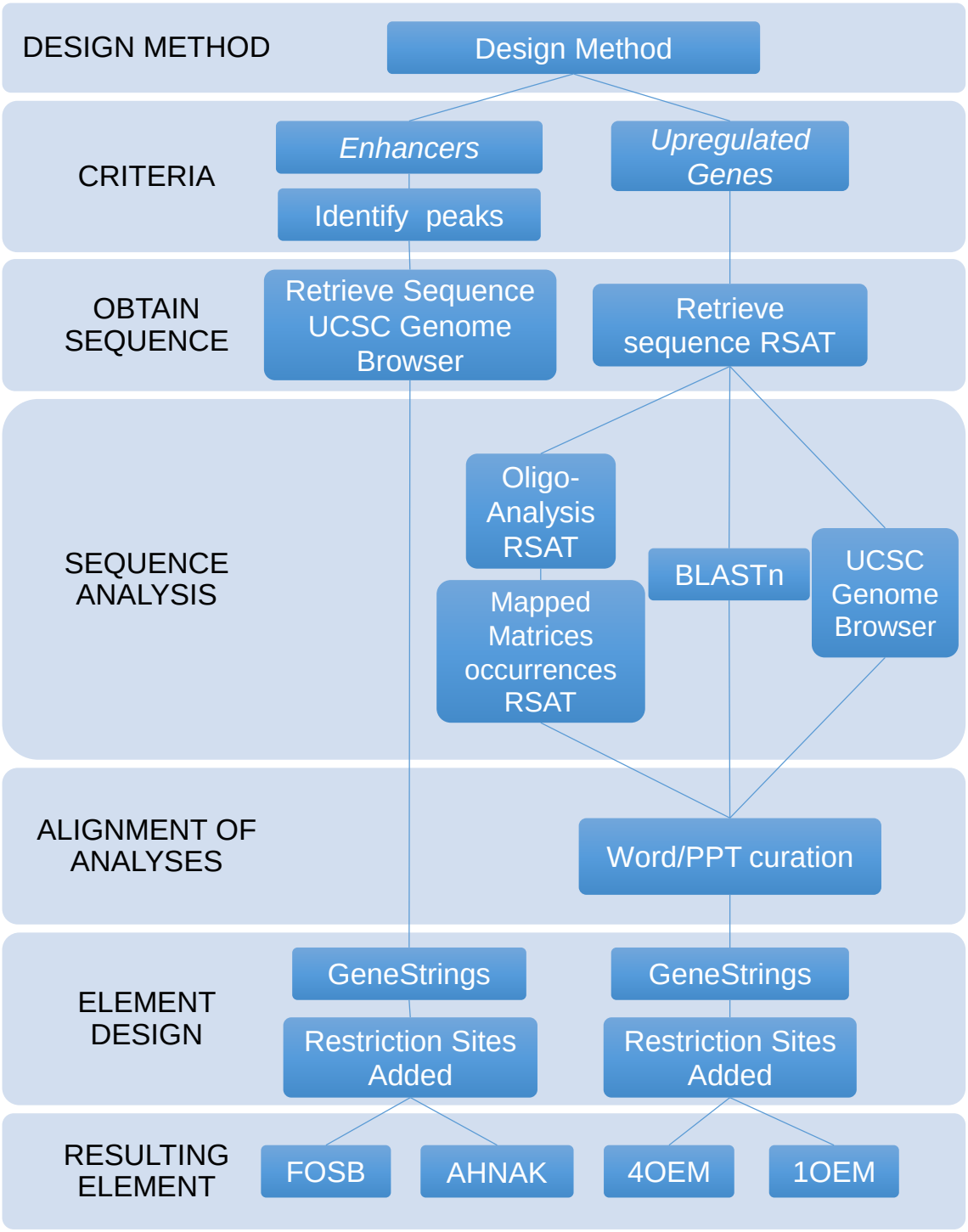


Figure 3.1. **Diagram of how the different elements were designed and produced.** UCSC acronym for University of California Santa Cruz genome browser. RSAT acronym for Regulatory Sequence Analysis Tools. BLASTn acronym for Basic Local Alignment Search Tool from NCBI. (See subsections 2.2.11.1 to 2.2.11.4 in Chapter 2.2.11 Computational methods).

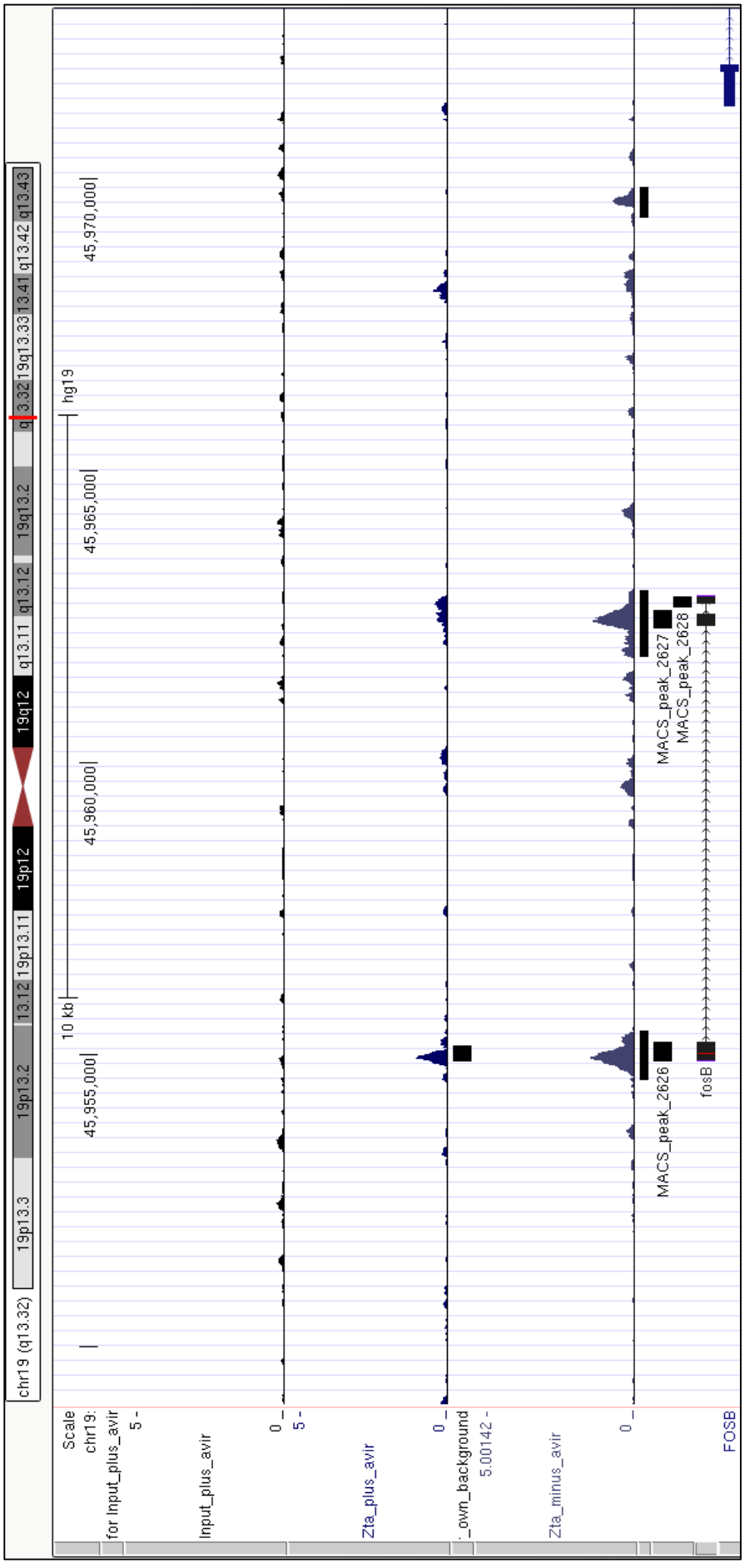


Figure 3.2. Clusters of peaks proximal to FOSB gene proposed for super-enhancer activity. The alignment of the FOSB construct designed (labeled fosB), can be appreciated in the lower middle part of the image, with the stretches of sequence beneath the peaks used to design said construct. Note that the actual FOSB gene is only shown partially in the lowest rightmost section of the map.

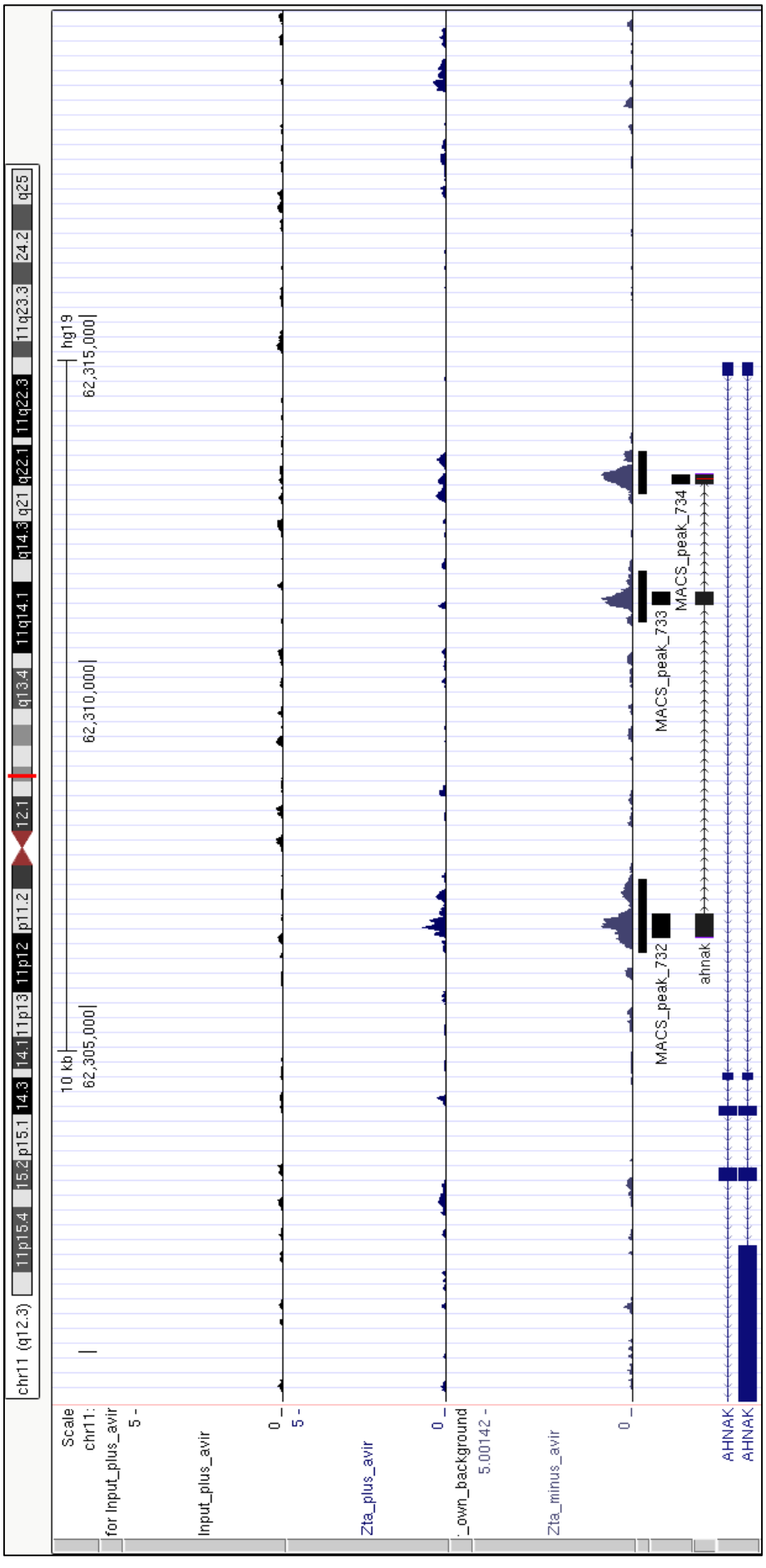


Figure 3.3 Clusters of peaks proximal to AHNAK gene proposed for super-enhancer activity. The alignment of the AHNAK construct designed, can be appreciated in the lower middle are of the image (labeled "ahnak"), with the stretches of sequence beneath the peaks used to design said construct. The actual AHNAK gene can be appreciated in the lowest part of the figure.

A

Peak 732
>hg19_dna range=chr11:62305993-62306348 5'pad=0
3'pad=0 strand=+ repeatMasking=lower 356bp
GCCTACAGCCGTTTACATCCGGAAAGTATCAACAGTCAATCAATGCATGTATA
CCGGCTGCCTCCATATGCCCAACACTATACTTTGCACTACAGACGGAACCAGC
CAGGGATGGACACAGCTGGGCCCAGCACCTAATTACACGCTGCGTTAGACGGT
GACCTCACATCACTCACGCACCCATGTGCCAGACAGGGAGATTTAAGACTCAG
TTCCTGTCACAGATACATATGATGAGCAAAGTGCACTAGCTGTGACCGATTTT
GCCGCAGAATATTCCAGAAGGCTTGGGAGCCTTCTAGGGGAAATGAAATATAC
ATGGGGTCTTGAAGGTGCTACAGGTGACCTCTGAGGAT

B

Peak 733
>hg19_dna range=chr11:62310818-62310998 5'pad=0
3'pad=0 strand=+ repeatMasking=lower 181bp
CGCAGGCAGAGGTGACACATCCAGGCACGCCACAAGGGGAAGACCTTCATTCA
CAACCAGTGGGATACTCTCTCACAGGCAATGACACACATCTCACTCAGAGACA
AGCTGTGACATGGTGGTCACAAGGACCTAAAGCCACAGTGGGACGACACCTGC
AATGCTATCAGTGACCTACCTA

C

Peak 734
>hg19_dna range=chr11:62312555-62312701 5'pad=0
3'pad=0 strand=+ repeatMasking=lower 147bp
AGGTCTAAGGACCTGACTGACTTGGGAGGCTCCCACACTGCCTTTCCACAAAT
CCCCTCCACAGAGCAGGGCACTCTGCAGCGACCGCACGCTCCCTGGGAGCACC
TCACTGCCACGGTCGCCGCGGCGGGAGCCACACCCTGCGAA

Figure 3.4. Sequences corresponding to peaks proximal to AHNAK. The DNA sequences are coming from the human assembly GRCh37, also known as hg19. The range of the sequences is shown, coming from the 11th chromosome, each from a different range. Any repetitive sequences would be shown in lower case letters. (**A**) Sequence for peak 732. (**B**) Sequence for peak 733. (**C**) Sequence for peak 734.

(continues from page 45) Any repetitive sequence was excluded from the design process, therefore sequence under two peaks close to FOSB was discarded, leaving us with sequence under 3 peaks close to AHNAK (Figure 3.4) and 3 peaks close to FOSB (data not shown).

3.2.3 *Upregulated genes* method of identification of genomic sequences

The other process of selection for which sequences to analyze, the *Upregulated genes* method, involved a different criteria. Instead of only considering nearby Zta peaks, the focus was on finding some shared or perhaps ubiquitous trait or component with enhancing capabilities, pervasive to several sequences; therefore, the analysis needed to encompass many sequences.

Since previous ChIP-seq experiments provided many different genes that have a Zta interaction within 2 kb from their transcription site, the idea of ranking them based on the fold increase of mRNA levels, helped select sequences to analyze (Ramasubramanyan et al. 2015). A list was then generated, composed of the top 30 genes with a higher than 5 fold increase of mRNA levels; which was used to retrieve the sequence 2 kb upstream from the transcription start site of those: ANXA2, FAM213A, WFIKKN2, AHNAK, LAMA3, C16orf73, GRIN1, LRP2, CD109, EMP1, PDGFB, BPIFB2, HAND1, EPHB3, FAM83C, GDF2, CRMP1, CTSW, USP50, RASGRF2, SCIMP, FCER1G, BCL2A1, FOSB, CPNE6, BOLA1, RASA3, GPRC5C, ARRB1, PDE1B.

3.2.4 *Upregulated genes* method of analysis of sequences and element design

In the *Upregulated Genes* method, the aforementioned list of 30 genes of interest was submitted to the Regulatory Sequence Analysis Tools (RSAT) website as described in the Materials and Methods section. The multiple sequences retrieved underwent three different analyses; through the use of the Basic Local Alignment Search Tool (BLASTn), matching short sequences across the analyzed sequences were found. The RSAT oligo-finder allowed to find a consensus sequence across the sequences conceptualized as matrices. The UCSC Genome Browser helped integrate the information found across platforms. Each will be expanded on, in the next section.

Given that previous research performed in a master's project showed through qPCR that Zta was interacting with the promotor regions of FOSB, SCIMP, BCL2A1 as well as peaks 2574, 2575 and 2576 (close to FOSB), the comparisons made between the sequences and the recently mentioned targets became a focal point of the design.

3.2.4.1 Derivation of matrices of Zta binding sites

3.2.4.1.1 Identification of genomic sequence of proposed interactions

Since the aim was to find interesting elements in the sequence upstream of the TSS of the genes, the sequences were requested to encompass from -2000 bp to +2 upstream of the TSS. The sequences came from the EnsEMBL database version 89, prompted by the RSAT – retrieve EnsEMBL sequence tool. The sequences were retrieved in FASTA format.

3.2.4.1.2 Basic Local Alignment Search Tool

Since part of the aim of analysis was to find some pervasive element found across all the different sequences, the sequences retrieved from the list of genes of interest were aligned against each other using BLASTn. This produced an extensive comparison between different sequences, resulting in finding various short segments of sequence to be found across sequences (Figure 3.5).

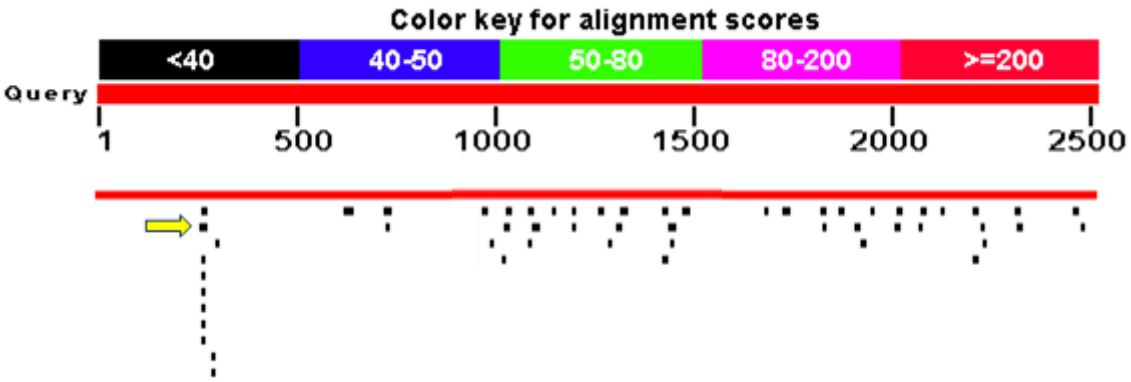
Although the alignment scores were small, that is only a reflection of the size of the fragment, and not the Expect value, which takes into consideration the chance of getting a false positive given the background of the analysis. In other words, the smaller the alignment and the bigger the list of and size of sequences to compare, the larger the expect value. Alignments of a larger span give a smaller expect value, compared to alignments of a shorter sequence.

This showed that there were indeed short pervasive sequences found across several of the different sequences analyzed.

3.2.4.1.3 Regulatory Sequence Analysis Tools

Although some alignments were found across several sequences through BLASTn, no ubiquitous and pervasive element was found in most of the sequences, therefore the analysis shifted towards (continues in page 54)

A



B

Homo_sapiens-ENSG00000170236-USP50-1

Sequence ID: lcl|166108 Length: 2003 Number of Matches: 6

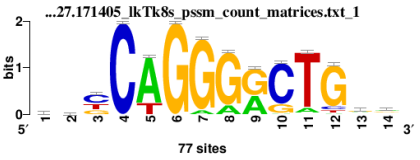
Range 1: 137 to 153 Graphics Next Match Previous Match

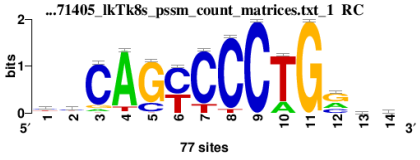
Score	Expect	Identities	Gaps	Strand
26.5 bits(28)	0.051	16/17(94%)	0/17(0%)	Plus/Plus

```
Query  263  TGTGAGCCAAGATTGCA  279
            ||||||||||||| |||
Sbjct  137  TGTGAGCCAAGATCGCA  153
```

Figure 3.5. BLASTn alignment of sequences with FOSB finds several matches. The sequence for FOSB goes from -2500 to +2 in order to incorporate the region corresponding to peak 2576. (**A**) Illustrative miniature**.** The thick red line represents the query sequence against which the subject sequences were compared. Black vertical squares/lines represent sequence similarity or matches of a given region against the query, in this case FOSB. Since the analysis was made between all sequences, the thin red line represents an alignment of the sequence of FOSB against itself, therefore it is shown in red, illustrating a very high alignment score. The color boxes represent the coloration the sequences would get based on the alignment score; the alignments found and used in subsequent analysis are in black showing that the length and therefore alignment score is low. The yellow arrow added to the figure points towards an element detected to be present in many of the sequences analyzed. (**B**) One of several matching sequences found, referred to in (A) with the yellow arrow, is found in the sequences of USP50, GRIN1, LAMA3, MEIOB, SCIMP, BOLA1 (the rest of the matching sequences are not shown).

A

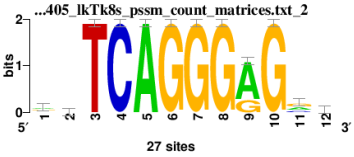


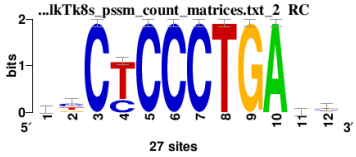


B

```
Matrix 1
Consensus        csyCAGGGrCTGrr
consensus.rc     YYCAGYCCCTGRSG
a   19   13   0    0    58   0    5    11   30   2    6    2    22   25
c   26   22   34   74   0    0    0    0    0    61   1    6    13   13
g   19   24   19   0    0    74   69   63   44   11   1    62   30   27
t   10   15   21   0    16   0    0    0    0    0    66   4    9    9
//
```

C

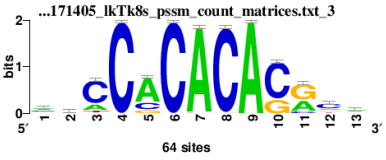


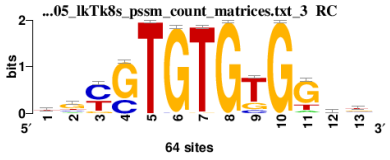


D

```
Matrix 2
consensus        rhTCAGGGRGrw
consensus.rc     WYCYCCCTGADY
a    10   7    0    0    27   0    0    0    20   0    9    9
c    6    8    0    27   0    0    0    0    0    0    6    3
g    9    5    0    0    0    27   27   27   7    27   11   5
t    2    7    27   0    0    0    0    0    0    0    1    10
//
```

E





F

```
Matrix 3
consensus        vcCCACACASrca
consensus.rc     TGYSTGTGTGGGB
a    22   15   12   0    44   0    63   0    64   0    23   11   27
c    21   25   45   64   13   64   1    63   0    46   6    34   13
g    16   14   2    0    7    0    0    1    0    18   35   10   15
t    5    10   5    0    0    0    0    0    0    0    0    9    9
//
```

Figure 3.6. Three Matrices were generated through RSAT analysis. In order to obtain these matrices, the sequence 2 kb upstream of the transcription start site of the list of genes of interest was analysed. (**A**) logo plot of Matrix 1, followed by its reverse complementation. The relative height of the letter represents the estimated probability of such base to be found in such position. (**B**) Matrix 1, each number represents the percentage of probability for the base to be found in a particular position. (**C**) Logo plot of Matrix 2, followed by its reverse complementation. (**D**) Matrix 2. (**E**) Logo plot of Matrix 3 followed by its reverse complementation. (**F**) Matrix 3.

(continues from page 51) detecting a consensus sequence of an over-represented sequence, which was also added to the analysis used to find the Matrices.

Then, in order to identify the places within the sequences that best reflected the consensus sequences of the generated matrices; the “Matrix scan” tool of RSAT was used. This found the incidences of each matrix present on any sequence. This analysis also gave each match with a “weight” score that reflects how much the matrix gets represented in the matched sequence outside of chance brought by the matrix composition or the background model.

This analysis was performed scrutinizing the presence of the 3 matrices in the upstream sequences used to generate the matrices. Out of the found incidences of the matrices on the sequences, we focused on the ones found in the upstream sequence of FOSB (Figure 3.7), SCIMP, BCL2A1, and peaks 2574 and 2575.

	A	B	C	D	E	F	G	H	I	J
1	seq_id	ft_type	ft_name	strand	start	end	sequence	weight	start	end
2	FOSB76	site	matrix-scan_2013-10-31.3	R	-1691	-1680	ACTCAGGGAGGT	12	809	820
3	FOSB76	site	matrix-scan_2013-10-31.2	R	-310	-297	CCCCAGGGGCAGAA	10.3	2190	2203
4	FOSB76	site	matrix-scan_2013-10-31.4	R	-235	-223	GCCCCCACAGGAT	8.2	2265	2277
5	FOSB76	site	matrix-scan_2013-10-31.3	D	-1818	-1807	TTTCAGGGAGTT	8	682	693
6	FOSB76	site	matrix-scan_2013-10-31.2	D	-305	-292	CCCCTGGGGGAGCA	7.1	2195	2208
7	FOSB76	site	matrix-scan_2013-10-31.2	D	-565	-552	CGGCAGGGACACGC	6.9	1935	1948
8	FOSB76	site	matrix-scan_2013-10-31.4	D	-1835	-1823	AACCCCAGACACA	6.8	665	677
9	FOSB76	site	matrix-scan_2013-10-31.3	D	-1316	-1305	GCTCAGAGAGGG	6.6	1184	1195
10	FOSB76	site	matrix-scan_2013-10-31.2	D	-2392	-2379	ACCCAGGAGATGTA	6.3	108	121
11	FOSB76	site	matrix-scan_2013-10-31.2	R	-996	-983	CCCCGGGGGCTGGT	6.3	1504	1517
12	FOSB76	site	matrix-scan_2013-10-31.2	D	-2165	-2152	GGCCAGGCGCTGTG	5.9	335	348
13	FOSB76	site	matrix-scan_2013-10-31.2	D	-2417	-2404	AGGCTGAGGCAGGA	5.7	83	96

Figure 3.7. Matrices incidences in FOSB (fragment). The incidences were found on the sequences used to generate the matrices. Each line corresponds to an instance of one of the matrices found against the sequence of FOSB. The sequence for FOSB goes from -2500 to +2 in order to incorporate the region corresponding to peak 2576. The weight value referring to how much the compared sequence resembles each matrix outside of chance, the higher the number the less probable the match happened by chance

The matrices incidences were also mapped against the sequences upstream of the genes of interest as part of the RSAT tools, resulting in a visual representation of the sequence and each time a sequence reflected one of the matrices. The results are shown for FOSB which clearly depicts several incidences of the 3 matrices (Figure 3.8).

3.2.4.1.4 Alignments and verifications

In order to find components of interest within the retrieved sequences through the *Upregulated Gene* method, a macros command was programmed and executed on Word. Sequences were retrieved with a masking of repetitive sequence in lower case characters, and the sequence found to be part of a peak to be presented in bold text (Figure 3.9 A). This allowed to find interesting alignments of the matrices incidences mapped in close proximity to ZREs or sequence matches found through BLASTn (Figure 3.9 B).

To unify the analysis obtained through BLASTn and RSAT as well as the visual advantages of the UCSC ChIP-seq map, the merging of the visual maps helped create an alignment of analysis across platforms. This was helpful to visualize where the interesting components of each analysis could fall in alignment relative to components of different analysis; strongly corroborating the presence of consensus sequence across platforms of analysis, some in areas of high conservation (See Figure 3.10, 3.11 and 3.12).

3.2.4.2 Reporter genes with Zta binding sites

The visual alignment of analyses, allowed to generate a table with the representations of each matrix according to its weight score, selected because of their pervasive presence across sequences, found at the upstream sequences of SCIMP, BCL2A1, FOSB and peaks 2574 and 2575 (Table 3.1). The weight score represents how much the matrix is represented and with resemblance in the sequence outside chance, the higher the number the less probable the match happened by chance. This illustrates that there were instances where some of matrices could not be found in some sequences, for instance no matrix had a match in the sequence for peak 2575. (continues in page 62)

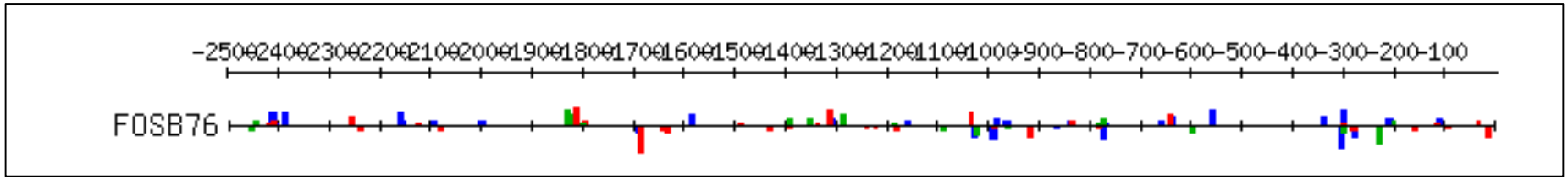


Figure 3.8. Mapped matrices occurrences in FOSB. On this Feature map each bar represents an incidence of one of the 3 matrices found. The height of each bar represents the significance of that incidence. On top a horizontal graded line functions as a ruler that shows the length of the sequence map. The horizontal line below the ruler, represents the sequence against which the matrices incidences were mapped to. Bars above represent incidences on the plus strand of the sequence, and bars below represent incidences on the minus strand. Blue bands correspond to Matrix 1, Red to Matrix 2, and Green to Matrix 3. The sequence for FOSB goes from -2500 to +2, relative to TSS, in order to incorporate the region corresponding to peak 2576.

A

>Homo_sapiens-ENSG00000125740-FOSB-ENST00000585836-ENSP00000467497 ENSG00000125740-ENST00000585836-ENSP00000467497; upstream from -2500 to 2; size: 2503; location: chromosome:GRCh37:19:1:59128983:1 45969345 45971847 D

aactctgtctcaaaaaacaaaacagaacaaaacaaaaattagccaggtgtggtggcgcac
acctgtaatcccagctactcgggaggctgaggcaggagaatcacttgaacccaggagatg
taggctgcagtgagccgagattgtggcactgcactccagcctgggcaacagagtgagact
ctgtcaaaaaaaaaaaaaaaaaaCGTCAGATTCATGAACCCCATTTGCTAAGAAGATTTG
CTCAGGTAGTCACCTGTACCCCTGTGAGCCAAGATTGCAACAGTACATGCTAAGAGCAGT
AAGAAGGAGATTGTAAAGCTTAGAAAGACCTGACAggccaggcgctgtggctcacgcctg
taatctcgacacttcgggaggccgaggcaggcggatcaccaggtcaggagttcaagacca
tcctgaccaatatggtgaaaccccgtctgtactaaacatataaaaattagcttgacgtgg
tggcccgcgcctgtagtccagcaactcgagaggctaaggtgaggcggaggttgcagtgag
ccgagatcgccactgcactccagcctggcggcagagcaagactccgtctcaaaaaaaaaa
aaaaaaaaaaaaaaaaaaaaGACCTGACAAGTGAAAGCTACTAATATTGCCATTATCATTT
AAAAAAACCCCAGACACAGGTTTTTCAGGGAGTTTCATCCAACCAGGCAGGTCTCAGAAA
TCAGAAAAGAaatggaaagggtaggaaatctggagtttgacaattctgagtttgaatttc
ttgtgatgggatcttgggcaagtcatttaacctccctgagtatcatttttttctttttata
aaatgaagatttttctctcttaaccttccagagctgttttaaggattacaaatcttctac
tgaaaggcgtagcacaggagctgtgcttggcaagtgccaaatacaaggcattaattattA
TTATTAGAATTAATAATAATATCCCCTCCCTCTTACACATTCTTTGTCTCCGGGTGGATT
AAAAGGTGGAAGGAGAGGCTACCAACACCATCAGAAGAGAGGCTTCTCTCTAAGTTTCAT
TTCCCATCTCCTCCAAATCCGCATCCCTCCCAAACGCCGGACCTGTAAGGCCAGCAGGGT
CCAAGACACACATCCTTTGCCCAGCGGGGAAGATTAaagctaaagctcagagagggaaaa
catttcctaagctcgcacagcgaatcaggacagaaaccaggacgagcctcggaatccctC
CATTACCTCCACTTTCACCTGAGCATCACAGCCCGCTCGGGACTCAGTTTCCCCACCTAC
GTGACCACACCACACTAATCAGGGTCTCCTTTTGGAGATCTGCTCTTCTTCTCGAATGGG
GGCGCTGCACCATCGGTAGAACAGGGTAGGTGGGGGCGCCAGAGGTGAAGGGGACCTGCA
GGCTGGGGTCTTCCCCGCCCGGGTCAGCGGGGTCCCTGCGGGGCTAGTCTAAGCGCCTAT
TATTACCAGCCCCCGGGGCGGCGTTGCACTGCGCAggcgcgggcggggcgcgggcgcgcg
cgcgAGCGAGCGAGGGATTCCCTCTGACGTCATTGCTAGGATACCAAACAAACACTCCGC
CGCGCCGGCCGAGCTCCTTATATGGCTAATTGCGTCACAGGAACTCCGGGAAGGCGGGGC
CGGGATCCCCTCCCGCCGAGTGGCCCGGAACGCAACCCCCGAGACCCCCAGGGCCCCGAG
GGTCATGCAAGTGACCAGATCGAGTCTAGAACAGACCTCTTGCTGGACAGTGCGGGACTC
GATTTGGCGGGGCCGGAGATTTGGGGAAGTTTGTCCAGCAAGGGGCGGGTGACGTAAGCA
GGGGGGCGGGTCCCGGGCATATAAATACAGGCTGGCGGGTCTGTGCTTCATTCATAAGAC
TCAGAGCTACGGCCACGGCAGGGACACGCGGAACCAAGACTTGGAAACTTGATTGTTGTG
GTTCTTCTTGGGGGTTATGAAATTTCATTAATCTTTTTTTTTCCGGGGAGAAAGTTTTTG
GAAAGATTCTTCCAGATATTTCTTCATTTTCTTTTGGAGGACCGACTTACTTTTTTTGGT
CTTCTTTATTACTCCCCTCCCCCCGTGGGACCCGCCGGACGCGTGGAGGAGACCGTAGCT
GAAGCTGATTCTGTACAGCGGGACAGCGCTTTCTGCCCCTGGGGGAGCAACCCCTCCCTC
GCCCCTGGGTCCTACGGAGCCTGCACTTTCAAGAGGTACAGCGGCATCCTGTGGGGGCCT
GGGCACCGCAGGAAGACTGCACAGAAACTTTGCCATTGTTGGAACGGGACGTTGCTCCTT
CCCCGAGCTTCCCCGGACAGCGTACTTTGAGGACTCGCTCAGCTCACCGGGGACTCCCAC
GGCTCACCCCGGACTTGCACCTTACTTCCCCAACCCGGCCATAGCCTTGGCTTCCCGGCG
ACCTCAGCGTGGTCACAGGGGCCCCCCTGTGCCCAGGGAAATG

B

WYCYCCCTGADY

R -2243 -2232 ACCCCTGTGAGC 1.5

AGTCACCTGTACCCCTGTGAGCCAAGATTGCAACAGTACATG

Figure 3.9. Annotation and alignment of sequence upstream of FOSB. (**A**) Sequence upstream of FOSB with annotations in text font. Sequence in bold corresponds to the bases that constitute the peak of interaction between Zta and the human genome. Sequence underlined with dashes, double lines, single lines and bold lines correspond to the sequence of known ZREs. Sequence underlined with dots marks sequence found to be shared across all the sequences found in the BLASTn analysis. Sequence boxed represent iterations of where the matrices were found in RSAT. Sequence in lower case represent repetitive sequence. (**B**) Direct alignment of matrix 2 with FOSB upstream sequence. The top line in red is the consensus sequence of Matrix 1. The second line comes from the RSAT matrix scan analysis, where it found an instance of sequence from FOSB upstream sequence, in the Reverse strand, that spans from -2243 to -2232. The value at the rightmost part of the second line is the weight value referring to how much the compared sequence resembles matrix 2 outside chance, the higher the number the less probable the match happened by chance. The third line is part of the sequence retrieved from the UCSC genome browser. The boxed sequence represents the alignment to the matrix, the sequence dotted represents sequence found to be shared across the list of genes of interest through BLASTn (See Fig 3.5 and Fig 3.10). The sequence for FOSB goes from -2500 to +2 in order to incorporate the region corresponding to peak 2576.

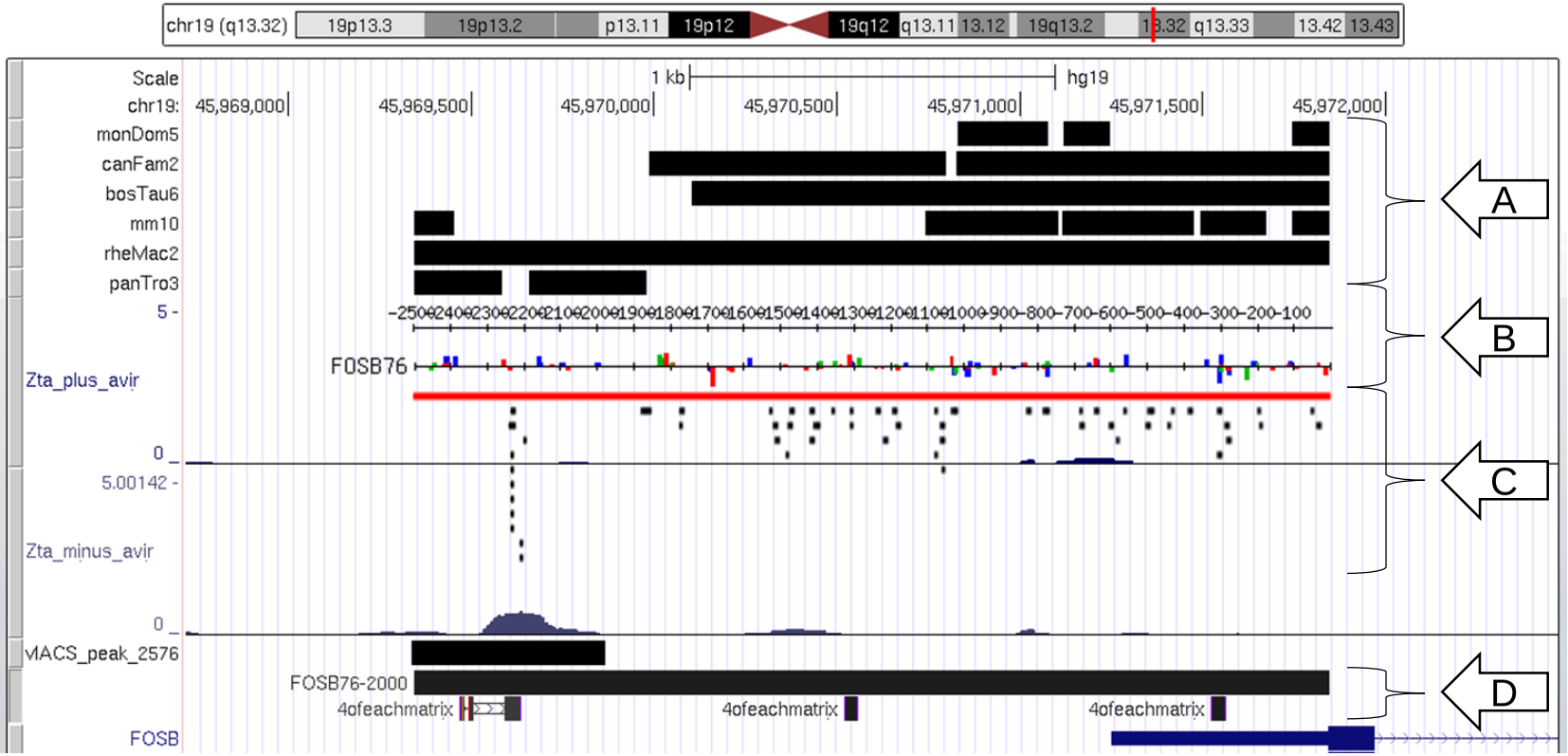


Figure 3.10. Alignment of analysis maps for the region upstream of FOSB. (**A**) sequence found to be conserved across different species. (**B**) Visual map of matrices incidences from RSAT (See Fig 3.8). (**C**) Visual map of alignments from BLASTn. (**D**) This corresponds to the alignment of the sequence retrieved from RSAT for upstream of the TSS of FOSB, named "FOSB76-2000"; notice how it aligns with the FOSB gene in the bottom of the image in blue. The alignment of the sequence of the element 4OEM is also aligned, named "4ofeachmatrix". It is interesting to see how the string of alignments at (C) which is referred to in Figure 3.5, aligns with the Zta peak 2576, and the incidence of Matrix 2 from (B) in red, below the line.

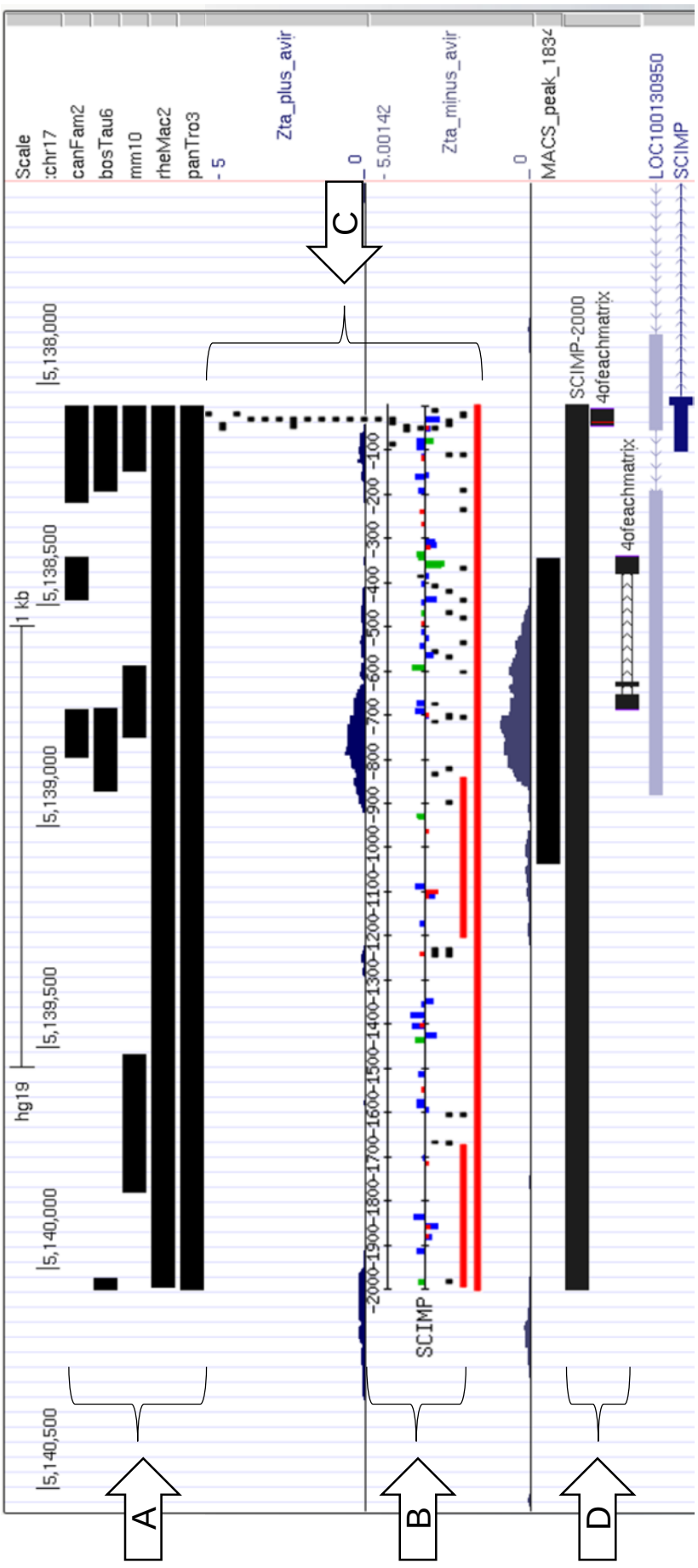


Figure 3.11. Alignment of analysis maps for the region upstream of SCIMP. (**A**) sequence found to be conserved across different species. (**B**) Visual map of matrices incidences from RSAT. (**C**) Visual map of alignments from BLASTn. (**D**) This corresponds to the alignment of the sequence retrieved from RSAT for upstream of the TSS of SCIMP, named "SCIMP-2000"; notice how it aligns with the SCIMP gene in the bottom of the image in blue. The alignment of the sequence of the element 4OEM is also aligned, named "4ofeachmatrix". It is interesting to see how the string of alignments at (C) aligns with the incidence of Matrix 1 from (B) in blue, below the line.

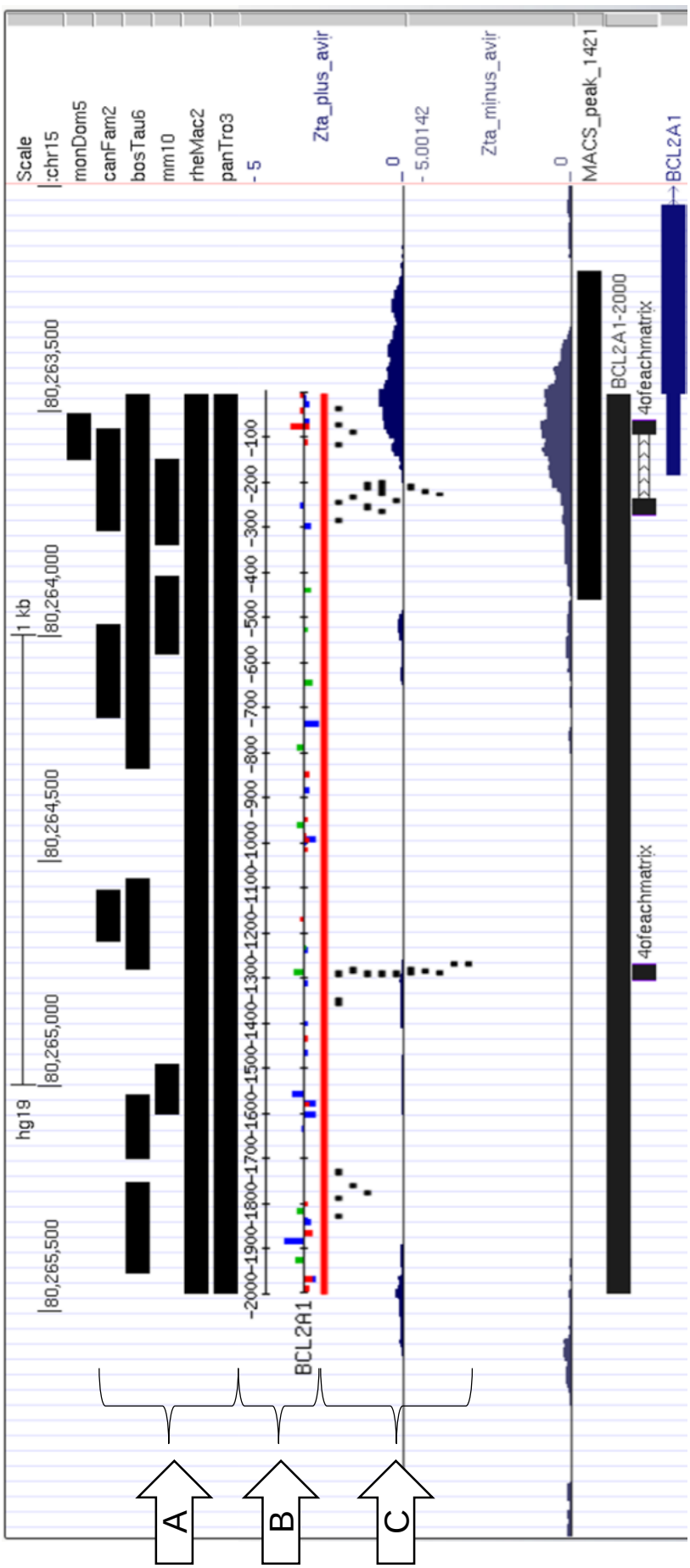


Figure 3.12. Alignment of analysis maps for the region upstream of BCL2A1. (**A**) sequence found to be conserved across different species. (**B**) Visual map of matrices incidences from RSAT. (**C**) Visual map of alignments from BLASTn. (**D**) This corresponds to the alignment of the sequence retrieved from RSAT for upstream of the TSS of BCL2A1, named "BCL2A1-2000", notice how it aligns with the BCL2A1 gene in the bottom of the image in blue. The alignment of the sequence of the element 4OEM is also aligned, named "4ofeachmatrix". It is interesting to see how the string of alignments at (C) aligns with the incidence of Matrix 3 from (B) in green, above the line.

Table 3.1. Table of matrices incidences and weight scores.

	MATRIX1 Weight Scores	MATRIX2 Weight Scores	MATRIX3 Weight Scores
SCIMP	6.1	*	7.3
	3.8		8.5
BCL2A1	1.7	6.6	4.9
FOSB	10.3	1.5	*
		6.6	
P 2574	*	4.4	4.4
			6
P 2575	*	*	*

(continues from page 55) Two different designs were used to compile the incidences of the matrices. One design compiled the top 4 instances of each matrix (according to its weight score) found across SCIMP, BCL2A1, FOSB and peak 2574. This was referred as 4OEM as an acronym for 4 of each matrix (Figure 3.13 A). The other design compiled the top instance of each matrix, referred as 1OEM (Figure 3.13 B).

3.2.4.3 Synthetic generation of elements through GENESTRINGS

Aiming to produce a synthetic DNA sequence composed of the elements found in the previous analyses, synthesis of the designed elements was carried out by the GENESTRINGS service. In order to have the capacity to be inserted into different vectors, two pairs of restriction sites, as well as GC clamps, were added to the 5' and 3' ends of the sequences (Figure 3.13 and Figure 3.14).

3.2.5 Element enhancing effect

Once all 4 elements were designed with restriction sites to the ends of their sequence, they were synthetized through the GENESTRINGS service of Life Technologies. (continues in page 65)

A

>4OEM

GGGCCCCTGCAGGATCCCTGAGCGGCAGCACTCCAGCCACTGCTGTCTC

CTTAGCGAGGCCGTGGGCCTCGCTGGGAGTTCTGTCTCACAGTGAATTAA

TCACAGGATGAGGAAGTGGCTTCTCTGAAACAGGACAGCGCTTTCTGCCC

CTGGGGGAGCAACCCCTCCACTCAGTCAAGCTCAGTGAGCATTCTCAGCA

CTAaagctaaagctcagagagggaaaaacatttcctaaAGTCACCTGTACC

CCTGTGAGCCAAGATTGCAACAGTACATGAGCAAAACTGCAGGACTCAGT

GAGTGGTTTCTGTGGATCTAGGAATGCTCACGGCAAACTCGTCTGTGGGG

ACACCAAACCAGGCCCTGTTCTCTGCCACCCAGGCTGTTTTGTAGTCACA

ACTCATCCCCAACACACACTCCCAACACCCTGTGTAGATATGAATGTGTG

TGCAGTTGTGAGTCTGTCGACACTAGTGGGCCC

B

>1OEM

GGGCCCCTGCAGGATCCGGACAGCGCTTTCTGCCCCTGGGGGAGCAACC

CCTCCAGTCACCTGTACCCCTGTGAGCCAAGATTGCAACAGTACATGAGG

CCCTGTTCTCTGCCACCCAGGCTGTTTTGTAGTCGTCGACACTAGTGGGC

CC

Figure 3.13. Compiled Matrices instances into 4OEM and 1OEM. These sequences were found to be the most pervasive matrix instances. Sequence is annotated as in Figure 3.9. As spacers, there are 10 nucleotides to each side of the either boxed sequence, (it being the matrix alignment (RSAT)) or dotted sequence (shared sequence fragments (BLASTn)). Sequence highlighted in blue is a GC clamp, sequence highlighted in red marks a PstI restriction site, sequence highlighted in green marks a BamHI restriction site, sequence highlighted in purple marks a SalI restriction site and sequence highlighted in orange marks a SpeI restriction site. (**A**) The 4 top instances found for matrix 1, 2 and 3 compiled into 4OEM. (**B**) The top instances found for matrix 1, 2 and 3 compiled into 1OEM.

A

>FOSB

GGGCCCCTGCAGGGATCCGGGCGAATCTCCGAGCAGAGAGAAGAAATTGAGTG
TGTGACAGTGTGAATGGATGTGACTATAGGTGAGCAACAGTCGGTGACAATGT
GCACATCTGGTATTTCCTGTGTGAGAGCAAAACTACAGGACTCAGTGAGTGGT
TTCTGTGGTTTGCACTTGTGTATCTGGGTGTATGGGAGTGTGTAGCCCTGCTG
ACTGTGTAGATATGAATGTGTGTGCAGTTGTGAGTCTTGCATGTGGTCATCTA
TGTGAAACAATGCCTGTATAAGTACCCATGGATGTGCACGGTTTTGCCTCTAG
CCGTGCATTTCTGGTTCTGTGTGTGTGTGTTTCTGGGACATTCTTCAAGTCCA
CAACTCTGAGGGTATCACTGTGAGAGCAGGGGTCCTAGCTCTACCCATTTGTG
TGTGTGTGGAGATGTGTAAGTCTATGTAGACAAAAGTGTGTGTCATTTACTGT
GTTTGGGGTGTGAACACCTATGTGATGTGTTTGCACAACTGCACGTGTTTTGT
TCTGTGTGTGTGATCGTGTGTTCAAATAAGTCATCTTGTCTAGTCATCAGGTG
CCTGCACAGACAAAGGTGAAAGGTGTCTGCCTGTTGAGATCTGTGGATAGGGT
GTATATATGGACATCTCAGCCTGTCTACGTGTGTATCTGTCTCTGTCCTCACG
GCAAAAGAGAGGTCGACACTAGTGGGCCC

B

>AHNAK

GGGCCCCTGCAGGGATCCGCCTACAGCCGTTTACATCCGGAAAGTATCAACAG
TCAATCAATGCATGTATACCGGCTGCCTCCATATGCCCAACACTATACTTTGC
ACTACAGACGGAACCAGCCAGGGATGGACACAGCTGGGCCCAGCACCTAATTA
CACGCTGCGTTAGACGGTGACCTCACATCACTCACGCACCCATGTGCCAGACA
GGGAGATTTAAGACTCAGTTCCTGTCACAGATACATATGATGAGCAAAGTGCA
CTAGCTGTGACCGATTTTGCCGCAGAATATTCCAGAAGGCTTGGGAGCCTTCT
AGGGGAAATGAAATATACATGGGGTCTTGAAGGTGCTACAGGTGACCTCTGAG
GATCGCAGGCAGAGGTGACACATCCAGGCACGCCACAAGGGGAAGACCTTCAT
TCACAACCAGTGGGATACTCTCTCACAGGCAATGACACACATCTCACTCAGAG
ACAAGCTGTGACATGGTGGTCACAAGGACCTAAAGCCACAGTGGGACGACACC
TGCAATGCTATCAGTGACCTACCTAAGGTCTAAGGACCTGACTGACTTGGGAG
GCTCCCACACTGCCTTTCCACAAATCCCCTCCACAGAGCAGGGCACTCTACAG
CGACCGCACGCTCCCTGGGAGCACCTCACTGCCACGGTCGCCGCGGCGGGAGC
CACACCCTGCGAAGTCGACACTAGTGGGCCC

Figure 3.14. Elements designed from the peak clusters proximal to FOSB and AHNAK genes. Any repetitive sequence found in the peaks was excluded. The sequence in blue represents a GGGCCC clamp added. On the 5' end, sequence in red marks a PstI restriction site, while sequence in green marks a BamHI restriction site. On the 3' end, sequence in purple marks a SalI restriction site and sequence in orange marks SpeI restriction site. (**A**) Sequence of FOSB element. (**B**) Sequence of AHNAK element.

3.2.5.1 Transfections and Reporter Assays

Three different vectors with different properties were used aiming to test different aspects of the possible enhancer activity. The vectors used were:

1. A vector composed of a pCpGL-basic backbone and a BHLF1 promoter region was selected to be used (Hammerschmidt and Sugden 1988; Ramasubramanyan et al. 2015). This choice was made, since previous empirical research, proved that Zta upregulates the activity of the BHLF1 gene through its interactions with ZREs present in the BHLF1 promoter sequence. For this reason, the vector was employed to test if any enhancer activity of the elements reflected changes in luciferase activity (Figure 3.15 A).
2. Another vector composed of a pCpGL basic backbone but containing a mutated BHLF1 promoter (BHLF1-Mut) sequence inserted was also selected to be used (Klug and Rehli 2006; Ramasubramanyan et al. 2015). This choice was made, since previous experiments showed that when the ZREs in the sequence are mutated, a loss of the activation caused by the interaction with Zta can be observed. Therefore, this tested if the supposed enhancer activity of the designed elements are present on a promoter not regulated by Zta (Figure 3.15 B).
3. A third vector composed of a pGL3 enhancer backbone and a minimal promoter sequence of the gene CIITA was selected (CIITA -214+52). This choice was made since previous research showed that the CIITA promoter activity does not greatly change when Zta is present in EBV negative BL DG75 cells, however if ZREs are inserted in tandem distally from the TSS, then activation by Zta is observed (Ramasubramanyan et al. 2015). However, in EBV negative Akata cells, this same vector shows repression when Zta is present (Balan et al. 2016). The restriction sites BamHI and SalI are distant to the transcription start of the luciferase gene, therefore, to test if the proposed enhancer sequence can change the level of repression by Zta when inserted far away from the luciferase transcription start site, this vector (addressed from here onwards simply as CIITA) was used (Figure 3.15 C). (continues in page 69)

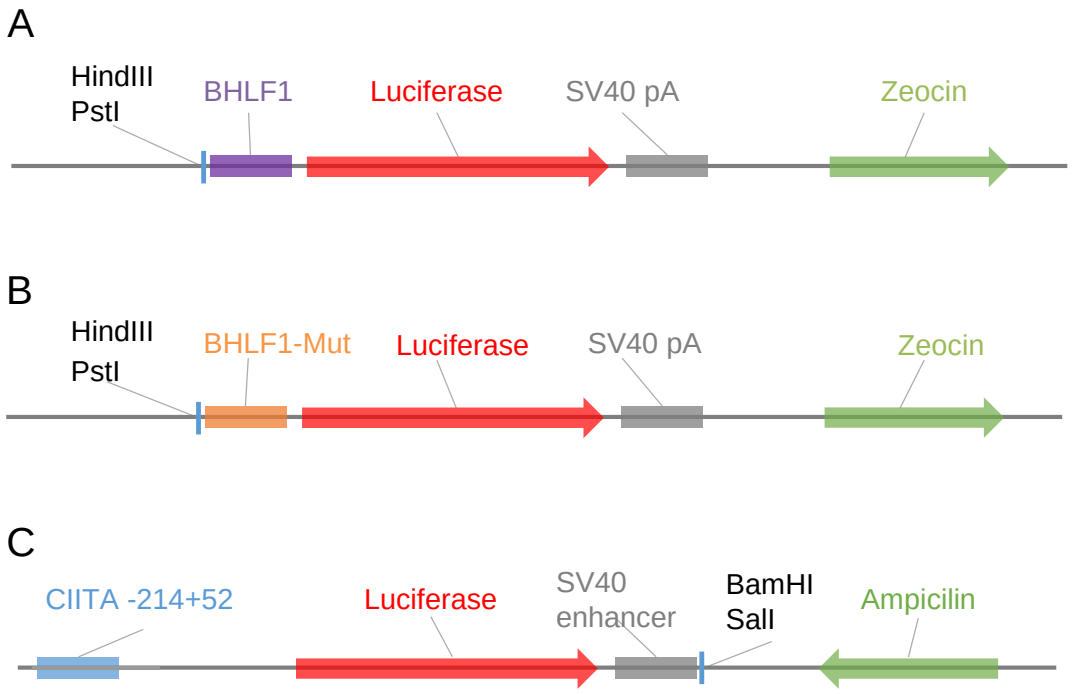
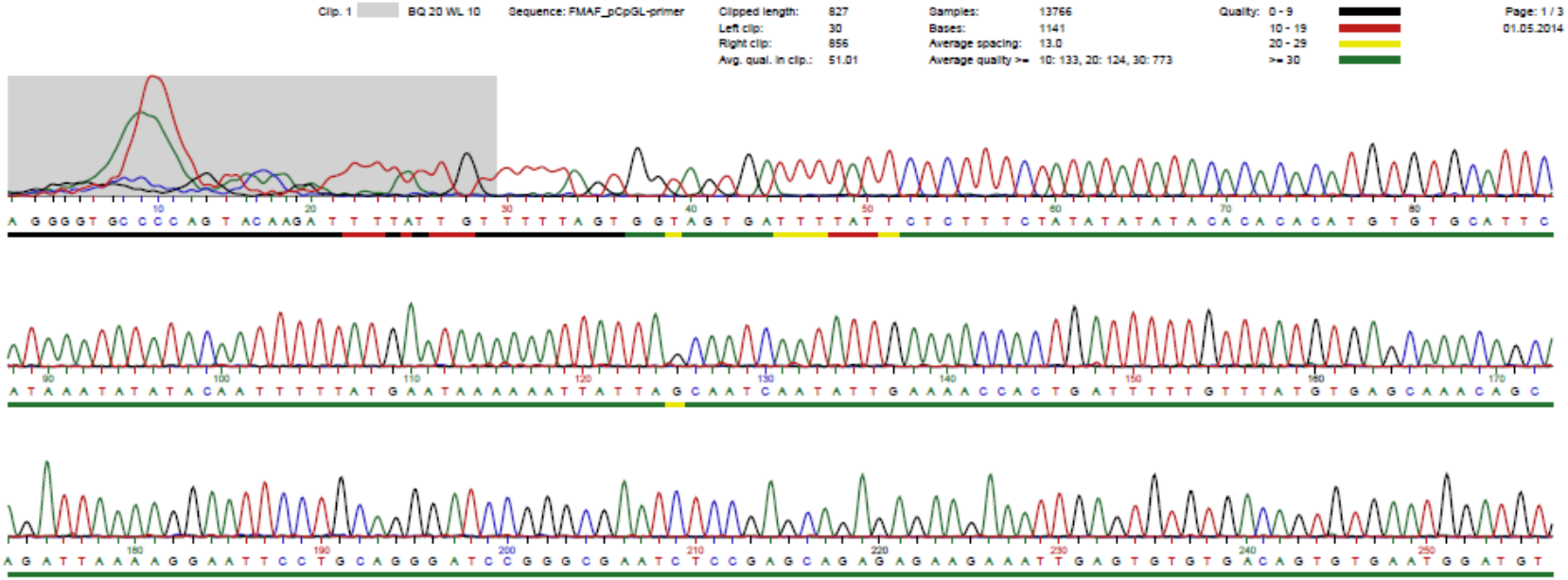


Figure 3.15 Three vectors were used to test the effect of the presence of the elements. (**A**) Schematic of the BHLF1 vector. The Luciferase gene is represented as a red arrow and the Zeocin resistance cassette in green. In dark grey the sv40 polyA terminator. In purple the promoter sequence of BHLF1. Restriction sites in dark blue for HindIII and PstI. (**B**) Schematic of the BHLF1-Mut vector. As above, the luciferase gene and the Zeocin resistance cassette are represented in red and in green respectively. In dark grey the sv40 polyA terminator. In orange the promoter sequence of BHLF1 with mutated ZREs. Restriction sites in dark blue for HindIII and PstI. (**C**) Schematic of the CIITA vector. Same as above, except the Ampicilin resistance cassette in green, and in dark grey the sv40 enhancer downstream of the luciferase gene. In blue the promoter sequence of CIITA. Restriction sites in dark blue for BamHI and SalI.

A



B

```
fosB_bis                CAACAGTCGGTGACAATGTGCACATCTGGTATTTCCTGTGTGAGAGCAAAACTACAGGAC      147
FMAF_pCpGL-primer       CAACAGTCGGTGACAATGTGCACATCTGGTATTTCCTGTGTGAGAGCAAAACTACAGGAC      300
                        ************************************************************

fosB_bis                TCAGTGAGTGGTTTCTGTGGTTTGCACTTGTGTATCTGGGTGTATGGGAGTGTGTAGCCC      207
FMAF_pCpGL-primer       TCAGTGAGTGGTTTCTGTGGTTTGCACTTGTGTATCTGGGTGTATGGGAGTGTGTAGCCC      360
                        ************************************************************
```

Figure 3.16 Sample sequencing and comparison for element FOSB. (**A**) Fragment of sequence chromatogram reading the sequence for one of the samples containing the FOSB element. (**B**) Fragment of comparison through clustal omega of the designed element FOSB (fosB_bis) against the sequence detected in the sample (FMAF_pCpGL-primer). The sequenced sample cleary consists of the designed sequence.

Table 3.2. Summary Table of transfections.

ELEMENT NAME	Design Criteria	Element length bp	Plasmid Name		
			CIITA	BHLF1-Mut	BHLF1
1OEM	The sequence of most resemblance to Each of the Matrices.	152	CIITA 1OEM	BHLF1-Mut 1OEM	BHLF1 1OEM
4OEM	The 4 sequences with the most resemblance to each of the matrices.	483	CIITA 4OEM	-	-
FOSB	3 Repeat-free clustered peaks near FOSB	718	-	BHLF1-Mut FOSB	BHLF1 FOSB
AHNAK	3 Repeat-free clustered peaks near AHNAK	720	-	BHLF1-Mut AHNAK	BHLF1 AHNAK

(continues from page 65) The elements were digested and ligated with the vectors, then transformed into *E. coli* competent cells as described in Materials & Methods. The cells were plated in selective agar and the selected samples were re-cultured in LB broth overnight, to extract DNA through a QIAGEN MINIPREP extraction kit. The samples were sequenced and compared against the designed elements (Figure 3.16).

Following the sequence verification, the samples were then cultured at a larger scale to extract larger amounts of DNA. The vectors containing the elements were transfected into EBV negative, BL DG75 suspension cells as well as in to EBV negative 293T epithelial adherent cells. A summary of the element transfections can be found in Table 3.2.

The transfections of constructs into each cell type and consecutive luciferase, BCA and western blot analyses were initially performed once. After that, 2 more transfections were attempted for a total of 3 different repetitions for each condition. However, when analysing the second and third repetitions, the values of the control conditions showed highly irregular behaviour, as well as finding data points off scale in most tested conditions. This lead us to think that human error occurred in the repetitions, due probably to the high number of samples for each experiment; therefore, the data and subsequent analysis of repetition #2 and #3 are omitted from the present document, and only the data for the first transfection is shown and discussed.

Each transfection was executed using either an expression vector for His tagged Zta (referred to simply as Zta) or an empty vector as a control (pcDNA). Western blots were performed verifying the expression of Zta for each transfection, the harvested cells were divided into two. One half was used to perform the Luciferase assays, measuring Relative Light Units, a reflection of the constructs activity; each sample was read by triplicate. The BCA assays were also performed from the same half of the sample as the luciferase assays by triplicate. The BCA assay reflects the protein content in the samples; the values from this assay were averaged and used to normalize the values from the luciferase assays (See section 2.2.7.2 in Materials and Methods). The other half of the samples was used for Western blots assays.

3.2.5.1.1 Is element AHNAK an enhancer in DG75 cells?

Due to the composition of the images and in the interest of the most possible clarity and the least confusion, figures 3.17 to 3.32 will be found at the end of the results section (pp 76-107), before the discussion of this chapter in page 104. The luciferase activity for BHLF1, increased 1.53 times when AHNAK was present compared to when it was absent (Figure 3.17 C first column). BHLF1 was activated by Zta as expected (155 times) (Figure 3.17 C second column), however the presence of AHNAK reduced activation by Zta to 38 times (Figure 3.17 C third column).

There is a significant reduction (0.38 times) in activity when we compare the activity of Zta when AHNAK is present (Figure 3.17 C last column). This reduced activation could be due to differences in Zta expression levels as there is slightly more Zta in cells with less luciferase activity (Figure 3.17 D and E).

In the context of BHLF1-Mut Zta has no activation, as expected (Figure 3.18 C second column). The presence of AHNAK in BHLF1-Mut made no difference on this lack of activation by Zta (Figure 3.18 C third column).

But the addition of AHNAK to BHLF1-Mut, enhanced promoter activity 5.28 fold (Figure 3.18 C last column). Although there is no significant change in the luciferase activation by Zta when AHNAK is present, the Zta band signal detected in the Western blot shows that the Zta levels in BHLF1-Mut AHNAK are slightly lower than the control sample (Figure 3.18 D and E).

3.2.5.1.2 Is element AHNAK an enhancer in 293T cells?

The luciferase activity in 293T cells for BHLF1 when AHNAK is present is 5.56 times higher than when it is absent (Figure 3.19 C first column).

The activation by Zta when there is no AHNAK element, is 518.72 (Figure 3.19 C second column) but when AHNAK is present in BHLF1, activation by Zta increases to 1654.08 (Figure 3.19 C third column).

When AHNAK is present in BHLF1, the activity with expression of Zta is 17.75 times higher than when AHNAK is not present (Figure 3.19 C last column). There is a little more Zta present in the sample (25.41 / 13.65 = 1.86 times more) in the Western blot assays compared to in the control sample (Figure 3.19 D and E).

However, when the ZREs are mutated, the enhancing effect disappears. The luciferase activity is slightly reduced when AHNAK is present in BHLF1-Mut (Figure 3.20 C first column).

There is a modest (0.67) activation by Zta when AHNAK is present in BHLF1-Mut (Figure 3.20 C last column), this is different from what is observed in DG75 cells. Expression of Zta in the Western Blot assay is similar.

3.2.5.1.3 Is element FOSB an enhancer in DG75 cells?

The possible enhancer effect of FOSB was tested by after transfections in DG75 cells through luciferase assays. The impact off adding FOSB to BHLF1 is a slight (0.35) decrease in activity (Figure 3.21 C first column).

As previously observed in section 3.2.5.1.1, Zta increases the activity of BHLF1 (155.53 fold). Albeit slightly less (117.9), this is also observed when FOSB is present (Figure 3.21 C third column). The abundance of Zta is the same in these experiments (Figure 3.21 D and E).

For BHLF1-Mut, the luciferase activity increased 1.94 times when the FOSB element was present compared to when it wasn't (Figure 3.22 C first column). Zta decreased activity in BHLF1-Mut (0.27) which was not expected (Figure 3.22 C second column). The presence of FOSB makes no difference over the activation of Z (1.10) (Figure 3.22 C third column).

The luciferase activity when Zta was expressed increased 7.9 times when FOSB was present compared to when it wasn't (Figure 3.22 C last column). However, the difference shown in the signal of the Zta bands in the Western blot is 1.5 times higher in the sample with the FOSB element than compared to the control (Figure 3.22 D and E), signaling for a higher presence of Zta in the sample which would be reflected in the increased luciferase activity.

3.2.5.1.4 Is element FOSB enhancing in 293T cells?

FOSB enhances promoter BHLF1 activity 2.37 times (Figure 3.23 C first column).

Zta increases the activity of BHLF1 (431.63), however when FOSB is present with BHLF1 the activation by Zta is 336.19 (Figure 3.23 C second and third column).

With Zta being expressed, the activity gets 1.73 times better when FOSB is present than without (Figure 3.23 C last column). The signal from the Zta band in the Western blot is 1.86 stronger in the sample with FOSB than the control sample; which could account for the difference in activity (25.4/13.64 = 1.86) (Figure 3.23 D and E).

For BHLF1-Mut the luciferase activity is 10.97 times higher with FOSB (Figure 3.24 C first column).

Luciferase activation by Zta, is 5.44 in BHLF1-Mut (Figure 3.24 C second column); but when FOSB is present it lowers to 2.90 (Figure 3.24 C third column), despite increased levels of Zta (Figure 3.24 D and E). This might point towards FOSB lowering the ratio of Zta activation 0.53 times compared to when the element is not present (2.90/5.44 = 0.53).

The luciferase activity when Zta is expressed, increases 5.85 times with the presence of the FOSB element (Figure 3.24 C last column). However, the signal from Zta in the Western blot is 4.35 times higher in the sample with FOSB than the control sample (73.31/16.82 = 4.35) (Figure 3.24 D and E).

3.2.5.1.5 Is 1OEM enhancing in DG75 cells?

The potential enhancing effect of 1OEM was tested through luciferase assays in DG75 cells. The luciferase activity of promoter BHLF1 is slightly reduced when 1OEM is present (Figure 3.25 C first column).

The activation of by Zta without 1OEM is 63.97 times higher than when Zta is not being expressed (Figure 3.25 C second column). This changes with the presence of 1OEM to 43 (Figure 3.25 C third column).12, yet with similar Zta levels in the western blot (Figure 3.25 D and E).

The luciferase activity when Zta is expressed is lowered 0.36 times when 1OEM is present compared to when it is absent (Figure 3.25 C last column).

The effect of 1OEM over the luciferase activity was also measured when the ZREs of the BHLF1 promotor were mutated. The luciferase activity found in BHLF1-Mut when the 1OEM element is present, is 1.97 times higher with 1OEM (Figure 3.26 C first column).

The luciferase activation by Zta in BHLF1-Mut is not changing (0.91, N.S. Figure 3.26 C second column), consistent with the mutation of the ZREs in BHLF1. However, when 1OEM is present with BHLF1-Mut, this is halved (0.51 Figure 3.26 C third column). Importantly, the signal detected in the bands of the WB (Figure 3.26 D and E); show a lower level of expression of Zta in the sample containing the 1OEM element.

The luciferase activity, when Zta is expressed, does not change in BHLF1-Mut with or without 1OEM (Figure 3.26 C last column).

When 1OEM is analysed on the promoter CIITA in DG75 cells, the luciferase activity is similar (0.75) than when the element is not present (Figure 3.27 C first column).

As expected and reflecting what was observed by Ramasubramanyan et al, there is no activation by Zta of CIITA (Figure 3.27 C second column). There is also no luciferase activation by Zta when 1OEM is present (Figure 3.27 C third column).

The luciferase activity when Zta is expressed is 0.43 times lower when 1OEM is present (Figure 3.27 C last column).

The Western blot shows a higher signal for Zta in CIITA 1OEM than in CIITA (Figure 3.27 D and E).

3.2.5.1.6 1OEM enhances in 293T cells?

There is 34.34 times more luciferase activity of promoter BHLF1 when 1OEM is present than compared to when it is absent (Figure 3.28 C first column).

There is activation by Zta in BHLF1 (71.59) as expected (Figure 3.28 C second column). When 1OEM is present, the activation by Zta is 45.66 (Figure 3.28 C third column), however it is important to see that there is a higher Zta signal in the WB sample that contains 1OEM (Figure 3.28 D and E). This can also be related to the 21.90 fold change of activation by Zta when 1OEM is present (Figure 3.28 C last column).

For BHLF1-Mut the luciferase activity when Zta is not being expressed increases 3.41 times when 1OEM is present (Figure 3.29 C first column).

Activation of BHLF1-Mut by Zta without 1OEM is similar (5.61) to when the 1OEM element is present (5.80), almost not changing (Figure 3.29 C second and third column).

In EBV negative 293T cells, when 1OEM is analyzed on the promoter CIITA, the luciferase activity is similar (0.98) to without 1OEM (Figure 3.30 C first column).

When Zta is present, the activity of CIITA is repressed (0.38) as expected and reflecting the findings from Balan et al (Figure 3.30 C second column). This repression by Zta is reversed when 1OEM is present (3.04) (Figure 3.30 C third column). Showing that when 1OEM is in the sequence, the activation by Zta increases 8 fold (3.04/0.38 = 8).

Luciferase activity when Zta is being expressed is considerably higher (7.92) in the control sample (Figure 3.30 C last column).

3.2.5.1.8 Is 4OEM enhancing in DG75 cells?

The luciferase activity of CIITA is 0.55 times lower in the sample with the 4OEM element (Figure 3.31).

The activation by Zta when there is no element, is 0.81 (reflecting the findings of Ramasubramanyan et al) (Figure 3.31 C second column), and the activation by Zta is almost the same to when the element is there (Figure 3.31 C third column).

The luciferase activity when Zta is expressed is decreased discreetly 0.80 when the element 4OEM is present. Overall, no impact can be observed when 4OEM is present (Figure 3.31 C last column).

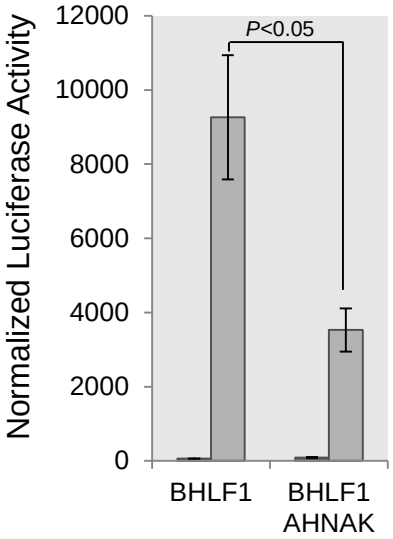
3.2.5.1.9 Is 4OEM enhancing in 293T cells?

The luciferase activity of CIITA is slightly higher (1.84 times) when the 4OEM element is present (Figure 3.32 C first column).

As expected and reflecting the findings by Balan et al, there is a 0.38 repression by Zta to CIITA (Figure 3.32 C second column). There is no repression by Zta when 4OEM is there (Figure 3.32 C third column). However, when taking into consideration the signal of Zta in the Western blot, we can see a stronger band for Zta in CIITA 4OEM than compared to the control, this points again towards 4OEM not altering the ratio of activation caused by Zta.

The luciferase activity with Zta is 5.19 times higher when 4OEM is there compared to when it is absent (Figure 3.32 C last column). Here is important to keep in mind that this might be an artificial increase of luciferase activity reflected in the higher signal of Zta found in the sample with the 4OEM construct in the Western blot assay.

A



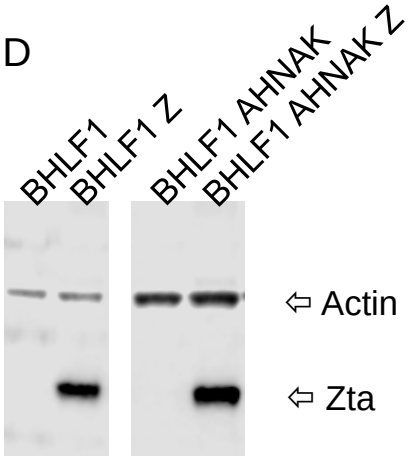
B

Transfection	AVERAGE	
	pcDNA	Z
BHLF1	59.54	9261.02
BHLF1 AHNAK	91.39	3528.72

C

	BHLF1 pcDNA	BHLF1 pcDNA	BHLF1 AHNAK pcDNA	BHLF1 Z
	BHLF1 AHNAK pcDNA	BHLF1 Z	BHLF1 AHNAK Z	BHLF1 AHNAK Z
Fold change	1.53	155.53	38.61	0.38
Sig.	P<0.05	P<0.05	P<0.05	P<0.05

D



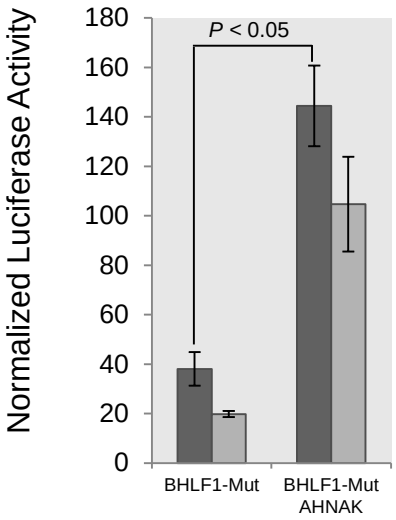
E

Transfection	Signal Value
BHLF1 Z	31.01
BHLF1 AHNAK Z	52.85

Figure 3.17 Influence over BHLF1 activity by AHNAK in DG75 cells. Comparison of normalized luciferase activity in DG75 cells with and without the AHNAK element. “BHLF1” represents (Figure legend continues…)

the control sample which contains ZREs. “BHLF1 AHNAK” represents the sample containing the construct AHNAK as well as ZREs. (**A**) Luciferase activity normalized to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student’s T-test provided P value, with an $\alpha=0.05$, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the AHNAK element is 1.7 times the value of the sample without the element.

A



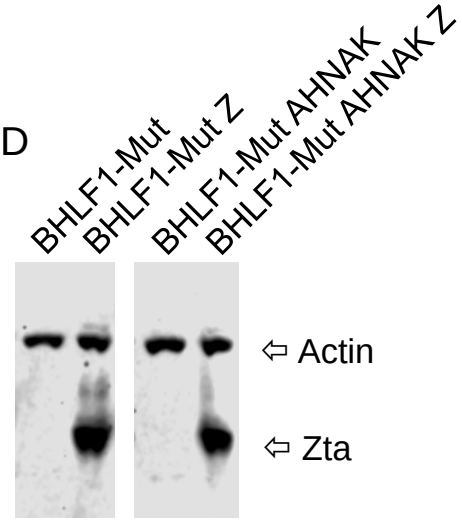
B

Transfection	AVERAGE	
	pcDNA	Z
BHLF1-Mut	38.08	19.82
BHLF1-Mut AHNAK	144.48	104.67

C

	BHLF1-Mut pcDNA	BHLF1-Mut pcDNA	BHLF1-Mut AHNAK pcDNA	BHLF1-Mut Z
	BHLF1-Mut AHNAK pcDNA	BHLF1-Mut Z	BHLF1-Mut AHNAK Z	BHLF1-Mut AHNAK Z
Fold change	3.79	0.52	0.72	5.28
Sig.	P<0.05	P<0.05	N.S.	P<0.05

D



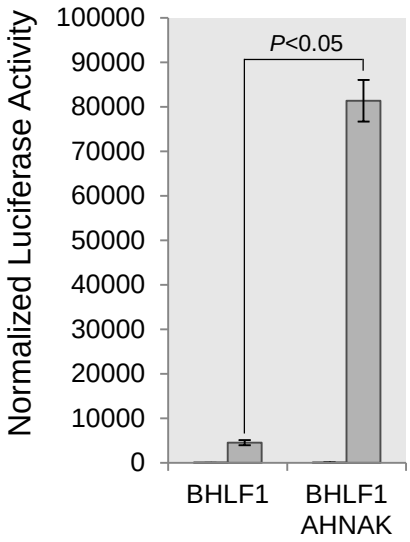
E

Transfection	Signal Value
BHLF1-Mut Z	139.34
BHLF1-Mut AHNAK Z	99.30

Figure 3.18. Influence over BHLF1 activity by AHANK when ZREs are mutated in DG75 cells. Comparison of normalized luciferase activity in DG75 (Figure legend continues…)

(cont…) after with and without the AHNAK element in absence of ZREs. “BHLF1-Mut” represents the sample without ZREs, used as control. “BHLF1-Mut AHNAK” represents the sample without ZREs yet containing the construct AHNAK. (**A**) Luciferase activity normalized to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student’s T-test provided P value, with an $\alpha=0.05$, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the AHNAK element is 0.71 times the value of the sample without the element.

A



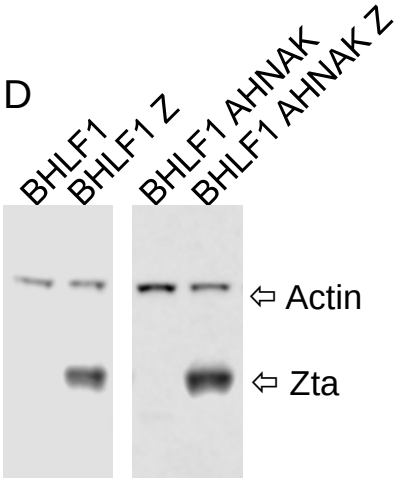
B

Transfection	AVERAGE	
	pcDNA	Z
BHLF1	8.83	4580.62
BHLF1 AHNAK	49.17	81336.01

C

	BHLF1 pcDNA	BHLF1 pcDNA	BHLF1 AHNAK pcDNA	BHLF1 Z
	BHLF1 AHNAK pcDNA	BHLF1 Z	BHLF1 AHNAK Z	BHLF1 AHNAK Z
Fold change	5.56	518.72	1654.08	17.75
Sig.	P<0.05	P<0.05	P<0.05	P<0.05

D



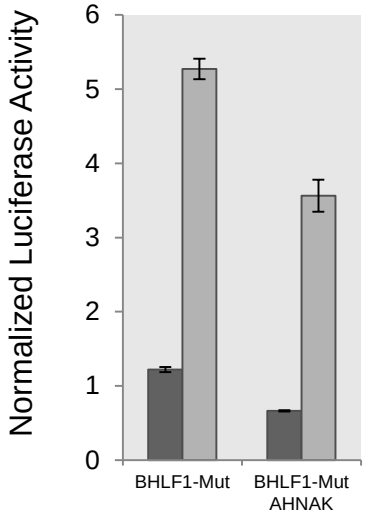
E

Transfection	Signal Value
BHLF1 Z	13.65
BHLF1 AHNAK Z	25.41

Figure 3.19. Influence over BHLF1 activity by AHNAK in 293T cells. Comparison of normalized luciferase activity in 293T cells with and without the AHNAK element. (Figure legend continues…)

(Cont…) “BHLF1” represents the control sample which contains ZREs. “BHLF1 AHNAK” represents the sample containing the construct AHNAK as well as ZREs. (**A**) Luciferase activity normalized to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student’s T-test provided P value, with an $\alpha=0.05$, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the AHNAK element is 1.86 times the value of the sample without the element.

A



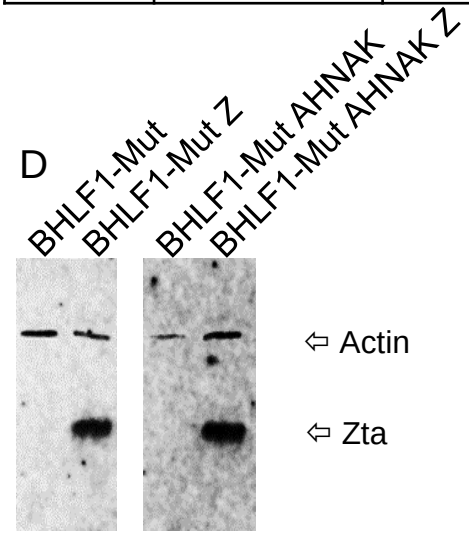
B

Transfection	AVERAGE	
	pcDNA	Z
BHLF1-Mut	1.22	5.27
BHLF1-Mut AHNAK	0.66	3.57

C

	BHLF1-Mut pcDNA	BHLF1-Mut pcDNA	BHLF1-Mut AHNAK pcDNA	BHLF1-Mut Z
	BHLF1-Mut AHNAK pcDNA	BHLF1-Mut Z	BHLF1-Mut AHNAK Z	BHLF1-Mut AHNAK Z
Fold change	0.54	4.32	5.38	0.67
Sig.	P<0.05	P<0.05	P<0.05	P<0.05

D



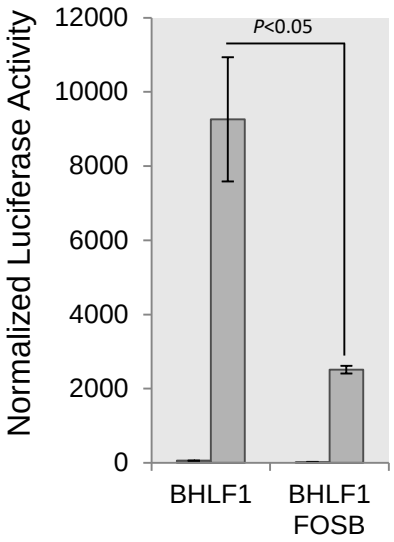
E

Transfection	Signal Value
BHLF1-Mut Z	0.62
BHLF1-Mut AHNAK Z	0.92

Figure 3.20. Influence over BHLF1 activity by AHNAK in 293T cells when ZREs are mutated. Comparison of normalized luciferase activity in 293T cells (Figure legend continues)

(Cont…) after transfection of the AHNAK element in absence of ZREs. “BHLF1-Mut” represents the sample without ZREs, used as control. “BHLF1-Mut AHNAK” represents the sample without ZREs yet containing the construct AHNAK. (**A**) Luciferase activity normalized to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student’s T-test provided P value, with an α=0.05, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the AHNAK element is 1.48 times the value of the sample without the element.

A



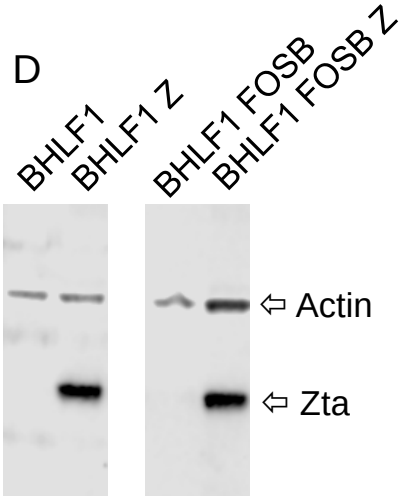
B

Transfection	AVERAGE	
	pcDNA	Z
BHLF1	59.54	9261.02
BHLF1 FOSB	21.32	2513.31

C

	BHLF1 pcDNA	BHLF1 pcDNA	BHLF1 FOSB pcDNA	BHLF1 Z
	BHLF1 FOSB pcDNA	BHLF1 Z	BHLF1 FOSB Z	BHLF1 FOSB Z
Fold change	0.35	155.53	117.90	0.27
Sig.	P<0.05	P<0.05	P<0.05	P<0.05

D



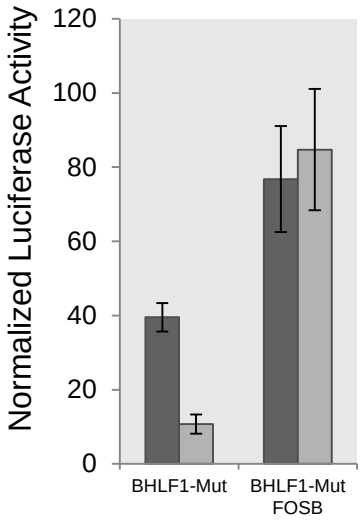
E

Transfection	Signal Value
BHLF1 Z	31.02
BHLF1 FOSB Z	32.53

Figure 3.21. Influence over BHLF1 activity by FOSB in DG75 cells. Comparison of normalized luciferase activity in DG75 cells with and without (Figure legend continues…)

(cont…) the FOSB element. “BHLF1” represents the control sample which contains ZREs. “BHLF1 FOSB” represents the sample containing the construct FOSB as well as ZREs. (**A**) Luciferase activity normalized to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student’s T-test provided P value, with an $\alpha=0.05$, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the FOSB element is 1.04 times the value of the sample without the element.

A



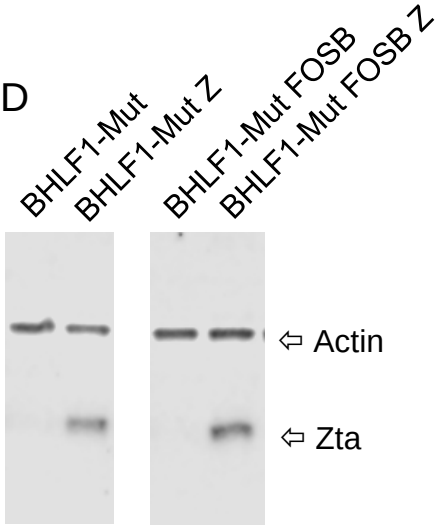
B

Transfection	AVERAGE	
	pcDNA	Z
BHLF1-Mut	39.54	10.72
BHLF1-Mut FOSB	76.79	84.71

C

	BHLF1-Mut pcDNA	BHLF1-Mut pcDNA	BHLF1-Mut FOSB pcDNA	BHLF1-Mut Z
	BHLF1-Mut FOSB pcDNA	BHLF1-Mut Z	BHLF1-Mut FOSB Z	BHLF1-Mut FOSB Z
Fold change	1.94	0.27	1.10	7.90
Sig.	P<0.05	P<0.05	N.S.	P<0.05

D



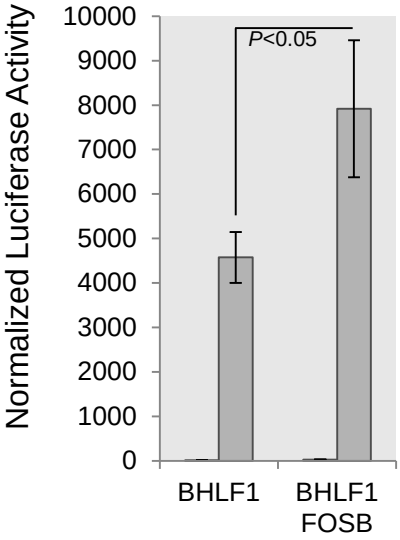
E

Transfection	Signal Value
BHLF1-Mut Z	6.86
BHLF1-Mut FOSB Z	10.06

Figure 3.22. Influence over BHLF1 activity by FOSB in DG75 cells when there are no ZREs available. Comparison of normalized luciferase activity in DG75 cells with and without the FOSB element in absence of ZREs. (Figure legend continues…)

(cont.…) “BHLF1-Mut” represents the sample without ZREs also used as a control. “BHLF1-Mut FOSB” represents the sample without ZREs yet containing the construct FOSB. (**A**) Luciferase activity normalized to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student’s T-test provided P value, with an α=0.05, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the FOSB element is 1.46 times the value of the sample without the element.

A



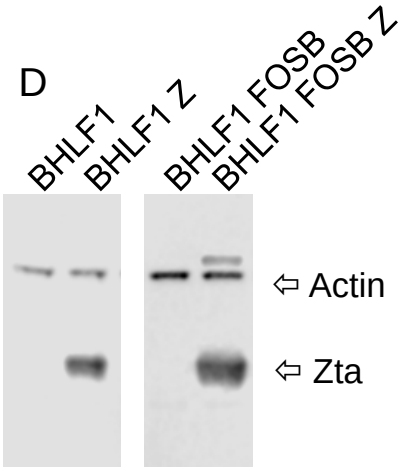
B

Transfection	AVERAGE	
	pcDNA	Z
BHLF1	9.91	4573.32
BHLF1 FOSB	23.55	7917.81

C

	BHLF1 pcDNA	BHLF1 pcDNA	BHLF1 FOSB pcDNA	BHLF1 Z
	BHLF1 FOSB pcDNA	BHLF1 Z	BHLF1 FOSB Z	BHLF1 FOSB Z
Fold change	2.37	461.63	336.19	1.73
Sig.	P<0.05	P<0.05	P<0.05	P<0.05

D



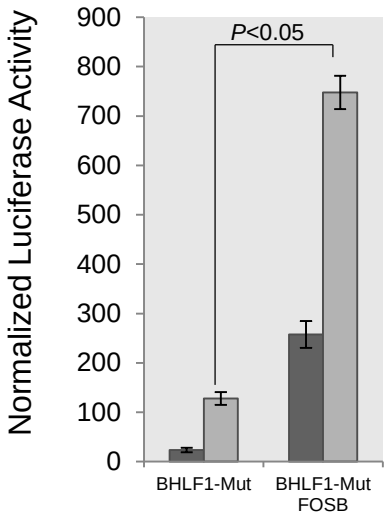
E

Transfection	Signal Value
BHLF1 Z	13.64
BHLF1 FOSB Z	25.4

Figure 3.23. Influence over BHLF1 activity by FOSB in 293T. Comparison of normalized luciferase activity in 293T cells with and without the FOSB element. (Figure legend continues…)

(cont.…) “BHLF1” represents the control sample which contains ZREs. “BHLF1 FOSB” represents the sample containing the construct FOSB as well as ZREs. (**A**) Luciferase activity normalized to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student’s T-test provided P value, with an $\alpha=0.05$, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the FOSB element is 1.86 times the value of the sample without the element.

A



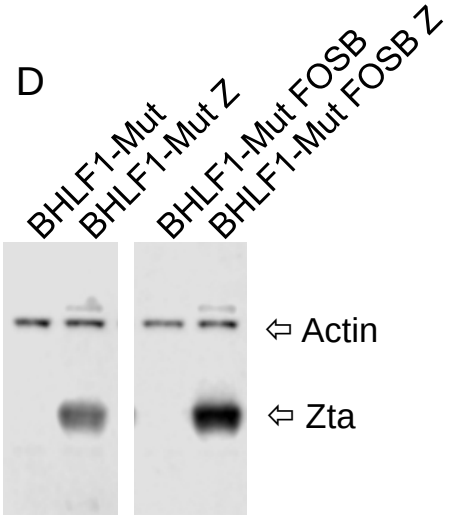
B

Transfection	AVERAGE	
	pcDNA	Z
BHLF1-Mut	23.49	127.75
BHLF1-Mut FOSB	257.73	748.01

C

	BHLF1-Mut pcDNA	BHLF1-Mut pcDNA	BHLF1-Mut FOSB pcDNA	BHLF1-Mut Z
	BHLF1-Mut FOSB pcDNA	BHLF1-Mut Z	BHLF1-Mut FOSB Z	BHLF1-Mut FOSB Z
Fold change	10.97	5.44	2.90	5.85
Sig.	P<0.05	P<0.05	P<0.05	P<0.05

D



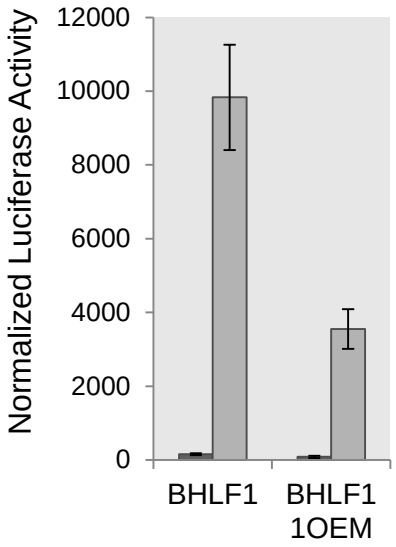
E

Transfection	Signal Value
BHLF1-Mut Z	16.82
BHLF1-Mut FOSB Z	73.31

Figure 3.24. Influence over BHLF1 activity by FOSB when ZREs are mutated in 293T cells. Comparison of normalized luciferase activity in 293T cells with and without (Figure legend continues…)

(Cont.…) the FOSB element in absence of ZREs. “BHLF1-Mut” represents the sample without ZREs also used as a control. “BHLF1-Mut FOSB” represents the sample without ZREs yet containing the construct FOSB. (**A**) Luciferase activity normalized to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student’s T-test provided P value, with an $\alpha=0.05$, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the FOSB element is 4.35 times the value of the sample without the element.

A



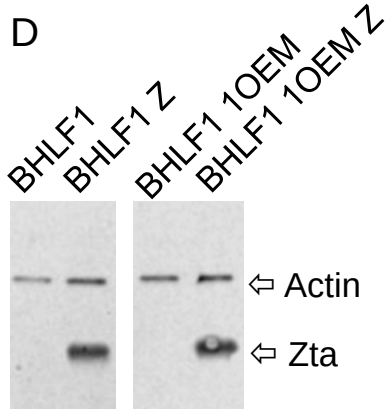
B

Transfection	AVERAGE	
	pcDNA	Z
BHLF1	155.89	9831.70
BHLF1 1OEM	85.86	3548.59

C

	BHLF1 pcDNA	BHLF1 pcDNA	BHLF1 1OEM pcDNA	BHLF1 Z
	BHLF1 1OEM pcDNA	BHLF1 Z	BHLF1 1OEM Z	BHLF1 1OEM Z
Fold change	0.55	63.97	43.12	0.36
Sig.	P<0.05	P<0.05	P<0.05	P<0.05

D



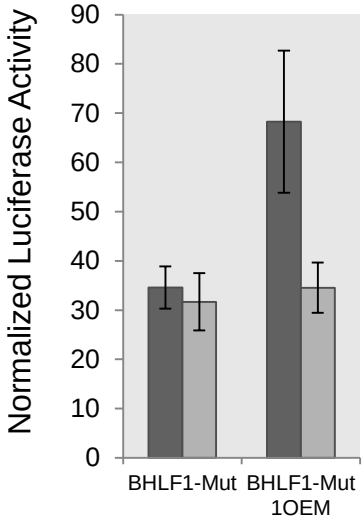
E

Transfection	Signal Value
BHLF1 Z	8.75
BHLF1 1OEM Z	9.19

Figure 3.25. Influence over BHLF1 activity by 1OEM in DG75 cells. Comparison of normalized luciferase activity in DG75 cells with and without the 1OEM element. “BHLF1” represents the control sample which contains ZREs. (Figure legend continues…)

(Cont.…) “BHLF1 1OEM” represents the sample containing the construct FOSB as well as ZREs. (**A**) Luciferase activity normalized to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student’s T-test provided P value, with an α=0.05, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the 1OEM element is 1.05 times the value of the sample without the element.

A



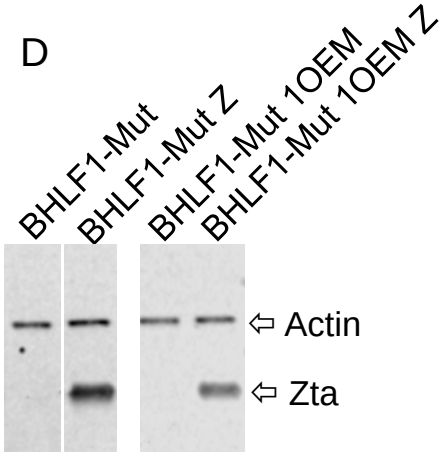
B

Transfection	AVERAGE	
	pcDNA	Z
BHLF1-Mut	34.59	31.67
BHLF1-Mut 1OEM	68.26	34.54

C

	BHLF1-Mut pcDNA	BHLF1-Mut pcDNA	BHLF1-Mut 1OEM pcDNA	BHLF1-Mut Z
	BHLF1-Mut 1OEM pcDNA	BHLF1-Mut Z	BHLF1-Mut 1OEM Z	BHLF1-Mut 1OEM Z
Fold change	1.97	0.91	0.51	1.09
Sig.	P<0.05	N.S.	P<0.05	N.S.

D



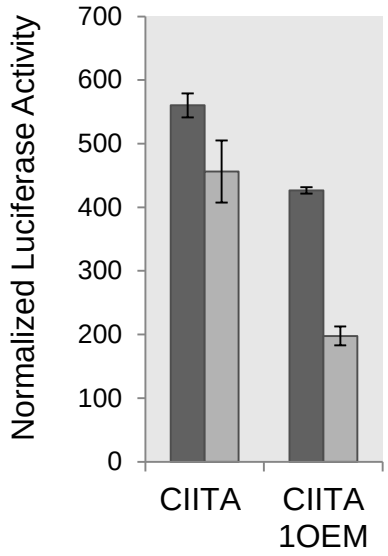
E

Transfection	Signal Value
BHLF1-Mut Z	13.08
BHLF1-Mut 1OEM Z	3.59

Figure 3.26. Influence over BHLF1 activity by 1OEM in DG75 when ZREs are mutated. Comparison of normalized luciferase activity in DG75 cells with and without the 1OEM element in absence of ZREs. “BHLF1-Mut” represents the sample (Figure legend continues…)

(cont…) without ZREs also used as a control. “BHLF1-Mut FOSB” represents the sample without ZREs yet containing the construct FOSB. (**A**) Luciferase activity normalized to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student’s T-test provided P value, with an $\alpha=0.05$, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the 1OEM element is 0.27 times the value of the sample without the element.

A



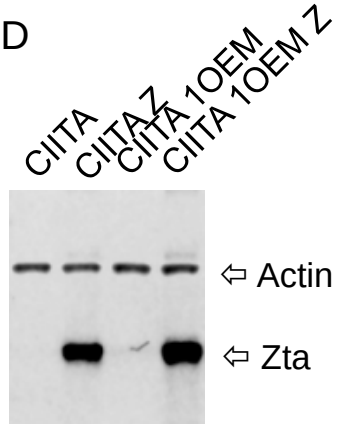
B

Transfection	AVERAGE	
	pcDNA	Z
CIITA	560.24	456.30
CIITA 1OEM	426.53	197.76

C

	CIITA pcDNA	CIITA pcDNA	CIITA 1OEM pcDNA	CIITA Z
	CIITA 1OEM pcDNA	CIITA Z	CIITA 1OEM Z	CIITA 1OEM Z
Fold change	0.75	0.81	0.46	0.43
Sig.	P<0.05	P<0.05	P<0.05	P<0.05

D



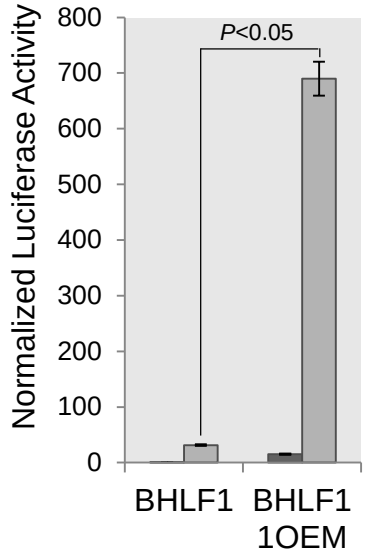
E

Transfection	Signal Value
CIITA Z	51.92
CIITA 1OEM Z	96.82

Figure 3.27. Influence over CIITA activity by 1OEM in DG75 cells. Comparison of normalized luciferase activity in DG75 cells with and without the 1OEM element. “CIITA” represents the sample used as a control. “CIITA 1OEM” represents the sample containing the construct 1OEM. (**A**) Luciferase activity normalized (Figure legend continues…)

(cont….) to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student's T-test provided P value, with an α=0.05, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the 1OEM element is 1.86 times the value of the sample without the element.

A



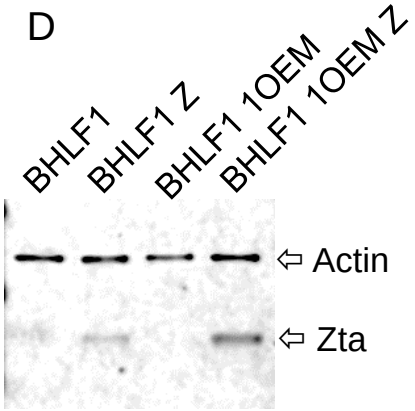
B

Transfection	AVERAGE	
	pcDNA	Z
BHLF1	0.44	31.50
BHLF1 1OEM	15.11	689.95

C

	BHLF1 pcDNA	BHLF1 pcDNA	BHLF1 1OEM pcDNA	BHLF1 Z
	BHLF1 1OEM pcDNA	BHLF1 Z	BHLF1 1OEM Z	BHLF1 1OEM Z
Fold change	34.34	71.59	45.66	21.90
Sig.	P<0.05	P<0.05	P<0.05	P<0.05

D



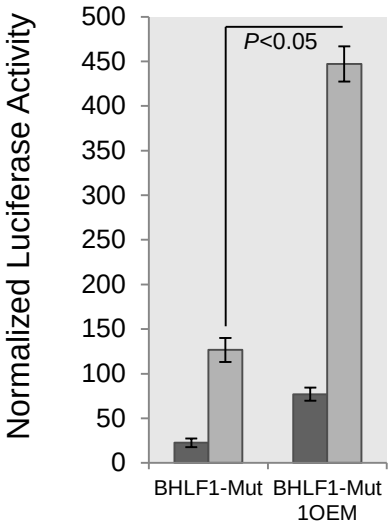
E

Transfection	Signal Value
BHLF1 Z	0.28
BHLF1 1OEM Z	0.62

Figure 3.28. Influence over BHLF1 activity by 1OEM in 293T cells. Comparison of normalized luciferase activity in 293T cells with and without the 1OEM element in absence of ZREs. “BHLF1-Mut” represents the sample without ZREs also used as a control. “BHLF1-Mut FOSB” represents the (Fig legend continues…)

(Cont.…) sample without ZREs yet containing the construct FOSB. (**A**) Luciferase activity normalized to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student's T-test provided P value, with an α=0.05, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the 1OEM element is 2.21 times the value of the sample without the element.

A



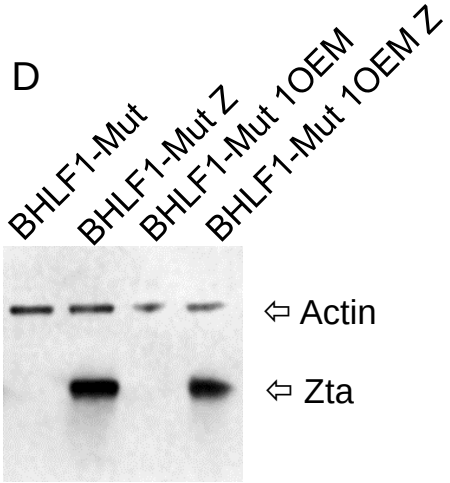
B

Transfection	AVERAGE	
	pcDNA	Z
BHLF1-Mut	22.55	126.65
BHLF1-Mut 1OEM	77.02	447.11

C

	BHLF1-Mut pcDNA	BHLF1-Mut pcDNA	BHLF1-Mut 1OEM pcDNA	BHLF1-Mut Z
	BHLF1-Mut 1OEM pcDNA	BHLF1-Mut Z	BHLF1-Mut 1OEM Z	BHLF1-Mut 1OEM Z
Fold change	3.41	5.61	5.80	3.52
Sig.	P<0.05	P<0.05	P<0.05	P<0.05

D



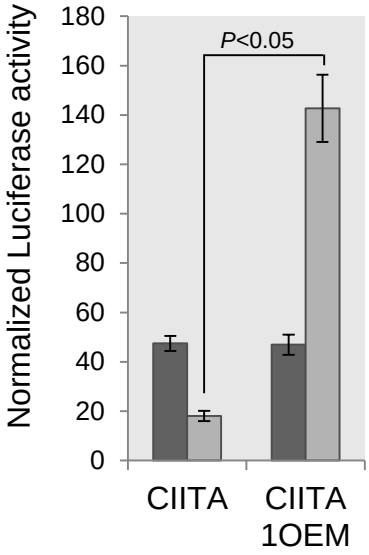
E

Transfection	Signal Value
BHLF1-Mut Z	223.56
BHLF1-Mut 1OEM Z	110.74

Figure 3.29. Influence over BHLF1 activity by 1OEM in 293T when ZREs are mutated. Comparison of normalized luciferase activity in 293T cells with and without the 1OEM element in absence of ZREs. “BHLF1-Mut” represents the sample (Figure legend continues…)

(cont.…) without ZREs also used as a control. “BHLF1-Mut FOSB” represents the sample without ZREs yet containing the construct FOSB. (**A**) Luciferase activity normalized to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student’s T-test provided P value, with an $\alpha=0.05$, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the 1OEM element is 0.49 times the value of the sample without the element.

A



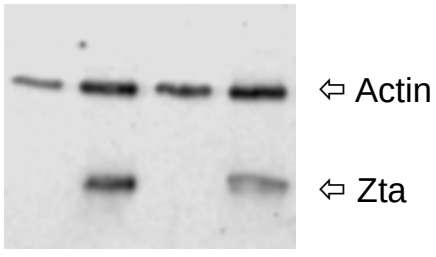
B

Transfection	AVERAGE	
	pcDNA	Z
CIITA	47.45	18.01
CIITA 1OEM	46.93	142.74

C

	CIITA pcDNA	CIITA pcDNA	CIITA 1OEM pcDNA	CIITA Z
	CIITA 1OEM pcDNA	CIITA Z	CIITA 1OEM Z	CIITA 1OEM Z
Fold change	0.98	0.38	3.04	7.92
Sig.	N.S.	P<0.05	P<0.05	P<0.05

D



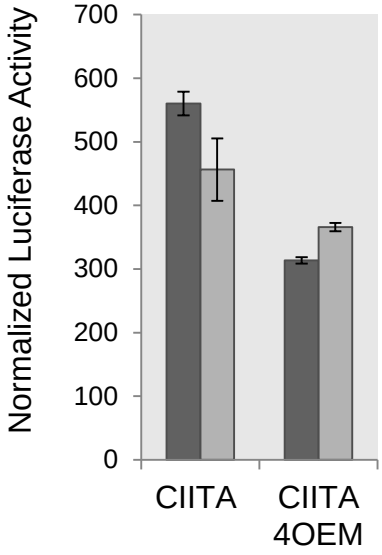
E

Transfection	Signal Value
CIITA Z	4.26
CIITA 1OEM Z	2.25

Figure 3.30 Influence over CIITA activity by 1OEM in 293T. Comparison of normalized luciferase activity in 293T cells with and without the 1OEM element. “CIITA” represents the sample used as a control. “CIITA 1OEM” represents the sample containing the construct 1OEM. (**A**) Luciferase activity (Figure legend continues…)

(Cont.…) normalized to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student's T-test provided P value, with an α=0.05, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the 1OEM element is 0.53 times the value of the sample without the element.

A



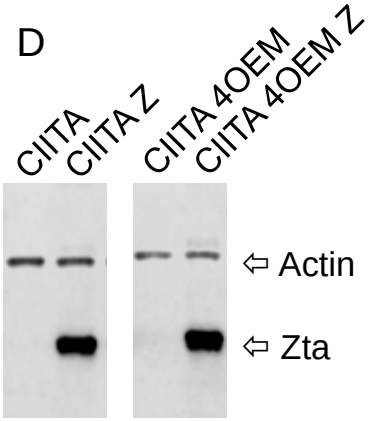
B

Transfection	AVERAGE	
	pcDNA	Z
CIITA	560.24	456.30
CIITA 4OEM	313.71	365.89

C

	CIITA pcDNA	CIITA pcDNA	CIITA 4OEM pcDNA	CIITA Z
	CIITA 4OEM pcDNA	CIITA Z	CIITA 4OEM Z	CIITA 4OEM Z
Fold change	0.55	0.81	1.16	0.80
Sig.	P<0.05	P<0.05	P<0.05	P<0.05

D



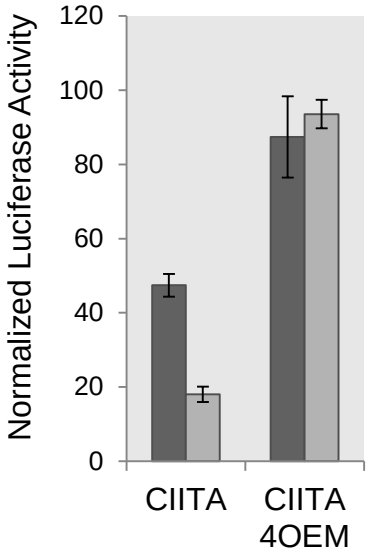
E

Transfection	Signal Value
CIITA Z	51.92
CIITA 4OEM Z	65.41

Figure 3.31. Influence over CIITA activity by 4OEM in DG75 cells. Comparison of normalized luciferase activity in DG75 cells with and without the 4OEM element. “CIITA” represents the sample used as a control. “CIITA 4OEM” represents the sample containing the construct 1OEM. (**A**) Luciferase activity normalized (Figure legend continues)

(cont.…) to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student's T-test provided P value, with an α=0.05, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the 4OEM element is 1.25 times the value of the sample without the element.

A



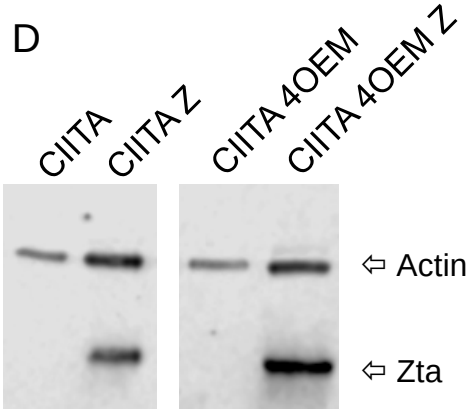
B

Transfection	AVERAGE	
	pcDNA	Z
CIITA	47.45	18.01
CIITA 4OEM	87.41	93.58

C

	CIITA pcDNA	CIITA pcDNA	CIITA 4OEM pcDNA	CIITA Z
	CIITA 4OEM pcDNA	CIITA Z	CIITA 4OEM Z	CIITA 4OEM Z
Fold change	1.84	0.38	1.07	5.19
Sig.	P<0.05	P<0.05	N.S.	P<0.05

D



E

Transfection	Signal Value
CIITA Z	4.26
CIITA 4OEM Z	15.11

Figure 3.32. Influence over CIITA activity by 4OEM in 293T. Comparison of normalized luciferase activity in 293T cells with and without the 4OEM element. “CIITA” represents the sample used as a control. “CIITA 4OEM” represents the sample containing the construct 4OEM. (**A**) Luciferase activity normalized (Figure legend continues…)

(Cont…) to BCA levels. Dark grey bars represent the normalized luciferase value for samples transfected with plasmid DNA not containing the Zta gene. Light grey represent the value for samples transfected with plasmid containing the Zta gene. Error bars calculated from the Standard deviation from 3 readings for each sample. (**B**) Table of values for (A). (**C**) Table showing the difference in the fold change between samples and its significance; values from (B). Student's T-test provided P value, with an $\alpha=0.05$, in equal variances. (**D**) Western blot of transfected samples presented in (A). (**E**) Signal from the bands corresponding to Zta in (D). The signal of the Zta band corresponding to the sample with the 1OEM element is 3.55 times the value of the sample without the element.

3.3 Discussion

Research performed previously in our lab showed that Zta binds to the human genome across many regions. Some of these interactions can be found proximal (< 2kb) to the transcription start site of genes that have been shown to be upregulated in lytic cycle when analyzed through RNA-Seq analysis, and ranked based on the change in upregulation (Ramasubramanyan et al. 2015). This suggested that there might be elements in the sequences to which Zta is bound, which enhance its activity; sequences apart from ZREs. Therefore, the sequences upstream of the top 30 upregulated genes (found in the previous research by Ramasubramanyan et al), were studied, and short regions that are shared across the analyzed sequences were found. Two elements (1OEM and 4OEM) were designed and synthetized based in this, and their ability to enhance expression was studied.

Parallel to this, Zta peaks from the ChIP-Seq analysis were tested to be enhancers; the sequence of these peaks was close to genes known to be upregulated in lytic cycle, therefore two elements were designed with the sequence of these peaks and named after the gene close to those peaks (FOSB and AHNAK).

Different analyses of the aforementioned sequences were carried out to generate plasmids that would be composed of the elements with a potential enhancing activity. The exploration on this enhancer activity was performed by testing *in vitro* the change in activity in Luciferase assays.

It is important then to consider that different molecular environments exist in the cell lines used for the experiments, and this in turn can be observed in the activity of the control samples. In DG75 cells, there was no repression of CIITA by Zta which is something observed as well by Ramasubramanyan et al (2015) in the same cell line. When Balan et al (2016) tested the capacity of Zta to repress the activation of CIITA in EBV negative BL Akata cells, a clear repression was present in the promotor even when there were no ZREs available; the same was found in this research in 293T cells. A similar cell line difference can be seen in the lack of activation by Zta of BHLF1-Mut in DG75 cells (Figs 3.18, 3.22 and

3.26), and a small activation by Zta of BHLF1-Mut in 293T cells (Figs 3.20, 3.24 and 3.29).

Regarding the designed elements enhancing activity; when the AHNAK element was incorporated into promoters and transfected, an enhancer effect was found in both DG75 and 293T cells when the element was adjacent to the BHLF1 promoter. There was no additional increase in expression with AHNAK when Zta was expressed in DG75 cells, but a small Zta driven enhancement was seen in 293T cells.

The FOSB element showed cell specific enhancer activity in 293T cells when placed next to BHLF1-Mut, but this was not the case in DG75 cells; there was no additional change when Zta was expressed. In contrast, the element showed repression in DG75 cells.

The 1OEM element had little impact on the BHLF1 promoter in DG75 cells. However, in 293T cells the element acted as an enhancer on the BHLF1 promoter, independently of Zta, and was able to reduce the Zta driven repression of the CIITA promoter in these cells.

Element 4OEM was also able to overcome the repression of the Zta promoter by Zta in 293T cells.

Overall, these findings show that the elements representative of peaks where Zta is binding to (for AHNAK and FOSB), as well as the elements representative of conserved sequences and motifs (for 1OEM and 4OEM), both have cell specific enhancer ability; however, if the effects of each element are considered collectively across the experiments, the enhancing activity found is minor and not substantial. Albeit this, the computational approach devised to generate the elements worked as an appropriate tool for the design and analysis of sequences involved with enhancing activity, given that 4 different constructs were designed and tested.

It is also important to acknowledge that the experiments were not successfully repeated and therefore the observed behavior of the enhancers was not reproduced, however, the activity of the controls in the presented data mirror findings from the works of Balan et al as well as Ramasubramanyan et al.

It bears considering that the analysis used to generate the two of the elements tested, did not consist of sequences exclusively beneath Zta peaks (in 1OEM and 4OEM). This means that the consensus sequences found to be pervasive might be representing an enhancer sequence that responds to other transcription factors than Zta. This speculation could address the increases in luciferase activity when there were lower levels of Zta being expressed (Fig 3.26 and 3.29), or with no expression of Zta at all (Fig 3.18 and 3.22); or when the presence of an enhancing element in 293T did not enhance in DG75, perhaps other transcription factors are occupying the element sequence but not increasing activation (Fig 3.17 and 3.21).

Further analysis could be carried out by testing if the enhancer capability remains after removing components of the elements 1OEM and 4OEM; since those elements are composed of compilations of the most representative sequences reflecting the matrices, it would be useful to test the enhancer capability of single matrix instances free from ZREs, similar to how Balan et al showed how Zta had an effect over CIITA promoter without key ZREs.

Chapter 4. Potential looping interactions of Zta with promoters

4.1 Introduction

The knowledge and number of the different factors that play an important role in transcriptional regulation in general has been increasing and expanding as technology facilitates the study of different models of regulation. Although it has been a subject of research in the last decades, the full extent of knowing how enhancers work and regulate gene expression has not been reached.

The understanding of the importance of; cis acting transcription factors, protein-protein modifications to TFs, the composition of binding motifs near to TSS, has been complemented with the advent of findings showing a higher level of complexity in the transcriptional regulation by; acetylation or methylation of histones, DNA looping or inter-chromosomal interactions (reviewed by (Lelli et al. 2012) and (Mercer and Mattick 2013)).

Chromosome conformation capture is one of the new techniques that have expanded the understanding of the three-dimensional events taking part in transcriptional regulation; it was first used to study interactions between distant regions in in *Saccharomyces cerevisiae* (Dekker et al 2002). It requires the production of a template DNA library and a control DNA library which will be used to compare and test for the presence of particular PCR products of a set length (this is expanded in the introduction of chapter 5). However, since the first necessary step is to postulate which interactions are occurring, in this chapter we will focus solely in these potential interactions.

Genomic elements that are located far away from each other come in close contact by the formation of chromatin loops, these loops can be detected by 3C experiments (reviewed by (Denker and de Laat 2016)). Since in previous analysis it was found that Zta can regulate gene expression distally (Ramasubramanyan et al. 2015), I asked whether Zta binding at distal sites promotes the formation of chromatin loops to neighbouring promoters.

Previously, using data from ChIP-seq analysis as a point of departure, research by McClellan et al showed binding sites of EBNA2 and EBNA3 genes to be looping to promoter regions in EBV positive BL tumour cells (MutuIII) (McClellan et al. 2012; McClellan et al. 2013). Over 25% of the binding sites of EBNA2 and

EBNA3s were found to be shared; high levels of acetylation of the 27th lysine residue of histone 3 (H3K27ac) helped distinguish EBNA2 binding sites as active enhancer sites, while low levels of acetylation showed low level of gene activation. This is something that resonated through the analysis of Wood et al, where they found that EBNA2 targets tumour-associated super-enhancers at MYC. These super-enhancers are large lineage-specific regulatory elements typically comprising multiple transcription factors binding sites; and the super-enhancers upstream of MYC were distinguished by marks of acetylation in H3K27, regions which were both occupied by EBNA2 and looped to the promotor region of MYC (McClellan et al. 2013; Wood et al. 2016).

The mark of H3K27ac has also been linked to an active EBV C promoter (Cp) stimulated by EBNA2 (reviewed by (H. S. Chen et al. 2013)). Perhaps more importantly, these marks of acetylation were consistent with looping events between regions found through 3C experiments, making it clear that marks of H3K27ac are important when considering potential loops (McClellan et al. 2013; Wood et al. 2016).

Although the 3C technologies have evolved recently, the technique is still used to explore “one to one” type of interactions (reviewed by (Denker and de Laat 2016)). The initial step was to select “one to one” interactions of Zta to promoters that could potentially be forming loops (this will be expanded in Section 4.1.1). When selecting which interactions to explore, the position of restriction sites had to be regarded, this since it is one of the most important considerations of the 3C experiment design. If the fragment sizes were smaller than 1Kb or greater than 10 Kb in length, differences in the efficiency of ligation/amplification could bring biases to the data. In a similar way, the position of the restriction sites should not digest within the sequence of the promoter or potential enhancer (reviewed by (Naumova et al. 2012)).

Finally, as the analysis progressed, the idea arose of selecting which interactions to include in the 3C experiments based on grouping the genes into possible clusters. This idea borrows from diverse studies that detected groups of genes to be found in relative proximity to each other, while sharing a specific regulated activity (Fraser and Bickmore 2007; McClellan et al. 2013; Montavon and

Duboule 2013; Osborne et al. 2004; Shopland et al. 2003). Part of the analysis employed data previously generated in the Sinclair lab, mRNA-seq analysis found genes to be highly regulated in lytic cycle in Akata cells (Ramasubramanyan et al. 2015); this allowed to rank the genes based on the fold change in regulation levels.

4.2 Results

4.1.1 Identification of potential looping targets between Zta binding sites and promoters

The Zta peaks found in the ChIP-seq experiments were sometimes found distally from genes; this helped formulate the question; Is Zta forming loops to promoter regions of genes known to be upregulated in lytic cycle? Which Zta interactions could be candidates for looping events?

In order to answer the first question, it is necessary to answer the second one. Aiming to discriminate possible Zta interactions that would be good candidates for looping activity, two different strategies were implemented. One selection criteria analysed the Zta interactions with the highest regulated genes. Then, another selection criteria grouped the aforementioned upregulated genes into clusters.

4.1.1.1 Selection Criteria

4.1.1.1.1 Highest regulated genes

Initially, the Zta interactions considered to be looping were selected by a direct and simple criteria. Genes found to be upregulated more than twofold in previous mRNA-seq analysis (Ramasubramanyan et al. 2015) were sorted in a list based on the fold change of transcription in lytic cycle.

From this list of genes, the 30 genes that are most regulated were visualized through the UCSC genome browser, with a view that included; any Zta peaks found from 2kb to 200 kb away from the canonical TSS, the promoter sequence of each gene, and the restriction sites for BamHI and HindIII. Acetylation of histone 3 in the 27^{th} lysine residue was also included to help consider potential enhancers regions if Zta peaks were found in alignment with such feature.

Following this, three genes were selected based on restriction sites not fragmenting promoter sequences or peaks, and the particular characteristics of each gene:

Gene LHX1, upregulated 5546 fold in lytic cycle, with 3 Zta peaks found in NPC cells, two of them in places with marks of acetylation, and no gene between the

peaks and LHX1 promoter. The peak closest to LHX1 was detected in NPC cells, while the farthest two from LHX1 was detected in both NPC and BL cells, however Figure 4.1 A shows all the peaks considered for the selection, in NPC cells.

Gene SLC6A7, upregulated 2510 fold in lytic cycle, with two Zta peaks found in both NPC cells and BL cells, one of them over a modest mark of acetylation (Figure 4.1 B).

Gene FZD10, upregulated 1420 fold in lytic cycle, found with no other genes in relative proximity to it, and with 3 Zta peaks found in both NPC cells and BL cells surrounding it, one of them aligned to acetylation marks (Figure 4.1 C).

4.1.1.2.1 Upregulated gene clusters

Subsequently, a different selection criteria was devised for selecting the Zta interactions to identify potential loops. Using the 500 highest regulated genes, a BED file (Browser Extensible Data) was retrieved from the UCSC genome browser; this file was composed of each gene's canonical name, chromosomal position, position start, and position end (Figure 4.2).

This BED file was very useful to group the genes in clusters of two or more, with a maximum distance between them of 100 kb; this choice was made considering any distance larger than 100 kb too distant to have a shared influence or activity. This discrimination was accomplished by first sorting the genes by chromosomal position, then by gene start position. After this the value of the end position of gene A (Figure 4.2 B, "*b*") was subtracted from the value of the start position of gene B (Figure 4.2 B "*c*"); if the resulting value was less than or equal to 100 kb (c-$b \leq 100$ kb), the genes were considered a cluster. This analysis resulted in finding 29 clusters, 6 of them with a distance between genes closer than 10 kb and 23 clusters with a distance between 10 and 100 kb between genes. (Figure 4.2 C).

Even though this last analysis helped greatly to focus which genes to study, a subsequent analysis was devised to search if there were any Zta peaks in proximity (up to 100 kb away) to the clusters. After obtaining the BED file which mark the location of the Zta peaks in the genome (Akata BL (Ramasubramanyan et al. 2015)), the same sorting (continues in page 118)

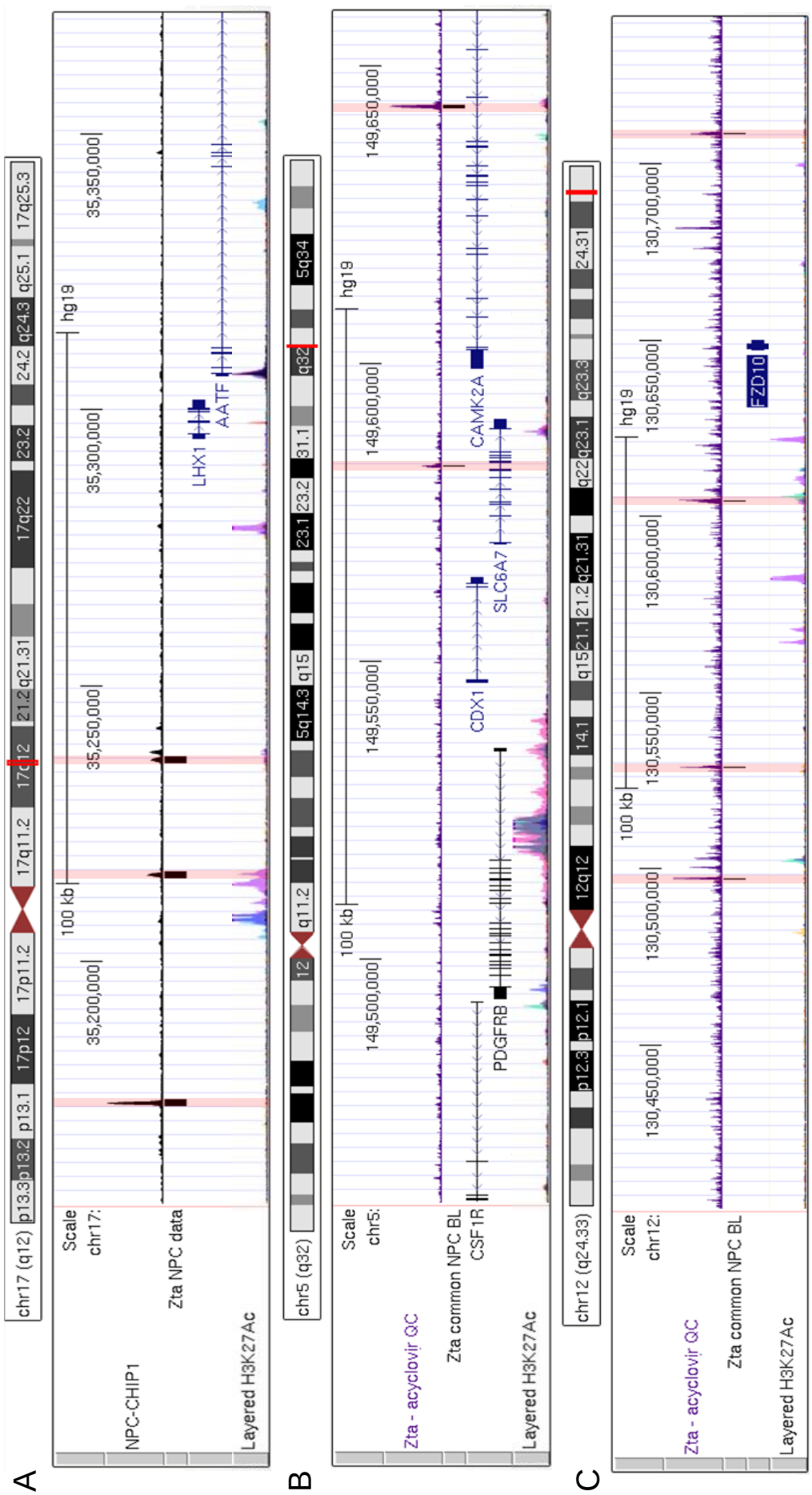
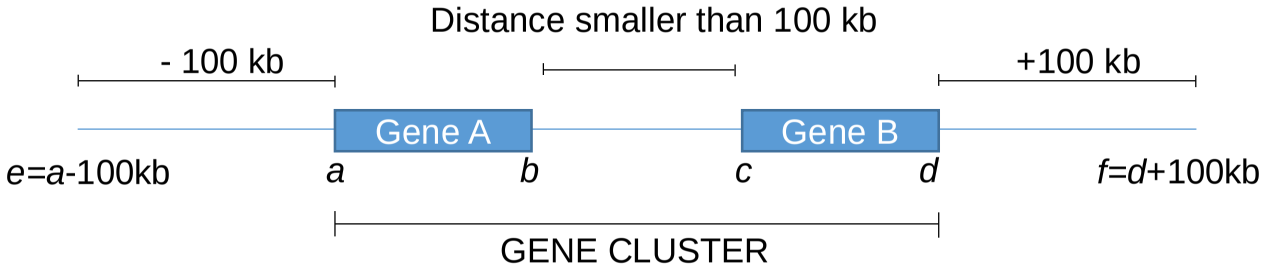


Figure 4.1. Genes selected for analysis. Visualization on the genome browser of genes with adequate Zta peaks, restriction sites can not be appreciated at this scale. The peaks are highlighted in pink (**A**) LHX1 (**B**) SLC6A7 (**C**) FZD10.

A

	A	B
2	geneID	fold change
318	HSPA6	58.11
319	KIAA1549	57.99
320	PLIN2	57.92
321	VWF	57.66
421	CD9	30.22
422	FAM105A	30.21
423	STXBP5L	30.19
424	PTPN21	30.13
425	SDSL	29.66

B



C

	A	B	C	D	G	H	J	M	N	O
1	chromosome	gene start	gene end	gene name	Highest value	Lowest value	distance (bp) between neighbouring genes	combination	lowest value of cluster -100000	highest value of cluster + 100000
99	chr12	1901122	2027870	CACNA2D4	2027870	1901122	4,030,169			
100	chr12	6058039	6233836	VWF	6233836	6058039	75,645	VWF and CD9	5958039	6447437
101	chr12	6309481	6347437	CD9	6347437	6309481	2,872,866			

Figure 4.2. Genes were grouped into clusters. In order to exemplify the process, figure panes will follow genes VWF and CD9. (**A**) Fragment of the list of highly regulated genes. (**B**) Visual representation of a gene cluster, *a* represents the start of Gene A, *b* represents the end of gene A, *c* and *d* represent the start and end of gene B respectively. See section 4.1.1.2.1 in text for details. (**C**) Fragment of BED file of genes position and details arranged in a single row for each gene. The distance between genes was obtained by subtracting the highest value (*b*) of VWF to the lowest value (*c*) of CD9. The column titled combination performed an "if check" for the distance between genes to be equal or lower than 100 kb; if yes, the name of the cluster was generated by combining the gene names, if not the cells were left blank. In column N a similar if check was performed, where the start of gene A (*a*) was subtracted 100 kb (*e*), and the end of gene B (*d*) was added 100kb(*f*).

(continues from page 115) of features was programmed as in the previous section. The position of each single Zta peak was listed in a vertical list, sorted by ascending order of chromosome position, then, in the same sheet the previously found clusters were listed horizontally by ascending order of chromosome position (Figure 4.3). The resulting shared area contained cells that were programmed to perform an “if” logical test checking for the Zta peak chromosome position to match the chromosome position of the cluster on top. If they were indeed the same, then another “if” logical test was performed, if the midpoint of the peak could be found within the span (±100kb) of a gene cluster (*e* > peak midpoint < *f*) then the cell would be filled with the “HERE” message, otherwise it would be left blank; cell content =IF(AND(peak chromosome = cluster chromosome, *e*>peak midpoint, peak midpoint<*f*),"HERE",""). Then each cluster column was filtered in order to display only cells signalling for a match, in order to explore the peaks near the cluster in the genome browser (Figure 4.3).

This selection criteria helped focus the Zta interactions suspected to be interacting with gene clusters to only eight gene clusters (Table 4.1). To select which cluster genes to carry further in the analysis, the peak height, the alignment with acetylation and the presence of the Zta peak in ChIP-seq in a different cellular environment than BL cells was considered. Three different clusters were selected: SIRPA and PDYN, PDGFRB and SLC6A7, and VWF and CD9 (Figure 4.4). None of the other genes visualized in Figure 4.4 are regulated by Zta. (continues in page 121)

Table 4.1. Eight gene clusters with Zta peaks less than 100 kb away from TSS.

7 gene clusters with Zta peaks less than 100 kb away		
Name of cluster	# of peaks	Peak score value
VWF and CD9	2	219, 667
ANXA2 and RORA	3	77, 196, 125
SIRPA and PDYN	1	145
PLCB4 and LAMP5	1	158
ATP9A and SALL4	7	104, 98, 73, 154, 76, 145, 77
PDGFRB and SLC6A7	2	219, 730
SHROOM2 and WWC3	1	176

1 gene cluster of a span smaller than 10 kb with peaks less than 100 kb away		
Name of cluster	# of peaks	Peak attributed value
FCGR2A and HSPA6	1	106

A

	A	B	C	D	E	F	G	K	L	M	N
2	peak position						peak	chr1	chr11	chr12	chr13
3	chromosome	peak start	peak end	peak name	peak value	peak half spa	mid point	FCGR2A and HSPA6	ZC3H12C and RDX	VWF and CD9	METTL21C and KDELC1
900	chr12	5905266	5905735	MACS_peak_897	387.99	235	5905501				
901	chr12	6032303	6032589	MACS_peak_898	218.78	143	6032446			HERE	
902	chr12	6141807	6142330	MACS_peak_899	667.13	262	6142069			HERE	
903	chr12	6588286	6588519	MACS_peak_900	253.42	117	6588403				

B

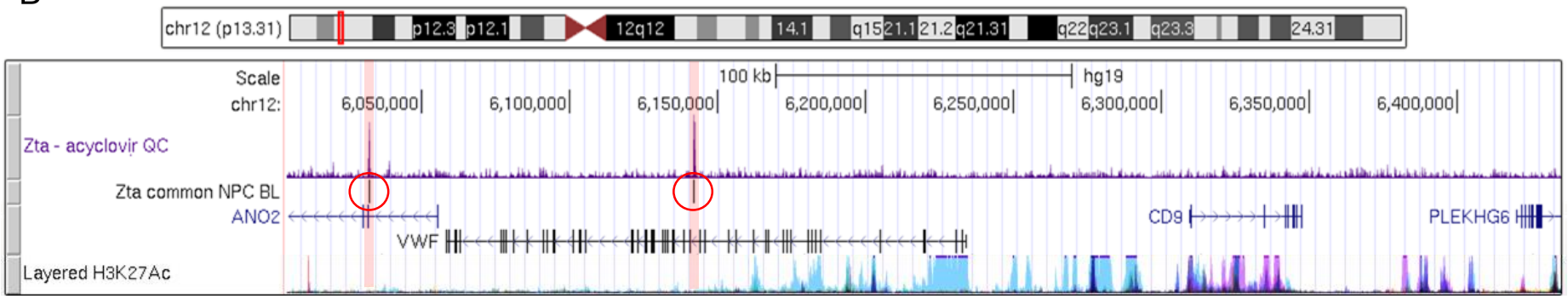


Figure 4.3 . **Zta peaks were found within 100 kb from gene clusters.** The position of the Zta peaks was compared to the position of the clusters. (**A**) Fragment of the data in excel, filtered and sorted. The column labelled peak value shows a numerical value reflecting the reads and height of the peak. Column labeled “peak half span” shows a value obtained from dividing between two, the subtraction of the peak end minus the peak start. Column labeled “peak mid point“ was obtained by adding the peak half span to the peak start. From column H onwards the clusters were listed however only data surrounding peaks in cluster VWF and CD9 is shown; two peaks within 100 kb of cluster VWF and CD9 were found. See section 4.1.1.2.1 in text for details (**B**) Both Zta peaks can be visually located in the UCSC genome browser, circled in red.

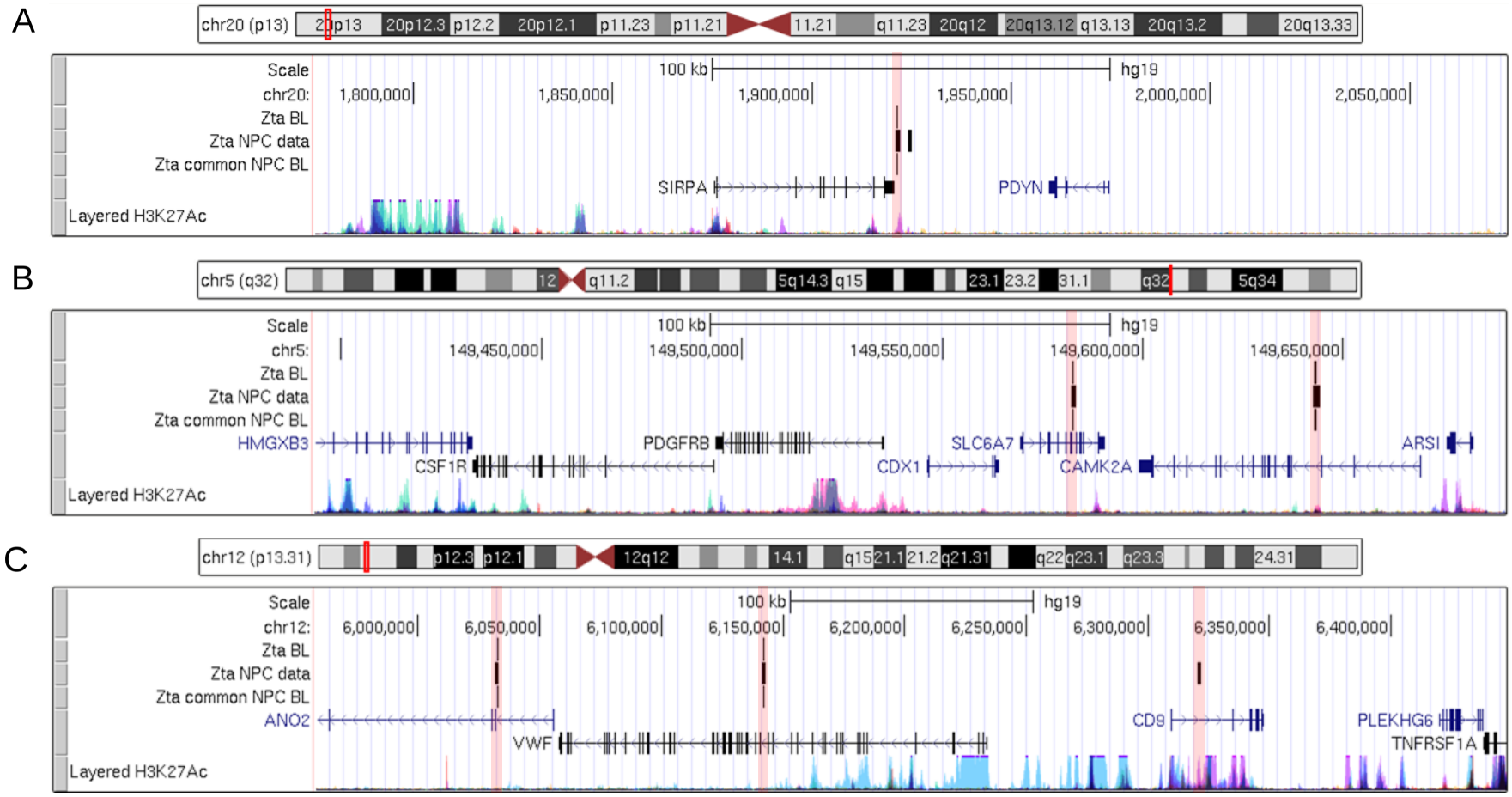


Figure 4.4 . Clusters selected. View on the UCSC Genome browser of the clusters selected. The track “Zta common NPC BL” shows the peaks found both in NPC cells and BL cells, peaks highlighted in pink. (**A**) SIRPA and PDYN present a peak aligned to acetylation. (**B**) PDGFRB and SLC6A7 present Zta peaks with high scoring value . (**C**) VWF and CD9 have two Zta peaks common to NPC and BLs of high scoring value.

(continues from page 118) The restriction sites HindIII, BamHI, EcoRI and AciI were displayed in the UCSC genome browser. I examined all peaks and promoter regions to find a restriction digestion that would generate fragments smaller than 10 kb in size, without cutting the Zta peaks or the promoter sequence of the genes composing the cluster (visualized in Chapter 5). This proved successful only for cluster VWF and CD9 when digested with EcoRI, therefore it was selected to be analysed by 3C (Table 4.2).

Table 4.2. Fragment sizes for the clusters

Feature of Interest		Fragment size (bp) after digestion with R.E.			
		HindIII	BamHI	EcoRI	AciI
SIRPA and PDYN	SIRPA Promoter	10500	11500	3270	I. F.
	PDYN Promoter	28760	28620	16490	I. F.
	Peak3027	19740	2260	I.F.	665
PDGFRB and SLC6A7	PDGFRB Promoter	11200	11300	3200	I. F.
	SLC6A7 Promoter	12000	7320	16900	I. F.
	Peak4066	16480	26000	8800	
	Peak4065	16240	3300	3600	
VWF and CD9	VWF Promoter	I.F.	I.F.	5100	
	CD9 Promoter	13000	11600	4700	
	Peak898	6450	2800	5100	
	Peak899	5230	5100	2200	

Inadequate fragmentation abbreviated as “I.F.”. Green represents fragments of an acceptable size, yellow represents fragment too close to the size limit. Orange and red represent fragments larger than adequate.

4.3 Discussion

In summary, 3 genes highly regulated by Zta and 29 clusters of genes regulated by Zta were identified. This analysis obtained targets to be followed by 3C experiments, testing for possible looping events happening between Zta and promoter regions of genes FZD10, SLC6A7, LHX1 as well as cluster VWF and CD9. Different restriction enzymes were analysed and digestion regimens were selected for the subsequent 3C experiments.

The method presented in this thesis of finding gene clusters as well as searching for Zta peaks to be encompassed within a cluster span (±100kb) is a simple method of scrutinizing and relating BED data files. This process would have been tedious and very time consuming if only using the UCSC genome browser; and although Galaxy provides a similar service, the input files require a particular format that limits the accessibility of the service.

It has been shown that Zta can bypass methylation of CpGs, although not H3K27me; and many of the lytic promoters of genes in EBV are repressed by H3K27me (Woellmer et al. 2012), therefore, it would be plausible for Zta to bypass H3K27me silencing. Perhaps in future analysis of which potential loops to follow, H3K27me would be an interesting mark to gather information about and then consider.

The gene SLC6A7, is reported to share several enhancers with PDGFRB (GeneCards), which was found as a cluster of regulated genes in our results. It also presents transcription factor binding sites for AP1 and c-Jun in vicinity to its promoter region (GeneCards). The presence of the AP1, c-Fos, c-Jun TF sites is also true for gene LHX1, and given that Zta shares homology to AP1 and c-Fos and that it binds to the canonical site of AP1 (Farrell et al. 1989; Lieberman et al. 1990) it would seem plausible that Zta could be binding and regulating these genes; and since there are no Zta binding peaks in the promoter sequence of these genes it is plausible for this regulation to be happening through looping to distal regions. Finally, cluster VWF and CD9 is an interesting target, since those two genes share several enhancers (GeneCards), making a stronger case for a possible co-regulation by Zta.

To finalize, it seems that the targets found are adequate for an initial exploration of possible looping between Zta and promoter regions.

Chapter 5. Chromatin Conformation Capture for the proposed Zta loops

5.1 Introduction

Linked to Chapter 4, the possibility of Zta distally enhancing the expression of genes could be studied by testing the presence of looping events between regions. In this chapter we will focus on the design and execution of 3C experiments devised to explore the previously proposed Zta loops.

The design of 3C experiments is a complex process, many different critical steps in the design have to be carefully considered and finely tuned in order to obtain solid results. In the original work by (Dekker et al. 2002) a novel approach that analysed frequencies of interactions in *Saccharomyces cerevisiae* between pairs of genomic sequences, also validated known qualitative features of chromosome organization. This 3C protocol consisted of cross linking DNA, digesting it, and re-ligating the cross-linked fragments, followed by a reversal of cross linking and subsequent amplification of the ligated products by PCR; then finally the detected products of PCR were compared to products coming from control template (Dekker et al. 2002). The design of the 3C experiments presented here follows the original concept presented by Dekker et al, and the particular protocol for the generation of the 3C libraries was based on the protocols used by; (McClellan et al. 2013) to study loops formed with EBNA2 and 3s in EBV positive latency II BL Mutu cells, and by (Tempera et al. 2011) to study loops formed between OriP and promoters of Qp and Cp in the EBV genome in EBV positive Mutu I cells. Overall, the schematic of the generation of the 3C libraries can be seen in (Figure 5.1).

The central part of the 3C technique is the detection of the PCR amplified products; however, this corresponds only to the last part of the assay; which in turn, is a reflection of an adequate experiment design. This design should take into account considerations like; choosing regions of interest and a control region, selecting a digestion regimen for an adequate fragmentation, choosing an appropriate restriction enzyme and a desired resolution, proper position of restriction sites, digestion efficiency, primer design in region of interest, primer design in control region, and production of a control template (reviewed by (Naumova et al. 2012)).

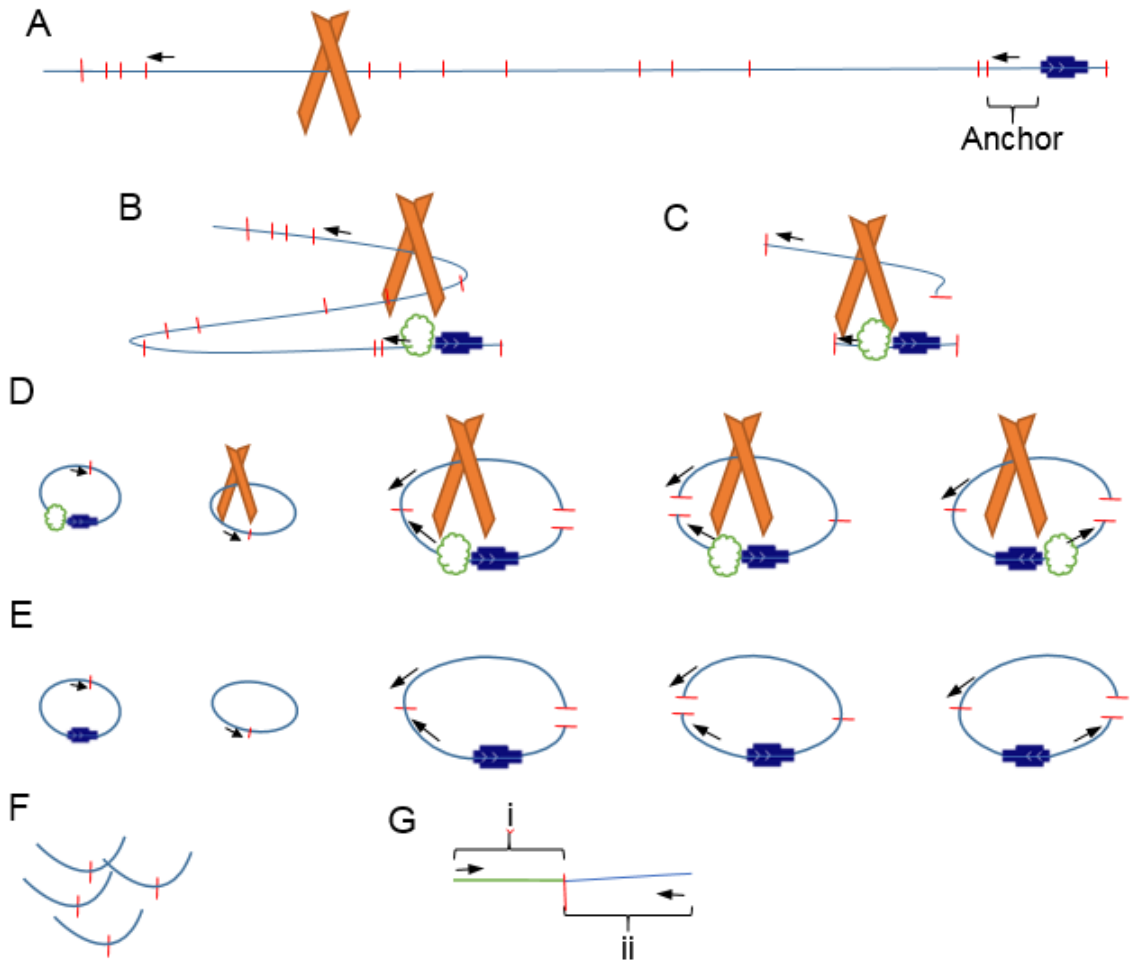


Figure 5.1. Schematic of 3C detection library. The horizontal blue line represents DNA. Small vertical red lines represent binding sites of a specific restriction enzyme. Small black arrows represent where primers would bind to, pointing to the 3' direction of their own sequence. The orange diagonal bars represent Zta as a dimer binding to DNA. The blue box represent a gene of interest. (**A**) Design of primers between a region that binds to the protein of interest and a region containing the promotor sequence of a gene of interest (anchor). (**B**) Loop formation and Crosslinking, the formed loops are chemically fixed in place. (**C**) Enzymatic digestion carried out with the specific restriction enzyme. (**D**) Possible ligation products are shown. (**E**) In this step proteins are digested, the crosslink reverted and finally the DNA is purified. (**F**) Then DNA can be amplified, and subsequent detection can occur. (**G**) For the generation of synthetic controls, the sequence that would result from the linking of the proposed sequence interactions caused by loopings was generated. Sequences were composed by two segments united at the restriction site; the reverse-complemented sequence targeted by the primer in the anchor (i) and the sequence targeted by the primer in the proposed looping fragment (ii).

One of the initial considerations in the design is an adequate fragmentation, the position of restriction sites for different restriction enzymes had to be analysed, this was mostly presented in Chapter 4, where the selection of the targets had to consider the fragment size and keeping Zta peaks and promoter regions undigested.

Once the fragments were identified, primers had to be designed. The amplification target of the primers must be the sequence 80 to 150bp away from the restriction site, “inside” of the restriction fragment of interest, if the primers are too close or far from the restriction site, the amplification might not be efficient. All of the different primers are designed in the same strand making sure that any amplification comes from a “head to head” ligation of the intended fragments, otherwise, the primers could amplify a product that is not representative of a looping event in the case of an incomplete digestion, giving a false positive interaction.

Regarding ligation, if it happens between regions different than the proposed ones in the looping, there will not be a primer pair formed for amplification; the same would happen if the halves swapped in orientation, if self-circularization occurs or if the ligation is incomplete, there will be no amplification of a product. If an incomplete digestion occurs, the subsequent ligation could unite fragments of sequence that separate the primers demanding a longer amplification step in the PCR, and therefore could be undetectable.

Typically, for the generation of control libraries, bacterial artificial chromosomes (BACs) containing all possible ligation products relevant for genomic regions of interest were originally used, to assess PCR efficiency, in particular primer efficiency (Palstra et al. 2003; Tolhuis et al. 2002); however generating a BAC library template for each target would be complex and time consuming, instead we opted to synthetically generate controls from the sequence hypothesized to result from the amplification of each primer pair.

In order to increase the chance of testing chromatin that is representative of looping events pertaining to lytic cycle, different cell lines were used, both to generate a comparison of cellular environmental backgrounds and to compare if any possible looping event captured is present only in lytic cycle (Table 5.1).

Table 5.1. Cell lines used to generate 3C library.

Cell lines used	Attributes	Induction	Reference	3C experiment			
				Highest fold mRNA			Cluster
				SLC6A7	LHX1	FZD10	VWF & CD9
				BamHI	Hindiii		EcoRI
LCL#3	LCL, EBV+, Spontaneously lytic 6% to 20%	None	Sinclair et al 1994	+	+	+	-
GM2188	LCL, EBV+, Tightly latent	None	Stiff et al 2004	+	+	+	-
Akata	BL EBV+, Anti IgG inducible	Mouse anti-IgG	Takada et al 1989	+	-	-	+
Akata-Zta	BL EBV+, BZLF1 expression vector Doxycycline inducible	Doxycycline	Ramasubramanyan et al 2015	+	-	-	+

On the first column the name of the cell line is displayed. On the Attributes column, any pertinent details of the corresponding cell lines are displayed. On the induction column, the method of induction, if any, is displayed. In the column with the header “3C experiment” we can find the genes tested depending on the method of design, either from top ranking genes with a high fold change in mRNA seq, or through cluster formation of up to 100kb of distance between genes and Zta peaks. Beneath the genes names, the restriction enzyme needed to generate the library can be found. A + symbol represents that such library was created in that cell line, a – symbol represents that no library was created in that cell line for those genes.

5.2 Results

Following the idea that Zta could be regulating gene expression distally by looping to the promoter region of genes, a design for an assay of Chromatin conformation capture had to be devised. The selected targets were the promoter region of genes SLC6A7, LHX1, FZD10 as well as clustered genes VWF and CD9.

5.2.1 Preparation of the Template

5.2.1.1 Lytic Cycle confirmation

Since the potential looping events were based on the interactions found in the ChIP-seq of lytic cells, the cells used to generate the 3C template needed to reflect this. Four different cell lines were used aiming to generate two different comparisons; and each was verified to be in lytic cycle. Cell lines LCL#3 and GM2188 are compared in a western blot, showing Zta as expected in lytic LCL and an absence of it in GM2188 (Figure 5.2. Data showing lytic cycle. A). In a similar way, we can see a higher viral genome load, reflective of lytic cycle, in the induced cells of the Akata line (Figure 5.2. Data showing lytic cycle. B). The Akata-Zta cells were a kind gift from Anja Godfrey, (Data not shown).

5.2.1.2 Enzymatic digestion of Chromatin

As discussed in Chapter 4, in order to obtain the fragments proposed to be looping, the fixated chromatin had to be digested with different restriction enzymes. After an overnight digestion, the chromatin was analysed in agarose gels, where a smear running down reflects digestion of chromatin (Figure 5.3. Enzymatic Digestion of Chromatin analysed on 1.2% Agarose Gel.). The template from induced and non-induced Akata and Akata–Zta cells can be seen in (Figure 5.3 A) for the digestion of BamHI and (Figure 5.3 B) for the digestion of EcoRI. The main band on top of each lane is above the markers band of 10 kb and with a tight band on top. While some lanes present minimal smear, others do show a smear running down the lane; this is a reflection of the quality of the 3C DNA template, indicative of a template of good quality when there is no smear, and of passable quality when the smear is more prevalent (Naumova et al. 2012). The quality of the template coming from LCL#3 cells and GM2188 is lower (Figure 5.3 C and D).

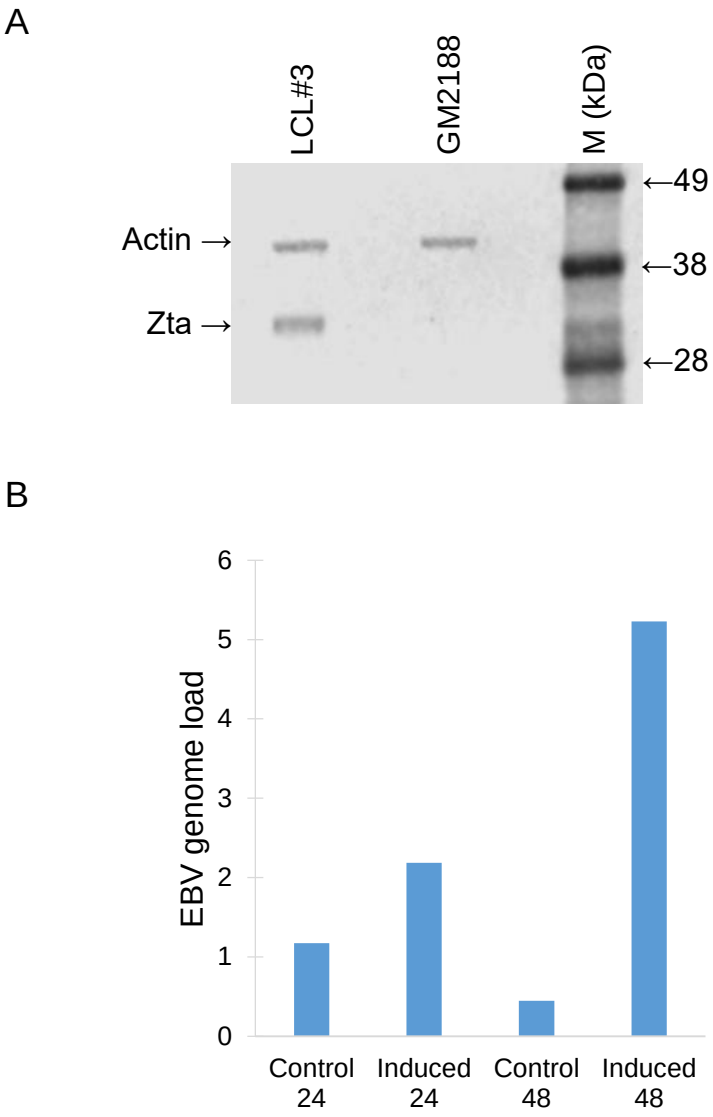


Figure 5.2. Expression of Zta only in lytic cell lines and higher expression of EBV viral load in induced cell lines. (**A**) Western blot staining showing bands of an expected size corresponding to Actin and Zta in LCL#3 cells and no band in the latent GM2188 cells. (**B**) Relative amount of Viral load from Akata IgG induced and non induced (control) cells, this was obtained through qPCR.

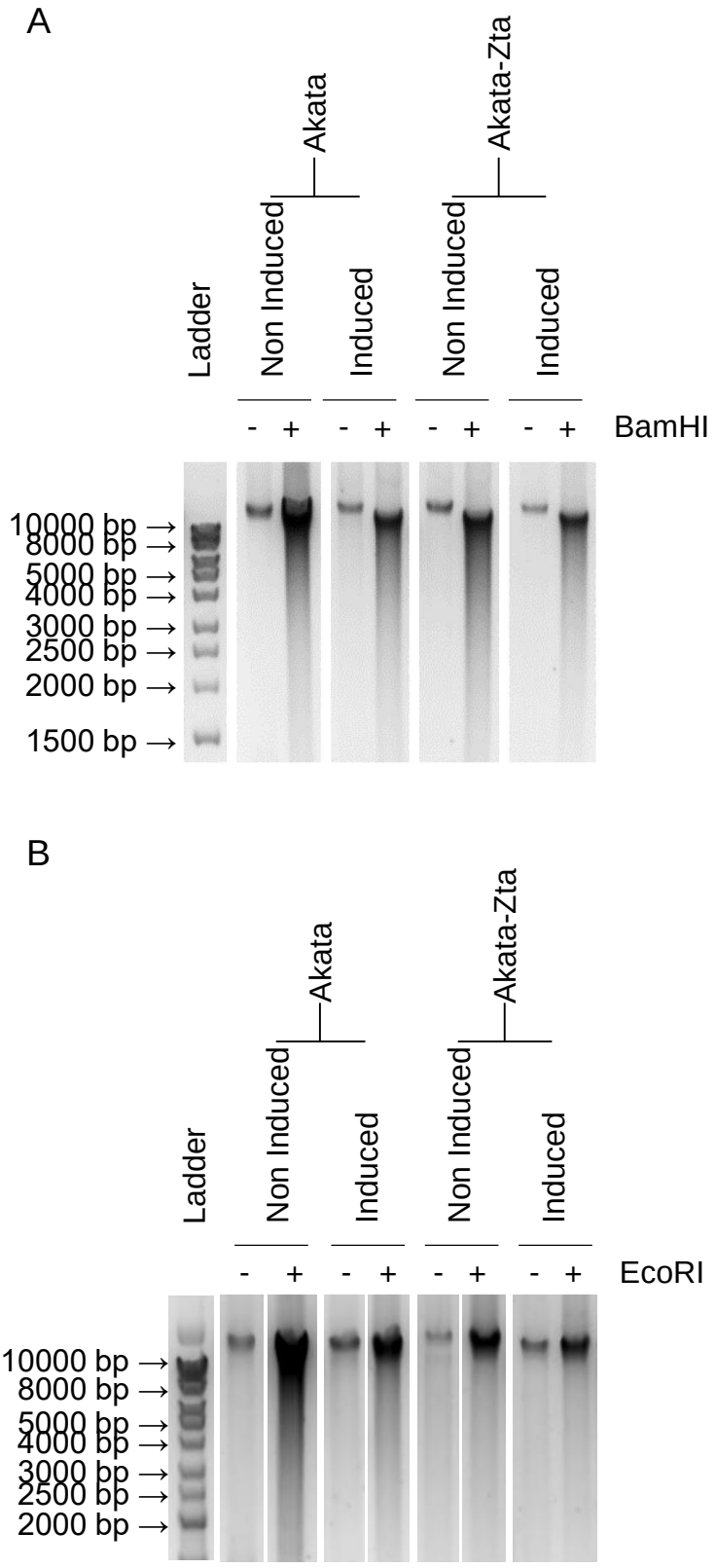
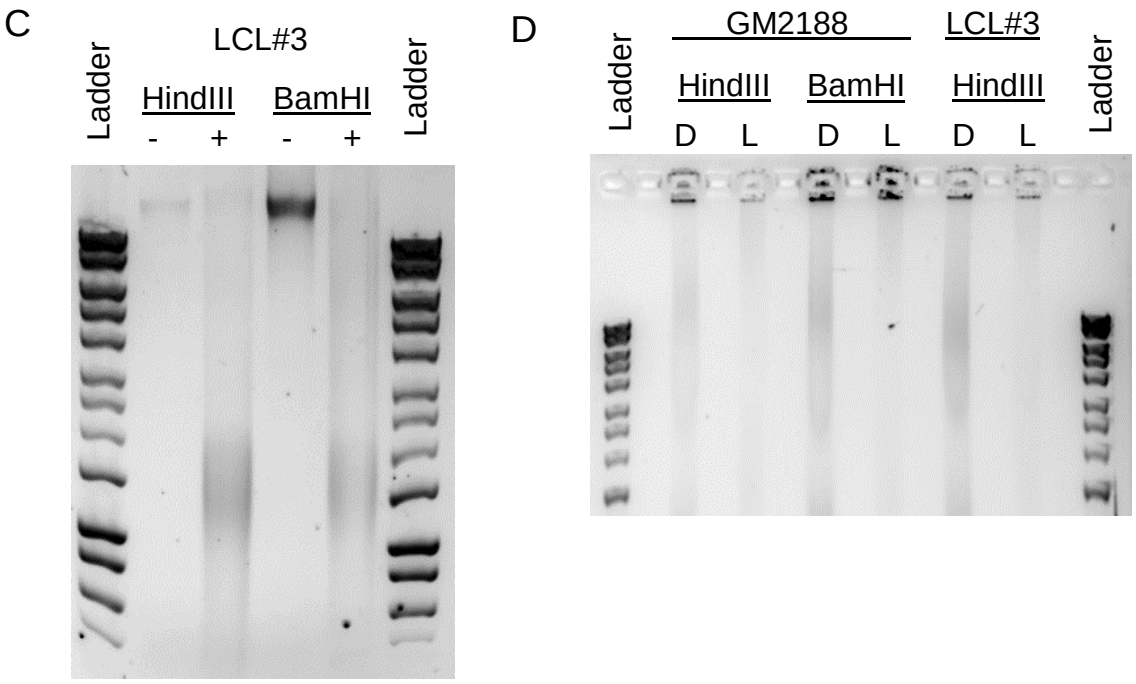


Figure 5.3. Enzymatic Digestion of Chromatin analysed on 1.2% Agarose Gel. (**A**) Digestion of the induced and non induced cells by BamHI. (**B**) Digestion of the induced and non induced cells by EcoRI. (**C**) Digestion of LCL#3 cells by HindIII and BamHI. (**D**) Digestion ("D") and ligation ("L") of GM2188 and LCL#3 cells by HindIII and BamHI.



5.2.2 Zta and promoter sequence of SLC6A7

Based on the interactions of Zta found in the vicinity of SLC6A7 in Chapter 4, four potential loops were considered; two loops from fragments with Zta peaks to the promoter of SLC6A7 (Loop A and loop B), and two loops from fragments without Zta peaks to the promoter of SLC6A7 (Loop C and loop D). These last two loops, were designed to be considered negative (not looping) controls (Figure 5.4).

The sequence of the fragments was used to produce primers through Primer-blast. The primers were selected if they amplify towards the restriction site, so that if looping events were happening the primers would amplify towards each other (Figure 5.5).

The hypothesised resulting amplified sequences were assembled together virtually, using the sequences from the fragments, so that they could be synthetically generated to be used as positive controls. This resulted in 4 positive controls, one corresponding to each loop (Figure 5.6).

Aiming to test both the primers and the synthetic controls as template, an initial PCR was set up with and without polymerase, at 60 °C, controlling for external contamination. This temperature was chosen since the design of the primers set the Tm to 60°C. (Continues in page 138)

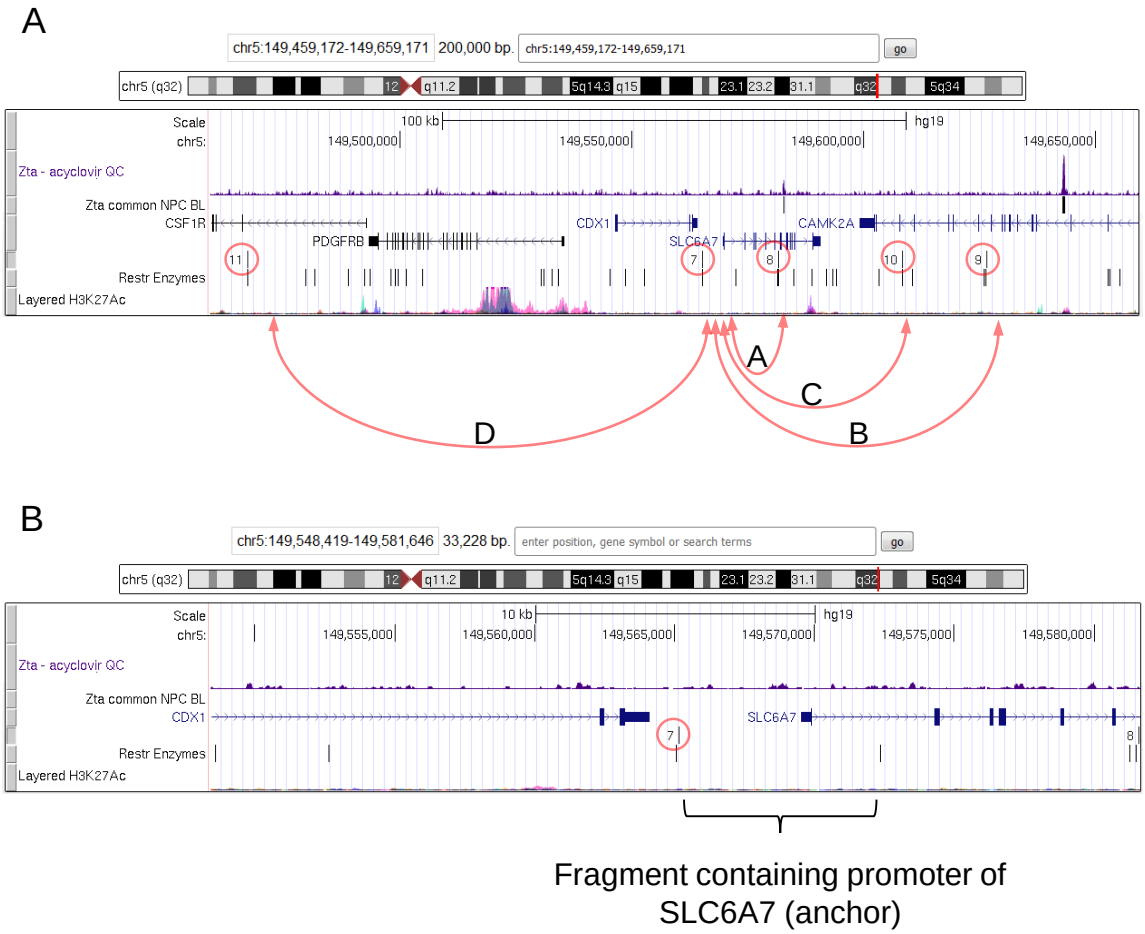
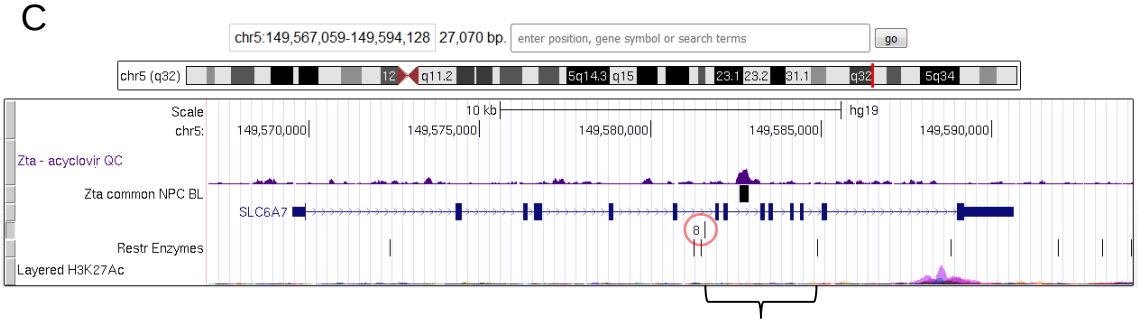


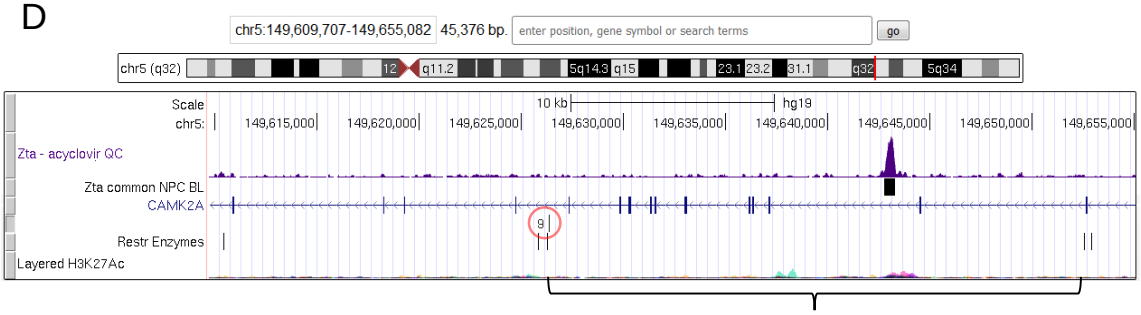
Figure 5.4 Maps of potential looping regions at SLC6A7. (**A**) Red arrows represent the potential loops forming across regions. The track labelled as “Zta – acyclovir QC” shows the stacked reads of Zta alignments to the hg19 release of the human genome. This data was previously generated in the lab, from ChIP-sequencing in Akata cells. The track labelled “Zta common NPC BL” corresponds to the peaks called by MACS from previous ChIP-Seq experiments performed in both HONE1 NPC cells and Akata BL cells. The genes displayed come from the UCSC Genes database. The following track shows circled in red, the alignments and names of the primers designed to amplify possible loops. The track labelled as “Restr Enzymes” shows the sites where BamHI aligns to. The track labelled as “Layered H3K27Ac” shows the regions with heavy acetylation on cell lines GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK, NHLF, signalling for active enhancing regions. (**B**) Local map of the region containing the promoter region of gene SLC6A7 (anchor region). (**C**, **D**) Local maps of regions containing a Zta peak that could be looping to the region with the promoter of SLC6A7. (**E, F**) Local maps of regions with low Zta signal, used to design negative controls of regions not looping.

C



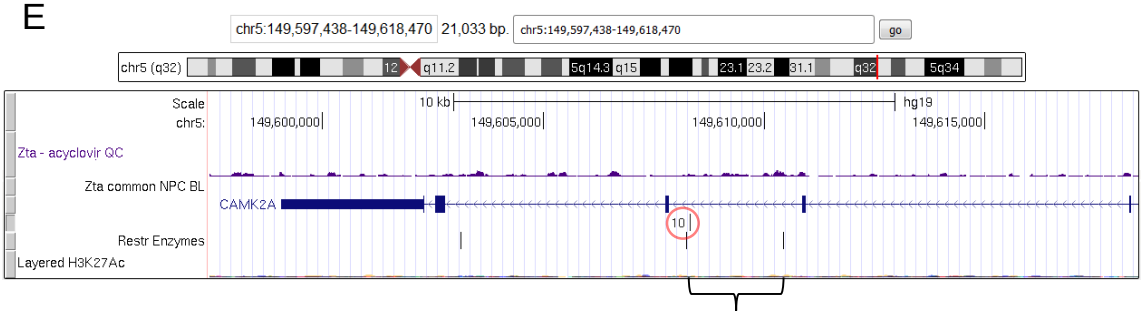
Fragment containing peak
for loop A

D



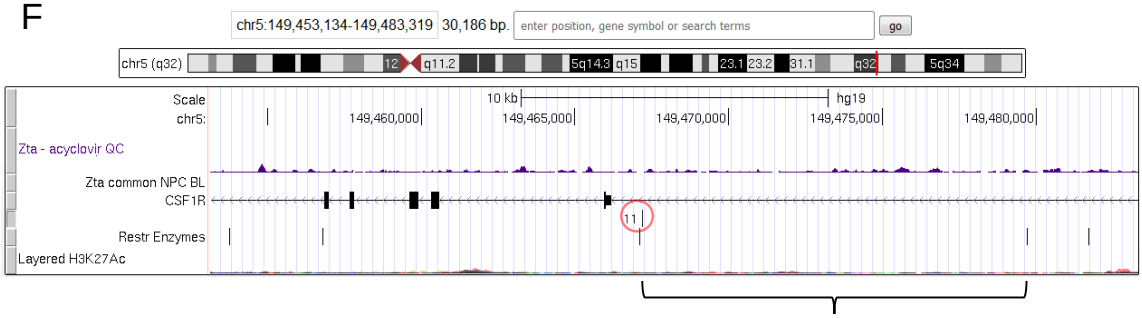
Fragment containing peak
for loop B

E



Fragment of interest
for loop C

F



Fragment of interest
for loop D

A

CCGTGCTGGCGGTGGCCAACCCGAGGTCTGGCTGCCTGGCCCCAGGCTGGACACGCCCTCCCC
ACCCTGGCCCTCTCTAATTTTAGCCCCATGCTGCTGCGGAGGTCATGTGTCAGCCCTGCCTAC
TCACTCCGGACAGCTGCAGGCTGCGTCTTGCCAGTGACAAGAAGCAGATGCCCCCACTGGGGC
CCACCCTGCCAGAGCACCCATCCCCTGCCTCTCTCCTCAGGCCATATAGGCCTCCCCAACCTC
AGACTCCAGAGCCCCATATTGGCACTGGCTATGGCCCGGGTGGCAGGGACAACCAGAGCAACA
GCTCTTTGGGGGCTTCCCCTGCTGTCATTTCCCCTTACCTCCCTCCTCACCTCCTGGCCAGAG
CCCTGGAAGTGGGCATCAGCCAGGAAATGGCCTGAATGTCCAGCCCTGCCCACAGGCATCCGG
CTGCTTCAGACTTAAGCAAATACCttaaatgaaacaaggtatctgtgaagtacttagcaaagc
acctgacccataagccgcacatgctcaggaaatggtagcCATGATAAGACTTCTCCTGGTGGC
CTTGAGAGGGGGAGTTGGGAGGAGGGTCATATGAGCCACTGATGCCCCAGGGGCTCTGAGGCA
TGGGGCTCATGGCAGGTCACCTTCCCAGTCACCCCCACATTTCCACCTATGACCCCTTCGAGG
CTCTCATTCCAAGTTCCCTACTTGCAAACCAAGGCAGGGGCCACCTCTGGATCCCGAAGCTGA
CAGGAATGGCTGTCCATTCACCACAGGCTTTGGAGGGTTGGACTAGGATGTGCAGTGGGGCAG
ggggagtgctttctcctcgagcctcaaggtagacgcccctctcagctccagcctactgtcaca
ggatcgtgggtccagtgaggccacagcgtctaattttcaagagacgacagcaatctggggttt
tctatgaaatttcctaatttttaagtgttagaaactcattcaaaaaaatttttaaaacacccc
tacctccacaaagcatatcaagagattcaatccagctcacaggccgccattttgtgagctcAG
GACTAGAGTTCTTTAGGGTGCTTCAGTTGTTAAAAAAAGACTCTGACTATCTGGGTAGCTCAT
CTTTCAGTTTACGGAATGGCAGGATCCTTGTCCAGGGAGAGGGTGGGAAGGCAGTTTAACGAG
GTAGGAAAAGCCCAGTCAGAAGTTCTAGTATCTGCCCACCTCTGGGTCTCACTCCTCCCAACT
GGTTACATGGGCAGGTGGACCCACTCAATGCAGCCAGCTGCCCCCTTCTCAGAGCTGCCCTCT
TGCTGTGGTGCTGGCCACAAATCTGCAGCAGCTTTCTCGGCTGGCCCCAGCCCCAGTCAGAGG
GCCCCTCGGGTGAGCTTTGCTCTATCTGCTATCTATATCTCCTCTACTGCCTGCCAGCGGCCT
GCGCTCCCCAGAAGGAAGCCTCTGGAGGGCACTGGCCCCAGCCCTCTCTGATGTGCCAAGAGG
AGCTTCTTGCTAAACCCTGACAATAATCCCTTCCCCTTCTCTCTGCAAAGCCACATGCCTCAG
GTGATTGGGCACTCCCTGCCTTAACCAGCCCAGTTTGGCCAAGGGCCTGGGACGCTCCTGCCC
TGGACTCACTGGTGACCCAATCCAGTTaaaggcacttgccctgggtttggtcaaacctgtgct
gaatgcaggctctgcctctctagccatgtgacctgggacaagtcagagctcaccttcctgagc
ttctgcttcctcatcccagaaatagggatggcagcacccccatcatcccctctgctgccttgt
catgtaggccacctgagtggaagcacgggactgcttagcacagggcTCCAATCATGGTGGGCA
ATATTGTTACTCCC

Figure 5.5. Sequences used to design primers for SLC6A7. These were obtained from the UCSC Genome browser and correspond to the upstream and downstream sequence relative to the restriction site that flanks each fragment of interest. The sequence was analysed by primer-BLAST to generate primers downstream of the restriction site which would amplify towards the restriction site and that would align to sequence with no repetitive elements. Primers that would align ~80 to ~150 bp away from the restriction site were selected. The sequence highlighted in green corresponds to BamHI. (**A**) Sequence in red corresponds to primer 9, designed for amplification of potential loop B (**B**) Sequence used to obtain primer 7 in anchor region, designed for amplification of all potential loops in SLC6A7. Sequence highlighted in green corresponds to BamHI. Sequence in blue corresponds to primer 7. (**C**) Sequence used to obtain primer 8, designed for amplification of potential loop A. Sequence highlighted in green corresponds to BamHI. Sequence in pink corresponds to primer 8. (**D**) Sequence used to obtain primer 10, designed for amplification of loop C. Sequence highlighted in green corresponds to BamHI. Sequence in teal corresponds to primer 10. (**E**) Sequence used to obtain primer 11, designed for amplification of loop D. Sequence highlighted in bright green corresponds to BamHI. Sequence in dark green corresponds to primer 11.

B

GAATGCATTCAGAGTCCATGTGGTCTGGGCAGCTGGAGCAGGAGAGAAGCCACAAAGCCCCCC
TCAGGGTCTGACTTCAGGGGTCCTGAATtgctctgcagaagtccccagtccagtggggaagac
tgctatgcacacagatgacactaatataacaggataagggccaatctgagcgtcaggggagcc
agggcatccaacgagagcagagcacatgcctagcttatcctggaaggcttcctggagtaggta
atgtttgaactgagactgaaaaggactagcgagtatggatcaggccaagaagggaataaagga
gaaaaggaacttgagcaatatggactgaaagtataaTTTTTTTTGAACTCGGGAGTCTACTAC
CCGGGGCCTTCTCTCCCCGGGGGCTCTGGTGCCAGCTGCCTGCCCCTCTGTTTACTGTTCTGG
GAGGTGGCCAGGCCATAAAAAGTGTGCACAGCCTGCCTGACATCGCCCTTTGCTCCCATGATA
ACAGTGCACAAGGTGGATCCTGGCCTGGCACAGGCCCCTGAAGTGCAGCAGGCACTTCTCTCT
CTCCTCCTACCAGTGGCCTGGGAGACAGCGGGCATTCATTAGTAATGGTGGACACCTGAGGCT
TGGGAAAGAACCAGGGTTCCAAGAGTAGAGGGGGAGGGGCTGGTGAAAAGGCGGCTGGCCCGG
ACTGAATAGAGTTTTCTGATCTTCAAAAAAATATgctctgtgaccctagggtaactgcttccc
ctctttgggcccgcattgccccatctgtaaacttggggaagttgTCTACCAGAAAGAAATAAC
TCtggtcaagtaagtttgggaaatcctgctgaagccacttctctgatagagtcacagtgggcc
attagcacatcaaagtctctgacaagccccacagcaaagaaacccctgaattttctttcatc
tggccttccccgtgctttgctacctaggaggcccctactcacacctccgagcatcccactgaa
ctCCACACACTC

C

ttgataggtaaaattttatgtatttatcatgtacaacatgatgttttgaagcatatgtacctg
ggggaactgttaaatctagctaattagcatatgtattatctcacataattatcatttttgtag
tgagaacacttaacatccactctcttagcattctccaagaatacgatatattgtcactaactg
catattatacTtcatttattattattaacaatcatggaaacactatcctcattttacagagga
ataaactgaggctcaaaagggttcttatcaagcttacggtcatacagctaggggatcgaagaa
caggatccagcccagatttgtgtgactgtaaaccccAAGGAGGGCATTAGTGACAATGGTG
ATAACAATGGTACTGACAGCtaatgttaatatttattgactggggactgtgcctgaccttgcc
ctagtgttttacatgcatgtgtgtagtgatttaacccaccctcctcccatgaagtaggcacta
ttcctatcccctcgttacagataggatcctgcggtccagaggggtgacctggcttgttcagg
ctgcacagctagaaaatggtgaagctggggtccaaatccaggtgccaggattccagaAAGGAA
GTAAAGGGGTTATGCAGGCTTACTGCCAATACAAATCTCCTGCACACTTTACATTAGGCTCTG
TGTCCTtcagaactggaccaagcttagagaccatcttctagatcctgggtgggaaactgaagc
ccgcagaggggaagggacttgcccatgtcccaccgtgtaaatgttgggggaggctttgctcct
ggtccacgtcacttgtgagccaggctcatgaccacgccatccctcGTGACCACGCCATCCCTG
GAGCTGTCACACCATCTTCATTCACTGCCTTCCTTCTCTGTCCTTGACCATCCGCACACCCTC
CCTTTCCATCCTTCCCGGCACAGGTGTGGATTGAAGCTGCTCTTCAGATCTTCTATTCCCTGG
GTGTGGGCTTCGGGGGGCTCCTCACCTTTG

D

CCTGGGGCCCTGCCCGTCACCCTGCTGTGCCATCCAGCAGGACATGGAACGAGAGCAGACAA
CAGGCACCACAGAGAGAAGGAATTGCTCACGGTTTTCAAAATAGAATCGATGGAAGTCCAGG
CCCTCAACCAGGTTCCCCAGGGCCTCAGGTTCGAAGGCTGTCATGCCAGGGTCGCACATCTT
CCTGGGGGAAAGAAGCCAGAGGGAAGAGGGACTGGGGCGGCTTCTCATTGGACCCTTGGAGG
GTGAGCctgtgaaataggaatgattgcacctacctcacaggatcattgaagttagtagatgg
gaAGGGGCCAGGGGCTTGCAAGTTATAAAGTATGGGATATCACTGACTGGTTGGTGCCatct
agatccgggaagcctgggttcaggccctggctctgccaatggctgtgagaccccaagcaagt
ggcttaatccctttaagcaagagcaccatgaatagtgtacagggctgtgcagtgcactaaag
gtgctggtggggaggaaattcagcttgctctctactctccaaattgtactccctggttcagg
gcagtggatccagaggggctttttctaatttgTCTTCCTGTGGGGTCGGGATGGGGGACAC
TGCCTTTTCCTTCACATAAAAGTGACTTCTGCTCATCAAGCCAGGTATATACCTggaagggg
cgagttttctaatttgaacaaaggtgccctattgttaacagcagccttgGctgtgagcctca
gtttcctcttctgtaagatgagaaaaatagcattagatgaGGTAATATCTGGGAGGAGAAGC
ACAGGGCATGACGGAGGCTGTGTATTTTCAGCTGTTATTCCCTTCCAGGTCTTAGTCTTCTC
GGAGATCCTGACAGCCAGGCAGGGGAAGGGAACAAGCTGTCTTTCCCATCTCCAGGTCTGTG
GGATCTTTTGTGGTATAACATGGACCCCTGTGACAATAGCAATGAAGCCAGGGGCAGGAAAT
GTCCCAGGACAAGCAAAGCTCCCAACCCAGGGTAGAATTGGAGACACCCAGGAGAAGGTGTC
AGGCCCAGCACCGTCTGAGCTGGAGAAAGCATTCACAGGGTGCACTGTCCCTGCTCTGTCCT
TTCTGTTTGGGTGACCCCAGACAACTTGCTTCTCTACTCATGACCCCAAATTCTCCAGATGG
CTTAAGCCTGGGGCTGGGGGACTGTGCTCAGACCAAACGGGATGCCCAGGACCCTTCCAGCT
CCAGAATCCCCCAAGAGTGGGGGTCTGGACATGTGGGAGACTGAGCGCCCCCACTGAAGAAG
CCCCAACCCTGGAGGCACCTGGGTGTTCCTGCCTCCTAGAGTgaagggacttggaggccatc
tgccccaactccattgtacatatggggaaactgaggctcagagagatgaaggaacatgttca
agattacacagGGCATCCTGGACCCCAGAGGTCATGACAAGGGGTCAGAACCTAGAGCAGGC
TGCAGATGTCAGAGGAGGTGACATCACATAGGTGGGAGGGGTTCCCCAGGACCC

E

Agtttcactctgttgcccaggctggagctctgtggctcgatcatagctcactgcagcctcaa
actcctggctcaagccttcaagctgtgagatctgacgttatctctaggtaaatatagtgttc
aaattgaatcggagggcaccagattcgtgtctgctgcagaatctgaacaattggttgattgc
ttggtgtatggggaaatatgccccatcctctacatctggggtcagaagtattttgtgttgtg
agactataggagaaacggagtttgttttcttCCTGTATACAGGTTGGTCTGGGAGAGCCTGA
TAATATTGGCaataagcttccttttacaggacatttactagccaaacctttacattcactg
tctcctttagtactgcaaaactttatgtaggtaagtactattgtcattctcatttcacagat
gggaaaactaaagctcagagagatcaaatgccttaagcttgtccatagtcaaacagctacca
aatggtggagataggatttggatccagcagtctggttataaaacTATCTTTTGTGCGTTAC
TTACATTGTTTATTTATTTTTACAAAGAATCTAATTGTTCTGCATAATTCTTAAAATATATA
GCATTCctttgctccctctttcctggaggcaattgcttttacctcttttagttgattcgct
tagtattttacttcttatctctaaagaacatggtcatttcttcttctttttttctttttttt
tttctgagacagggtttggttctgtcacccaggctggagtgcagtagtgtgatctcggctca
ctgcaacctccacctcccggactcaagccatcctcctacctcatcctcctgagtagttggga
ctacaggggcatgtgccaccattcctagcaaatttctgtatttttggtagacatggggtttc
accatgttgtccaggctggtctcaaacccctgagctcaagtgattcgcataccttggcttcc
caaagtactgggattataggtatgagcc

A

Promoter Vs Potential Enhancer1

>A

cagttaccctagggtcacagagcATATTT
TTTTGAAGATCAGAAAACTCTATTCAGTC
CGGGCCAGCCGCCTTTTCACCAGCCCCTC
CCCCTCTACTCTTGGAACCCTGGTTCTTT
CCCAAGCCTCAGGTGTCCACCATTACTAA
TGAATGCCCGCTGTCTCCCAGGCCACTGG
TAGGAGGAGAGAGAGAAGTGCCTGCTGCA
CTTCAGGGGCCTGTGCCAGGCCAggatcc
tgcggtccagaggggtgacctggcttgtt
caggctgcacagctagaaaatggtgaagc
tggggtccaaatccaggtgccaggattcc
agaAAGGAAGTAAAGGGGTTATGCAGGCT
TACTGCCAATACAAATCTCCTGCACACTT
TACATTAGGCTCTGTGTCCTtcagaactg
gaccaagcttagagaccatcttctagatc
ctgggtgggaaactgaagccc

B

Promoter Vs Potential Enhancer2

>B

cagttaccctagggtcacagagcATATTT
TTTTGAAGATCAGAAAACTCTATTCAGTC
CGGGCCAGCCGCCTTTTCACCAGCCCCTC
CCCCTCTACTCTTGGAACCCTGGTTCTTT
CCCAAGCCTCAGGTGTCCACCATTACTAA
TGAATGCCCGCTGTCTCCCAGGCCACTGG
TAGGAGGAGAGAGAGAAGTGCCTGCTGCA
CTTCAGGGGCCTGTGCCAGGCCAGGATCC
TTGTCCAGGGAGAGGGTGGGAAGGCAGTT
TAACGAGGTAGGAAAAGCCCAGTCAGAAG
TTCTAGTATCTGCCCACCTCTGGGTCTCA
CTCCTCCCAACTGGTTACATGGGCAGGTG
GACCCACTCAATGCAGCCAGCTGCCCCCT
TCTCAGAGCTGCCCTCTTGCTGTGGTGCT
GGCCACAAATCTGCAGCAGCTTTCTCGGC
TGGCCCCAGCCCCAGT

C

Promoter Vs Negative Control1

>C

cagttaccctagggtcacagagcATATTT
TTTTGAAGATCAGAAAACTCTATTCAGTC
CGGGCCAGCCGCCTTTTCACCAGCCCCTC
CCCCTCTACTCTTGGAACCCTGGTTCTTT
CCCAAGCCTCAGGTGTCCACCATTACTAA
TGAATGCCCGCTGTCTCCCAGGCCACTGG
TAGGAGGAGAGAGAGAAGTGCCTGCTGCA
CTTCAGGGGCCTGTGCCAGGCCAggatcc
agaggggggctttttctaatttgTCTTCCT
GTGGGGTCGGGATGGGGGACACTGCCTTT
TCCTTCACATAAAAGTGACTTCTGCTCAT
CAAGCCAGGTATATACCTggaaggggcga
gttttctaatttgaacaaaggtgccctat
tgttaacagcagccttgGctgtgagcctc
agtttcctcttctgtaagatgaga

D

Promoter Vs Negative Control2

>D

cagttaccctagggtcacagagcATATTT
TTTTGAAGATCAGAAAACTCTATTCAGTC
CGGGCCAGCCGCCTTTTCACCAGCCCCTC
CCCCTCTACTCTTGGAACCCTGGTTCTTT
CCCAAGCCTCAGGTGTCCACCATTACTAA
TGAATGCCCGCTGTCTCCCAGGCCACTGG
TAGGAGGAGAGAGAGAAGTGCCTGCTGCA
CTTCAGGGGCCTGTGCCAGGCCAggatcc
cagcagtctggttataaaacTATCTTTTG
TGCGTTACTTACATTGTTTATTTATTTTT
ACAAAGAATCTAATTGTTCTGCATAATTC
TTAAAATATATAGCATTCctttgctccct
ctttcctggaggcaattgcttttacctc
ttttagttgattcgcttagtattttactt
cttatctctaaagaacatggtcatttctt
cttctttttttcttttt

Figure 5.6. Sequences used as template for synthetic controls of SLC6A7. These were generated by uniting at the restriction site; the reverse-complemented sequence of the anchor amplification target, and the sequence from the proposed loops amplification targets. Sequence highlighted in green represents BamHI. Sequence highlighted in blue represents primer 7 from the anchor region. (**A**) Cyan represents primer 8. (**B**) Red represents primer 9. (**C**) Teal represents primer 10. (**D**) Dark green represents primer 11.

(continues from page 131) Subsequently, the PCR products were loaded in an agarose gel, and an electrophoresis was run, resulting in clear bands of the intended size in lanes containing primers, template and polymerase. No amplification was observed in the lanes with primers and template, but without polymerase; indicative of no contamination (Figure 5.7).

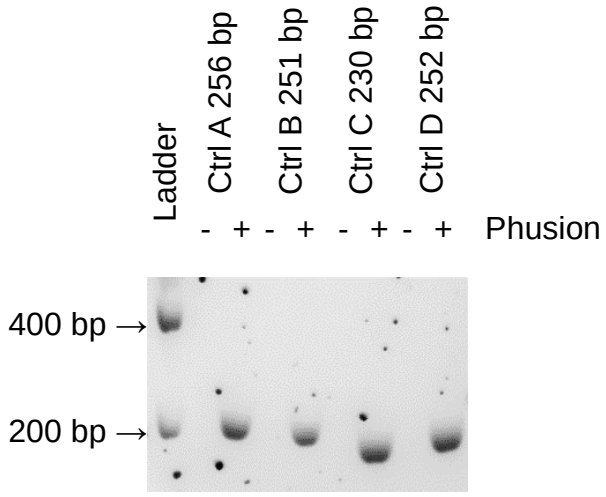


Figure 5.7. Amplification of PCR products from synthetic positive controls for loops with promoter of SLC6A7. A set of PCR reactions were performed using the synthetic positive control as template at 60°C. The reactions were performed with and without Phusion polymerase to corroborate that there was no contamination in any of the components of the reaction, the expected size of amplification to be correct, and that such polymerase would function as expected.

Trying to find an optimal temperature of amplification, a gradient of temperatures was used in PCR using the template from LCL#3 cells and the primers for loops A and B. The products were loaded in an agarose gel, and after an electrophoresis was run, it was observed that lower temperatures resulted in non-specific amplification, and higher temperatures reversed this (Figure 5.8). Synthetic positive controls were amplified at the closest temperature to 60°C, found in the gradient

Aiming to test if the presence of products representing looping events was observed, the PCR for amplification of loops between Zta and promoter region of SLC6A7 were performed at 64 °C, since this temperature showed no non-specific bands in the previous gradient of temperatures. The template from induced and non-induced Akata cells showed no amplification (continues in page 140)

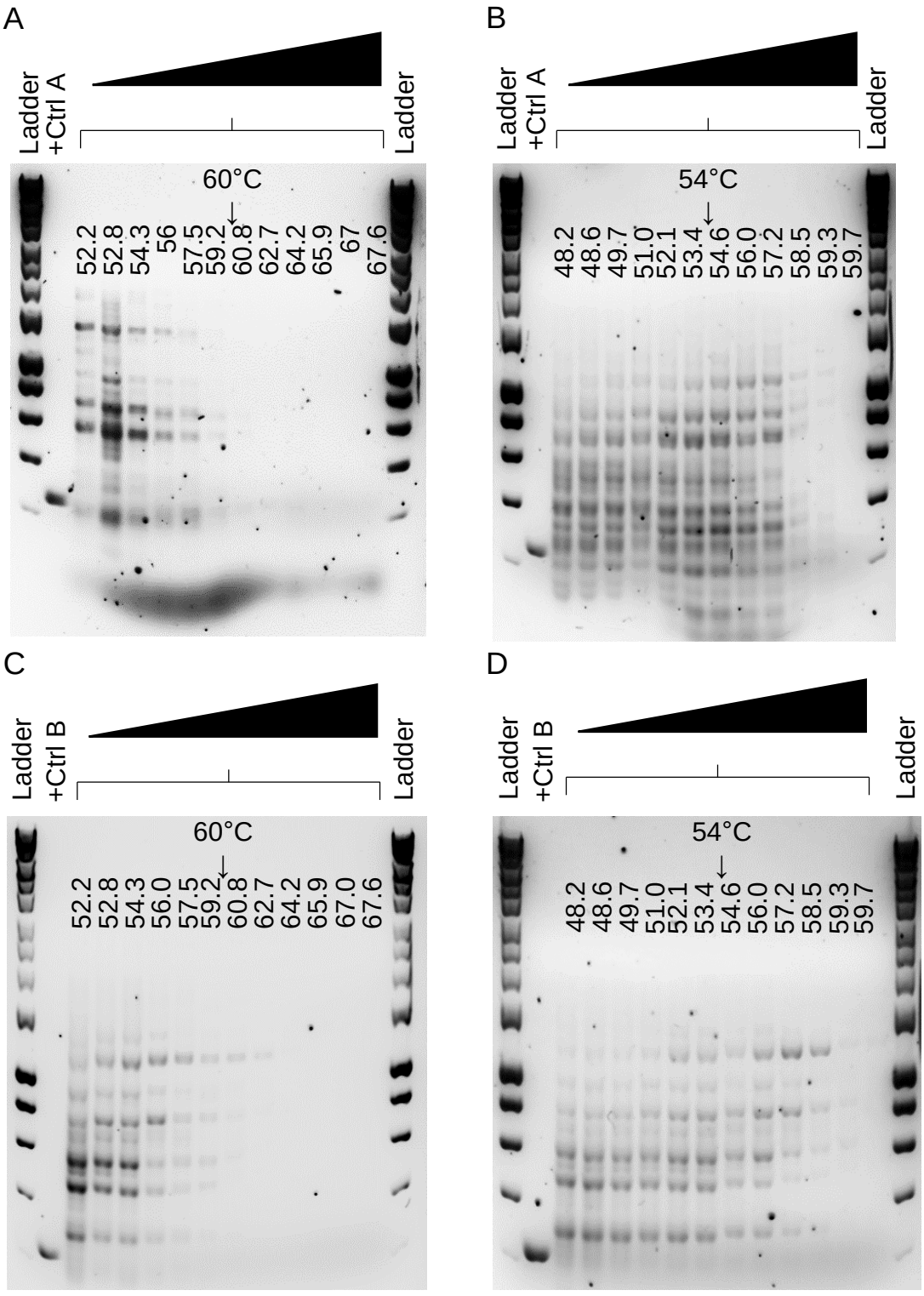


Figure 5.8. Temperature gradients used to optimise melting temperatures for loops with SLC6A7. Using template from LCL#3 cells, four gradients were tested. Black triangle on top represents the increasing temperature in the gradient. (**A**) Performed at a 15.4°C gradient with 60°C as a midpoint testing for loop A. (**B**) Performed at a 11.5°C gradient with 54°C as midpoint testing for loop A. (**C**) Performed at a 15.4°C gradient with 60°C as a midpoint testing for loop B. (**D**) Performed at a 11.5°C gradient with 54°C as midpoint testing for loop B. Each control was amplified at whichever temperature closest to 60°C was found in the gradient.

(continues from page 138) of a product of similar size as the control (Figure 5.9 A and C)**.** In a similar way no product amplification could be detected in latent cells GM2188 or lytic LCL#3 (Figure 5.9 B and D). The positive controls were run in the same PCR reaction at the same temperature, time and day as the tested samples.

5.2.3 Zta and promoter sequence of LHX1

Based on the interactions of Zta found in the vicinity of LHX1 in Chapter 4, five potential loops were considered; three loops from fragments with Zta peaks to the promoter of LHX1 (Loops E, F, G), and two loops from fragments without Zta peaks to the promoter of LHX1 (Loop H and I). These last two loops, designed to be considered negative (not looping) controls (Figure 5.10).

The sequence of the fragments was used aiming to produce primers through Primer-blast. The primers were designed to amplify towards the restriction site, so that if looping events were happening the primers would amplify towards each other. The hypothesised resulting amplified sequences were assembled together virtually, using the sequences from the fragments, so that they could be synthetically generated to be used as positive controls. This resulted in 5 positive controls, one corresponding to each loop (Figure 5.11).

Aiming to test both the primers and the synthetic controls as template, an initial PCR was set up with and without the polymerase at 60°C, controlling for external contamination. This temperature was chosen since the design of the primers set the Tm to 60°C. Subsequently, the PCR products were loaded in an agarose gel, and an electrophoresis was run, resulting in clear bands of the intended size in lanes containing primers, template and polymerase. No amplification was observed in the lanes with primers and template, but without polymerase; indicative of no contamination (Figure 5.12).

In a PCR with a gradient of temperatures using the template from LCL#3 cells, a band of similar size to the positive control of loop E could be appreciated (Figure 5.13 A). Aiming to reproduce this, a subsequent PCR assay was performed with template from latent and lytic cells at 55°C. The products were loaded in an agarose gel, and after an electrophoresis was run, the band of similar size was observed (Figure 5.13 B). Asking whether the amplified (continues on page 149)

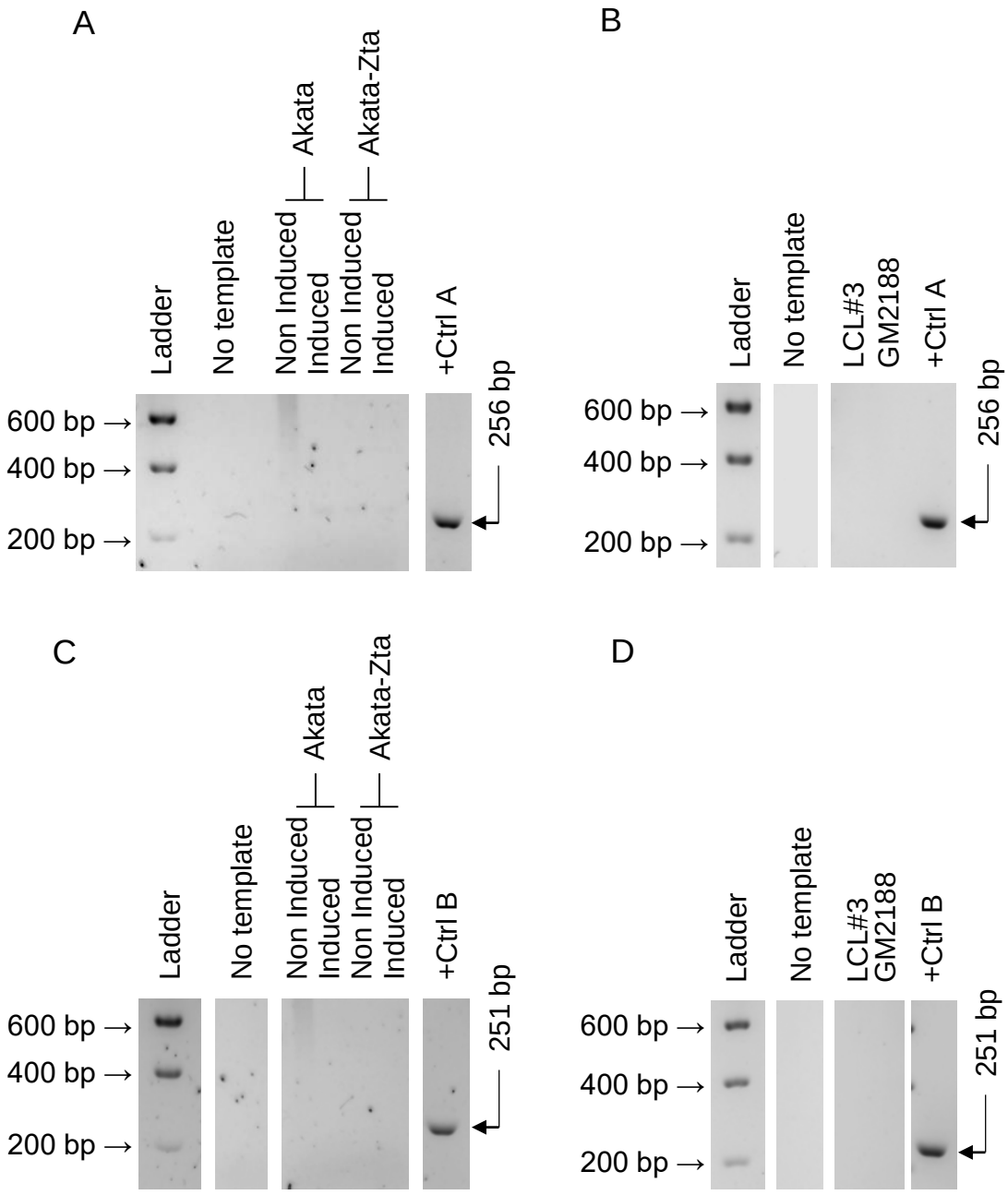


Figure 5.9. Chromosome conformation capture analysis of the SLC6A7 promotor region. Image captures of 1.2% Agarose gels after electrophoresis of PCR amplification products. PCR amplification was performed at 64°C. Each panel (A-D) is composed of sections (lanes) coming from the same gel. (**A**) The image shows no band of similar size to the synthetic positive control for loop A in lanes with the Akata and Akata-Zta templates. (**B**) The image shows no product amplification of comparable size to the synthetic positive control for loop A in lanes with LCL#3 and GM2188 templates. (**C**) The image shows no band of similar size to the synthetic positive control for loop B in lanes with the Akata and Akata-Zta templates. (**D**) The image shows no product amplification of comparable size to the synthetic positive control for loop B in lanes with LCL#3 and GM2188 templates. The same experiments were performed for negative controls C and D, not showing amplification (data not shown).

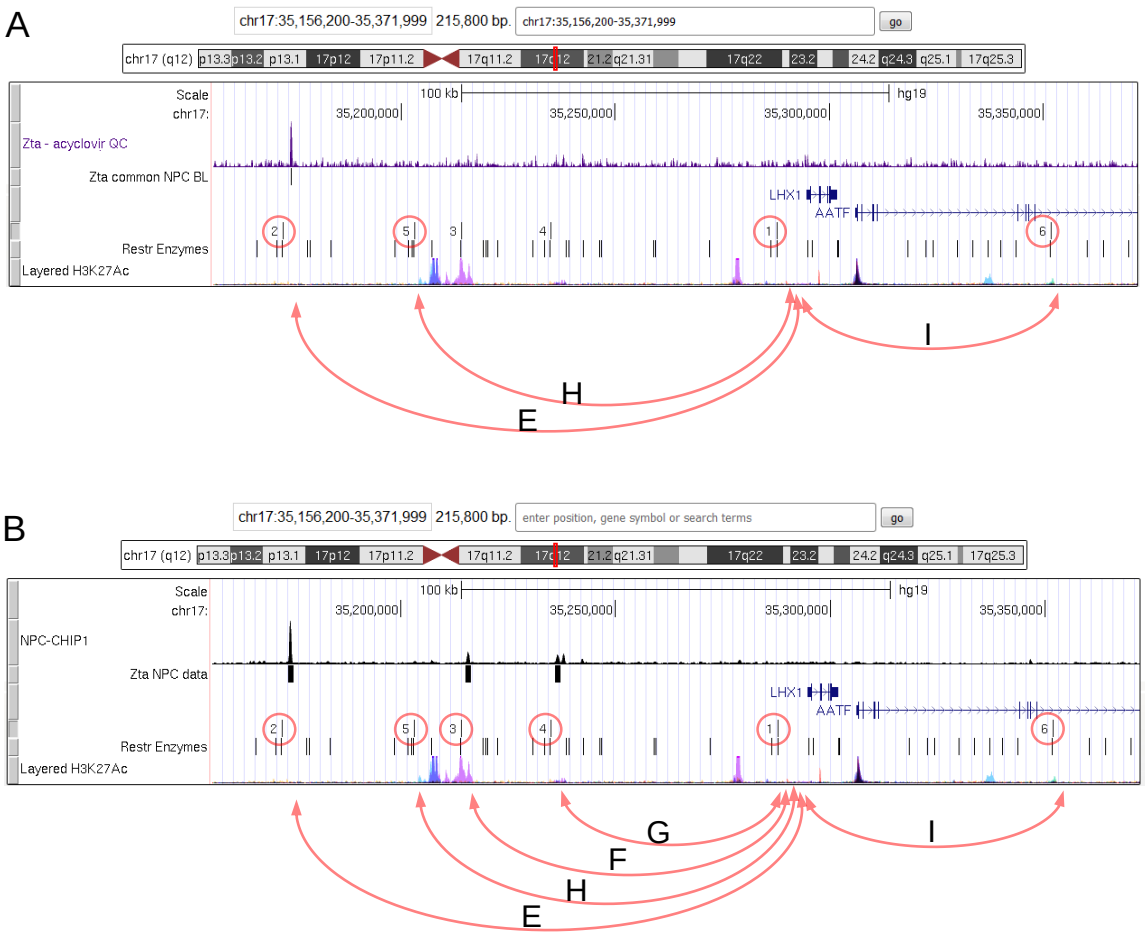
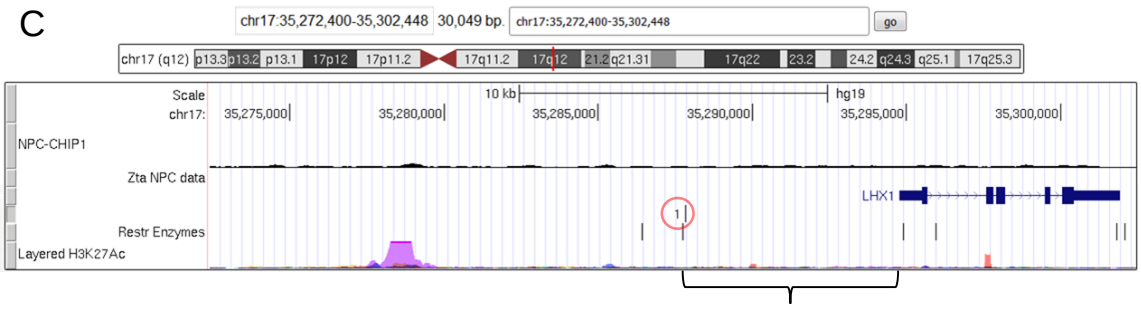


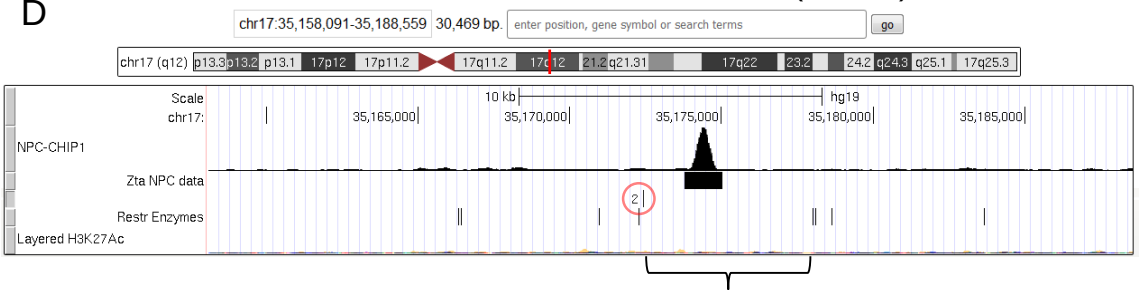
Figure 5.10. Maps of potential looping regions at LHX1. (**A**) Red arrows represent the potential loops forming across regions. The track labelled as “Zta – acyclovir QC” shows the stacked reads of Zta alignments to the hg19 release of the human genome. This data was previously generated in the lab, from ChIP-sequencing in Akata cells. The track labelled “Zta common NPC BL” corresponds to the peaks called by MACS from previous ChIP-Seq experiments performed in both HONE1 NPC cells and Akata BL cells. The genes displayed come from the UCSC Genes database. The following track shows circled in red, the alignments and names of the primers designed to amplify possible loops. The track labelled as “Restr Enzymes” shows the sites where HindIII aligns to. The track labelled as “Layered H3K27Ac” shows the regions with heavy acetylation on cells lines GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK, NHLF, signalling for active enhancing regions. (**B**) In the track labelled “NPC-CHIP1”, ChIP data from HONE1 NPC cells displays two more Zta peaks which were promising targets due to their proximity to the promoter region of the gene of interest and their alignment to a region with heavy acetylation. This is also reflected in the track labelled “Zta NPC data” in which the peaks from ChIP in NPC were called by MACS, therefore two more potential loops were considered. (**C**) Local map of the region containing the promoter region of gene LHX1. (**D**, **E**) Local maps of regions containing a Zta peak that could be looping to the region with the promoter of LHX1. (**F**) Local map of a region with both a Zta peak used to generate a potential loop and a region with low Zta signal, used to design a negative control. (**G**) Local map of a region with a low Zta signal, used to design another negative control.

C



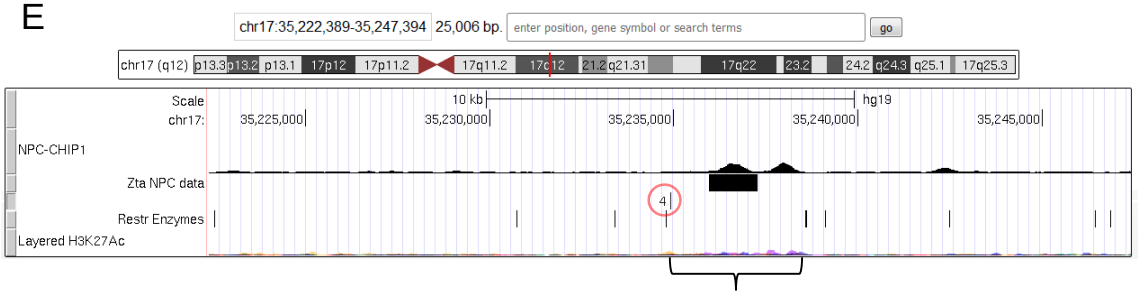
Fragment containing promoter of LHX1 (anchor)

D



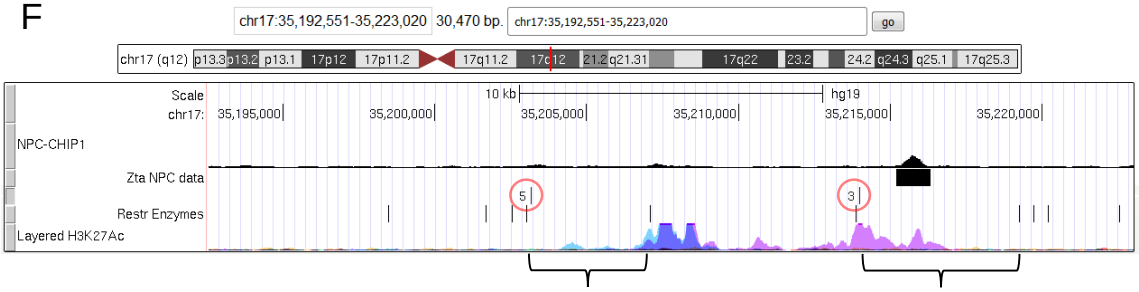
Fragment containing peak for loop E

E



Fragment containing peak for loop G

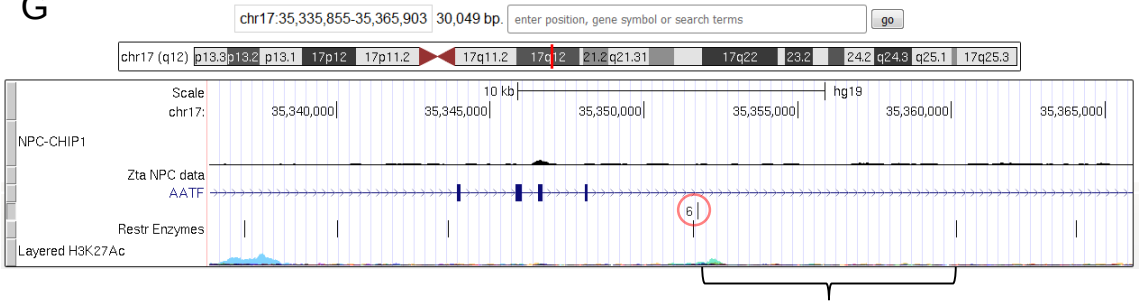
F



Fragment of interest for loop H

Fragment containing peak for loop F

G



Fragment of interest for loop I

A

Promoter Vs Potential Enhancer1

>E

GCAGGGAGGGGGGAGAGACTATAATAAAA
GAGGGTGAAGGCGGCCCCAGGAGGCTGGG
CCTCGTCCCAAACCATCCGTCTTCATTGT
CTCCGTGTGTGCAAAGGCAACTCTGGCAT
AAGGAAAACCCACTATTGGAGCCGAGGCT
TTATTTTATATCCTATACATCAGGAGGGA
GGCAAAGCTTGGGTGGGGGGCGGATTCTG
GGGTCAGTCTGAGGAGGCCTTGGGGAAGA
AGTGAGCGCTAAGCAAGGCTCCCAAGAGG
AATGGAAGGGCCTGAACCAAAGCTGGCTT
GCCCTCAGCCAGGGGGCAGTGCCCACAGA
TGCCCTGGATGGAGCCTGTCCAAGGCGTC
CTGTGTCTGCTAGTCTGAGCCAGGGATGA
GCCTGAACGGGCCTCCCATGCTATCCATC
TTTCTGTCTGATGGGCTGTTGTCAGTATC
CTCAAT

B

Promoter Vs Potential Enhancer2

>F

AACCCAGAAAGGAAAGAAAGTGAAGAAGA
AAGAAAAAGGCAGGGAGGGGGGAGAGACT
ATAATAAAAGAGGGTGAAGGCGGCCCCAG
GAGGCTGGGCCTCGTCCCAAACCATCCGT
CTTCATTGTCTCCGTGTGTGCAAAGGCAA
CTCTGGCATAAGGAAAACCCACTATTGGA
GCCGAGGCTTTATTTTATATCCTATACAT
CAGGAGGGAGGCAaagcttggggttgata
ctgaaagcagcagggaaacgtttttaaa
tagaagaatgatatagtttacagacagtt
tactccaggacctggcaaacagcaggcat
ttaataaatacctcttgagtgaatgaatg
aatcaagattggaggaggagagaccaatt
aggcggccacttcagcgacccaggtatcc
catcatgagggctgaacccagggcagctg
ccatggagTTTTGGGGGGACCCTCAGCCA
GGCACTGGGC

C

Promoter Vs Potential Enhancer3

>G

AGAAGGGCATATAAAAGGCAAATCCAACC
CAGAAAGGAAAGAAAGTGAAGAAGAAAGA
AAAAGGCAGGGAGGGGGGAGAGACTATAA
TAAAAGAGGGTGAAGGCGGCCCCAGGAGG
CTGGGCCTCGTCCCAAACCATCCGTCTTC
ATTGTCTCCGTGTGTGCAAAGGCAACTCT
GGCATAAGGAAAACCCACTATTGGAGCCG
AGGCTTTATTTTATATCCTATACATCAGG
AGGGAGGCAAAGCTTATCCAAGGAAGTTA
GTGGCTGGCCCTGGACCCTAAGGCCTGAC
TGAGCTAGGGGGTCCCTCCGGCCTGCTAT
GCCAGAGGAAGGCCGGCCTCAcgtgatgg
aggggagtgagcaagggcttggaagtcaa
acagccttgggttcaaatcctggctttta
tgagccttgtaacctcaagcaagtcattg
cccttgctgagcatctgacgggctcctcc
ttggccaatagggatgataggaccaactt
cacaggg

D

Promoter Vs Negative Control1

>H

AACCCAGAAAGGAAAGAAAGTGAAGAAGA
AAGAAAAAGGCAGGGAGGGGGGAGAGACT
ATAATAAAAGAGGGTGAAGGCGGCCCCAG
GAGGCTGGGCCTCGTCCCAAACCATCCGT
CTTCATTGTCTCCGTGTGTGCAAAGGCAA
CTCTGGCATAAGGAAAACCCACTATTGGA
GCCGAGGCTTTATTTTATATCCTATACAT
CAGGAGGGAGGCAAAGCTTTTTATCAGTC
AGCTTTAAAGATCAGCATTCCTTAGAGGC
TGctccatcttcctctctctctgtttctc
tctcgcatcccctccttctcctggtctct
ctatctcccttcccatctcacggtcccttt
tttcGGCACCGCATTTGTCTGTGTGGTTT
TTCTGCTCCTGGAAATCCCTTGCCCTCTT
CTGCCTCACTGCTCGCTGACAACAGGCCA
GGCCCTGCTCCGTGTCTCCAGGTGGCCGT
GGGGACAGAGAGGGGAGAGGGCAGGGGCA
AGACTGGCAGAGCCACC

Figure 5.11. Sequences used as template for synthetic controls of LHX1. These were generated by uniting at the restriction site; the reverse-complemented sequence of the anchor amplification target, (figure legend continues…)

E

Promoter Vs Negative Control2

>I

AATCCAACCCAGAAAGGAAAGAAAGTGAAGAAGAAAGAAAAAGGCAGGGAGGGGGGAGAGACT
ATAATAAAAGAGGGTGAAGGCGGCCCCAGGAGGCTGGGCCTCGTCCCAAACCATCCGTCTTCA
TTGTCTCCGTGTGTGCAAAGGCAACTCTGGCATAAGGAAAACCCACTATTGGAGCCGAGGCTT
TATTTTATATCCTATACATCAGGAGGGAGGCAAAGCTTTATTTTCATAGAACCAGAAAAAGTA
TTTTGTGTTAGAAAGTAATTTTTTTCCTTAGAAGGCAAATGATTTATTGAGCTTTTGTTACTT
AAACAATTCTAATGAAAAAATTAGCTTTATAGTTTTGTTACCGGTTTGTTTGGAAACTGGAAT
TGCTGAAAGCATGAAGTATAGCCAATGGTGAAGGTTTTATGAGAGAAACTGTTCATCTGCTTC
AGAAAAGCATTGCAGGATTTTACTTCCAAACAGA

(Fig legend continues…) and the sequence from the proposed loops amplification targets. Sequence highlighted in yellow represents HindIII. Sequence highlighted in cyan represents primer 1 from the anchor region. (**A**) Purple represents primer 2. (**B**) Gray represents primer 3. (**C**) Gold represents primer 4. (**D**) red represents primer 5. (**E**) Dark gray represents primer 6.

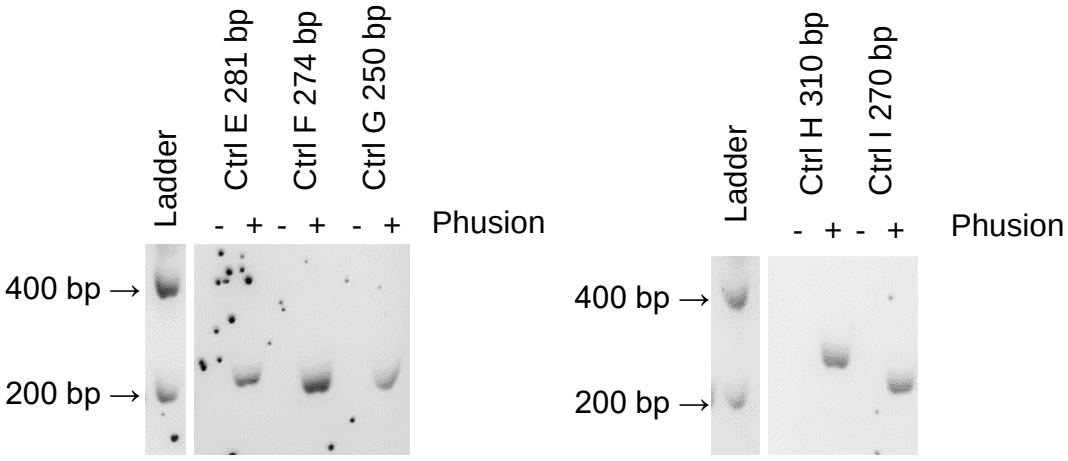


Figure 5.12. Amplification of PCR products from synthetic positive controls for loops with promoter of LHX1. A set of PCR reactions were performed using the synthetic positive control as template at 60°C. The reactions were performed with and without Phusion polymerase to corroborate that there was no contamination in any of the components of the reaction, the expected size of amplification to be correct, and that such polymerase would function as expected.

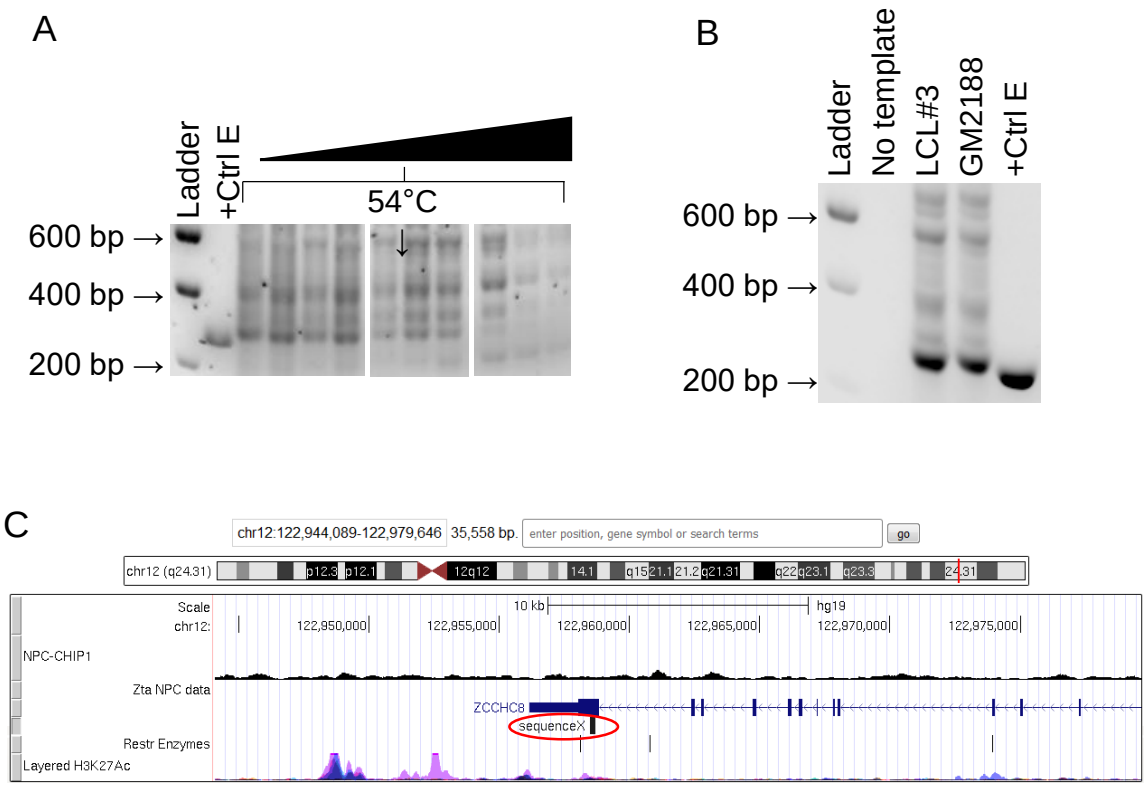
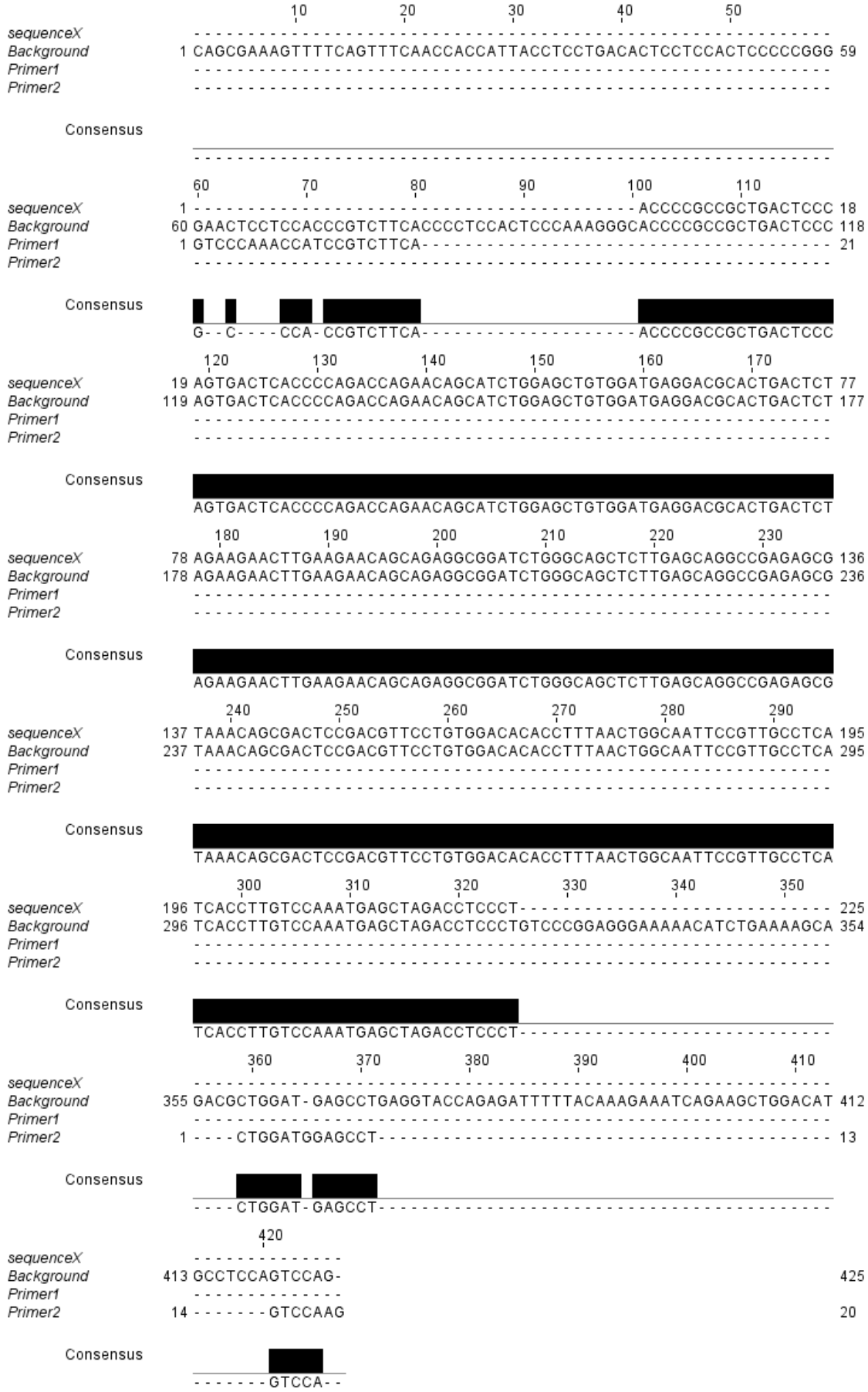


Figure 5.13. Non specific amplification of a false positive. (**A**) Initially, a gradient temperature PCR (11.5°C gradient with 54°C as midpoint) showed amplification of bands of similar size to the positive control in using LCL#3 template. (**B**) The PCR was performed with both LCL#3 and GM2188, at 55°C still showing a band of similar size to the control in both templates. The bands where excised and the product sequenced. The retrieved sequence (225bp) was labelled "sequenceX" and then aligned (BLAT) to the UCSC genome browser (**C**) where it was found not corresponding to the anchor region or restriction fragment intended to be amplified. (**D**) Alignment of sequences through ClustalOmega on Jalview. Black boxes in the Consensus track represent alignments of sequences. The sequenced product was labelled as "*sequenceX*" (circled in red). The region where the amplification of the sequenced product aligned-to (BLAT) was retrieved adding 100bp upstream and downstream "*Background*". The sequence of primer 1 "*Primer1*" and the reverse-complemented sequence of primer 2 "*Primer2*" are also in the alignment (See Figure 5.11 A). We can see a partial alignment from bp 60 to 80 of Primer 1, and on 359 to 371 we can see a partial alignment of Primer 2. This would explain why there is an amplification of a similar size to the positive control in both LCL#3 and GM2188 when the PCR amplification is performed at 55°C. It was because of this event, that every PCR was performed at 60°C to minimise any possibility of non-specific amplification.

D


sequenceX
Background
Primer1
Primer2
Consensus
1 CAGCGAAAGTTTTCAGTTTCAACCACCATTACCTCCTGACACTCCTCCACTCCCCCGGG 59
1 ACCCCGCCGCTGACTCCC 18
60 GAACTCCTCCACCCGTCTTCACCCCTCCACTCCCAAAGGGCACCCCGCCGCTGACTCCC 118
1 GTCCCAAACCATCCGTCTTCA 21
G--C----CCA-CCGTCTTCA----------ACCCCGCCGCTGACTCCC
19 AGTGACTCACCCCAGACCAGAACAGCATCTGGAGCTGTGGATGAGGACGCACTGACTCT 77
119 AGTGACTCACCCCAGACCAGAACAGCATCTGGAGCTGTGGATGAGGACGCACTGACTCT 177
AGTGACTCACCCCAGACCAGAACAGCATCTGGAGCTGTGGATGAGGACGCACTGACTCT
78 AGAAGAACTTGAAGAACAGCAGAGGCGGATCTGGGCAGCTCTTGAGCAGGCCGAGAGCG 136
178 AGAAGAACTTGAAGAACAGCAGAGGCGGATCTGGGCAGCTCTTGAGCAGGCCGAGAGCG 236
AGAAGAACTTGAAGAACAGCAGAGGCGGATCTGGGCAGCTCTTGAGCAGGCCGAGAGCG
137 TAAACAGCGACTCCGACGTTCCTGTGGACACACCTTTAACTGGCAATTCCGTTGCCTCA 195
237 TAAACAGCGACTCCGACGTTCCTGTGGACACACCTTTAACTGGCAATTCCGTTGCCTCA 295
TAAACAGCGACTCCGACGTTCCTGTGGACACACCTTTAACTGGCAATTCCGTTGCCTCA
196 TCACCTTGTCCAAATGAGCTAGACCTCCCT 225
296 TCACCTTGTCCAAATGAGCTAGACCTCCCTGTCCCGGAGGGAAAAACATCTGAAAAGCA 354
TCACCTTGTCCAAATGAGCTAGACCTCCCT
355 GACGCTGGAT-GAGCCTGAGGTACCAGAGATTTTTACAAAGAAATCAGAAGCTGGACAT 412
1 ----CTGGATGGAGCCT 13
----CTGGAT-GAGCCT
413 GCCTCCAGTCCAG- 425
14 -------GTCCAAG 20
-------GTCCA--

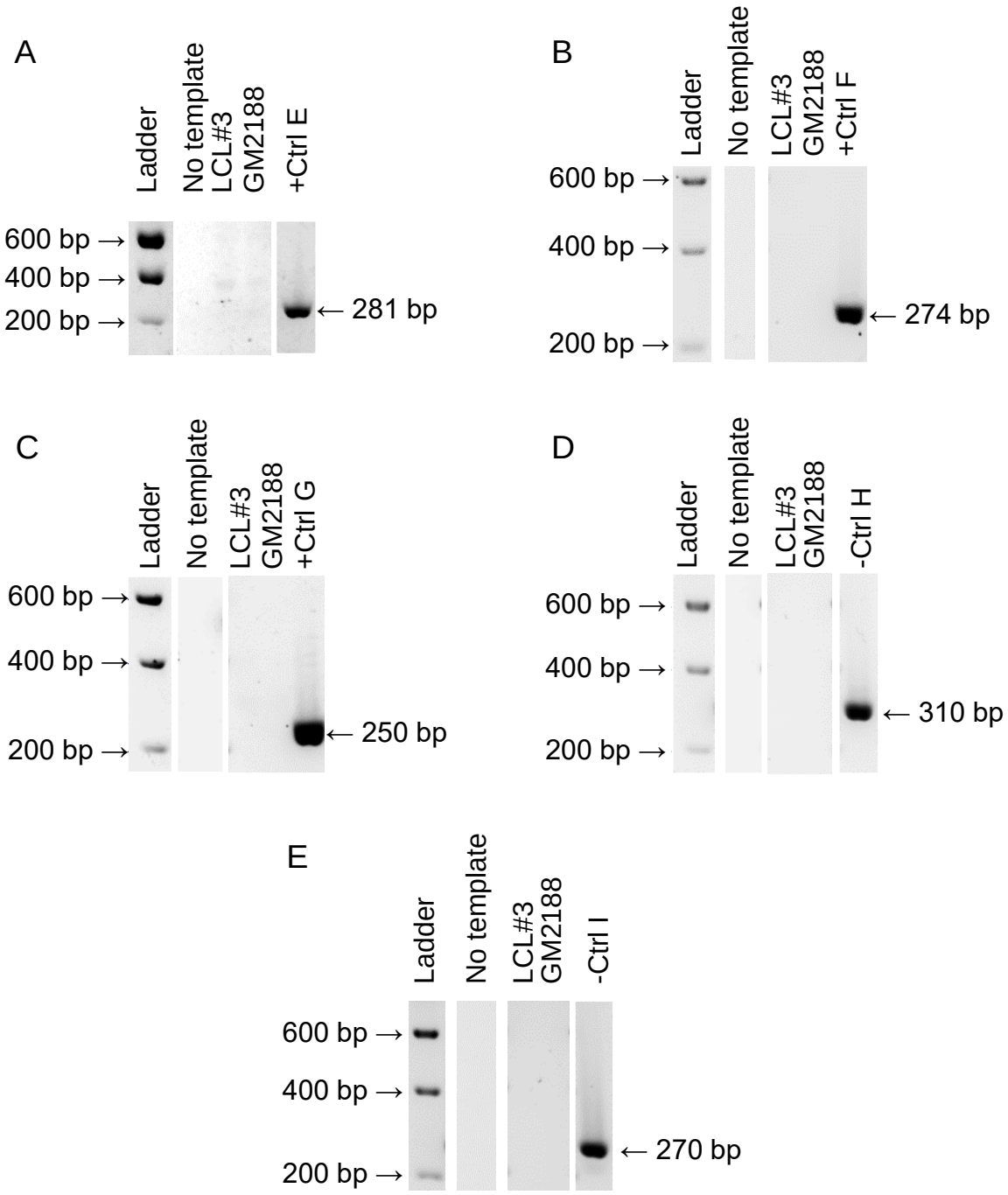


Figure 5.14. Chromosome conformation capture analysis of the LHX1 promotor region. Image captures of 1.2% Agarose gels after electrophoresis of PCR amplification products. The PCR amplification was performed at 64°C. Each panel (A-E) is composed of sections (lanes) coming from the same gel. (**A**) The image shows no product amplification of comparable size to the synthetic positive control for loop E in lanes with LCL#3 and GM2188 templates. (**B**) No product amplification of comparable size to the synthetic positive control for loop F in lanes with LCL#3 and GM2188 templates. (**C**) No product amplification of comparable size to the synthetic positive control for loop G in lanes with LCL#3 and GM2188 templates. (**D**, **E**) No product amplification of comparable size to the synthetic negative controls H and I.

(continues from page 140) product was the intended sequence or unintended amplification, the band was excised and sequenced.

Using the BLAT tool, the sequence retrieved was aligned in the genome browser, showing to be found at chromosome 12, a different location than the sequences considered when the 3C experiment was designed (Figure 5.13 C “sequenceX”). The sequence of the primers for the amplification of loop E (Primer 1 and Primer 2) were aligned to the sequence from the excised band, showing a partial alignment in the flanks of sequenceX (Figure 5.13 D). These results show that the band found is a product of non-specific amplification, probably due to the temperature used, and not the actual fragment of loop E.

Based on the previous experiment, all subsequent PCR experiments were performed with an amplification of 64 °C since this temperature showed no non-specific bands in previous gradient of temperatures. When executed with the latent and lytic template of GM2188 and LCL#3 cells, no bands of similar size to the positive control were found in the products of the PCR experiments (Figure 5.14). The positive controls were run in the same PCR reaction at the same temperature, time and day as the tested samples.

5.2.4 Zta and promoter sequence of FZD10

Based on the interactions of Zta found in the vicinity of gene FZD10, six potential loops were considered; four loops from fragments with Zta peaks to the promoter of FZD10 (Loops J, K, L, and M), and two loops from fragments without Zta peaks to the promoter of LHX1 (Loop N and O). These last two loops, designed to be considered negative looping controls (Figure 5.15).

The sequence of the fragments was used aiming to produce primers through Primer-blast. The primers were designed to be 20 nucleotides in length and to amplify towards the restriction site, so that if looping events were happening, the primers would amplify towards each other. The hypothesised resulting amplified sequences were assembled together virtually, using the sequences from the fragments, so that they could be synthetically generated to be used as template for positive controls. This resulted in 6 positive controls, one corresponding to each loop (Figure 5.16). (continues in page 152)

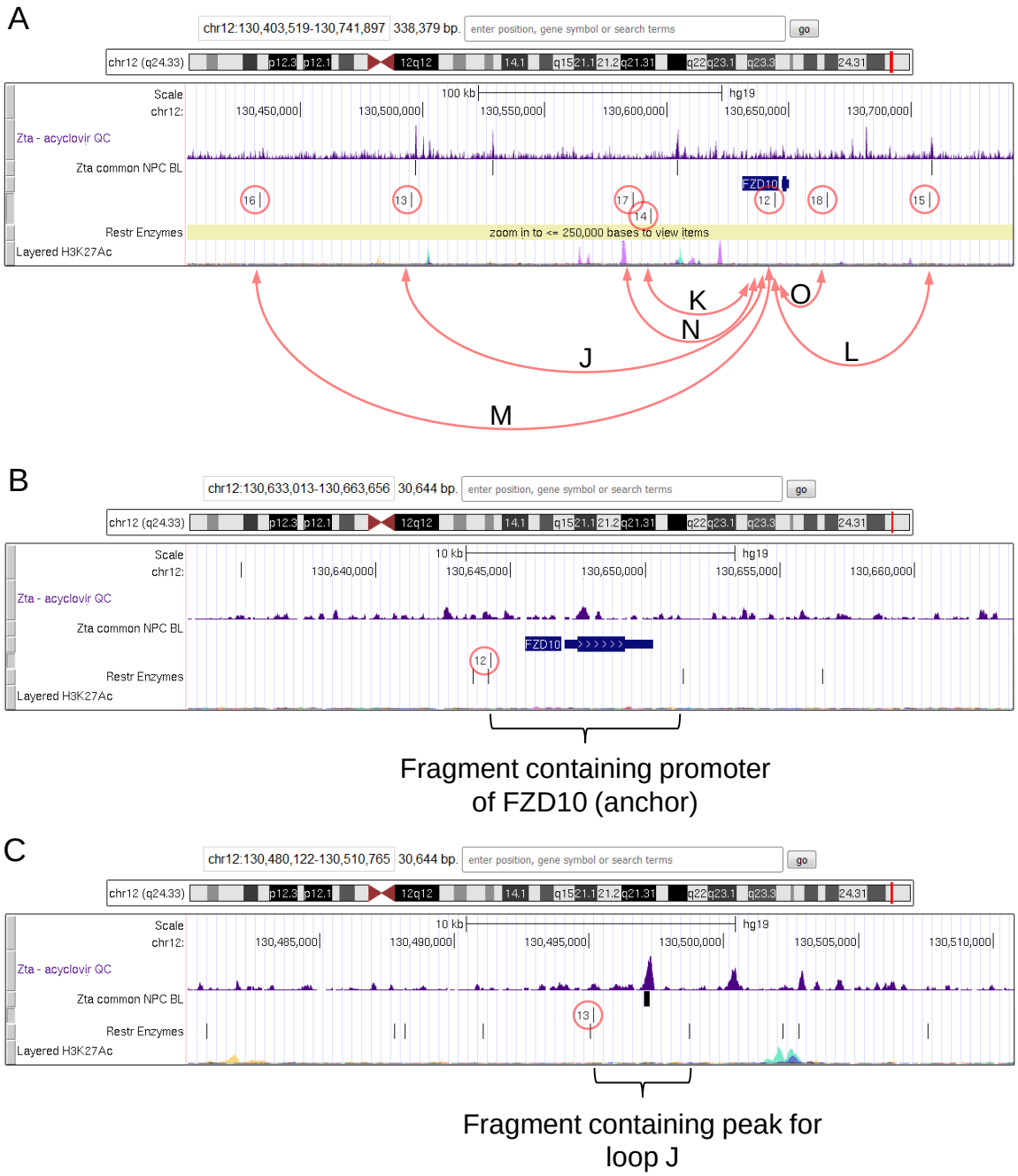
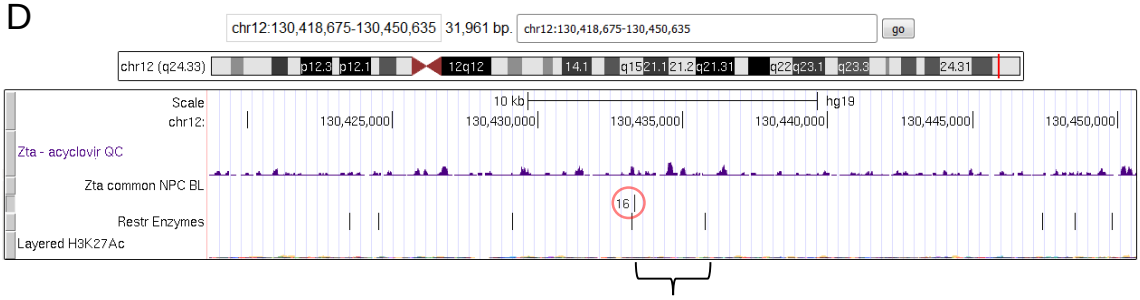


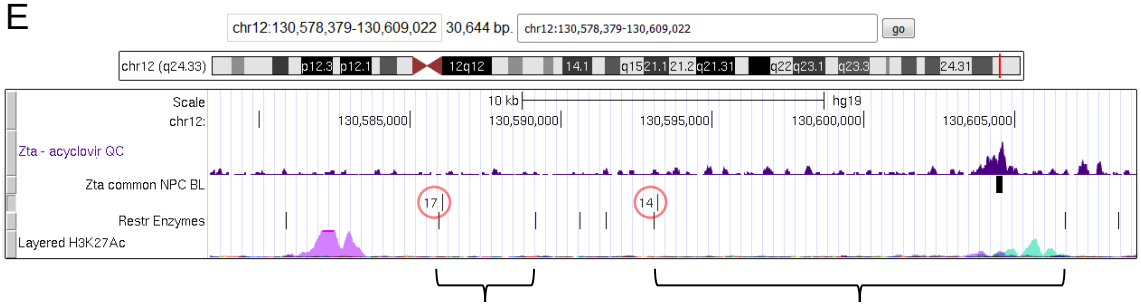
Figure 5.15. Map of potential looping regions at FZD10. (**A**) Red arrows represent the potential loops forming across regions. The track labelled as "Zta – acyclovir QC" shows the stacked reads of Zta alignments to the hg19 release of the human genome. This data was previously generated in the lab, from ChIP-sequencing in Akata cells. The track labelled "Zta common NPC BL" corresponds to the peaks called by MACS from previous ChIP-Seq experiments performed in both HONE1 NPC cells and Akata BL cells. The genes displayed come from the UCSC Genes database. The following track shows circled in red, the alignments and names of the primers designed to amplify possible loops. The track labelled as "Restr Enzymes" shows the sites where HindIII aligns to. The track labelled as "Layered H3K27Ac" shows the regions with heavy acetylation on cells lines GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK, NHLF, signalling for active enhancing regions. (**B**) Local map of the region containing the promoter region of gene FZD10 (anchor). (Figure legend continues…)

D



Fragment of interest
for loop M

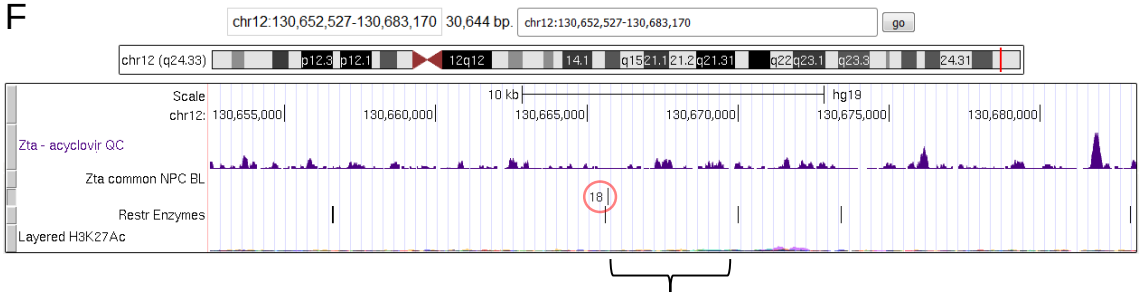
E



Fragment of interest
for loop N

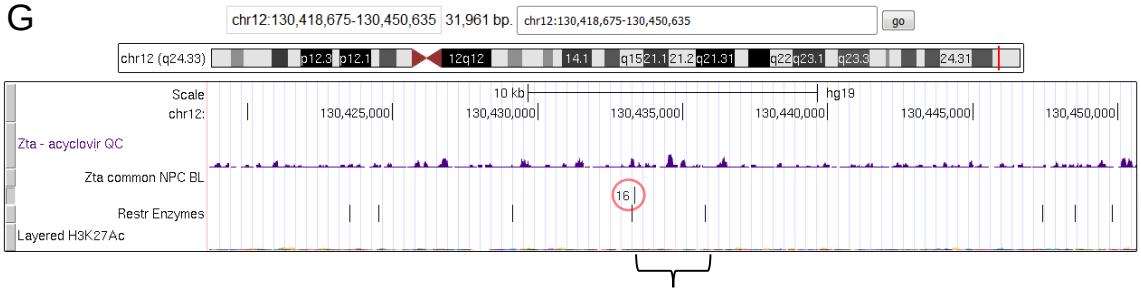
Fragment containing peak
for loop K

F



Fragment of interest
for loop O

G



Fragment of interest
for loop M

(cont…) (**C**, **D**, **E**) Local maps of regions containing a Zta peak that could be looping to the region with the promoter of FZD10. (**E**, **F**, **G**) Local maps of regions with a low Zta signal, used to design negative controls.

A

Promoter Vs Potential Enhancer1

>J

GTCCCCCGCTGCTCAGCCCTGGCTGGCAA
CCGACCCCCCTCTGGCAAATGTTATGGAG
AATTGGGAGGAATTTTATGGCTGTCTTTC
CCTCAGCTCTTTTTTCTTTTGTCTTCTTT
GAAGAATGACTTTAGACATTTTATTTACA
TATTTAACCATGGATGAAAGACGGAAAAG
AAAGCTTAGGCCACCAGCCAATGGTTTTC
TTTCTAAAGGAAAGGAGAGGAGAGGCGAA
CCTTAAAACGCCGTCTCACGGCTGATCTG
CCCTGCTTCCAGGTCAAGCCCAAGCGAGG
GCATGAGGGAGAAAGCAGGGGACGTAGAG
ACAGCCTGGCAAGAGCTGAGGCAGGAGGA
AACGAGCACCTCTTTCTCCTTCCTGAGCA
TT

B

Promoter Vs Potential Enhancer3

>K

AGCCGGGGTCCCCCGCTGCTCAGCCCTGG
CTGGCAACCGACCCCCCTCTGGCAAATGT
TATGGAGAATTGGGAGGAATTTTATGGCT
GTCTTTCCCTCAGCTCTTTTTTCTTTTGT
CTTCTTTGAAGAATGACTTTAGACATTTT
ATTTACATATTTAACCATGGATGAAAGAC
GGAAAAGAAAGCTTTTGAGGGCAAGGACT
GAATTTTTTACATTTATGCATGGAAGATG
GGAAAATGTTGAATATATTCACTTCAATA
AATGCCAAGAATACAGCAGTACCATAGCC
TGATACTCAAACTACTGCTTGGACCAACT
GTGGTAGATAAAAAATGACTACACATTct
ttgctccctctcccattgagaggtggaat
taaattccctttctcttgaacatgagctg
gccttggtgatgatt

C

Promoter Vs Potential Enhancer4

>L

AGCCGGGGTCCCCCGCTGCTCAGCCCTGG
CTGGCAACCGACCCCCCTCTGGCAAATGT
TATGGAGAATTGGGAGGAATTTTATGGCT
GTCTTTCCCTCAGCTCTTTTTTCTTTTGT
CTTCTTTGAAGAATGACTTTAGACATTTT
ATTTACATATTTAACCATGGATGAAAGAC
GGAAAAGAAAGCTTGCCATGCGCCCCGCG
TCATCCTGGCGTCTGCCTCCTTTGGCAGG
GGAATTAATGGTTTTTTAGAGGTTAATAA
CTTGTTGGTGTTTCCCCATTCCCTTGATG
GCGTTGTCATCTTCCATCTTTTTTCATTT
CCTAAGGAGGAAAATAACTTCTGCGGAAC
AAAGGCTGGGCAACTGCGCCAAAGGCAGC
CCTCAGAACCTTATCAAA

D

Promoter Vs Negative Control1

>M

AGCCGGGGTCCCCCGCTGCTCAGCCCTGG
CTGGCAACCGACCCCCCTCTGGCAAATGT
TATGGAGAATTGGGAGGAATTTTATGGCT
GTCTTTCCCTCAGCTCTTTTTTCTTTTGT
CTTCTTTGAAGAATGACTTTAGACATTTT
ATTTACATATTTAACCATGGATGAAAGAC
GGAAAAGAaagcttaagtaacttcccagg
gtcacctaagtagaaagtgtcaggacagg
aatcccaggcaggctttctagctcttgaa
tctgtgctctgagctgcaatgtcatcctg
tCTCTTGCAGGGCGTCATCTGGTTGTATC
TAACCAGGTGCATGAGACCGTGATCATCC
CATAAACCAAATAACATACAGCTTAGGAA
TGTTTCTCTTcaattctttttattatcc
agttccctttc

Figure 5.16. Sequences used as template for synthetic controls for FZD10. These were generated by uniting at the restriction site; the reverse-complemented sequence of the anchor amplification target, and the sequence from the proposed loops amplification targets. Sequence highlighted in yellow represents HindIII. Sequence highlighted in bright green represents promoter primer 12 from the anchor region. (**A**) Teal represents primer 13. (**B**) Blue represents primer 14. (**C**) Cyan represents primer 15. (**D**) dark green represents primer 16. (Figure legend continues…)

E

Promoter Vs Negative Control2

>N

AGCCGGGGTCCCCCGCTGCTCAGCCCTGG
CTGGCAACCGACCCCCCTCTGGCAAATGT
TATGGAGAATTGGGAGGAATTTTATGGCT
GTCTTTCCCTCAGCTCTTTTTTCTTTTGT
CTTCTTTGAAGAATGACTTTAGACATTTT
ATTTACATATTTAACCATGGATGAAAGAC
GGAAAAGAAAGCTTTCATATTCAAGAGCA
AAGTAACTGAATTTGAAACTTTTATTAAT
AATGTAAAAGGCTTGCGTGACACGCAGAA
GCTGGAAAAATATGGGGAAATCAGAGGAG
ACTTTGCTTTATCAAAGAGCAGCCAAATA
AAGGTTGGGCTTGCTTCTGCGGTGGGAGG
AGGGGAGAAACGCCTCCAGGGAGGAGGTC
AATGTCTTTTTCTCCACTTGAGTGCTAGG
ACTTTGGGACTGAGGACCACCAAT

F

Promoter Vs Negative Control3

>O

AGCCGGGGTCCCCCGCTGCTCAGCCCTGG
CTGGCAACCGACCCCCCTCTGGCAAATGT
TATGGAGAATTGGGAGGAATTTTATGGCT
GTCTTTCCCTCAGCTCTTTTTTCTTTTGT
CTTCTTTGAAGAATGACTTTAGACATTTT
ATTTACATATTTAACCATGGATGAAAGAC
GGAAAAGAAAGCTTATCAGGCATTTAATA
CATAAGATCTCCTTCCATCCCCACAAAAA
TATTGTTATTCCAGTTTTCCAGGAAGGAG
AAATGAGGGTTAAAGAGAAAGTAGGGAAA
AGAAAGGATGCTTGCATAGCCAAGGCCAC
CAATCAGGAGCAGAACAGTTATCAATCTT
CAGGTTGATTTTTCTTTAAATAATTCatt
gaggtgaaattcacacaacataaagtta

(cont…) (**E**) Red represents primer 17. (**F**) Dark grey represents primer 18.

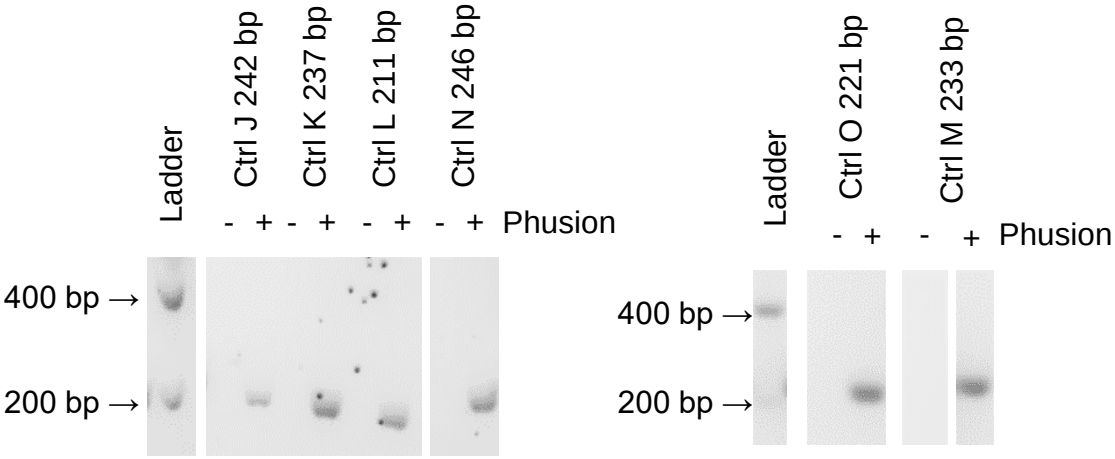


Figure 5.17. Amplification of PCR products from synthetic positive controls for loops with promoter of FZD10. A set of PCR reactions were performed using the synthetic positive control as template at 60°C. The reactions were performed with and without Phusion polymerase to corroborate that there was no contamination in any of the components of the reaction, the expected size of amplification to be correct, and that such polymerase would function as expected.

(continues from page 149) Aiming to test both the primers and the synthetic controls as template, an initial PCR was set up with and without the polymerase at 60°C, controlling for external contamination. This temperature was used since the design of the primers set the Tm to 60°C. Subsequently, the PCR products were loaded in an agarose gel, and an electrophoresis was run, resulting in clear bands of the intended size in lanes containing primers, template and polymerase. No amplification was observed in the lanes with primers and template, but without polymerase; indicative of no contamination (Figure 5.17).

A gradient of temperatures were used in PCR using the template from LCL#3 cells and the primers for loops A and B. The products were loaded in an agarose gel, and after an electrophoresis was run, it was observed that lower temperatures resulted in non-specific amplification, and higher temperatures reversed this (Figure 5.18). Synthetic positive controls were amplified at the closest temperature to 60°C, found in the gradient.

Aiming to test if evidence for looping events was observed, the PCR amplifications were performed at 64°C; and the products run in an agarose gel electrophoresis. No evidence of loops forming between fragments with Zta interactions and promoter region of FZD10 can be observed. (Figure 5.19). The positive controls were run in the same PCR reaction at the same temperature, time and day as the tested samples.

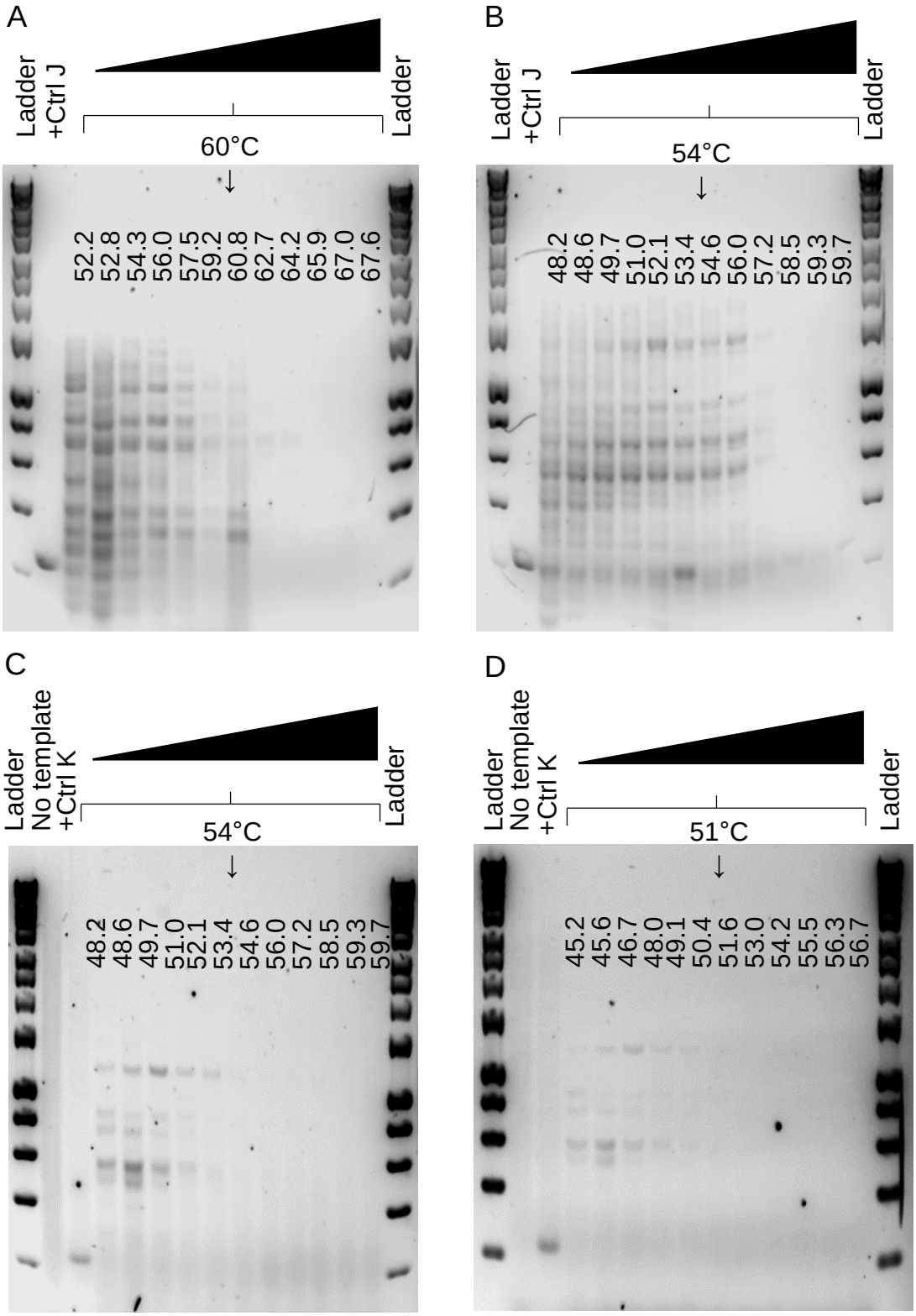
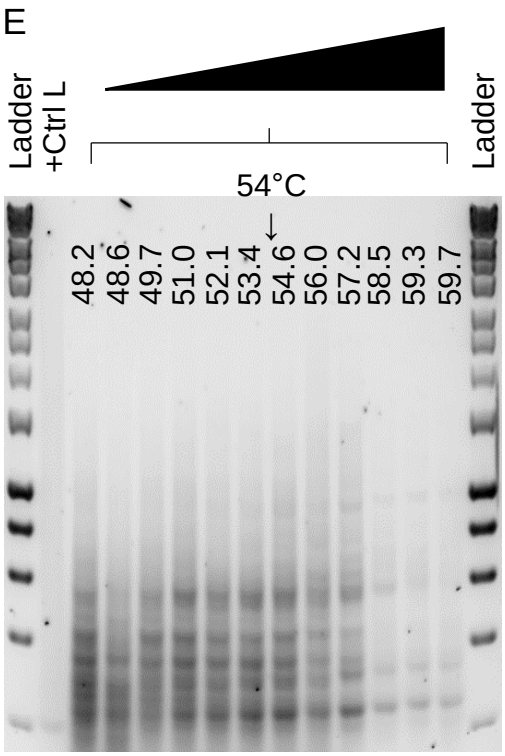


Figure 5.18. Temperature gradients used to optimise melting temperatures for loops to FZD10. Using template from LCL#3 cells, four gradients were tested searching for amplification of loops. Black triangle on top represents the increasing temperature in the gradient. (**A**) Performed at a 15.4°C gradient with 60°C as a midpoint testing for loop J. (**B**) Performed at a 11.5°C gradient with 54°C as midpoint testing for loop J. (**C**) Performed at a 11.5°C gradient with 54°C as a midpoint testing for loop K. (figure legend continues…)



(cont…) (**D**) Performed at a 11.5°C gradient with 51°C as midpoint testing for loop K.(**E**) Performed at a 11.5°C gradient with 54°C as midpoint testing for loop L. Each control was amplified at whichever temperature was closest to 60°C in the gradient.

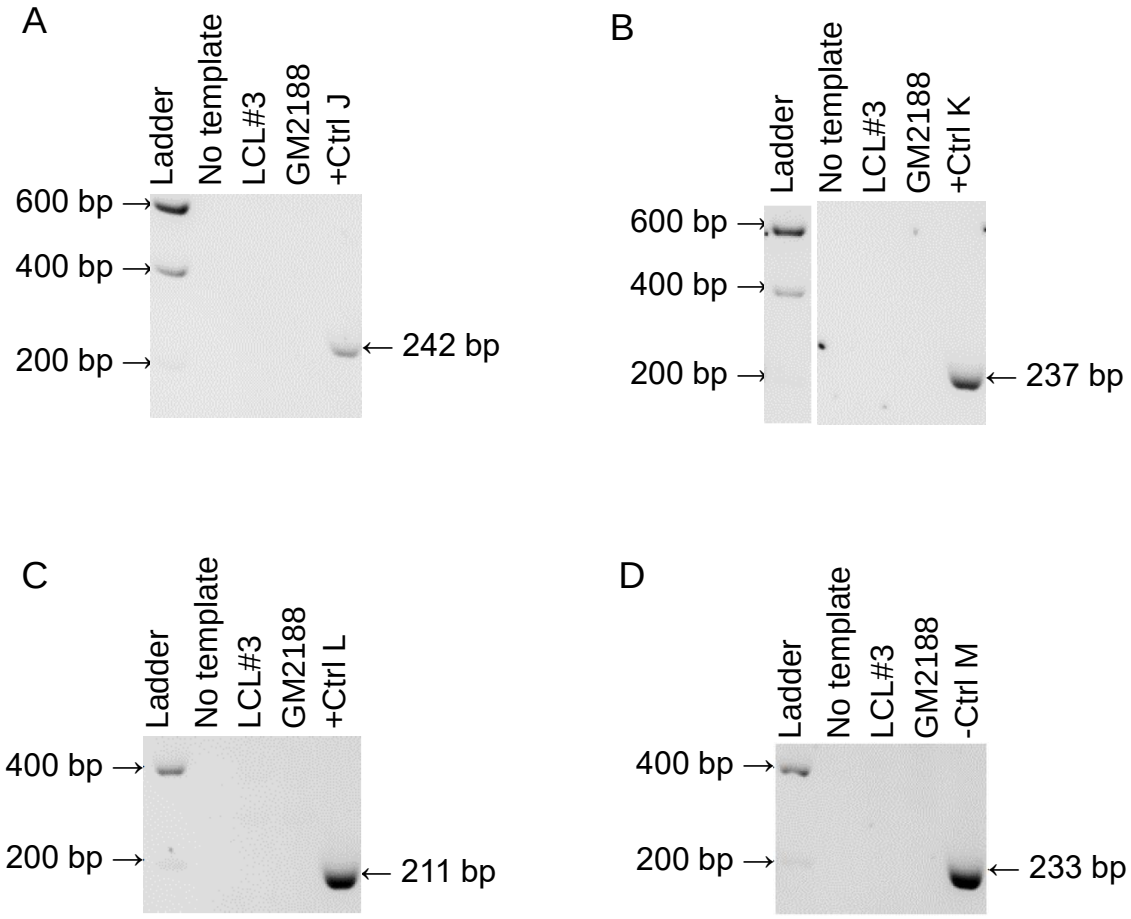


Figure 5.19. Chromosome conformation capture analysis of the FZD10 promotor region. Image captures of 1.2% Agarose gels after electrophoresis of PCR amplification products. The PCR amplification was performed at 64°C. Each panel (B) is composed of sections (lanes) coming from the same gel. (**A**) There is no band of similar size to the synthetic positive control for loop J in lanes with the LCL#3 and GM2188 templates. (**B**) The image shows no product amplification of comparable size to the synthetic positive control for loop K in lanes with LCL#3 and GM2188 templates. (**C**) There is no band of similar size to the synthetic positive control for loop L in lanes with the LCL#3 and GM2188 templates. (**D**) The image shows no product amplification of comparable size to the negative control for loop M in lanes with LCL#3 and GM2188 templates.

5.2.5 Zta and promoter sequence of cluster VWF and CD9

Since two different gene promoters were found with Zta interactions in its vicinity, two sets of loops were considered for cluster VWF and CD9 (Fig 5.20). For the promoter region of VWF loops CA and CB correspond to loops between the anchor fragment and fragments with Zta peaks. Given the specific properties found in the sequence of the anchor fragment of the promoter region of VWF, two primers instead of one had to be designed. The sequence close to the restriction site finds repetitive sequence surrounding a small “island” of unique sequence, this raises problems when trying to design primes that amplify regions 80 to a 150 bp away from the restriction site following the recommendations by Naumova et al (2012). Being aware that it could cause problems, rather than designing a primer in a repetitive sequence that might bind somewhere else, two primers were designed in the unique sequence, one being relatively close to the restriction site, and another one relatively far (Fig 5.21 A and B sequence highlighted in light grey and dark grey). This in turn duplicates the primer pairs used to amplify a single loop, in other words; CAs and CBs, reflect the same looping event as CA and CB respectively. Loops CE and CG were considered negative looping controls since they are formed from fragments without Zta peaks (Fig 5.20 A).

For the promoter region of gene CD9 two loops, CC and CD, were hypothesised based on the presence of Zta peaks in vicinity. Loops CF and CH were considered negative looping controls since they don't contain Zta peaks (Figure 5.20 B).

The virtually assembled sequences hypothesised to be forming through looping, were formed using the sequences from the fragments, so that they could be synthetically generated to be used as template for positive controls. This resulted in 8 positive controls, one corresponding to each loop (Figure 5.21)

Aiming to test both the primers and the synthetic controls template, an initial PCR was set up with and without the polymerase at 60°C, controlling for external contamination. This temperature was chosen since the design of the primers set the Tm to 60°C. Subsequently, the PCR products were loaded in an agarose gel, and an electrophoresis was run, resulting in clear bands of the intended size in lanes containing primers, template and polymerase. (continues in page 164)

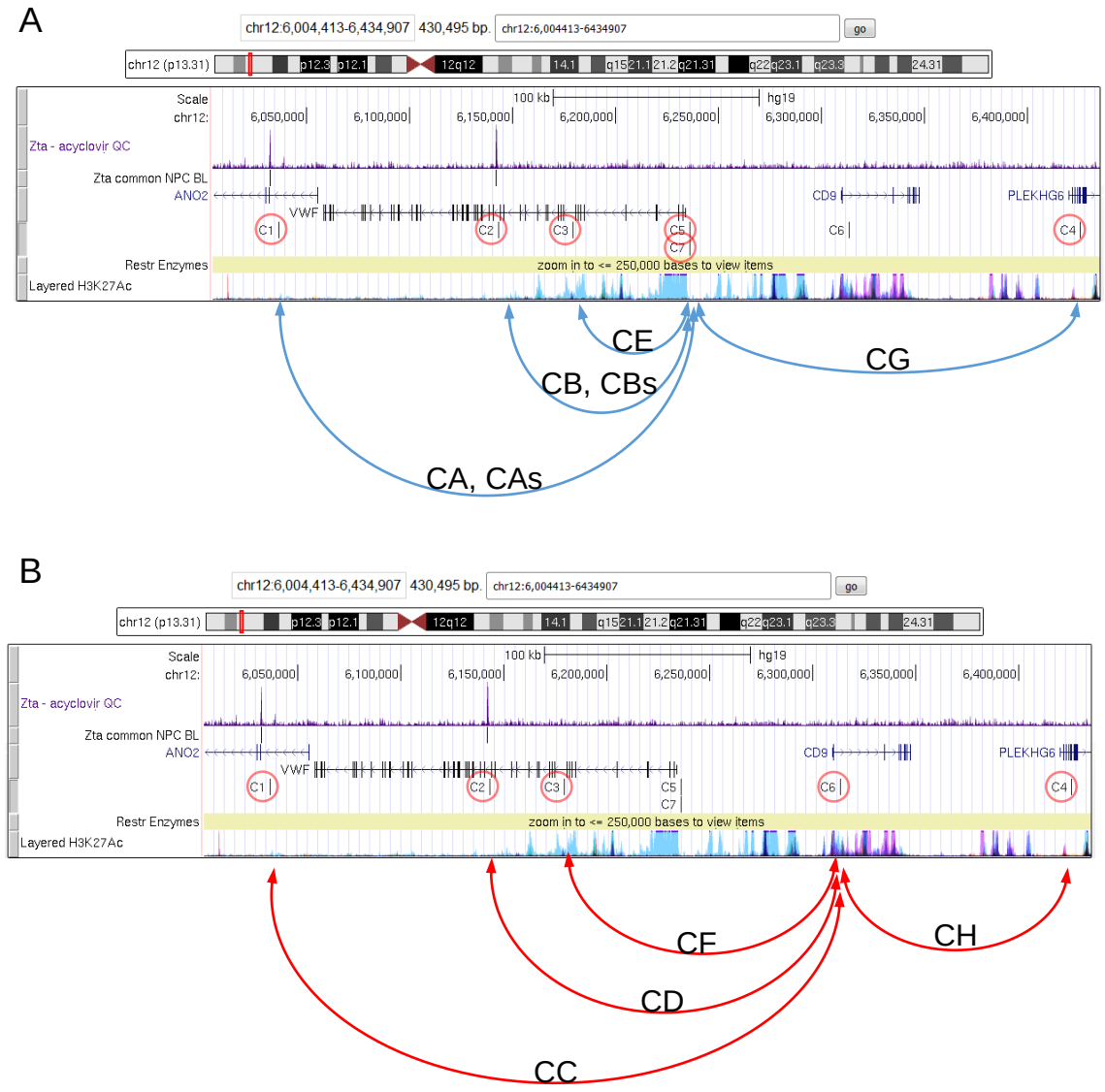
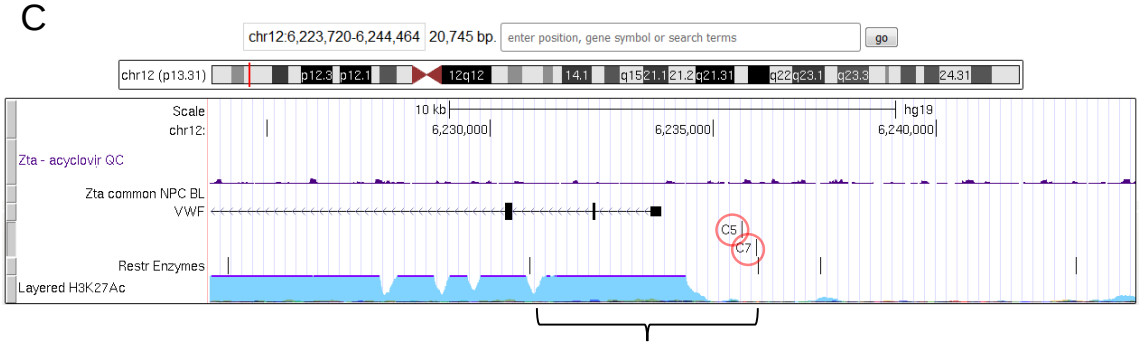


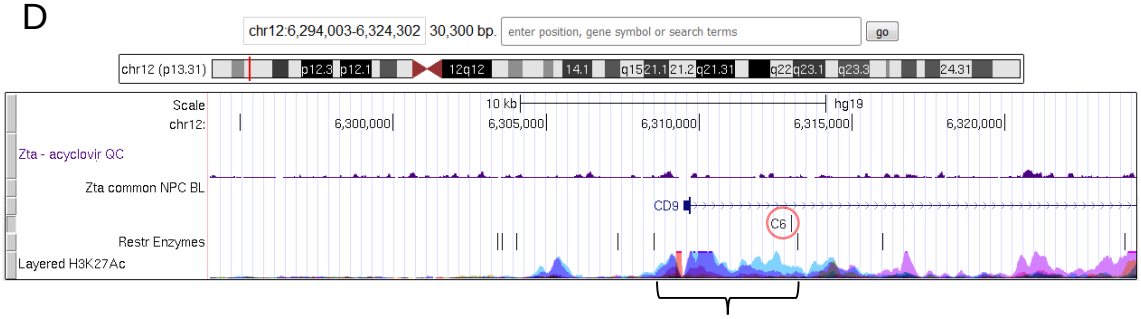
Figure 5.20. Maps of potential looping regions at cluster VWF-CD9. The Zta peaks in proximity to a cluster of genes increase the level of complexity. However this can be broken down into the potential loops to each gene. (**A**) Blue arrows represent the potential loops forming across regions to the VWF promoter fragment (anchor). The track labelled as “Zta – acyclovir QC” shows the stacked reads of Zta alignments to the hg19 release of the human genome. This data was previously generated in the lab, from ChIP-sequencing in Akata cells. The track labelled “Zta common NPC BL” corresponds to the peaks called by MACS from previous ChIP-Seq experiments performed in both HONE1 NPC cells and Akata BL cells. The genes displayed come from the UCSC Genes database. The following track shows circled in red, the alignments and names of the primers designed to amplify possible loops. The track labelled as “Restr Enzymes” shows the sites where EcoRI aligns to. The track labelled as “Layered H3K27Ac” shows the regions with heavy acetylation on cell lines GM12878, H1-hESC, HSMM, HUVEC, K562, NHEK, NHLF, signalling for active enhancing regions. Circled in red, are the alignments and names of the primers designed to amplify possible loops. (Figure legend continues…)

C



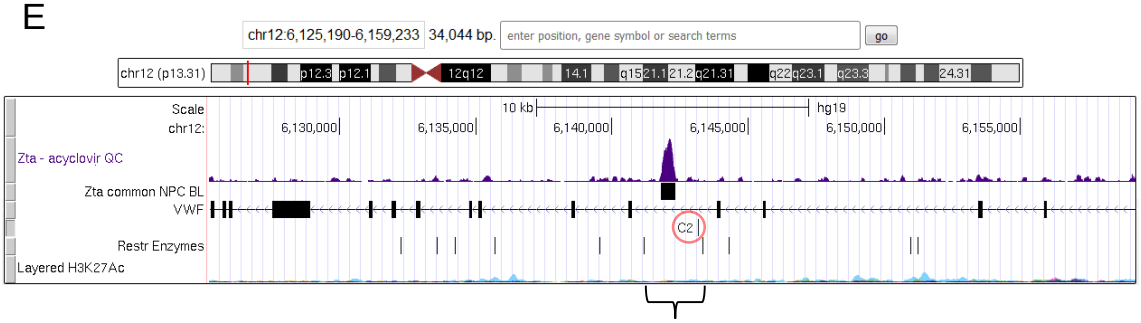
Fragment containing promoter of VWF (anchor)

D



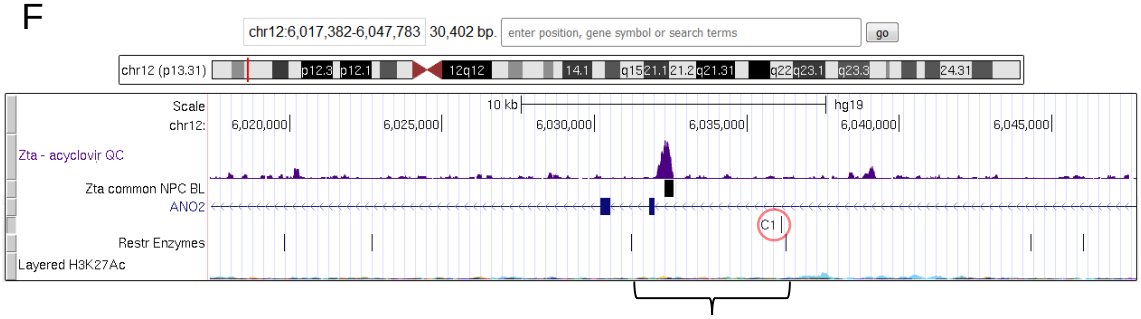
Fragment containing promoter of CD9 (anchor)

E



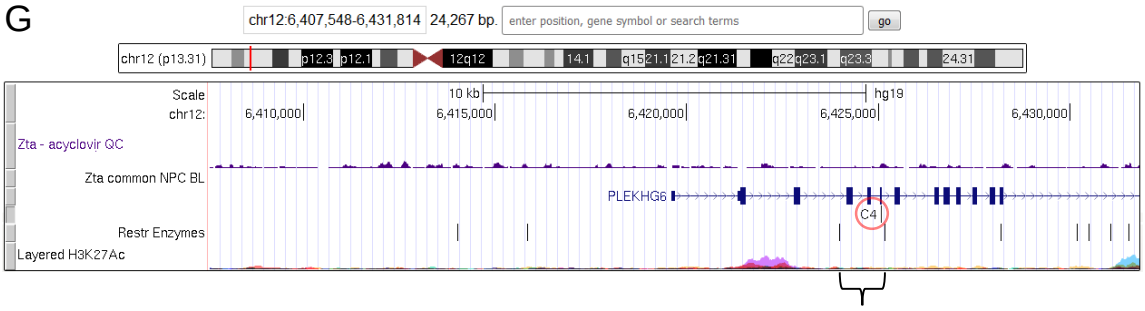
Fragment containing peak for loop CB and loop CD

F



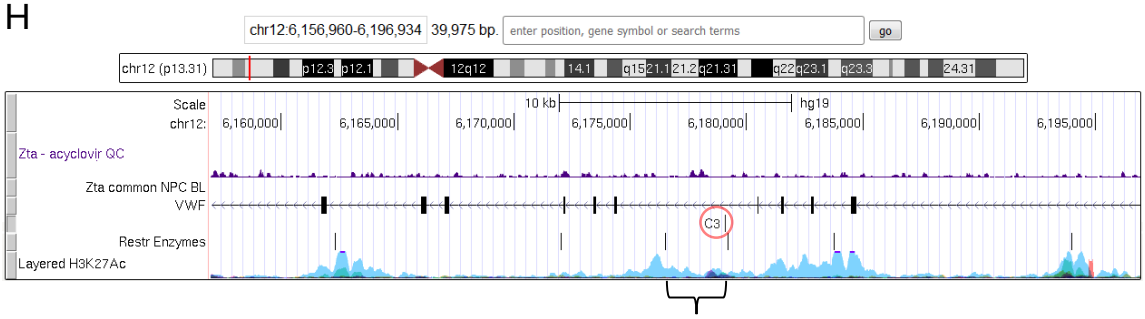
Fragment containing peak for loop CA and loop CC

G



Fragment of interest for loop CG and loop CH

H



Fragment of interest for loop CE and loop CF

(cont..) (**B**) Red arrows represent the potential loops forming across regions to the CD9 promoter fragment (anchor). (**C**) Local map of the region containing the promoter region of gene VWF. Due to the specific properties of the sequence in this fragment, two primers had to be designed even though they have the same target of amplification. (**D**) Local map of the region containing the promoter region of gene CD9. (**E**, **F**) Local maps of regions containing a Zta peak that could be looping to the anchors of VWF or CD9. (**G, H**) Local maps of regions with a low Zta signal, used to design negative controls.

A Promoter1 Vs Potential Enhancer1

>CA

AACATTCCTTGGAAGTACATGAGCTATGA
AAAATGCACAGGATAATTATCAAACCCTA
CTTCAGGACGTCCATGCTGCCTCCTTCAT
CCGTCTCTTCTCCACTTCATTCCAGCTGC
TCTGCTTCAAGTCCCTGGaattcatcgtc
aagagagctttatttgCATGAGTGCAAAG
GATGAAAATTCTAgactgggcgtggtggc
tcacgcctgtaatcccagcactttgggag
accgaggtgggcagatcacgaggtcagga
gtttgagaccagcctggctaacatagtgg
aaccccatctctactaaaaatacaaaaaa
ttagctgggtgtagtggtgtgtgcatgta
atcccagctacttgggaggctgaggcagg
agaattgcttgaagccgggaggcagaggt
tgcagtgagccatgattgcatcactgcac
tccagcccagcggacagtgcgagactcca
tctcaaaaaaaaaaaaagaaagaaaagaa
TATTCTAAAAAAAGACTTAATTCC

B Promoter1 Vs Potential Enhancer2

>CB

AAACTTACAAGAATCCACGCAGTGGGGAG
TGGGAGGCAGAGGGGCTGCTGCCCACACT
CACCGTGAGGATTCCTGGATCGCTCTCAT
CTGCTGGGCAGGATGACTTTCCTGTGTCC
CATCCTCCCACTCCTCCCATTCTGAATTG
TACCAAGTTAGACATTAAAATGGCGGAGa
attcatcgtcaagagagctttatttgCAT
GAGTGCAAAGGATGAAAATTCTAgactgg
gcgtggtggctcacgcctgtaatcccagc
actttgggagaccgaggtgggcagatcac
gaggtcaggagtttgagaccagcctggct
aacatagtggaaccccatctctactaaaa
atacaaaaaattagctgggtgtagtggtg
tgtgcatgtaatcccagctacttgggagg
ctgaggcaggagaattgcttgaagccggg
aggcagaggttgcagtgagccatgattgc
atcactgcactccagcccagcggacagtg
cgagactccatctcaaaaaaaaaaaaaga
aagaaaagaaTATTCTAAAAAAAGACTTA
ATTCC

C Promoter2 Vs Potential Enhancer1

>CC

AACATTCCTTGGAAGTACATGAGCTATGA
AAAATGCACAGGATAATTATCAAACCCTA
CTTCAGGACGTCCATGCTGCCTCCTTCAT
CCGTCTCTTCTCCACTTCATTCCAGCTGC
TCTGCTTCAAGTCCCTGGaattccagtg
taggtctctgtgtttcgcACCAGCCTCCC
CTAGGTCTAGGGGATATGGTTTCTTGGGA
AAGAGACCTCCCAGCCCCACACTGCTGTT
TCCATCAGTGAGCGAGGCATGAGAGACAG
TAAGCAAAGGGGCTGAGGGAGGTGGGGAG
GGGATTGGGCTCTGTCAGTATAGCCAAAC
CTTTCCCTCTTTTGCCCTAAGGAGTTCCT
CCCCTTCTCTGACCAGAAAGAGGGCAGAC
ATGCATGTGTCAGACTGGACTGGCCCTAA
CCCTGGTTCCCCATTTC

D Promoter2 Vs Potential Enhancer2

>CD

AAACTTACAAGAATCCACGCAGTGGGGAG
TGGGAGGCAGAGGGGCTGCTGCCCACACT
CACCGTGAGGATTCCTGGATCGCTCTCAT
CTGCTGGGCAGGATGACTTTCCTGTGTCC
CATCCTCCCACTCCTCCCATTCTGAATTG
TACCAAGTTAGACATTAAAATGGCGGAGa
attccagtgtaggtctctgtgtttcgcA
CCAGCCTCCCCTAGGTCTAGGGGATATGG
TTTCTTGGGAAAGAGACCTCCCAGCCCCA
CACTGCTGTTTCCATCAGTGAGCGAGGCA
TGAGAGACAGTAAGCAAAGGGGCTGAGGG
AGGTGGGGAGGGGATTGGGCTCTGTCAGT
ATAGCCAAACCTTTCCCTCTTTTGCCCTA
AGGAGTTCCTCCCCTTCTCTGACCAGAAA
GAGGGCAGACATGCATGTGTCAGACTGGA
CTGGCCCTAACCCTGGTTCCCCATTTC

Figure 5.21. Sequences used as template for synthetic controls for cluster VWF and CD9. These were generated by uniting at the restriction site; the reverse-complemented sequence of the anchor amplification target, and the sequence from the proposed loops amplification targets. Sequence highlighted in green represents EcoRI. Please note that there is a big a zone of repetitive sequence in the fragment of the VWF anchor, therefore 2 primers had to be devised. (**A** to **H**) Sequence highlighted (figure legend continues…)

E

Promoter1 Vs Negative Control1

>CE

GGAAGGATGATGGACGGAAGATTCACCCT
ATAGTAAGTACTTGTTAAAACTTCCTAAT
CTGGTGGTTACAAACGTCTAAAAAAACAC
ACACAATAAACTGaattcatcgtcaagag
agctttatttgCATGAGTGCAAAGGATGA
AAATTCTAgactgggcgtggtggctcacg
cctgtaatcccagcactttgggagaccga
ggtgggcagatcacgaggtcaggagtttg
agaccagcctggctaacatagtggaaccc
catctctactaaaaatacaaaaaattagc
tgggtgtagtggtgtgtgcatgtaatccc
agctacttgggaggctgaggcaggagaat
tgcttgaagccgggaggcagaggttgcag
tgagccatgattgcatcactgcactccag
cccagcggacagtgcgagactccatctca
aaaaaaaaaaaagaaagaaaagaaTATTC
TAAAAAAAGACTTAATTCC

F

Promoter2 Vs Negative Control1

>CF

GGAAGGATGATGGACGGAAGATTCACCCT
ATAGTAAGTACTTGTTAAAACTTCCTAAT
CTGGTGGTTACAAACGTCTAAAAAAACAC
ACACAATAAACTGaattccagtgtaggt
ctctgtgtttcgcACCAGCCTCCCCTAGG
TCTAGGGGATATGGTTTCTTGGGAAAGAG
ACCTCCCAGCCCCACACTGCTGTTTCCAT
CAGTGAGCGAGGCATGAGAGACAGTAAGC
AAAGGGGCTGAGGGAGGTGGGGAGGGGAT
TGGGCTCTGTCAGTATAGCCAAACCTTTC
CCTCTTTTGCCCTAAGGAGTTCCTCCCCT
TCTCTGACCAGAAAGAGGGCAGACATGCA
TGTGTCAGACTGGACTGGCCCTAACCCTG
GTTCCCCATTTC

G

Promoter1 Vs Negative Control2

>CG

GACTGCTGATGGAAGTGAGTGGGTGCTCA
GGAGGGGACCCTGGCACAGCCCGACCTCT
GAGCCTGGGGATGAGGGTGGGCCCTCCAG
CCATCCCTGTCTGaattcatcgtcaagag
agctttatttgCATGAGTGCAAAGGATGA
AAATTCTAgactgggcgtggtggctcacg
cctgtaatcccagcactttgggagaccga
ggtgggcagatcacgaggtcaggagtttg
agaccagcctggctaacatagtggaaccc
catctctactaaaaatacaaaaaattagc
tgggtgtagtggtgtgtgcatgtaatccc
agctacttgggaggctgaggcaggagaat
tgcttgaagccgggaggcagaggttgcag
tgagccatgattgcatcactgcactccag
cccagcggacagtgcgagactccatctca
aaaaaaaaaaaagaaagaaaagaaTATTC
TAAAAAAAGACTTAATTCC

H

Promoter2 Vs Negative Control2

>CH

GACTGCTGATGGAAGTGAGTGGGTGCTCA
GGAGGGGACCCTGGCACAGCCCGACCTCT
GAGCCTGGGGATGAGGGTGGGCCCTCCAG
CCATCCCTGTCTGaattccagtgtaggt
ctctgtgtttcgcACCAGCCTCCCCTAGG
TCTAGGGGATATGGTTTCTTGGGAAAGAG
ACCTCCCAGCCCCACACTGCTGTTTCCAT
CAGTGAGCGAGGCATGAGAGACAGTAAGC
AAAGGGGCTGAGGGAGGTGGGGAGGGGAT
TGGGCTCTGTCAGTATAGCCAAACCTTTC
CCTCTTTTGCCCTAAGGAGTTCCTCCCCT
TCTCTGACCAGAAAGAGGGCAGACATGCA
TGTGTCAGACTGGACTGGCCCTAACCCTG
GTTCCCCATTTC

(**cont…**) in light gray represents the alignment region of the promoter primer C5. Sequence highlighted in dark gray represents the alignment region of the promoter primer C7. Sequence in yellow represents promoter primer C6. Dark green represents primer C1. purple represents primer C2, dark red represents primer C3, sequence underlined represents primer C4.

(continues from page 158) No amplification was observed in the lanes with primers and template, but without polymerase; indicative of no contamination (Figure 5.22).

To test if evidence for looping events was observed, the PCR amplifications were performed at 64°C since this temperature minimized the non-specific amplification in previous experiments. Then, the products were run in an agarose gel electrophoresis. No evidence of loops forming between fragments with Zta interactions and promoter regions of VWF or CD9 can be observed. (Figure 5.23). The positive controls were run in the same PCR reaction at the same temperature, time and day as the tested samples.

5.2.6 Template controls

Since no product of comparable size to the positive control was found in the agarose electrophoresis in the previous experiments across the several different cell templates; a subsequent experiment was performed to corroborate that indeed the different templates contained DNA. Primers aiming to amplify a fragment of DNA were used in PCR reactions with the same concentration of template and amplifying temperature. This resulted in clear bands of expected size in 6 out of 8 templates tested, reflecting that those templates used in previous experiments were capable of producing bands, and that if an absence of bands is noted, is not due to an inadequate template production (Figure 5.24). The template for Akata-Zta shows faint bands, however, in Fig 5.3 it is clear that there is Chromatin present in those samples.

Finally, to validate the quality of the 3C process, primers for the MYC loop proved to be present in ER-EB cells (Wood et al. 2016) were used in a PCR (Figure 5.25 B). A synthetic template was used as control; in the lane with template from ER-EB cells, as expected, a band of similar size to the control can be observed. In the lanes with template from LCL#3 cells and GM2188 cells, a band of similar size can be found (Figure 5.25 A). Although the band can be seen fainter than in template coming from ER-EB cells, it cannot be expected to be of the same intensity, since different cells lines provide a different context. The presence of the band reflects a looping event being detected; this in turn means that the 3C

method employed to generate the template of LCL#3 cells and GM2188 cells is valid and that looping events can be detected using these templates (Figure 5.25).

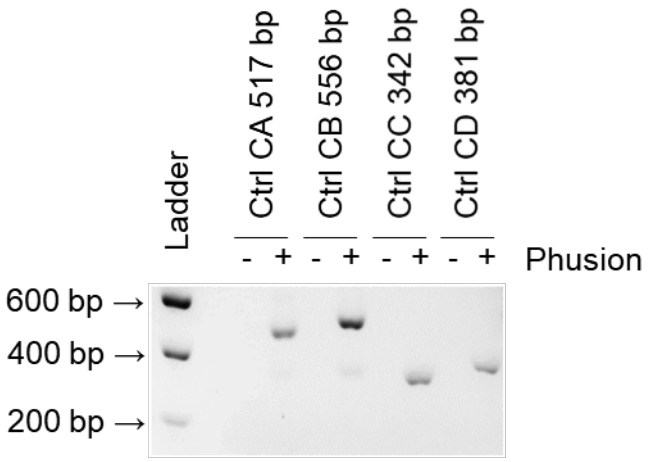


Figure 5.22. Amplification of PCR products from synthetic positive controls for loops with cluster VWF and CD9. A set of PCR reactions were performed using the synthetic positive control as template at 60°C. The reactions were performed with and without Phusion polymerase to corroborate that there was no contamination in any of the components of the reaction, the expected size of amplification to be correct, and that such polymerase would function as expected.

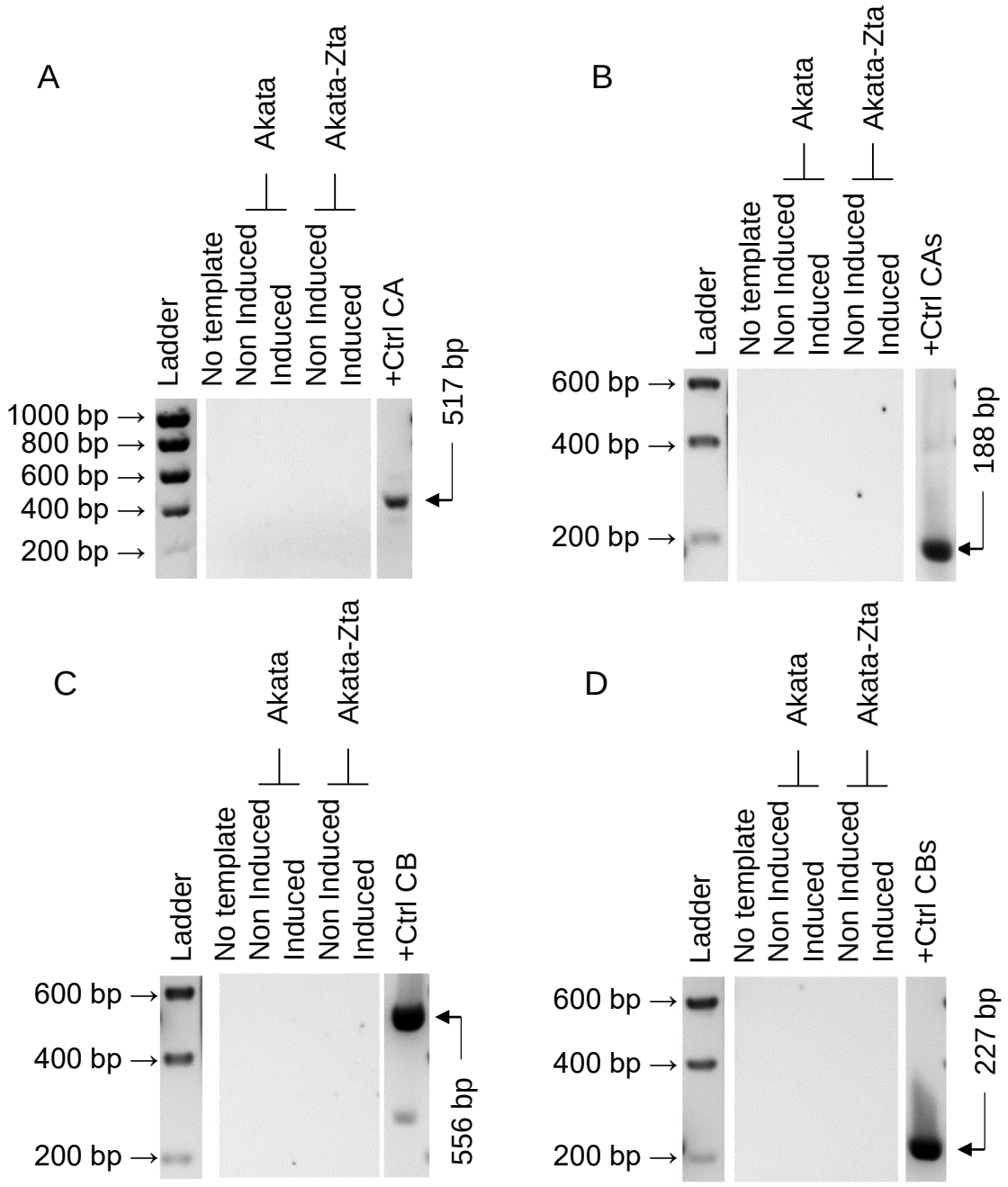
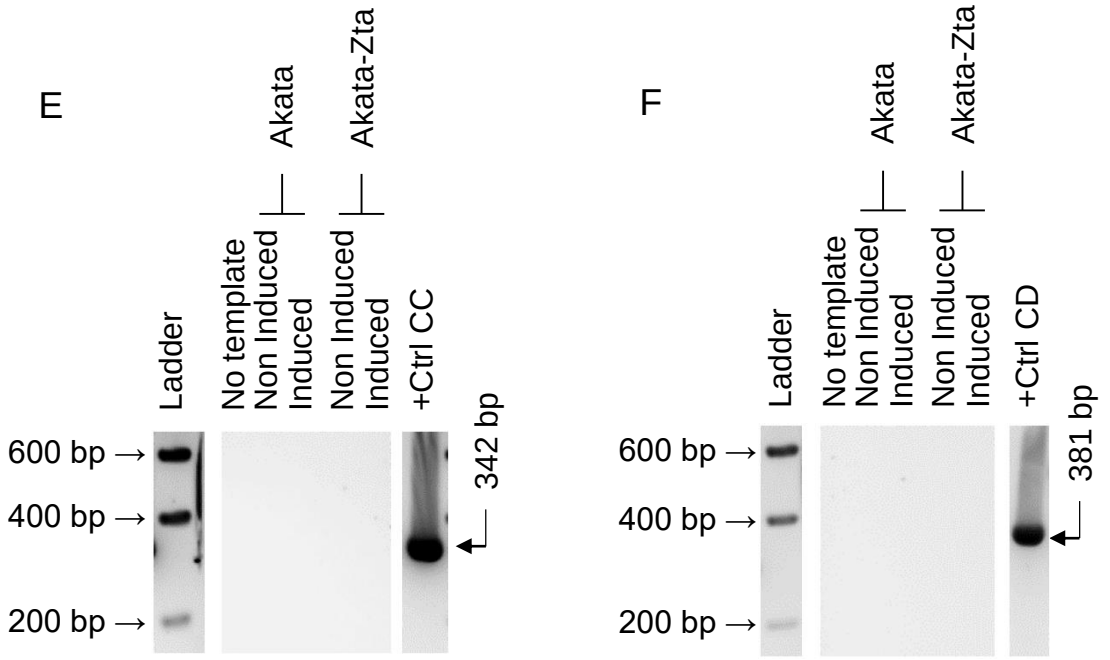


Figure 5.23. Chromosome conformation capture analysis of cluster VWF-CD9. Image captures of 1.2% Agarose gels after electrophoresis of PCR amplification products. The PCR reactions were performed at 64 °C. Each panel (A-F) is composed of sections (lanes) coming from the same gel. (**A**) There is no band of similar size to the synthetic positive control for loop CA in lanes with the Akata and Akata-Zta templates. (**B**) No product amplification of comparable size to the synthetic positive control for loop CAs in lanes with the Akata and Akata-Zta templates. (**C**) There is no band of similar size to the synthetic positive control for loop CB in lanes with the Akata and Akata-Zta templates. (**D**) No product amplification of comparable size to the synthetic positive control for loop CBs in lanes with the Akata and Akata-Zta templates. (figure legend continues…)



(cont…) (**E**) There is no band of similar size to the synthetic positive control for loop CC in lanes with the Akata and Akata-Zta templates. (**F**) No product amplification of comparable size to the synthetic positive control for loop CD in lanes with the Akata and Akata-Zta templates.

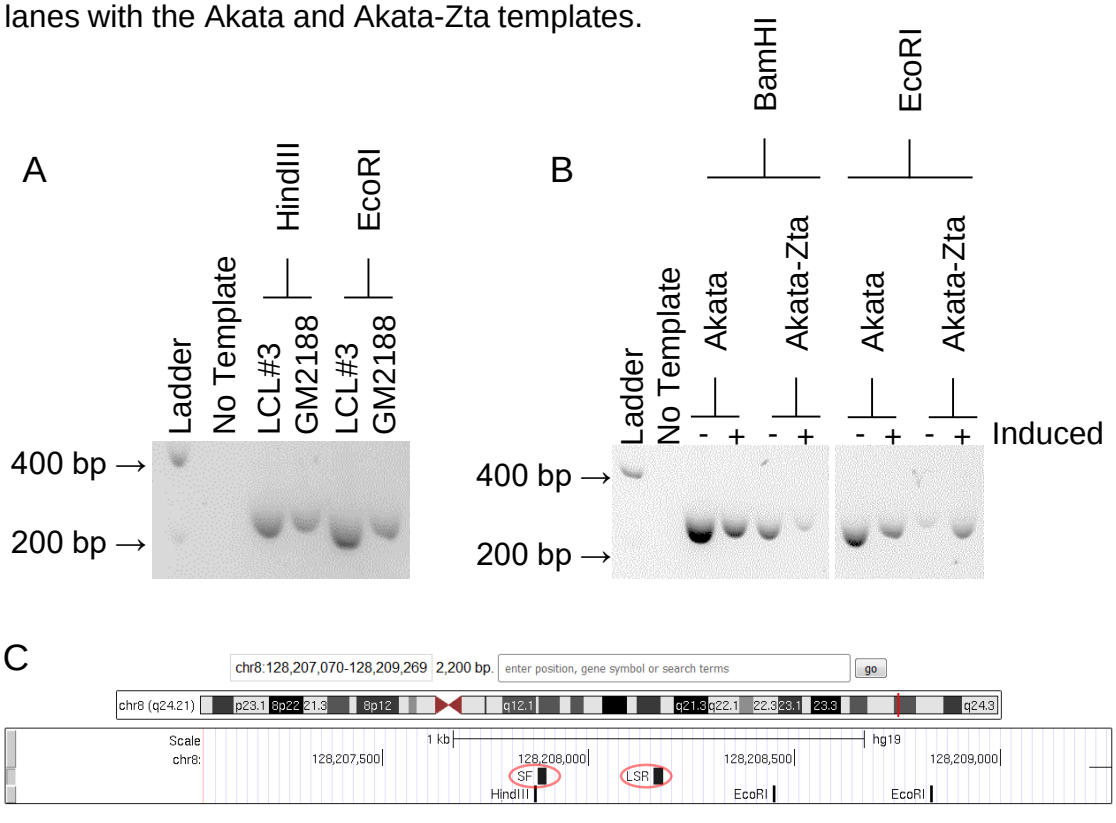


Figure 5.24. Amplification of DNA in 3C templates. Primers were designed for the amplification of a contiguous region of 304 bp, this to show that indeed there was DNA in the template used in the previous experiments. PCR reaction was performed at 64°C. (**A**) LCL#3 and GM2188 templates (**B**) Akata and Akata-Zta templates. (**C**) Primers designed to amplify a 304bp product, independently from digestion, in chr8:128,206,689-128,208,688.

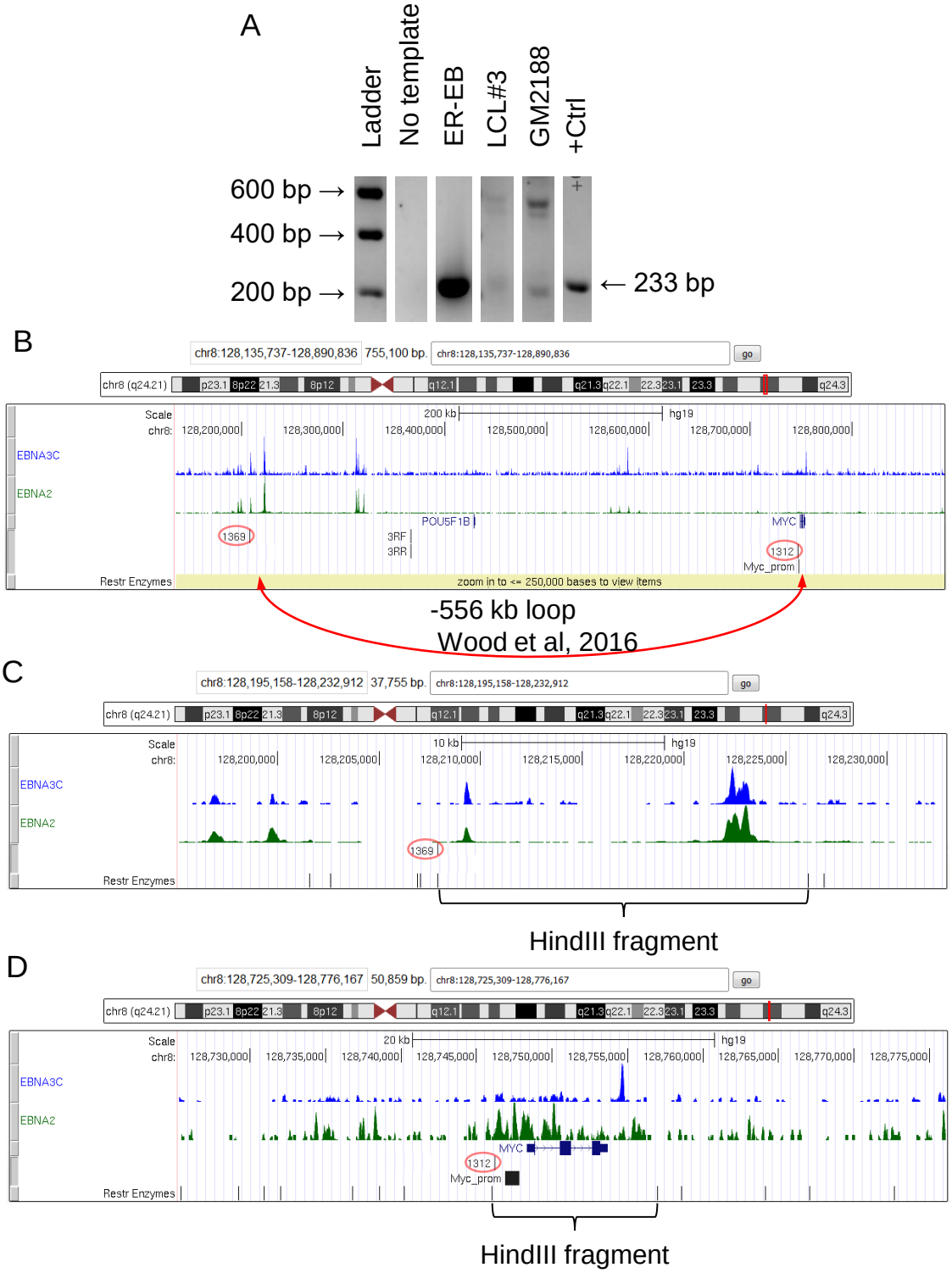


Figure 5.25. Template amplification of control loop MYC. To show the validity of the 3C process, DNA digested with HindIII was used to amplify the control loop MYC, known to loop in ER-EB cells. (**A**) Template coming from ER-EB cells show the amplification desired. Template from LCL#3 and GM2188 show a faint band of the expected size. On the far right, a synthetic positive control show the 233 bp band. PCR amplification was performed at 64°C (**B**) Map of the MYC loop to the EBNA2 ChIP peaks detected by Wood et al, performed using a HindIII digestion. Fragments containing the peaks (**C**) and MYC promotor (**D**) and the respective primers circled in red.

5.3 Discussion

The design of the 3C experiments was completed for each target and executed in cells that displayed lytic activity or, for comparison, cells which stayed in latency. The cross-linked chromatin from the cells was digested with specific restriction enzymes and then reverse cross-linked. Synthetic controls were designed and the designed primers were used, this resulted in the amplification of a product of the expected size. Initially the amplification of the controls was performed at 60°C, the Tm used when designing the primers. However, the temperature was then increased to 64°C in all subsequent experiments, since this increase of temperature reduced the background caused by non-specific amplification present in the temperature gradients, but maintained the signal for positive controls.

It was disappointing that we found no evidence of looping between the proposed Zta interactions and the promoter regions of SLC6A7, LHX1, FZD10 and cluster VWF and CD9 in our results. It is important to keep in mind that given the complex process of undertaking 3C experiments, it is inherently difficult to interpret correctly and reliably negative results obtained by any 3C assay; however, since we tested the capability of the templates to detect loops with the MYC loop, the observed results from the experiments seem to hold validity. The reason why the MYC loop could not be analysed in Akata and Akata-Zta cells was that EcoRI and BamHI were used in the digestion, and the primers used by Wood et al to detect the MYC loop were designed on a digestion by HindIII. Even if new primers were to be designed based on an EcoRI and BamHI digestion for a MYC loop, we couldn’t be certain that we would capture the MYC loop.

Not finding evidence of looping between the proposed Zta interactions and the studied promoter regions is not enough evidence to claim that the loops are not forming. Since the execution of 3C experiments is so detailed and specific, it is necessary to consider alternative explanations for our results. It is possible that the fixation time might have affected the efficiency of the crosslinking by capturing protein complexes that occupy and block restriction sites, therefore affecting the subsequent restriction digestion, and in turn producing partial digestion and large DNA fragments (Naumova et al. 2012). Parallel to this, some of the templates in section 5.2.1.2 show a smear, signifying degradation of the 3C template, this was

tolerated since according to the guidelines of Naumova et al, the quality is passable; however, the quality reflected for LCL#3 and GM2188 was low.

Being aware of the limitations of the executed experiments, it would be valuable to point out where future design could improve the assay if more time and resources were available. Besides altering the fixation time, the primer design could be more specific and with a higher Tm. By obtaining primers longer than 20 bp, any non-specific amplification like we found with “sequenceX” when testing for loop E would be minimized.

As the 3C assay technologies evolve, new improvements come over the quantity of interactions found and information gathered on these interactions. One of these evolved 3C assays is called 4C, where a second digestion is followed by a second ligation, allowing small DNA circles to be formed (reviewed by (Denker and de Laat 2016). Then, after using inverse PCR with primers designed to bind in the region of interest (a prominent Zta peak in our case) this region gets sequenced and mapped. As suggestions for future research, this approach could be taken; since interactions have already been proposed, the use of 4C would allow to find in a single run several looping interactions between the selected Zta binding sites and many other possible regions, this could be subsequently corroborated by 3C.

Chapter 6. Discussion

In this project, a possible enhancer influence by central EBV factor Zta, over gene regulation was explored in different cellular contexts. Two different methods of analysis were used to probe different aspects of the interactions between Zta and the human genome.

Research previously performed in the Sinclair Lab studied the change of mRNA levels in the transcriptome of lytic B cells, as well as a genome wide analysis of Zta interactions found in the human genome; this identified genes that are regulated and have Zta binding sites associated with them, which allowed us to prioritize which Zta interactions to investigate (Ramasubramanyan et al. 2015).

Sequences corresponding to the upstream region of 30 highly regulated genes were analysed and, through computational analysis, conserved sequences were found. Interestingly, the identified sequences did not resemble the well characterized 7 base motif of AP1 (TGAGTCA) to which Zta is known to bind, or its variants in ZREs.

Two elements (1OEM, 4OEM) consisting of the most representative sequence of the conserved regions were designed, under the supposition that those sequences could present enhancer activity since they were found in highly regulated genes with Zta peaks close to its promoters.

Parallel to this, two other elements (AHNAK, FOSB) were designed based on the presence of Zta peaks near the promoter region of highly regulated genes AHNAK and FOSB observed in the ChIP-seq performed previously in the Sinclair Lab.

The designed elements were synthesized and then cloned into 3 different vectors each containing a different promotor reactive to Zta of known behaviour in different cell environments. After this, the vectors containing the elements were transformed into bacterial cells, and cultured overnight for a subsequent DNA extraction. The obtained DNA samples were verified through a 3rd party sequencing service.

After this, the DNA was transfected into EBV negative BL cells (DG75) and EBV negative epithelial cells (293T) together with an expression vector for His-tagged

Zta or an empty vector as a control. The transfected cells were cultured and harvested. Then, the presence of Zta was confirmed in the samples through Western blot assays, and the Luciferase assays were executed.

The results obtained showed interesting but modest and diverse enhancer effects over the expression of luciferase in a cell dependent context. Still, although the controls reflected the behaviour expected of the vectors relative to the cell line, agreeing with previous literature; the enhancer effect of the elements was not substantial overall.

The constrained enhancer effect found in the experiments was also limited by the problematic reproducibility encountered in subsequent experiments. However, future analysis could benefit from this initial undertaking, perhaps with some changes and tweaks in the virtual sequence analysis.

In a different approach to study Zta and its regulation over gene expression, a chromosome conformation capture assay was designed. This was based on the observation that a large proportion of Zta interactions were found several kb away from the promotor region of any genes in vicinity.

As a first step, a selection criteria had to be designed to discriminate amongst many Zta peaks and the promoter region of many genes. One such criteria involved analysing the presence of any Zta peaks in distal vicinity of the promoter region of 30 genes previously identified as highly regulated.

Regions that allowed an appropriate digestion in the context of 3C experimental design allowed to postulate 3 candidates for looping re-arrangement to the promoter region of genes FZD10, SLC6A7 and LHX1.

In a parallel analysis, a computational selection criteria was devised that cross related genes known to be highly regulated that could be clustered together, and the presence of Zta interactions in vicinity of the cluster. This analysis helped identify 8 gene clusters from which one cluster of 2 genes (VWF and CD9) was postulated as candidate of looping regulation.

This first step of finding which interactions to research was followed up by designing a digestion regimen that would cut the DNA into fragments that would contain the potential looping regions. After this, primers for the amplification of

looping events were designed and synthesized. Positive controls that would result from such amplification were also designed and synthesized.

The primers were tested as well as the positive controls, showing that an amplification of products of expected size based on the design of the assay was possible using the appropriate primer pairs and the temperature selected when designing the primers (60°C).

Chromatin coming from EBV-positive inducible BL cells, as well as chromatin from latent and lytic LCL cells, was used and the protein-DNA interactions were fixed with the use of formaldehyde. After this, the chromatin was digested and then re-ligated, followed by a reversal of the crosslink; lastly DNA was purified. A temperature gradient PCR was used to explore the presence of products of similar size to the control synthetic DNA; a band of similar size to the control in loop E to the promoter region of LHX1 was appreciated at 55°C in chromatin from latent and lytic LCLs. This band was excised from the electrophoresis gel and the DNA was purified for a subsequent 3rd party sequencing verification.

The retrieved sequence from the band did not match the sequence employed for the generation of the looping control DNA, however they shared a similar length. A virtual alignment allowed to see that this DNA band of similar size corresponded to a different place in the genome, and flanked by regions of similar sequence to the primers employed, showing that the band was a product of non-specific amplification. Based on this, and following the literature guidelines for troubleshooting 3C assays, a higher temperature was used in subsequent PCR reactions (64°C).

No evidence of loops forming in the studied regions was found in the chromatin employed. The capability of capturing any loops, in chromatin coming from the inducible BL cells employed, was tested by using primers employed by Wood in testing the presence of MYC loop; a band of similar size to the synthetic control was found in our template, confirming the potential of the chromatin to capture loops (Wood et al. 2016).

Despite the lack of evidence for looping events in the work presented here, no final claim that the looping events are not occurring can be postulated. One of the factors that limited the capacity to draw a final strong conclusion arose in the

quality of the digested chromatin used in the experiments. Still, in the undertaking of this research, a novel method of grouping gene clusters and relating them to Zta peaks was developed, and future analysis could benefit from this method.

The research presented here shows an initial attempt to analyse any enhancing effect found in sequences where Zta is binding to in the human genome, as well as an attempt to study any regulation of expression by Zta, performed through chromatin rearranging by looping. Although no breakthrough kind of finding was observed, during the presented research interesting computational methods of analysis were generated.

As a suggestion to consider for future research; a very interesting genome wide analysis of long-range interactions between transcription start sites and distal enhancer elements in several human cells lines, found a bias favouring the looping interactions with enhancers that were located ~120kb upstream of the TSS. Additionally, they observed that only 7% of looping events were occurring to the nearest gene, indicating that genomic proximity is not a simple predictor for long-range interactions (Sanyal et al. 2012). If taken into consideration, this could easily be implemented in the computational analysis presented in chapter 4 of this thesis by excluding any peak found closer than 100 kb away or farther than 140 kb from the found clusters.

The integration of next generation sequencing to the 3C technologies, has greatly broadened the information that can be gathered from chromatin conformation capture experiments. The use of 4C, 5C or Hi-C (amongst many others) present very viable options to study chromatin rearrangement and gene regulation by Zta over “classical” 3C.

The 4C assay follows a similar method as used in 3C assays, except that after the reversal of the crosslink a secondary digestion is performed, followed by a new ligation. This produces small DNA fragments that form circles, and through the use of inverse PCR, primers designed to bind to the postulated enhancer region, enrich the presence of products. These amplified products are analysed through sequencing and generate a map of interactions of a few kilobases of resolution. This represents an immediate advantage over 3C; using inverse PCR

an experiment can be designed without a priori hypotheses of both candidate regions forming long range interactions.

The use of 4C for the exploration of the interactions of Zta could bring some insights on particular looping regions. If an interaction of Zta was selected based on the value from the ChIP-seq peak, to design the primers for the inverse PCR; a profile of all the interactions to that single particular Zta interaction would be obtained, however, data from previous studies gave a large amount of interactions to study, therefore 4C might not be the best choice.

The 5C assay differs from the “classic” 3C in that after the reversal of the crosslink, primers with a particular 5’ “adaptor” tail are designed to bind to the fragments of interest, if the postulated interaction exists and the fragments are ligated, then the primers would amplify a product which can be subsequently amplified with universal primers complementary to the “adaptor” tail of the first round of amplification. After this, next generation sequencing would create an interaction frequency matrix.

The use of 5C for the exploration of Zta interactions could take advantage of data previously generated. The list of highly regulated genes presented in previous chapters could be used to generate primers for the promoter regions of the 10 most highly regulated genes; and if the value of the Zta peaks from the ChIP-seq was to be used to narrow the top 10 Zta interactions, then primers could be designed for those targets. This, however, would limit the study to only 10 genes out of hundreds that are known to be highly regulated in lytic cycle, and 10 Zta peaks out of hundreds in vicinity.

In the Hi-C assay, after a regular chromatin cross-link, a restriction digest that leaves an overhang in the 5’ end of the fragments takes place. Then, using biotin-labelled nucleotides the 5’ overhang gets complemented; this is followed by a blunt end ligation, which unites the biotin labelled ends. Then the reversal of the crosslink allows for a DNA purification, followed by a sonication and a streptavidin pulldown of the ligation junctions. Finally, using paired-end sequencing, a matrix of pair-wise interaction frequencies would be obtained. This wouldn’t require the use of previous data to find interactions in cells undergoing lytic cycle, however,

when mapped and compared to the mapped generated with ChIP-seq, a very interesting overlap could be seen in regions shown to be binding, and Zta peaks.

Finally, the emerging Capture-C technologies allow for a further target enrichment in the preparation of 3C libraries; this is achieved by using a digestion enzyme in the first digestion, followed by a second fragmentation by sonication. Then designed biotinylated RNA baits bind to the fragments of interest and a streptavidin bead pull down increase the presence of only the interactions of interest. This technique is still evolving (NG Capture-C), but if was to be employed in the study of Zta interactions with the genome, it would help obtain more detailed results than the initial first wave of 3C technologies (Dekker et al 2016).

As a final summation, the advent of new technologies for enrichment of products and interest; and the decline in the price of sequencing techniques will create an enormously good opportunity to vastly expand the knowledge of Zta regulation of gene expression. This, with much better ease than when using 3C as an initial exploration.

Bibliography

Abate, F., et al. (2015), 'Distinct Viral and Mutational Spectrum of Endemic Burkitt Lymphoma', *PLoS Pathog,* 11 (10), e1005158.

Adamson, A. L. and Kenney, S. C. (1998), 'Rescue of the Epstein-Barr virus BZLF1 mutant, Z(S186A), early gene activation defect by the BRLF1 gene product', *Virology,* 251 (1), 187-97.

Andersson, R., et al. (2014), 'An atlas of active enhancers across human cell types and tissues', *Nature,* 507 (7493), 455-+.

Arnold, C. D., et al. (2013), 'Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq', *Science,* 339 (6123), 1074-77.

Arnone, M. I. and Davidson, E. H. (1997), 'The hardwiring of development: Organization and function of genomic regulatory systems', *Development,* 124 (10), 1851-64.

Arvey, A., et al. (2012), 'An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions', *Cell Host Microbe,* 12 (2), 233-45.

Arvin, A. and Abendroth, A. (2007), 'VZV: immunobiology and host response', in A. Arvin, et al. (eds.), *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis* (Cambridge).

Avolio-Hunter, T. M. and Frappier, L. (1998), 'Mechanistic studies on the DNA linking activity of Epstein-Barr nuclear antigen 1', *Nucleic Acids Res,* 26 (19), 4462-70.

Baer, R., et al. (1984), 'DNA sequence and expression of the B95-8 Epstein-Barr virus genome', *Nature,* 310 (5974), 207-11.

Bailey, S. G., et al. (2009), 'Functional interaction between Epstein-Barr virus replication protein Zta and host DNA damage response protein 53BP1', *J Virol,* 83 (21), 11116-22.

Balan, N., Osborn, K., and Sinclair, A. J. (2016), 'Repression of CIITA by the Epstein-Barr virus transcription factor Zta is independent of its dimerization and DNA binding', *J Gen Virol,* 97 (3), 725-32.

Balfour, H. H., Jr., Dunmire, S. K., and Hogquist, K. A. (2015), 'Infectious mononucleosis', *Clin Transl Immunology,* 4 (2), e33.

Banerji, J., Rusconi, S., and Schaffner, W. (1981), 'EXPRESSION OF A BETA-GLOBIN GENE IS ENHANCED BY REMOTE SV40 DNA-SEQUENCES', *Cell,* 27 (2), 299-308.

Barton, E., Mandal, P., and Speck, S. H. (2011), 'Pathogenesis and host control of gammaherpesviruses: lessons from the mouse', *Annu Rev Immunol,* 29, 351-97.

Baumann, M., et al. (1998), 'Activation of the Epstein-Barr virus transcription factor BZLF1 by 12-O-tetradecanoylphorbol-13-acetate-induced phosphorylation', *J Virol,* 72 (10), 8105-14.

Ben-Bassat, H., et al. (1976), 'Concanavalin A receptors on the surface membrane of lymphocytes from patients with African Burkitt's lymphoma and lymphoma cell lines', *Int J Cancer,* 17 (4), 448-54.

Bergbauer, M., et al. (2010), 'CpG-Methylation Regulates a Class of Epstein-Barr Virus Promoters', *Plos Pathogens,* 6 (9).

Bhende, P. M., et al. (2004), 'The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome', *Nat Genet,* 36 (10), 1099-104.

Bhende, P. M., et al. (2005), 'BZLF1 activation of the methylated form of the BRLF1 immediate-early promoter is regulated by BZLF1 residue 186', *J Virol,* 79 (12), 7338-48.

Bioinformatics, SIB Swiss Institute of (2017), 'Herpesviridae virion', in Herpesviridae_virion.jpg (ed.), (www.expasy.org/viralzone).

Boerma, E. G., et al. (2009), 'Translocations involving 8q24 in Burkitt lymphoma and other malignant lymphomas: a historical review of cytogenetics in the light of todays knowledge', *Leukemia,* 23 (2), 225-34.

Bonn, S., et al. (2012), 'Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development', *Nature Genetics,* 44 (2), 148-56.

Burkitt, D. (1958), 'A sarcoma involving the jaws in African children', *Br J Surg,* 46 (197), 218-23.

Caldwell, R. G., et al. (1998), 'Epstein-Barr virus LMP2A drives B cell development and survival in the absence of normal B cell receptor signals', *Immunity,* 9 (3), 405-11.

Chang, Y. N., et al. (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 (7), 3358-69.

Chen, H. S., Lu, F., and Lieberman, P. M. (2013), 'Epigenetic regulation of EBV and KSHV latency', *Curr Opin Virol,* 3 (3), 251-9.

Chen, M. R. (2011), 'Epstein-barr virus, the immune system, and associated diseases', *Front Microbiol,* 2, 5.

Chevallier-Greco, A., et al. (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 (12), 3243-9.

Chi, T. and Carey, M. (1993), 'The ZEBRA activation domain: modular organization and mechanism of action', *Mol Cell Biol,* 13 (11), 7045-55.

Chua, M. L. K., et al. (2016), 'Nasopharyngeal carcinoma', *Lancet,* 387 (10022), 1012-24.

Countryman, J. and Miller, G. (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 (12), 4085-9.

Countryman, J. K., Gradoville, L., and Miller, G. (2008), 'Histone hyperacetylation occurs on promoters of lytic cycle regulatory genes in Epstein-Barr virus-infected cell lines which are refractory to disruption of latency by histone deacetylase inhibitors', *J Virol,* 82 (10), 4706-19.

Davison, A. J. (2007), 'Overview of classification', in A. Arvin, et al. (eds.), *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis* (Cambridge).

Davison, A. J. (2010), 'Herpesvirus systematics', *Vet Microbiol,* 143 (1), 52-69.

Dawson, C. W., Port, R. J., and Young, L. S. (2012), 'The role of the EBV-encoded latent membrane proteins LMP1 and LMP2 in the pathogenesis of nasopharyngeal carcinoma (NPC)', *Semin Cancer Biol,* 22 (2), 144-53.

De Santa, F., et al. (2010), 'A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers', *Plos Biology,* 8 (5).

de-The, G., et al. (1975), 'Sero-epidemiology of the Epstein-Barr virus: preliminary analysis of an international study - a review', *IARC Sci Publ,* (11 Pt 2), 3-16.

Dekker, J., et al. (2002), 'Capturing chromosome conformation', *Science,* 295 (5558), 1306-11.

Denker, A. and de Laat, W. (2016), 'The second decade of 3C technologies: detailed insights into nuclear organization', *Genes Dev,* 30 (12), 1357-82.

Dowen, J. M., et al. (2014), 'Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes', *Cell,* 159 (2), 374-87.

DuBridge, R. B., et al. (1987), 'Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system', *Mol Cell Biol,* 7 (1), 379-87.

Edwards, R. H., Marquitz, A. R., and Raab-Traub, N. (2008), 'Epstein-Barr virus BART microRNAs are produced from a large intron prior to splicing', *J Virol,* 82 (18), 9094-106.

Epstein, M. A., Achong, B. G., and Barr, Y. M. (1964), 'Virus Particles in Cultured Lymphoblasts from Burkitt's Lymphoma', *Lancet,* 1 (7335), 702-3.

Farrell, P. J., et al. (1989), 'Epstein-Barr virus BZLF1 trans-activator specifically binds to a consensus AP-1 site and is related to c-fos', *EMBO J,* 8 (1), 127-32.

Feederle, R., et al. (2000), 'The Epstein-Barr virus lytic program is controlled by the co-operative functions of two transactivators', *EMBO J,* 19 (12), 3080-9.

--- (2011), 'A viral microRNA cluster strongly potentiates the transforming properties of a human herpesvirus', *PLoS Pathog,* 7 (2), e1001294.

Fingeroth, J. D., et al. (1984), 'Epstein-Barr virus receptor of human B lymphocytes is the C3d receptor CR2', *Proc Natl Acad Sci U S A,* 81 (14), 4510-4.

Fixman, E. D., Hayward, G. S., and Hayward, S. D. (1992), 'trans-acting requirements for replication of Epstein-Barr virus ori-Lyt', *J Virol,* 66 (8), 5030-9.

Flemington, et al. (1992), 'Characterization of the Epstein-Barr virus BZLF1 protein transactivation domain', *J Virol,* 66 (2), 922-9.

Flemington, E. and Speck, S. H. (1990a), 'Autoregulation of Epstein-Barr virus putative lytic switch gene BZLF1', *J Virol,* 64 (3), 1227-32.

Flemington, E. and Speck, S. H. (1990b), 'Epstein-Barr virus BZLF1 trans activator induces the promoter of a cellular cognate gene, c-fos', *J Virol,* 64 (9), 4549-52.

Flower, K., et al. (2011), 'Epigenetic control of viral life-cycle by a DNA-methylation dependent transcription factor', *PLoS One,* 6 (10), e25922.

Fraser, P. and Bickmore, W. (2007), 'Nuclear organization of the genome and the potential for gene regulation', *Nature,* 447 (7143), 413-7.

Gao, X., et al. (2001), '12-O-tetradecanoylphorbol-13-acetate induces Epstein-Barr virus reactivation via NF-kappaB and AP-1 as regulated by protein kinase C and mitogen-activated protein kinase', *Virology,* 286 (1), 91-9.

Giresi, P. G., et al. (2007), 'FAIRE ((F)under-barormaldehyde-(A)under-barssisted (I)under-barsolation of (R)under-baregulatory (E)under-barlements) isolates active regulatory elements from human chromatin', *Genome Research,* 17 (6), 877-85.

Glickman, J. N., Howe, J. G., and Steitz, J. A. (1988), 'Structural analyses of EBER1 and EBER2 ribonucleoprotein particles present in Epstein-Barr virus-infected cells', *J Virol,* 62 (3), 902-11.

Goswami, R., et al. (2012), 'Protein kinase inhibitors that inhibit induction of lytic program and replication of Epstein-Barr virus', *Antiviral Res,* 96 (3), 296-304.

Gunnell, A., et al. (2016), 'RUNX super-enhancer control through the Notch pathway by Epstein-Barr virus transcription factors regulates B cell growth', *Nucleic Acids Res,* 44 (10), 4636-50.

Hah, N., et al. (2015), 'Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs', *Proceedings of the National Academy of Sciences of the United States of America,* 112 (3), E297-E302.

Hammerschmidt, W. and Sugden, B. (1988), 'Identification and characterization of oriLyt, a lytic origin of DNA replication of Epstein-Barr virus', *Cell,* 55 (3), 427-33.

Harada, S. and Kieff, E. (1997), 'Epstein-Barr virus nuclear protein LP stimulates EBNA-2 acidic domain-mediated transcriptional activation', *J Virol,* 71 (9), 6611-8.

Hardwick, J. M., Lieberman, P. M., and Hayward, S. D. (1988), 'A new Epstein-Barr virus transactivator, R, induces expression of a cytoplasmic early antigen', *J Virol,* 62 (7), 2274-84.

Heather, J., et al. (2009), 'The Epstein-Barr virus lytic cycle activator Zta interacts with methylated ZRE in the promoter of host target gene egr1', *J Gen Virol,* 90 (Pt 6), 1450-4.

Heintzman, N. D., et al. (2007), 'Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome', *Nature Genetics,* 39 (3), 311-18.

Heinz, S., et al. (2015), 'The selection and function of cell type-specific enhancers', *Nat Rev Mol Cell Biol,* 16 (3), 144-54.

Henle, G. and Henle, W. (1966), 'Immunofluorescence in cells derived from Burkitt's lymphoma', *J Bacteriol,* 91 (3), 1248-56.

Henle, G., et al. (1969), 'Antibodies to Epstein-Barr virus in Burkitt's lymphoma and control groups', *J Natl Cancer Inst,* 43 (5), 1147-57.

Henle, W., et al. (1967), 'Herpes-type virus and chromosome marker in normal leukocytes after growth with irradiated Burkitt cells', *Science,* 157 (3792), 1064-5.

Hicks, M. R., Al-Mehairi, S. S., and Sinclair, A. J. (2003), 'The zipper region of Epstein-Barr virus bZIP transcription factor Zta is necessary but not sufficient to direct DNA binding', *J Virol,* 77 (14), 8173-7.

Hicks, M. R., et al. (2001), 'Biophysical analysis of natural variants of the multimerization region of Epstein-Barr virus lytic-switch protein BZLF1', *J Virol,* 75 (11), 5381-4.

Hnisz, D., et al. (2013), 'Super-enhancers in the control of cell identity and disease', *Cell,* 155 (4), 934-47.

Hsieh, J. J., et al. (1996), 'Truncated mammalian Notch1 activates CBF1/RBPJk-repressed genes by a mechanism resembling that of Epstein-Barr virus EBNA2', *Mol Cell Biol,* 16 (3), 952-9.

Hsu, M., et al. (2008), 'Epstein-Barr virus lytic transactivator Zta enhances chemotactic activity through induction of interleukin-8 in nasopharyngeal carcinoma cells', *J Virol,* 82 (7), 3679-88.

Jaffe, E. S. and Pittaluga, S. (2011), 'Aggressive B-cell lymphomas: a review of new and old entities in the WHO classification', *Hematology Am Soc Hematol Educ Program,* 2011, 506-14.

Jenkins, P. J., Binne, U. K., and Farrell, P. J. (2000), 'Histone acetylation and reactivation of Epstein-Barr virus from latency', *J Virol,* 74 (2), 710-20.

Jochum, S., et al. (2012), 'RNAs in Epstein-Barr virions control early steps of infection', *Proc Natl Acad Sci U S A,* 109 (21), E1396-404.

Kagey, M. H., et al. (2010), 'Mediator and cohesin connect gene expression and chromatin architecture', *Nature,* 467 (7314), 430-35.

Karlsson, Q. H., et al. (2008), 'The reversal of epigenetic silencing of the EBV genome is regulated by viral bZIP protein', *Biochem Soc Trans,* 36 (Pt 4), 637-9.

Kaye, K. M., Izumi, K. M., and Kieff, E. (1993), 'Epstein-Barr virus latent membrane protein 1 is essential for B-lymphocyte growth transformation', *Proc Natl Acad Sci U S A,* 90 (19), 9150-4.

Kenney, S. C. (2007), 'Reactivation and lytic replication of EBV', *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, 403-33.

Kieff, E., et al. (1982), 'The biology and chemistry of Epstein-Barr virus', *J Infect Dis,* 146 (4), 506-17.

Kim, H. J., et al. (2017), 'Epstein-Barr Virus-Associated Lymphoproliferative Disorders: Review and Update on 2016 WHO Classification', *J Pathol Transl Med,* 51 (4), 352-58.

Kim, T. K., et al. (2010), 'Widespread transcription at neuronal activity-regulated enhancers', *Nature,* 465 (7295), 182-U65.

Kintner, C. R. and Sugden, B. (1979), 'The structure of the termini of the DNA of Epstein-Barr virus', *Cell,* 17 (3), 661-71.

Klug, M. and Rehli, M. (2006), 'Functional Analysis of Promoter CpG Methylation Using a CpG-Free Luciferase Reporter Vector', *Epigenetics,* 1 (3), 127-30.

Koch, F., et al. (2011), 'Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters', *Nat Struct Mol Biol,* 18 (8), 956-63.

Kohn, G., et al. (1967), 'Involvement of C group chromosomes in five Burkitt lymphoma cell lines', *J Natl Cancer Inst,* 38 (2), 209-22.

Kouzarides, T. (2007), 'Chromatin modifications and their function', *Cell,* 128 (4), 693-705.

Kropp, K. A., Angulo, A., and Ghazal, P. (2014), 'Viral Enhancer Mimicry of Host Innate-Immune Promoters', *Plos Pathogens,* 10 (2).

Kutok, J. L. and Wang, F. (2006), 'Spectrum of Epstein-Barr virus-associated diseases', *Annu Rev Pathol,* 1, 375-404.

Laichalk, L. L. and Thorley-Lawson, D. A. (2005), 'Terminal differentiation into plasma cells initiates the replicative cycle of Epstein-Barr virus in vivo', *J Virol,* 79 (2), 1296-307.

Lelli, K. M., Slattery, M., and Mann, R. S. (2012), 'Disentangling the many layers of eukaryotic transcriptional regulation', *Annu Rev Genet,* 46, 43-68.

Lerner, M. R., et al. (1981), 'Two small RNAs encoded by Epstein-Barr virus and complexed with protein are precipitated by antibodies from patients with systemic lupus erythematosus', *Proc Natl Acad Sci U S A,* 78 (2), 805-9.

Levine, P. H., et al. (1971), 'Elevated antibody titers to Epstein-Barr virus in Hodgkin's disease', *Cancer,* 27 (2), 416-21.

Li, G. L., et al. (2012), 'Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation', *Cell,* 148 (1-2), 84-98.

Li, H., et al. (2016a), 'Epstein-Barr virus lytic reactivation regulation and its pathogenic role in carcinogenesis', *Int J Biol Sci,* 12 (11), 1309-18.

Li, W., Notani, D., and Rosenfeld, M. G. (2016b), 'Enhancers as non-coding RNA transcription units: recent insights and future perspectives', *Nat Rev Genet,* 17 (4), 207-23.

Lieberman, P. M. (2013), 'Keeping it quiet: chromatin control of gammaherpesvirus latency', *Nat Rev Microbiol,* 11 (12), 863-75.

Lieberman, P. M., et al. (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 (3), 1143-55.

Liu, C. L., et al. (2005), 'Single-nucleosome mapping of histone modifications in S-cerevisiae', *Plos Biology,* 3 (10), 1753-69.

Liu, F. Y. and Zhou, Z. H. (2007), 'Comparative virion structures of human herpesviruses', *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, 27-43.

Liu, Z. J., et al. (2014), 'Enhancer Activation Requires trans-Recruitment of a Mega Transcription Factor Complex', *Cell,* 159 (2), 358-73.

Longnecker, R. and Neipel, F. (2007), 'Introduction to the human gamma-herpesviruses', *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, 341-59.

Lu, et al. (2016), 'EBNA2 Drives Formation of New Chromosome Binding Sites and Target Genes for B-Cell Master Regulatory Transcription Factors RBP-jkappa and EBF1', *PLoS Pathog,* 12 (1), e1005339.

Lu, T. X., et al. (2015), 'Epstein-Barr virus positive diffuse large B-cell lymphoma predict poor outcome, regardless of the age', *Sci Rep,* 5, 12168.

Lubas, M., et al. (2015), 'The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Ribonucleolysis', *Cell Reports,* 10 (2), 178-92.

Luzuriaga, K. and Sullivan, J. L. (2010), 'Infectious mononucleosis', *N Engl J Med,* 362 (21), 1993-2000.

Marschall, M., et al. (1989), 'Identification of proteins encoded by Epstein-Barr virus trans-activator genes', *J Virol,* 63 (2), 938-42.

McClellan, M. J., et al. (2012), 'Downregulation of integrin receptor-signaling genes by Epstein-Barr virus EBNA 3C via promoter-proximal and -distal binding elements', *J Virol,* 86 (9), 5165-78.

McClellan, M. J., et al. (2013), 'Modulation of enhancer looping and differential gene targeting by Epstein-Barr virus transcription factors directs cellular reprogramming', *PLoS Pathog,* 9 (9), e1003636.

McDonald, C. M., Petosa, C., and Farrell, P. J. (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 (7), 3397-401.

Menotti, L., et al. (2000), 'The murine homolog of human Nectin1delta serves as a species nonspecific mediator for entry of human and animal alpha herpesviruses in a pathway independent of a detectable binding to gD', *Proc Natl Acad Sci U S A,* 97 (9), 4867-72.

Mercer, T. R. and Mattick, J. S. (2013), 'Understanding the regulatory and transcriptional complexity of the genome through structure', *Genome Res,* 23 (7), 1081-8.

Miller, N. and Hutt-Fletcher, L. M. (1992), 'Epstein-Barr virus enters B cells and epithelial cells by different routes', *J Virol,* 66 (6), 3409-14.

Mocarski, E. S. (2007), 'Comparative analysis of herpesvirus-common proteins', *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, 44-58.

Mogensen, T. H. and Paludan, S. R. (2001), 'Molecular pathways in virus-induced cytokine production', *Microbiol Mol Biol Rev,* 65 (1), 131-50.

Mohrs, M., et al. (2001), 'Deletion of a coordinate regulator of type 2 cytokine expression in mice', *Nat Immunol,* 2 (9), 842-7.

Montavon, T. and Duboule, D. (2013), 'Chromatin organization and global regulation of Hox gene clusters', *Philos Trans R Soc Lond B Biol Sci,* 368 (1620), 20120367.

Montavon, T., et al. (2011), 'A regulatory archipelago controls Hox genes transcription in digits', *Cell,* 147 (5), 1132-45.

Moreau, P., et al. (1981), 'THE SV40-72 BASE REPAIR REPEAT HAS A STRIKING EFFECT ON GENE-EXPRESSION BOTH IN SV40 AND OTHER CHIMERIC RECOMBINANTS', *Nucleic Acids Research,* 9 (22), 6047-68.

Morrison, T. E., et al. (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 (1), 544-9.

Murata, T. (2014), 'Regulation of Epstein-Barr virus reactivation from latency', *Microbiol Immunol,* 58 (6), 307-17.

Murata, T., et al. (2013), 'Contribution of myocyte enhancer factor 2 family transcription factors to BZLF1 expression in Epstein-Barr virus reactivation from latency', *J Virol,* 87 (18), 10148-62.

Naumova, N., et al. (2012), 'Analysis of long-range chromatin interactions using Chromosome Conformation Capture', *Methods,* 58 (3), 192-203.

Nemerow, G. R. and Cooper, N. R. (1984), 'Early events in the infection of human B lymphocytes by Epstein-Barr virus: the internalization process', *Virology,* 132 (1), 186-98.

Nemerow, G. R., et al. (1985), 'Identification and characterization of the Epstein-Barr virus receptor on human B lymphocytes and its relationship to the C3d complement receptor (CR2)', *J Virol,* 55 (2), 347-51.

Neph, S., et al. (2012), 'An expansive human regulatory lexicon encoded in transcription factor footprints', *Nature,* 489 (7414), 83-90.

Nilsson, K., et al. (1971), 'The establishment of lymphoblastoid lines from adult and fetal human lymphoid tissue and its dependence on EBV', *Int J Cancer,* 8 (3), 443-50.

Nonoyama, M. and Pagano, J. S. (1972), 'Separation of Epstein-Barr virus DNA from large chromosomal DNA in non-virus-producing cells', *Nat New Biol,* 238 (84), 169-71.

Noordermeer, D., et al. (2011), 'The dynamic architecture of Hox gene clusters', *Science,* 334 (6053), 222-5.

Osborne, C. S., et al. (2004), 'Active genes dynamically colocalize to shared sites of ongoing transcription', *Nat Genet,* 36 (10), 1065-71.

Packham, G., et al. (1990), 'Structure and function of the Epstein-Barr virus BZLF1 protein', *J Virol,* 64 (5), 2110-6.

Palstra, R. J., et al. (2003), 'The beta-globin nuclear compartment in development and erythroid differentiation', *Nat Genet,* 35 (2), 190-4.

Pear, W. S., et al. (1993), 'Production of high-titer helper-free retroviruses by transient transfection', *Proc Natl Acad Sci U S A,* 90 (18), 8392-6.

Pennacchio, L. A., et al. (2013), 'Enhancers: five essential questions', *Nat Rev Genet,* 14 (4), 288-95.

Petosa, C., et al. (2006), 'Structural basis of lytic cycle activation by the Epstein-Barr virus ZEBRA protein', *Mol Cell,* 21 (4), 565-72.

Pfeffer, S., et al. (2004), 'Identification of virus-encoded microRNAs', *Science,* 304 (5671), 734-6.

Price, A. M. and Luftig, M. A. (2014), 'Dynamic Epstein-Barr virus gene expression on the path to B-cell transformation', *Adv Virus Res,* 88, 279-313.

Price, A. M. and Luftig, M. A. (2015), 'To be or not IIb: a multi-step process for Epstein-Barr virus latency establishment and consequences for B cell tumorigenesis', *PLoS Pathog,* 11 (3), e1004656.

Rabson, M., et al. (1982), 'Non-immortalizing P3J-HR-1 Epstein-Barr virus: a deletion mutant of its transforming parent, Jijoye', *J Virol,* 44 (3), 834-44.

Rada-Iglesias, A., et al. (2011), 'A unique chromatin signature uncovers early developmental enhancers in humans', *Nature,* 470 (7333), 279-+.

Ramasubramanyan, S., et al. (2012), 'Genome-wide analyses of Zta binding to the Epstein-Barr virus genome reveals interactions in both early and late lytic cycles and an epigenetic switch leading to an altered binding profile', *J Virol,* 86 (23), 12494-502.

Ramasubramanyan, S., et al. (2015), 'Epstein-Barr virus transcription factor Zta acts through distal regulatory elements to directly control cellular gene expression', *Nucleic Acids Res,* 43 (7), 3563-77.

Reisman, D. and Sugden, B. (1986), 'trans activation of an Epstein-Barr viral transcriptional enhancer by the Epstein-Barr viral nuclear antigen 1', *Mol Cell Biol,* 6 (11), 3838-46.

Rickinson, A. B. (2014), 'Co-infections, inflammation and oncogenesis: future directions for EBV research', *Semin Cancer Biol,* 26, 99-115.

Rowe, M., et al. (2009), 'Burkitt's lymphoma: the Rosetta Stone deciphering Epstein-Barr virus biology', *Semin Cancer Biol,* 19 (6), 377-88.

Rowe, M., et al (1992), 'Three pathways of Epstein-Barr virus gene activation from EBNA1-positive latency in B lymphocytes', *J Virol,* 66 (1), 122-31.

Sanyal, A., et al. (2012), 'The long-range interaction landscape of gene promoters', *Nature,* 489 (7414), 109-13.

Seibl, R., Motz, M., and Wolf, H. (1986), 'Strain-specific transcription and translation of the BamHI Z area of Epstein-Barr Virus', *J Virol,* 60 (3), 902-9.

Seto, E., et al. (2010), 'Micro RNAs of Epstein-Barr virus promote cell cycle progression and prevent apoptosis of primary human B cells', *PLoS Pathog,* 6 (8), e1001063.

Shaffer, A. L., et al. (2002), 'Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program', *Immunity,* 17 (1), 51-62.

Shaw, J. E., Levinger, L. F., and Carter, C. W., Jr. (1979), 'Nucleosomal structure of Epstein-Barr virus DNA in transformed cell lines', *J Virol,* 29 (2), 657-65.

Shinozaki-Ushiku, A., Kunita, A., and Fukayama, M. (2015), 'Update on Epstein-Barr virus and gastric cancer (review)', *Int J Oncol,* 46 (4), 1421-34.

Shlyueva, D., Stampfel, G., and Stark, A. (2014), 'Transcriptional enhancers: from properties to genome-wide predictions', *Nat Rev Genet,* 15 (4), 272-86.

Shopland, L. S., et al. (2003), 'Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods', *J Cell Biol,* 162 (6), 981-90.

Sinclair, A. J. (2003), 'bZIP proteins of human gammaherpesviruses', *J Gen Virol,* 84 (Pt 8), 1941-9.

Sinclair, A.J. (2006), 'Unexpected structure of Epstein-Barr virus lytic cycle activator Zta', *Trends Microbiol,* 14 (7), 289-91.

Sinclair, A. J., et al. (1994), 'EBNA-2 and EBNA-LP cooperate to cause G0 to G1 transition during immortalization of resting human B lymphocytes by Epstein-Barr virus', *EMBO J,* 13 (14), 3321-8.

Stanfield, B. A. and Luftig, M. A. (2017), 'Recent advances in understanding Epstein-Barr virus', *F1000Res,* 6, 386.

Stiff, T., et al. (2005), 'Nbs1 is required for ATR-dependent phosphorylation events', *EMBO J,* 24 (1), 199-208.

Takada, K. and Ono, Y. (1989), 'Synchronous and sequential activation of latently infected Epstein-Barr virus genomes', *J Virol,* 63 (1), 445-9.

Takada, K., et al. (1986), 'trans activation of the latent Epstein-Barr virus (EBV) genome after transfection of the EBV DNA fragment', *J Virol,* 57 (3), 1016-22.

Taylor, G. S., et al. (2015), 'The immunology of Epstein-Barr virus-induced disease', *Annu Rev Immunol,* 33, 787-821.

Tellam, J. T., et al. (2012), 'Messenger RNA sequence rather than protein sequence determines the level of self-synthesis and antigen presentation of the EBV-encoded antigen, EBNA1', *PLoS Pathog,* 8 (12), e1003112.

Tempera, I. and Lieberman, P. M. (2010), 'Chromatin organization of gammaherpesvirus latent genomes', *Biochim Biophys Acta,* 1799 (3-4), 236-45.

Tempera, I., Klichinsky, M., and Lieberman, P. M. (2011), 'EBV latency types adopt alternative chromatin conformations', *PLoS Pathog,* 7 (7), e1002180.

Thorley-Lawson, D. A. (2015), 'EBV Persistence--Introducing the Virus', *Curr Top Microbiol Immunol,* 390 (Pt 1), 151-209.

Thorley-Lawson, D. A. and Allday, M. J. (2008), 'The curious case of the tumour virus: 50 years of Burkitt's lymphoma', *Nat Rev Microbiol,* 6 (12), 913-24.

Thorley-Lawson, D. A., et al. (2013), 'The pathogenesis of Epstein-Barr virus persistent infection', *Curr Opin Virol,* 3 (3), 227-32.

Toczyski, D. P. and Steitz, J. A. (1991), 'EAP, a highly conserved cellular protein associated with Epstein-Barr virus small RNAs (EBERs)', *EMBO J,* 10 (2), 459-66.

Tolhuis, B., et al. (2002), 'Looping and interaction between hypersensitive sites in the active beta-globin locus', *Molecular Cell,* 10 (6), 1453-65.

Tomkinson, B. and Kieff, E. (1992a), 'Second-site homologous recombination in Epstein-Barr virus: insertion of type 1 EBNA 3 genes in place of type 2 has no effect on in vitro infection', *J Virol,* 66 (2), 780-9.

Tomkinson, B. and Kieff, E. (1992b), 'Use of second-site homologous recombination to demonstrate that Epstein-Barr virus nuclear protein 3B is not important for lymphocyte infection or growth transformation in vitro', *J Virol,* 66 (5), 2893-903.

Tsai, K., et al. (2011), 'EBV tegument protein BNRF1 disrupts DAXX-ATRX to activate viral early gene transcription', *PLoS Pathog,* 7 (11), e1002376.

Tse, E. and Kwong, Y. L. (2017), 'The diagnosis and management of NK/T-cell lymphomas', *J Hematol Oncol,* 10 (1), 85.

Turatsinze, J. V., et al. (2008), 'Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules', *Nat Protoc,* 3 (10), 1578-88.

Urier, G., et al. (1989), 'The Epstein-Barr virus early protein EB1 activates transcription from different responsive elements including AP-1 binding sites', *EMBO J,* 8 (5), 1447-53.

van Helden, J., Andre, B., and Collado-Vides, J. (1998), 'Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies', *J Mol Biol,* 281 (5), 827-42.

Verweij, M. C., et al. (2015), 'Viral inhibition of the transporter associated with antigen processing (TAP): a striking example of functional convergent evolution', *PLoS Pathog,* 11 (4), e1004743.

Visel, A., Rubin, E. M., and Pennacchio, L. A. (2009), 'Genomic views of distant-acting enhancers', *Nature,* 461 (7261), 199-205.

Vockerodt, M., et al. (2014), 'Epstein-Barr virus and the origin of Hodgkin lymphoma', *Chin J Cancer,* 33 (12), 591-7.

Wang, D., Liebowitz, D., and Kieff, E. (1985), 'An EBV membrane protein expressed in immortalized lymphocytes transforms established rodent cells', *Cell,* 43 (3 Pt 2), 831-40.

Wang, P., et al. (2005), 'A redox-sensitive cysteine in Zta is required for Epstein-Barr virus lytic cycle DNA replication', *J Virol,* 79 (21), 13298-309.

Wasil, L. R., et al. (2013), 'The effect of Epstein-Barr virus Latent Membrane Protein 2 expression on the kinetics of early B cell infection', *PLoS One,* 8 (1), e54010.

Weiss, L. M., et al. (1989), 'Detection of Epstein-Barr viral genomes in Reed-Sternberg cells of Hodgkin's disease', *N Engl J Med,* 320 (8), 502-6.

Westphal, E. M., et al. (2000), 'Activation of lytic Epstein-Barr virus (EBV) infection by radiation and sodium butyrate in vitro and in vivo: a potential method for treating EBV-positive malignancies', *Cancer Res,* 60 (20), 5781-8.

White, R. E., et al. (2012), 'EBNA3B-deficient EBV promotes B cell lymphomagenesis in humanized mice and is found in human tumors', *J Clin Invest,* 122 (4), 1487-502.

Woellmer, A., Arteaga-Salas, J. M., and Hammerschmidt, W. (2012), 'BZLF1 governs CpG-methylated chromatin of Epstein-Barr Virus reversing epigenetic repression', *PLoS Pathog,* 8 (9), e1002902.

Wood, C. D., et al. (2016), 'MYC activation and BCL2L11 silencing by a tumour virus through the large-scale reconfiguration of enhancer-promoter hubs', *Elife,* 5.

Wu, T. C., et al. (1990), 'Detection of EBV gene expression in Reed-Sternberg cells of Hodgkin's disease', *Int J Cancer,* 46 (5), 801-4.

Xue, Y., et al. (2003), 'The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies', *Proc Natl Acad Sci U S A,* 100 (19), 10635-40.

Yanez-Cuna, J. O., et al. (2012), 'Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding', *Genome Research,* 22 (10), 2018-30.

Yates, J. L., Warren, N., and Sugden, B. (1985), 'Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells', *Nature,* 313 (6005), 812-5.

Young, L. S., Yap, L. F., and Murray, P. G. (2016), 'Epstein-Barr virus: more than 50 years old and still providing surprises', *Nat Rev Cancer,* 16 (12), 789-802.

Young, L. S., et al. (1991), 'Differentiation-associated expression of the Epstein-Barr virus BZLF1 transactivator protein in oral hairy leukoplakia', *J Virol,* 65 (6), 2868-74.