Development of a novel system to investigate the temporal regulation of V(D)J recombination.

Alastair Luke Smith

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Faculty of Biological Sciences
School of Molecular and Cellular Biology

September 2018

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Joan Boyes, for all of her training and guidance during the course of this project. I would also like to thank Joan for the speed in which comments on thesis drafts were given, for highlighting the importance of asking relevant scientific questions and for maintaining my blood sugar at dangerous levels. I also thank my co-supervisor Dr Gina Doody for her guidance and support.

I thank all the members of the Boyes lab, in particular I would like to thank: Dr James Scott for help in the lab during the early parts of this project and for kindly allowing me to use his ChIP data, Danny Thwaites for performing inductions at unsociable hours, Zeqian Gao for help with the luciferase reporter assays. I would also like to thank Dr. Vijaya Mahalingam Shanmugiah for helpful discussions about Hi-C data analysis and Dr Sally Boxall for help with flow cytometry.

On a personal level I also wish to thank my family for all their support. Finally, I thank my long-suffering girlfriend Toni, for putting up with me throughout this project and providing unending love and support.

Abstract

V(D)J recombination is vital for the generation of an effective adaptive immune system. Since the reaction involves breaking and re-joining DNA, it must be tightly regulated. Initiation of recombination relies on non-coding transcription which orchestrates a transient increase in the accessibility of the recombination signal sequences. Although these gene segments are separated by up to two megabases, remarkably, V and J non-coding transcription appears to be coordinately upregulated. The *Igλ* locus offers an ideal model system to investigate how this coordinate activation is achieved as the overexpression of a single transcription factor, IRF4, in pro-B cells is sufficient to prematurely activate *Igλ* recombination. Transgenic mice, which express IRF4 fused to the estrogen receptor hormone binding domain, have been generated previously and V(D)J recombination can be induced in pro-B cells by addition of tamoxifen. To overcome problems arising from limited numbers of primary pro-B cells, I firstly describe the generation of a novel pro-B cell line that expresses IRF4-ER_{T2} and allows V(D)J recombination to be induced. I next investigate how the coordinate regulation of *Vλ1* and *Jλ1* non-coding transcription is achieved using published ATAC-seq, ChIP-seq and Hi-C data together with the temporal analysis of the *Igλ* locus in the generated cell line by ChIP-qPCR and 3C. My data support a model whereby the coordinate upregulation of *Vλ1* and *Jλ1* non-coding transcription is mediated by the parallel activation of sister enhancers via IRF4. This results in the interaction of the enhancers with a previously uncharacterised hypersensitive site, forming an enhancer hub. This hub mediates the interaction of *Vλ1* and *Jλ1* with the enhancers to coordinately upregulate non-coding transcription and potentially facilitates recombination. The work presented here provides the first temporal insights into immunoglobulin locus activation and furthers our current understanding of how distant elements may be coordinately regulated.

Contents

Acknowledgements ii
Abstract iii
Contents iv
List of Figures and Tables ix
Abbreviations xiii
Chapter 1 - Introduction 1
A) Adaptive and innate immunity 1
1.1 B and T lymphocytes 1
1.2 Immunoglobulin structure 3
B) V(D)J recombination 7
1.3 The generation of a diverse antigen receptor repertoire 7
1.4 The RAG recombinase 9
1.5 HMGB1/2 11
1.6 The mechanism of V(D)J recombination 12
C) V(D)J recombination is highly ordered 15
1.7 B cell development 15
1.8 IL-7 signalling 18
1.9 Pre-BCR signalling 20
D) Regulation of V(D)J recombination 23
1.10 Regulation of *Rag* expression 23
1.11 RSS accessibility regulates V(D)J recombination 25
1.12 Histone acetylation 28
1.13 Histone methylation 29
1.14 Histone ubiquitination 30
1.15 Non-coding transcription 31
1.16 Allelic exclusion 32
E) Regulation of the antigen receptor loci 35
1.17 Structure of the immunoglobulin heavy chain locus 35
1.18 Structure of the immunoglobulin κ locus 37
1.19 Structure of the immunoglobulin λ locus 40
1.20 Subnuclear positioning 43
1.21 Locus contraction 44
1.22 Transcription factors regulating V(D)J recombination 45
1.23 Regulation of *Igh* recombination 52

1.24 Regulation of *Igκ* recombination....58
F) Generation of additional antigen receptor diversity....60
1.25 Somatic hypermutation....60
1.26 Class switch recombination....63
G) V(D)J recombination and leukaemia....65
H) Model systems to investigate V(D)J recombination....68
I) Aims....70
Chapter 2 – Materials and Methods....73
A. Common buffers....73
B. Media....75
C. DNA based methods....78
2.1 Preparation of genomic DNA....78
2.2 Phenol/chloroform extraction of DNA....78
2.3 Ethanol precipitation of DNA....78
2.4 A typical PCR reaction....79
2.5 Agarose Gel Electrophoresis....79
2.6 Isolation of DNA fragments by gel extraction....80
2.7 Ligation of DNA....80
2.8 Preparation of competent E.coli using Calcium Chloride....80
2.9 Transformation of *E.coli*....81
2.10 Small scale extraction of plasmid DNA from *E.coli* (Miniprep)....81
2.11 Large scale extraction of plasmid DNA from *E.coli*....82
D. RNA methods....82
2.12 RNA extraction....82
2.13 Generation of cDNA....82
2.14 Quantitative PCR (qPCR) using SYBR green....83
2.15 Classical 5' Rapid Amplification of Complementary DNA Ends (RACE)....84
2.16 Modified 5' Rapid Amplification of Complementary DNA Ends (RACE)....85
E. Protein extraction and analysis....86
2.17 Luciferase reporter assay....86
2.18 Preparation of whole cell and nuclear extracts....86
2.19 Western blotting....87
F. Cell culture methods....88
2.20 Culture of adherent cells....88

2.21 Culture of 103/BCL-2 cells 88
2.22 Culture of 1D1-T215 cells 88
2.23 Preparation of IL-7 88
2.24 Culture of primary pro-B cells....89
2.25 Transfection of 293T cells 89
2.26 Transfection of 103/BCL-2 cells by electroporation....89
2.27 Tamoxifen treatment of cell lines 90
G. Viral methods....90
2.28 Preparation of retrovirus....90
2.29 Preparation of Lentivirus 90
2.30 Retroviral/Lentiviral infection by spin-infection 91
H. Flow cytometry....91
2.31 Cell sorting by flow cytometry 91
2.32 Cell surface analysis by flow cytometry....94
I. Generation of a PIP-ER pro-B cell line....94
2.33 IL-7 dependent immortalization of PIP-ER D pro-B cells....94
2.34 Generation of A-MuLV pro-B cell lines 96
2.35 Generation of synergistic activation mediator (SAM) cell lines....97
2.36 Generation of MSCV-IRF4-ER cell lines 99
2.37 Generation of IRF4-DD cell lines 100
2.38 Generation of MSCV-Irf4-ER (T2) cell lines 100
2.39 Knockout of the PU.1/IRF4 binding site in the HS6 and Eλ$_{3-1}$ enhancers in 1D1-T215 cells....101
J. ChIP and 3C 102
2.40 Chromatin Immunoprecipitation (ChIP)....102
2.41 Chromosome conformation capture (3C)....104
2.42 Preparation of BAC template 105
2.43 3C primer design....105
2.44 Analysis of 3C templates by qPCR 105
K. Bioinformatic methods 106
2.45 Motif analysis 106
2.46 Analysis of published ChIP-seq data....106
2.47 Analysis of published ATAC-seq data....106
2.48 RNA-seq analysis....107
2.49 Hi-C analysis....107
2.50 Igλ locus homology analysis 107

L. Software....108
M. Accession numbers....109
N. Oligonucleotides used....110
O. Antibodies....114
P. Plasmid maps....115
Chapter 3 – Generation of a cell line capable of inducing the *Igλ* locus....118
A. Introduction....118
B. Results....122
3.1 Immortalisation by long term growth in IL-7....122
3.2 Immortalisation by infection with A-MuLV....123
3.3 Upregulation of IRF4-ER by CRISPR activator....128
3.4 Re-introduction of IRF4-ER by MSCV-IRF4-ER....130
3.5 Generation of an estrogen insensitive IRF4-ER cell line by the use of a destabilisation domain.....137
3.6 Generation of an estrogen insensitive IRF4-ER cell line by mutagenesis of the ER domain in IRF4-ER.....139
C. Discussion....145
3.7 Immortalisation of pro-B cells from PIP-ER transgenic mice....145
3.8 IRF4-ER is not expressed at the correct level in PIP-ER pro-B cells 146
3.9 Generation of an IRF4-ER expressing pro-B cell line....147
3.10 1D1-T215 cells represent an ideal system to investigate gene segment activation....150
Chapter 4 – What are the factors regulating the activation of *Vλ1* and *Jλ1* gene segments?....153
A. Introduction....153
B. Results....155
4.1 Mapping of the *Vλ1* promoter....155
4.2 Identification of consensus transcription factor binding sites....160
4.3 Identification of additional regulatory elements in the *Igλ* locus....162
4.4 HS6 binds the same transcription factors as $E\lambda_{3\text{-}1}$....164
4.5 HS6 is not a recently diverged duplication of $E\lambda_{3\text{-}1}$....167
4.6 HS6 is an enhancer of *Vλ1* non-coding transcription in a transient transfection assay.....169
4.7 HS6 contacts *Vλ1 in vivo*....170
4.8 CTCF binding sites mediate the HS6/Vλ1 interaction....177
4.9 The *Igλ* locus is separated into domains by CTCF/cohesin....182

4.10 How is the enhancer Eλ3-1 involved in *Igλ* locus activation?.....188
4.11 Analysis of transcription factors binding to both enhancer elements 190
C. Discussion 192
4.12 Identification of the *Vλ1* promoter.....192
4.13 Analysis of transcription factor binding to the *Vλ1* and *Jλ1* promoters 193
4.14 Identification of an enhancer for the *Vλ1* promoter.....194
4.15 Which interactions mediate the coordinate activation of *Vλ1* and *Jλ1*? 196
4.16 How might the coordinate activation of *Vλ1* and *Jλ1* be regulated? 198
Chapter 5 – Temporal analysis of the coordinate activation of *Vλ1* and *Jλ1*...200
A. Introduction.....200
B. Results.....202
5.1 Deletion of the HS6 and Eλ3-1 enhancers.....202
5.2 Mutation of the PU.1/IRF4 binding sites in HS6 and Eλ3-1.....204
5.3 Mutation of the PU.1/IRF4 binding sites at HS6 and Eλ3-1 via small-scale deletions.....206
5.4 HS6 and Eλ3-1 are both required for the activation of *Vλ1* and *Jλ1* non-coding transcription 209
5.5 Eλ3-1 and HS6 are simultaneously bound by IRF4 211
5.6 CTCF/cohesin mediated interactions are present at the pro-B cell stage 213
5.7 Temporal analysis of interactions formed in the *Igλ* locus.....219
5.8 IRF4 binding to HS6 and Eλ3-1 is essential for correct *Igλ* locus contraction.....223
C. Discussion 226
5.9 IRF4 binding to the *Igλ* enhancers is crucial for *Igλ* activation 226
5.10 IRF4 is essential for *Igλ* locus contraction.....229
Chapter 6 – Discussion.....230
6.1 Generation of a system capable of inducing V(D)J recombination.230
6.2 Coordinate activation of *Vλ1* and *Jλ1* gene segments by the parallel activation of sister enhancers 235
Appendix.....247
References.....250

List of Figures and Tables

Figure 1.1: Structure of an immunoglobulin.. 5
Figure 1.2: Structure of a T cell receptor.....6
Figure 1.3: Simplified V(D)J recombination event at a hypothetical light chain locus to form the antibody light chain.....8
Figure 1.4: Structure of RAG1 and RAG2.. 10
Figure 1.5: Overview of the cleavage and joining phases of V(D)J recombination.. 13
Figure 1.6: Overview of B cell development.....17
Figure 1.7: Signal transduction pathways activated by IL-7 signalling.. 19
Figure 1.8: Simplified overview of pre-BCR signalling. …….21
Figure 1.9: Schematic representation of a di-nucleosome.. 26
Figure 1.10: The murine immunoglobulin heavy chain locus. 36
Figure 1.11: The murine immunoglobulin kappa chain locus. 38
Figure 1.12: A simplified overview of the immunoglobulin lambda locus.. 41
Figure 1.13: Long range interactions mediated by Eμ.....53
Figure 1.14: Long-range interactions mediating V_H to DJ_H recombination.. 56
Figure 1.15: Long-range interactions mediating Igκ recombination.....59
Figure 1.16: Simplified overview of somatic hypermutation. 62
Figure 1.17: Overview of class switch recombination. 64
Figure 1.18: Vλ1 and Jλ1 non-coding transcription is coordinately upregulated.. 71
Figure 2.1: Flow cytometry template for purification of pro-B cells……..96
Figure 2.2: Flow cytometry template for purification of pre-B cells.....93
Figure 2.3: Gating strategy to obtain a pure CD19+ population.. 95
Figure 2.4: Gating strategy for the isolation of the cells with highest transgene expression.....99
Table 2.1: Software utilised during the course of this project.....108
Table 2.2: Accession numbers of published ChIP-seq data analysed. 109
Table 2.3: Accession numbers of published ATAC-seq and RNA-seq data....109
Table 2.4: Oligonucleotides used for RT-qPCR. 110
Table 2.5: Oligonucleotides used for 5’ RACE.....111
Table 2.6: Oligonucleotides used for ChIP.....111
Table 2.7: sgRNA oligonucleotides. 112
Table 2.8: Oligonucleotides used for the confirmation of HS6 and $Eλ_{3-1}$ deletions. 112
Table 2.9: Oligonucleotides used for cloning. 113

Table 2.10: Antibodies together with their application and amount used per experiment.. 114

Figure 2.5: Map of MSCV-IRF4-ER-IRES-GFP. 115

Figure 2.6: Map of lenti-P2A-Puro (A) and lenti-IRF4-DD-P2A-Puro (B).....116

Figure 2.7: Map of a modified lenti-CRISPRv2 vector with two sgRNA cassettes (A). 117

Figure 3.1: IL-7 dependent cell lines have a pro-B cell phenotype.....123

Table 3.1: The number of A-MuLV infected cell lines generated immediately following extraction or after five/six days in culture.. 124

Figure 3.2: The generated cell lines have pro-B cell markers. 125

Figure 3.3: *Irf4* and *Pu.1* expression in v-abl immortalised pro-B cells. 126

Figure 3.4: Analysis of IRF4-ER expression in *v-abl* immortalised cell lines by Western blotting. 127

Figure 3.5: Schematic of the Synergistic Activation Mediator (SAM) system targeted to the λ5 cassette.....129

Figure 3.6: Analysis of transgene expression in MSCV-IRF4-ER-IRES-GFP infected 1D1 cells.....131

Figure 3.7: Alteration in *Vλ1* non-coding transcription in the MSCV-IRF4-ER infected pro-B cell lines after induction. 132

Figure 3.8: IRF4-ER expression is readily detectable in the 1D1-A4 cell line and Vλ1 non-coding transcription in this line mirrors PIP-ER primary pro-B cells. .134

Figure 3.9: IRF4-ER is present in the nucleus of untreated 1D1-A4 cells. 135

Figure 3.10: The estrogen receptor ligand binding domain of IRF4-ER is the G400V variant.. 136

Figure 3.11: IRF4-DD is stabilised in the presence of Trimethoprim (TMP).....139

Figure 3.12: Alteration in *Vλ1* non-coding transcription in the MSCV-IRF4-ER$_{T2}$ infected pro-B cell lines after induction. 140

Figure 3.13: *Irf4* is expressed at similar levels to pre-B cells in the 1D1-T215 cell line and IRF4-ER nuclear localisation is dependent on 4-OHT. 142

Figure 3.14: IRF4-ER$_{T2}$ is stably expressed in the 1D1-T215 cell line and the 1D1-T215 has no gross copy number alterations. 143

Figure 3.15: *Vλ1* and *Jλ1* non-coding transcription are upregulated at four hours post-induction in the 1D1-T215 cell line. 144

Table 3.2: Summary of the strategies used to generate a pro-B cell line capable of inducing the activation of IRF4.....148

Figure 4.1: Location of *Vλ1* non-coding start sites identified by 5' RACE. The location of transcription start sites are indicated by the black arrows.....157

Figure 4.2: Assembled transcripts from pre-B RNA-seq data indicate a TSS of 35-40 bp upstream of the Vλ1 gene segment.. 158

Figure 4.3: Identification of transcription factor binding sites at the *Vλ1* and *Jλ1* promoters by motif prediction. 161

Figure 4.4: Analysis of pro-B ChIP-seq data in the 3' half of the *Igλ* locus to identify elements with enhancer-like characteristics. 163

Figure 4.5: HS6 binds the same transcription factors as Eλ3-1. 165

Figure 4.6: ChIP-qPCR analysis of E2A, IRF4 and PU.1 binding to Eλ3-1 and HS6 in pre-B cells. 166

Figure 4.7 Analysis of consensus transcription factor binding sites at HS6 (Top) and Eλ3-1 (Bottom). 168

Figure 4.8: HS6 is an enhancer of the *Vλ1* promoter. 169

Figure 4.9: Schematic of Chromosome Conformation Capture (left) and Hi-C (right). 171

Figure 4.10: Analysis of pro-B Hi-C data performed using the restriction enzyme DpnII at a resolution of 20 kb. 173

Figure 4.11: 3C analysis of the 3' half of the *Igλ* locus in pre-B cells using HS6 as the viewpoint. 176

Figure 4.12: ChIP-seq data analysis of factors involved in long range chromatin interactions 178

Figure 4.13: ChIP-qPCR analysis of CTCF binding to the regions of accessible chromatin upstream of *Vλ1* (HSVλ1) and downstream of HS6 (HS7) in pre-B cells. 179

Figure 4.14: CTCF binding sites at HSVλ1 and HS7 are in a convergent orientation. 180

Figure 4.15: Diagram indicating the proposed structure of the 3' half of the *Igλ* locus if HS7 interacts with HSVλ1. 181

Figure 4.16: The CTCF binding pattern is mirrored in the 5' half of the locus. ChIP-seq analysis of the *Igλ* locus. 183

Figure 4.17: Analysis of pro-B Hi-C data performed using the restriction enzyme DpnII at a resolution of 20 kb. 184

Figure 4.18: Proposed long-range interactions in the Igl locus mediated by CTCF/cohesin. 185

Figure 4.19: Analysis of homologous regions between the 5' and 3' half of the *Igλ* locus. 187

Figure 4.20: 3C analysis of the 3' half of the *Igλ* locus in pre-B cells using Eλ3-1 as the viewpoint. 189

Figure 4.21: ChIP-qPCR analysis of E2A, IRF4 and PU.1 binding to Eλ3-1 and HS6 in pro-B and pre-B cells 191

Figure 5.1: Strategies used to mutate the enhancer HS6. 203

Figure 5.2: Identification of cell line pools containing the required deletion of HS6. 204

Figure 5.3: Example of a heterozygous HS6 PU.1/IRF4 binding site mutant..205

Figure 5.4: Analysis of monoclonal cell lines to identify PU.1/IRF4 binding site deletions at HS6 or Eλ3-1.....207

Figure 5.5: Sequence alignments of clones bearing deletions of the PU.1/IRF4 binding site at either HS6 or Eλ3-1. 208

Figure 5.6: *Vλ1* and *Jλ1* non-coding transcription is substantially decreased in the HS6 and Eλ3-1 PU.1/IRF4 binding site mutants.....210

Figure 5.7: IRF4 binding to the enhancers Eλ3-1 and HS6 is simultaneously increased upon induction. 212

Figure 5.8: CTCF is bound at its maximal level at the pro-B cell stage.....213

Table 5.1: Interactions formed during *Igλ* locus activation. 214

Figure 5.9: Analysis of the interactions formed by the HS6 and Eλ3-1 enhancers in pro-B and pre-B cells. 216

Figure 5.10: The level of YY1 bound at HS6 and HSCλ1 increases during *Igλ* activation.....218

Figure 5.11: Temporal analysis of interactions formed by the 3' *Igλ* enhancers. 221

Figure 5.12: Analysis of the interactions formed by HS6 and Eλ3-1 in the HS6/Eλ3-1 PU.1/IRF4 binding site mutant cell lines.. 228

Figure 6.1: Proposed model for the coordinate regulation of Vλ1 and Jλ1 non-coding transcription ……..……..…..239

Abbreviations

3C	Chromosome conformation capture
3'RR	3' regulatory region
4C	Circularized Chromosome Conformation Capture
4-OHT	4-hydroxytamoxifen
AID	Activation-induced cytidine deaminase
ALL	Acute lymphoblastic leukaemia
A-MuLV	Abelson murine leukaemia virus
APE1	Apurinic/apyrimidinic endonuclease 1
ATAC-seq	Assay for transposase-accessible chromatin with high-throughput sequencing
ATM	Ataxia telangiectasia mutated
BAC	Bacterial artificial chromosome
BCR	B cell receptor
BER	Base excision repair
BLAST	Basic local alignment search tool
BLINK	B cell linker protein
bp	Base pair
BTK	Bruton tyrosine kinase
CDK	Cyclin dependent kinase
CDKN2A	cyclin-dependent kinase inhibitor 2A
cDNA	complementary DNA
CDR	Complementarity determining region
Cer	Contracting element for recombination
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation with high-throughput sequencing
cRSS	Cryptic recombination signal sequence
CSR	Class switch recombination
CTCF	CCCTC-binding factor
CXCR4	C-X-C chemokine receptor type 4
DAPI	4′,6-diamidino-2-phenylindole
DD	destabilisation domain
dH2O	Deionised water
DMEM	Dulbeccos modified Eagles medium
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
DSB	Double strand break
E.coli	*Escherichia coli*
EBF1	Early B cell Factor 1
ecDHFR	E.coli Dihydrofolate reductase
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme linked immunosorbent assay
EM	Electron microscopy
EMCCD	Electron multiplying charge coupled device
ER	Estrogen receptor

ERK	extracellular signal-regulated kinase
ESC	Excised signal circle
Fab	Fragment antigen binding
FACS	Fluorescence activated cell sorting
Fc	Fragment crystallisation
FCS	Foetal calf serum
FISH	Fluorescence *in situ* hybridisation
FITC	Fluorescein isothiocyanate
GFP	Green fluorescent protein
H3K27Ac	Histone H3 lysine 27 acetylation
H3K4me3	Histone H3 lysine 4 trimethylation
HA	Human influenza hemagglutinin
HBD	Hormone binding domain
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMGB1/2	High mobility group protein B1/2
HS	Hypersensitive site
HSC	Haematopoietic stem cell
HSF1	Heat shock factor 1
Ig	Immunoglobulin
IGCR1	Intergenic control region 1
IKZF1	IKAROS Family Zinc Finger 1
IL-7	Interleukin 7
IL-7Rα	Interleukin 7 receptor α
IRF4	Interferon regulatory factor 4
IRF8	Interferon regulatory factor 8
ITAM	Immunoreceptor tyrosine-based activation motif
JAK3	Janus kinase 3
kDa	Kilodalton
KOAc	Potassium acetate
LCR	Locus control region
LMO2	LIM only 2
LMPP	lymphoid primed multi-potent progenitor
MACS	Model based analysis of ChIP-seq
MEF2C	Myocyte Enhancer Factor 2C
MHC	Major histocompatibility complex
MMR	Mismatch repair
MPP	Multipotent progenitors
MSCV	Murine stem cell virus
NaOAc	Sodium acetate
NBD	Nonamer binding domain
NHEJ	Non-homologous end joining
NTG	Non-transgenic
PAGE	polyacrylamide gel electrophoresis
PAIR	Pax-5 associated intergenic repeat
PAX5	Paired Box 5

PBS	Phosphate buffered saline
PcG	Polycomb group
PCR	Polymerase chain reaction
PE	Phycoerythrin
PEI	Polyethyleneimine
PHD	Plant homeodomain
PI3K	Phosphoinositide 3-kinase
PMSF	Phenylmethylsulfonyl fluoride
p-TEFb	Positive transcription elongation factor B
qPCR	Quantitative PCR
RACE	Rapid amplification of complementary (DNA) ends
RAG	Recombination activating gene
RING	Really interesting new gene
RNA	Ribonucleic acid
RPMI	Roswell Park Memorial Institute
RSS	Recombination signal sequence
RT-qPCR	Reverse transcription-qPCR
SAM	Synergistic activation mediator
SDS	Sodium dodecyl sulphate
sgRNA	Single guide RNA
SHM	Somatic hypermutation
Sis	Silencer in the intervening sequence
STAT5	Signal transducer and activator of transcription 5
SYK	Spleen tyrosine kinase
TAE	Tris acetate EDTA
TCR	T cell receptor
Tdt	Terminal deoxynucleotidyl transferase
TE	Tris EDTA
TEA	T early α
T_H	Helper T cell
Treg	Regulatory T cell
TSS	Transcription start site
UNG	Uracil DNA glycosylase
YY1	Yin Yang 1
ZAP70	ζ-chain-associated protein kinase of 70 kDa

Chapter 1 - Introduction

A) Adaptive and innate immunity

A vast number of potential pathogens are constantly encountered by multicellular organisms which must be recognised and eliminated by the immune system to avoid infection. The innate immune system provides the first line of defence against pathogens by non-specifically targeting their common features. These mechanisms include the phagocytosis of bacterial pathogens by phagocytic macrophages and the stimulation of a local inflammatory response by neutrophils and monocytes. These reactions from the innate immune system are the predominant response to an unknown pathogen within five to seven days of infection (Murphy and Weaver, 2017) but as many pathogens have evolved to evade the innate immune responses, vertebrates have evolved an adaptive immune system that provides a more specific defence against pathogens (Murphy and Weaver, 2017).

1.1 B and T lymphocytes

Central to the adaptive immune response are B (Bursal or Bone marrow-derived) and T (Thymus-derived) lymphocytes, which express clonally diverse and highly specific receptors on their cell surface that are capable of recognising a vast range of antigens. Each B and T lymphocyte expresses an antigen receptor that is specific to a single antigen and is able to detect single amino acid differences. B lymphocytes express a B-cell receptor (BCR) and secrete immunoglobulin(s) (Ig), whereas T lymphocytes express a T cell receptor (TCR). Whilst both B and T cells express similar receptors on the cell surface, both cell types mount different responses when encountering pathogens.

T-lymphocytes either express a TCR consisting of an αβ or γδ heterodimer. Remarkably, αβ T cells are unable to directly recognise antigens, but instead recognise processed peptide fragments bound to a cell surface MHC molecule. The response triggered by antigen binding is determined by the class of the T cell which is dependent on the expression of either the cell surface marker CD4 or CD8. Cytotoxic T cells, defined by CD8 expression, recognise virus infected host cells and induce apoptosis (Murphy and Weaver, 2017). In contrast, CD4+ T-cells, known as helper T cells, release cytokines to direct the activity of other immune cells upon recognition of MHC class II bound peptides on the surface of B cells, dendritic cell and macrophages (Murphy and Weaver, 2017).

Helper T cells are divided further into four distinct subsets, the best characterised are type 1 (T_H1) and type 2 (T_H2) helper cells (Parker, 1993). T_H1 cells activate the microbicidal properties of macrophages and induce B-cells to secrete IgG antibodies. T_H2 cells initiate the humoral immune response by stimulating naïve B cells to produce IgM isotype antibodies in addition to stimulating the production of other antibody isotypes e.g. (IgA, IgE) or IgG subtypes. Additional subsets of helper T cells include T_H17, which have a role in the immune response against bacteria and fungi, and regulatory T cells (T_{reg}) which supress the immune response (Yamane and Paul, 2013). The main function of T_{reg} cells is to limit the response to self-antigens to prevent autoimmune diseases, which can occur via a number of mechanisms including limiting the proliferation of CD4+ and CD8+ T cells (Murphy and Weaver, 2017).

B-cells are responsible for humoral immunity. Recognition of an antigen via B cell surface immunoglobulins results in the internalisation of the antigen which is processed into peptides that bind to a molecule of the major histocompatibility complex (MHC) class II. This complex is presented on the cell surface and recognised by helper T cells, resulting in the activation of the B-cell, leading to rapid proliferation and differentiation into plasma cells. B cell activation therefore results in the production of a large number of identical antibody secreting B cells (Parker, 1993). After the elimination of a pathogen, the majority of B and T cells die, leaving a subset of memory B and T cells. Upon re-exposure, these cells are rapidly activated to proliferate and differentiate into effector cells, resulting in a rapid response to the pathogen. Therefore, whilst the establishment of an adaptive immune response against an unknown pathogen requires several days, immunological memory enables pathogens to be rapidly eliminated on subsequent exposure.

1.2 Immunoglobulin structure

Antibodies are Y shaped structures consisting of two identical heavy (H) and light (L) chains, linked by disulphide bridges. The light-chain can consist of either a κ or λ chain, and although both have identical function, each immunoglobulin always contains two identical light chains of the same isotype.

The component heavy and light chains contain an N-terminal ‘variable’ (V) and one or more ‘constant’ (C) domains. Heavy chains contain three constant regions whereas light chains contain a single constant region. The variable region contains two identical antigen-binding sites, formed by the heavy and light chains, which consist of three hypervariable loops also known as complementarity determining regions (CDRs). The sequence of CDRs varies considerably between developing B cells, thus enabling a vast number of antigens to be detected. After folding, an immunoglobulin consists of three globular domains: two Fab (Fragment antigen binding) fragments, containing the antigen binding site, and a Fc (Fragment crystallisation) fragment which contains the C_H2 and C_H3 regions of the heavy chains (Murphy and Weaver, 2017; Figure 1.1).

There are five different isotypes of the heavy chain constant region which define the class of the immunoglobulin molecule: IgA, IgD, IgE, IgG and IgM. Furthermore, IgG can be divided into subclasses. In humans these are IgG_1, IgG_2, IgG_3 and IgG_4. Naïve B cells express membrane bound IgD and IgM. When stimulated appropriately, isotype switching occurs to allow the expression of alternative immunoglobulin classes. The switching of antibody class improves the ability of antibodies to remove the pathogen inducing the humoral response. For example, the Fc region of IgG_1 can bind to the Fcγ receptors present on macrophages and neutrophils to facilitate the phagocytosis of the bound pathogen (Murphy and Weaver, 2017). Furthermore, in addition to altering the properties of the antibody, the class of immunoglobulin also determines the capacity of the antibody to oligermerise as IgM is secreted as a pentamer, IgA as a dimer, whereas IgD, IgE and IgG are secreted as monomers.

The TCR is a heterodimer composed of αβ or γδ chains linked by disulphide bonds, Figure 1.2. T cells expressing αβ are the most prevalent in humans, with only 5% of T cells expressing γδ receptors. Similar to immunoglobulins, each polypeptide chain consists of a variable region and a constant region with hypervariable loops within the antigen binding site determining antigen specificity. Unlike immunoglobulins, TCRs cannot be secreted and are only expressed on the cell surface, where they interact with a transmembrane complex consisting of CD3 and CD247. These molecules act as signal transducers, as the TCR does not possess any enzymatic activity, activating intracellular signalling by immunoreceptor tyrosine-based activation motifs (ITAM) resulting in T cell activation (Brownlie and Zamoyska, 2013).

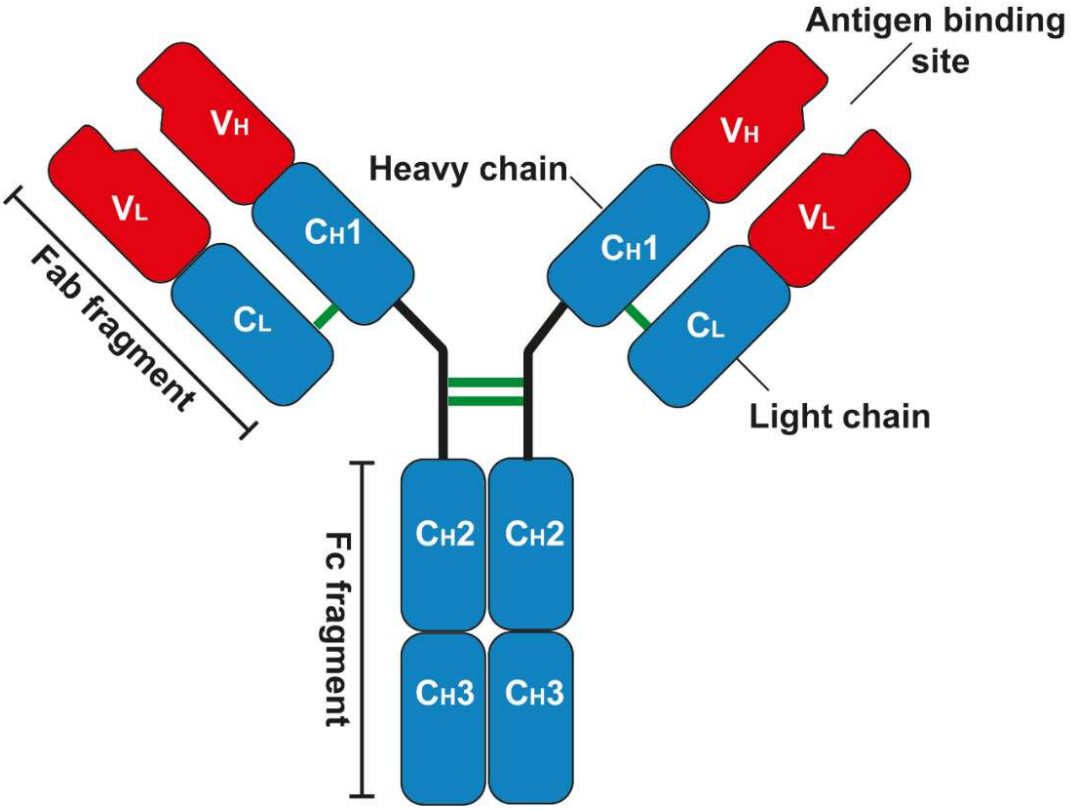


Figure 1.1: Structure of an immunoglobulin. An immunoglobulin molecule consists of two identical heavy chains and two identical light chains linked by disulphide bridges (green). A heavy chain consists of three constant regions (C_H1-3) and one variable region (V_H). The light chains consist of a single constant region (C_L) and a single variable region (V_L). The antigen binding site is formed from the variable regions of the heavy and light chains. The regions that comprise the fragment antigen binding (Fab) and fragment crystallisation (Fc) domains are indicated.

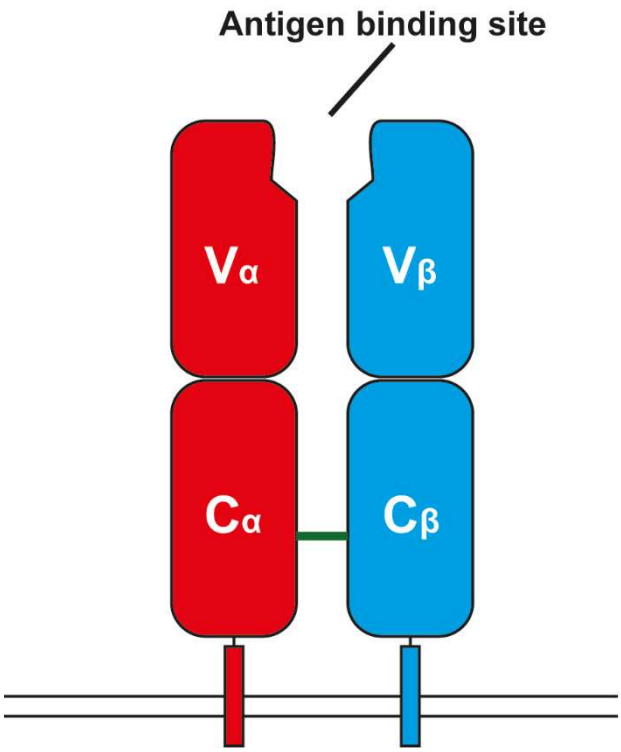


Figure 1.2: Structure of a T cell receptor. The T cell receptor (TCR) consists of two peptide chains, α and β (shown) or γ and δ, linked by disulphide bonds (green). Each polypeptide chain consists of a constant (C) and variable region (V) with the antigen binding site comprising the variable region of both polypeptide chains. TCRs are membrane bound and contain a transmembrane domain and a short intracellular tail.

B) V(D)J recombination

1.3 The generation of a diverse antigen receptor repertoire

To generate an adaptive immune system that is able to respond to a vast variety of potential pathogens, significant diversity in antigen receptors must be generated. With the number of unique immunoglobulins and T cell receptors exceeding one billion per individual, such magnitude of diversity would be impossible to produce, even if all the known protein coding genes in the human genome were used for this singular purpose. A substantial proportion of this diversity is generated by the somatic recombination of gene segments within the antigen receptor loci of B and T cells (Gellert, 2002), a process known as V(D)J recombination.

Mammals have seven antigen receptor loci, three immunoglobulin (Ig) loci – h, κ and λ, and four T-cell receptor (TCR) loci – α, β, γ and δ. Compared to the vast majority of vertebrate genes, the genes encoding immunoglobulin and T-cell receptors have a unique structure as they are arranged as arrays of gene segments (Figure 1.3). Historically named variable (V), diversity (D) and joining (J) gene segments, these segments are typically found as multiple copies lying upstream of a constant (C) region. Diversity (D) gene segments, however, are only located within the *Igh*, *TCRβ* and *TCRδ* loci (Gellert, 2002). During V(D)J recombination, one of each V, D (if present) and J gene segments are recombined together at random to form the variable exon of the antigen receptor which is spliced to a constant (C) region upon transcription. The randomness in the choice of gene segment, along with the imprecise joining of segments contributes to the diversity of the final antigen receptor (Lewis, 1994).

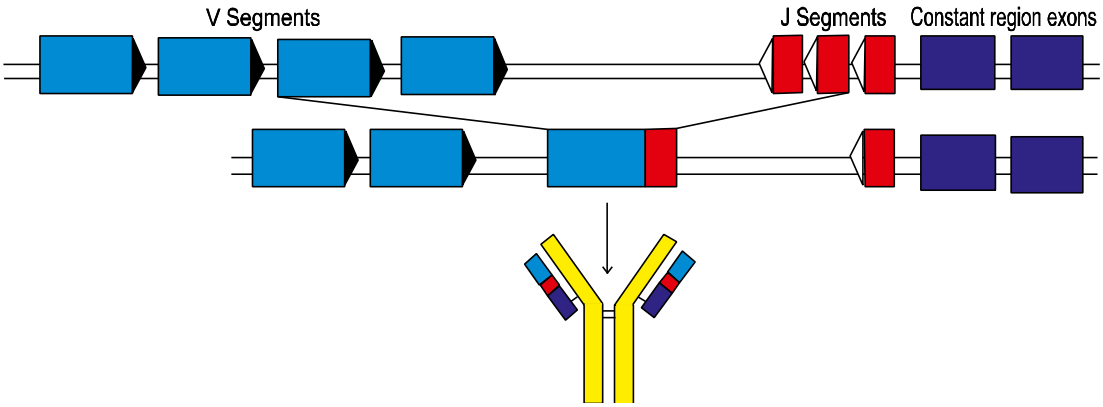


Figure 1.3: Simplified V(D)J recombination event at a hypothetical light chain locus to form the antibody light chain. A single V (blue) and J (red) gene segment are recombined by removing the intervening DNA. Splicing of the transcribed VJ gene segment with a constant region exon generates a complete light chain transcript. RSSs are represented by triangles, open triangles represent 12-RSS, filled triangles represent 23-RSS.

V(D)J recombination is initiated by the products of the lymphoid specific recombination activating genes 1 and 2 (*Rag1* and *Rag2*) (McBlane et al., 1995), which associate with High mobility group box (HMGB) 1 or 2 (van Gent et al., 1997) to recognise recombination signal sequences (RSSs) flanking V, D, and J gene segments. An RSS consists of conserved heptamer (CACAGTG) and nonamer (ACAAAAACC) elements, separated by 12±1bp or 23±1bp of non-conserved 'spacer' nucleotides. Importantly, efficient recombination only occurs between RSSs with dissimilar spacers, the '12/23 rule' (Tonegawa, 1983). At each antigen receptor locus, RSS spacer lengths are positioned so that recombination is directed to generate functional products (Figure 1.3) e.g. V-to-J, thus avoiding non-productive rearrangements such as V-to-V. Nevertheless, these non-productive re-arrangements do occur, albeit at low frequencies (Parkinson et al., 2014). Furthermore, not every pair of 12- and 23-RSSs are compatible, referred to as the "beyond 12/23" restriction. The mechanisms that govern this additional layer of regulation are incompletely understood, but include: catalytic incompatibilities, intra-locus long-range interactions and the distribution of chromatin marks (Banerjee and Schatz, 2014; Drejer-Teel et al., 2007; Hughes et al., 2003; Olaru et al., 2004; Sleckman et al., 2000).

1.4 The RAG recombinase

The RAG1 and RAG2 proteins alone are sufficient for the initiation of V(D)J recombination (McBlane et al., 1995). Remarkably, the transfection of *Rag1* cDNA into fibroblasts is able to initiate V(D)J recombination (Schatz et al., 1989). Whilst RAG1 is able to perform recombination inefficiently, co-transfection of *Rag2* results in a 1000-fold increase in recombination (Oettinger et al., 1990). Remarkably, deficiency in either RAG1 or RAG2 results in an early developmental block in lymphoid linages, demonstrating that the non-catalytic RAG2 protein is required for correct lymphocyte development (Mombaerts et al., 1992; Shinkai et al., 1992).

RAG1 and RAG2 (Figure 1.4A) are highly conserved in jawed vertebrates and appear to have been acquired by the horizontal gene transfer of an ancient transposase (Huang et al., 2016; Roth, 2003). Interestingly, whilst RAG1 is likely to have originated from a transposable element of the *Transib* family (Kapitonov and Jurka, 2005), RAG2 appears to have emerged later, to more efficiently catalyse V(D)J recombination (Huang et al., 2016). The genomic organisation of the *Rag* genes is also highly conserved. *Rag1* and *Rag2* are positioned 8 kb apart and are convergently transcribed from single exons and furthermore, a single enhancer *Erag* has been shown to upregulate the expression of both *Rag* genes (Hsu et al., 2003).

Recent structural studies have illustrated the architecture of the RAG heterotetramer, which contains two RAG1 and two RAG2 proteins (Schatz and Swanson, 2011). The paired RAG1 molecules form a Y-shaped structure with C-terminal ‘arms’ and an intercalated N-terminal ‘stalk, with the two RAG2 proteins occupying the ends of the ‘arms’ (Figure 1.4B).

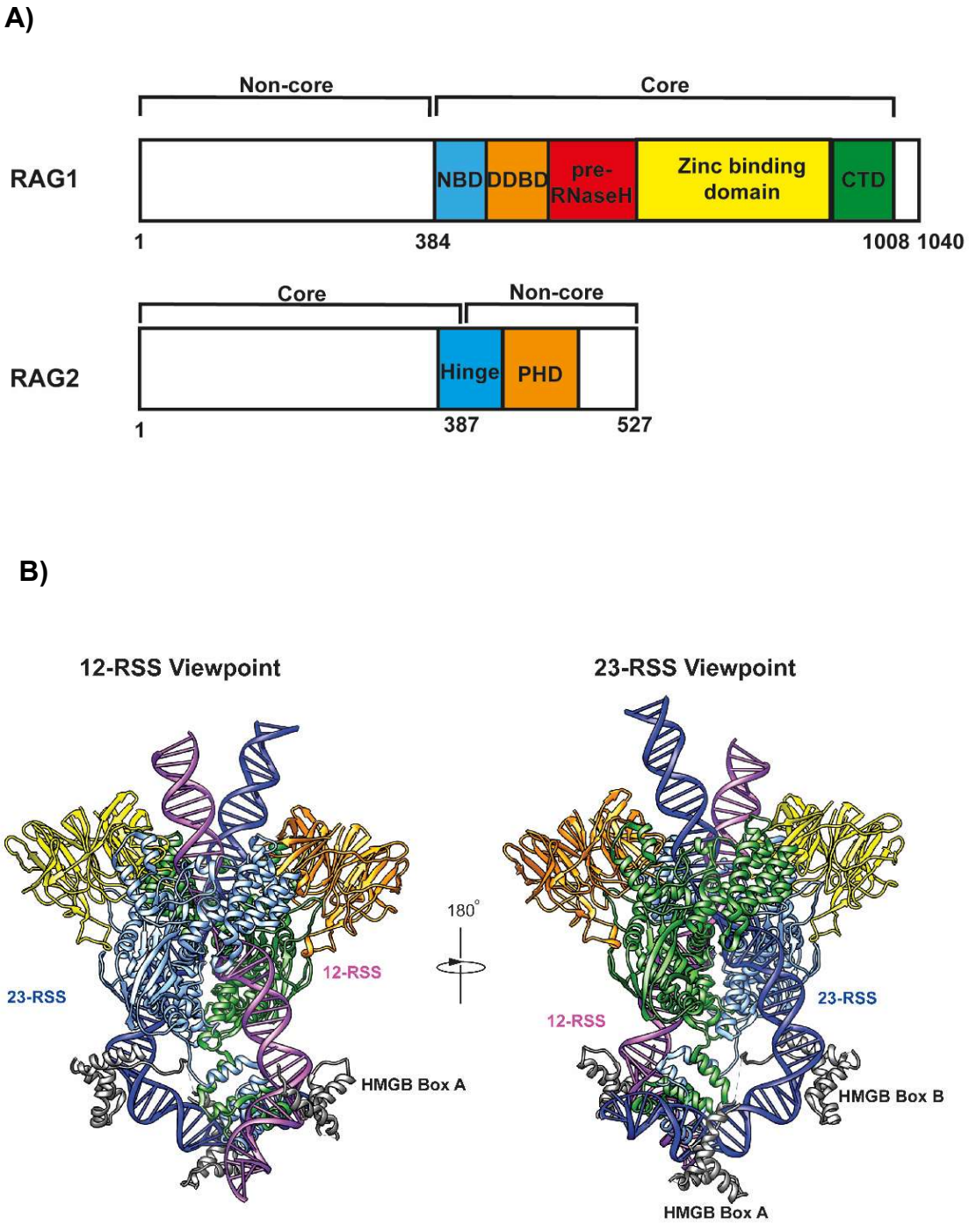


Figure 1.4: Structure of RAG1 and RAG2. A) Murine RAG1 and RAG2 can be functionally divided into core and non-core domains. Conserved domains are indicated. B) Structure of the RAG recombinase in complex with a 12- and 23-RSS. NBD – nonamer binding domain; DDBD – dimerization and DNA binding domain; CTD – carboxy-terminal domain; PHD – plant homeodomain.

Interestingly, the DNA cleavage and sequence specificity of the RAG recombinase is entirely contained within the 'core' (residues 384-1008) domain of RAG1. The sequence specificity of the recombinase is established by the RAG1 nonamer-binding domain (NBD; residues 389-464) (Difilippantonio et al., 1996; Yin et al., 2009) via the AT-hook motif which binds to the A-rich tract of the nonamer (Yin et al., 2009). Furthermore the nonamer binding domains of RAG1 make extensive contacts with each other and the nonamers, suggesting that the nonamer binding domains play a key roles in promoting synapsis between RSSs (Yin et al., 2009). The active site of RAG1 consists of three acidic residues (D600, D708 and E962) which fold into a metal-coordinating motif to initiate a hydrolysis-transesterification reaction (Fugmann et al., 2000; Kim et al., 1999; Landree et al., 1999).

RAG2 enhances the ability of RAG1 to bind and cleave DNA, possibly by assisting in the formation or the maintenance of the active site (Kim et al., 2015; Ru et al., 2015). Furthermore, RAG2 appears to have a role in directing the RAG recombinase to active chromatin through its plant homeodomain (PHD) finger which binds to histone 3 lysine 4 trimethylation (H3K4me3; Matthews et al., 2007), marking active transcriptional start sites. Interestingly, not only does the interaction with H3K4me3 direct the recombinase to active regions, H3K4me3 binding appears to have a stimulatory effect on RAG cleavage (Grundy et al., 2010; Shimazaki et al., 2009)

1.5 HMGB1/2

Whilst RAG1 and RAG2 are sufficient for RSS cleavage *in vitro* (McBlane et al., 1995), the high mobility group box proteins HMGB1 and HMGB2 increase RAG-mediated cleavage by 7-100 fold (van Gent et al., 1997). HMGB proteins bind non-specifically to the minor groove of DNA to facilitate DNA bending (Malarkey and Churchill, 2012). During V(D)J recombination, HMGB1/2 increases RAG binding to 23-RSSs by bending the 23 bp spacer to bring the heptamer and nonamer into closer proximity, in addition to helping position the RAG complex at a 12-RSS by stabilising the bent structure of the RSS (Ciubotaru et al., 2013; van Gent et al., 1997; Kim et al., 2018; Ru et al., 2015).

1.6 The mechanism of V(D)J recombination

The process of V(D)J recombination can be divided into two distinct phases: a cleavage phase and a joining phase (Schatz & Swanson, 2011; Figure 1.5). During the cleavage phase, the 12-RSS and the 23-RSS are brought into close proximity by the RAG proteins to form a stable synaptic complex, directed by the nonamer binding domains of RAG1 homodimers and HMGB1/2 (Yin et al., 2009). RAG1 and RAG2 assemble on either a 12- or 23-RSS as a tetrameric complex (Swanson, 2002) followed by the capture of the alternative RSS type. The cryo-EM structure of RAGs in complex with two RSSs provided the first structural explanation of how the 12/23 rule is enforced. This revealed that both RAG1 monomers are required for the binding of each RSS, as one monomer binds at the CAC motif of the heptamer, whereas the second monomer binds to more distal positions of the RSS (Ru et al., 2015). When a 12-RSS is bound at both the heptamer and nonamer, in an HMGB1-bent conformation, the NBD dimer tilts towards the 12-RSS (Ru et al., 2015). The NBD dimer therefore is only able to synapse with a 23-RSS. Conversely, when a 23-RSS is bound the NBD dimer is titled away from the 23-RSS, which restricts the ability of the complex to synapse with a 23-RSS (Ru et al., 2015).

Following the generation of a stable synaptic complex, analysis of DNA nicking by the RAG complex has indicated that the DNA is unwound at the CAC/GTG base pairs by the RAG complex (Ru et al., 2018). This facilitates the introduction of a single-strand nick between the heptamer-coding sequence boundary. The RAG proteins then catalyse the coupled cleavage of both RSSs by direct *trans*-esterification using the exposed free 3’ hydroxyl group (McBlane et al., 1995), yielding a covalently sealed hairpin at the coding ends, a pair of blunt signal ends and a DNA double strand break (DSB; Gellert, 2002).

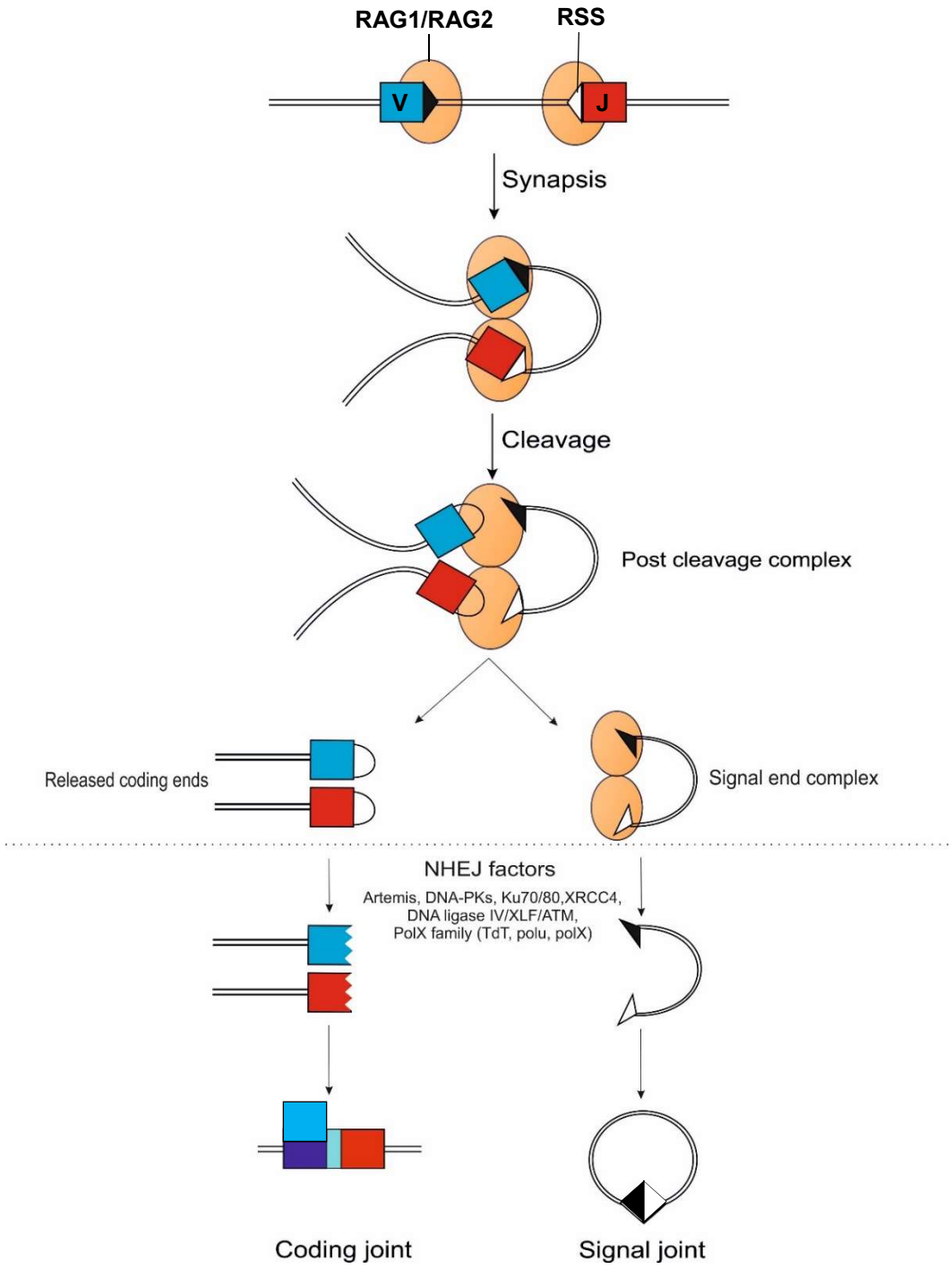


Figure 1.5: Overview of the cleavage and joining phases of V(D)J recombination. The cleavage phase begins with the binding of RAG proteins (orange ovals) to the RSSs (triangles), followed by synapsis with an appropriate partner RSS. Upon synapsis, RAG proteins complete nicking and hairpin formation, producing hairpinned coding ends and blunt signal ends. The signal ends remain bound to the RAG proteins in a signal end complex until repair. Coding and signal ends are repaired by the non-homologous end joining (NHEJ) machinery, but, coding ends are often repaired imprecisely (indicated by the light blue region present between recombined V and J gene segments) to increase the diversity of the variable region.

After cleavage, RAG proteins transiently retain the coding and signal ends in a post cleavage complex (Agrawal and Schatz, 1997; Hiom and Gellert, 1998). Coding ends are transferred to the end joining machinery (Schatz and Swanson, 2011). Signal ends remain tightly bound to the RAG proteins in a signal end complex. Within these complexes, pairs of coding ends and pairs of signal ends are joined to form coding and signal joints, respectively (Figure 1.5). However, as RAG proteins remain bound to the signal ends, downregulation of RAG expression is thought to be required before signal end repair can occur (Jones and Gellert, 2001), thus delaying the formation of signal joints compared to coding joints (Arnal and Roth, 2007).

Repair of coding ends is achieved by the non-homologous end joining (NHEJ) machinery, recruited to the site of recombination by Ku-70/Ku-80 binding. Ku70/Ku80 recruits and activates the catalytic subunit of DNA dependent-Protein Kinase (DNA-PKcs), which in turn recruits X-Ray Repair Cross Complementing 4 (XRCC4), XRCC4-Like Factor (XLF), and the nuclease Artemis to the site of the break (Schatz and Swanson, 2011). The endonuclease activity of Artemis opens the covalently sealed hairpin at a random site. Cleavage away from the apex of the hairpin results in overhangs, which are repaired by DNA polymerase μ or λ resulting in the addition of palindromic (P) nucleotides (Lieber, 2010). Coding junction diversity is further increased by *de novo* addition of N nucleotides by the pro-B/pro-T specific enzyme, terminal deoxynucleotidyl transferase (TdT) and/or the deletion of nucleotides from coding ends by exonucleases. Following end processing, ligation of the coding ends is achieved by the XRCC4/XLF/DNA ligase 4 complex (Lieber, 2010). Signal ends are typically precisely repaired by the NHEJ machinery, resulting in the joining of the RSSs in a head-to-head arrangement, forming an extrachromosomal circle (Gellert, 2002), termed the excised signal circle (ESC).

C) V(D)J recombination is highly ordered

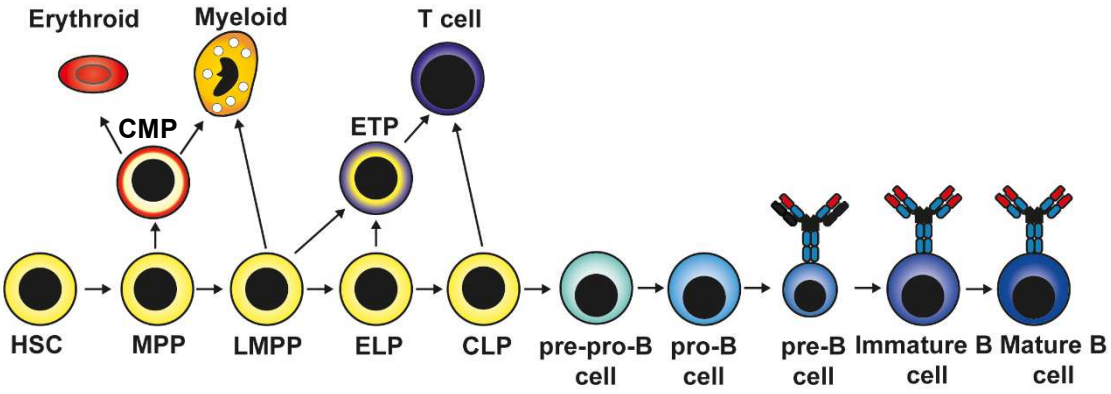
1.7 B cell development

Blood cells are categorised into either cells of the lymphoid (e.g. B and T cells) or myeloid lineage (e.g. megakaryocytes, macrophages and erythrocytes) which are generated from haematopoietic stem cells (HSC) present in the bone marrow. HSC give rise to lymphoid primed multi-potent progenitors (LMPPs) and then to common lymphoid progenitors (CLPs) which give rise to progenitors of the B, T or natural killer cell linages (Vale et al., 2015; Figure 1.6A). The basic helix-loop-helix transcription factor E2A together with interleukin-7 receptor (IL-7R) signalling in CLPs induces the expression of the B lineage specifying transcription factor Early B cell Factor 1 (EBF1), which together with FMS-like tyrosine kinase 3 (FLT3) signalling results in the differentiation of the CLP to the earliest committed B cell developmental stage, pre-pro-B cells (Boller et al., 2018; Zandi et al., 2008). Through the expression of B lineage genes by EBF1, further progression to the pro-B cell stage is permitted. The ability of EBF1 to direct uncommitted progenitors is due to its ability to activate the transcription of B cell specific genes such as *Pax5*, termed the guardian of B cell identity (Nutt et al., 1997), and also repress the drivers of alternative lineages (e.g. C/EBPα) by establishing a transcriptional network enforcing the B cell fate (Boller et al., 2018).

V(D)J recombination is intrinsically linked to B cell development as successful recombination events and the expression of a pre-BCR or BCR are required for the continued survival and proliferation of the B cell progenitor (Rolink et al., 2000; Figure 1.6B). The first rearrangement to occur is the joining of D_H and J_H gene segments at the intermediate pro-B cell stage, followed by V_H to DJ_H recombination at the late pro-B cell stage.

The successful, in frame, recombination of the heavy chain results in the expression of a μ chain which forms a complex with the components of the surrogate light chain (λ5 and VpreB) together with Igα and Igβ (Kitamura et al., 1991; Pelanda et al., 2002; Shimizu et al., 2002). Expression of this complex on the cell surface, termed the pre-B cell receptor (pre-BCR) results in a signalling cascade which promotes the proliferation of the large-pre-B cell. Light chain recombination, at the immunoglobulin κ (*Igκ*) or immunoglobulin λ (*Igλ*) loci, occurs on progression to the small-pre-B cell stage. At these loci, recombination occurs only between V and J gene segments. The protein product of a successful, in frame, light-chain rearrangement pairs with the μ chain and signal transducing proteins (Igα and Igβ) to form the B cell receptor (BCR) which is expressed on the surface of the immature B cell (Vale et al., 2015).

A)



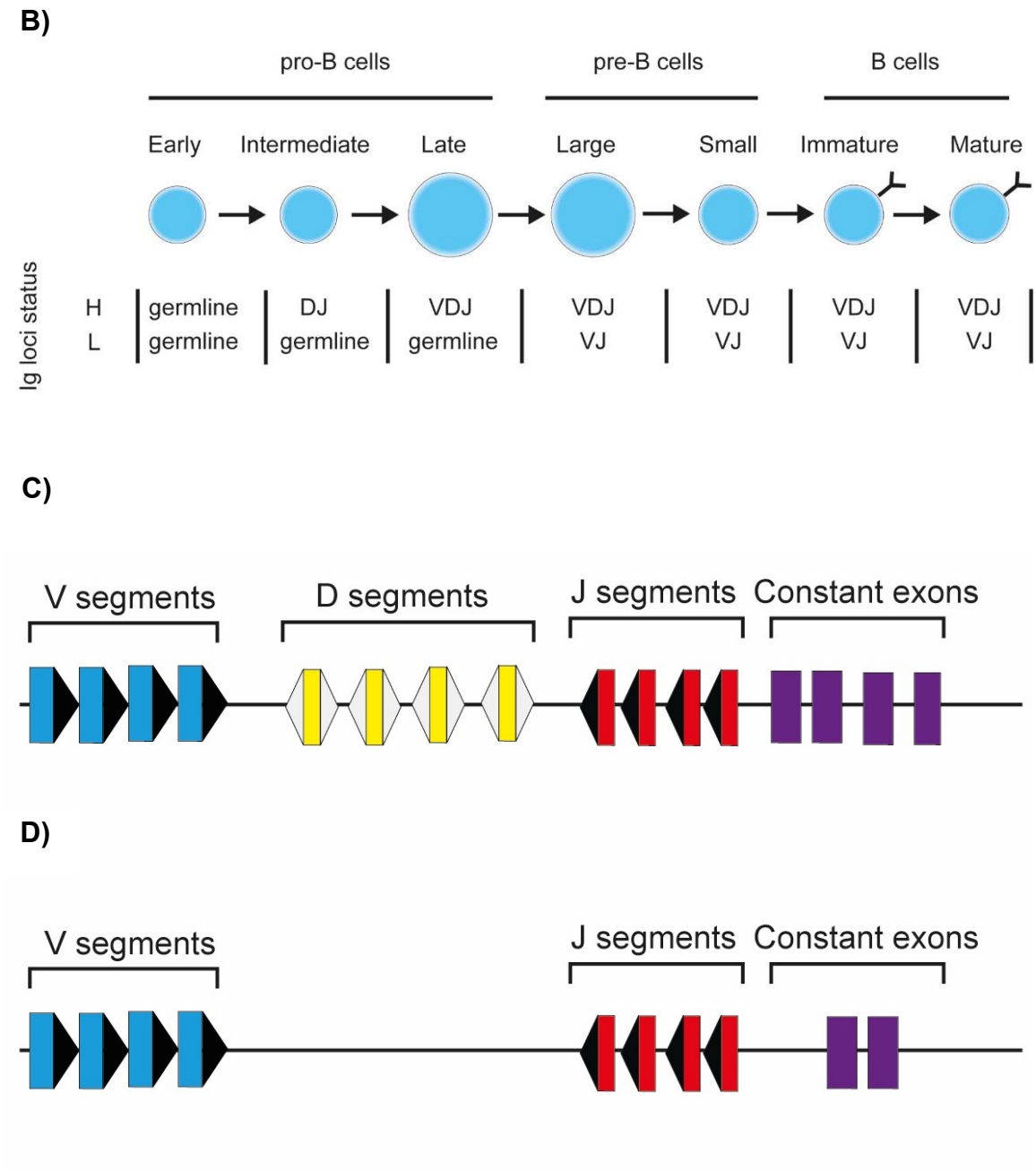


Figure 1.6: Overview of B cell development. A) The progression of cells from hematopoietic stem cells (HSC) is shown together with important branch points (arrows) indicating alternative developmental pathways. B) More detailed overview of B cell development from the pro-B cell stage. The state of the immunoglobulin heavy and light chain is shown at each developmental stage. A simplified overview of the germline configuration an immunoglobulin heavy chain locus (C) and immunoglobulin light chain (D) locus are provided for further clarity. HSC – hematopoietic stem cells, MPP – multipotent progenitors, CMP – Common myeloid progenitor, LMPP lymphoid-primed MPP, CLP – common lymphoid progenitors, ETP – early T lineage progenitors.

1.8 IL-7 signalling

The proliferation and survival of murine B cell progenitors is dependent on the IL-7R (Figure 1.7), composed of the IL-7Rα chain and the γc chain. Mutation of the gene encoding IL-7Rα or IL-7 results in a severe impairment of B cell lymphopoiesis, resulting in a developmental block at the pre-pro-B cell stage (Peschon et al., 1994). The IL-7R possesses no intrinsic kinase activity but instead Janus kinase 3 (JAK3) associates with γc and JAK1 associates with IL7-Rα. Binding of IL-7 results in the phosphorylation of these associated kinases and of Y449 in the IL-7Rα (O'Shea and Plenge, 2012). This serves to recruit signalling transducer and activator of transcription (STAT5A/STAT5B), which activates multiple genes involved in proliferation (e.g. *Ccnd3*), metabolism (e.g. *Glut1*), cell survival (e.g. *Bcl-2*) and represses pro-apoptotic genes (e.g. *Bim*) resulting in the continued proliferation and survival of B cell progenitors (Clark et al., 2014).

IL-7 signalling is essential for the specification of pro-B cells (Singh et al., 2005). Transient IL-7 signalling is believed to promote the expression of EBF1, potentially by regulating a distal promoter of *Ebf1* (Roessler et al., 2007), which in turn upregulates the expression of *Pax5,* resulting in the specification of B-cell fate (Clark et al., 2014). Consistent with this hypothesis, either the ectopic expression of EBF1 or constitutively activated STAT5 is able to restore normal B cell development in $IL7^{-/-}$ mice (Banerjee and Rothman, 1998; Kikuchi et al., 2005).

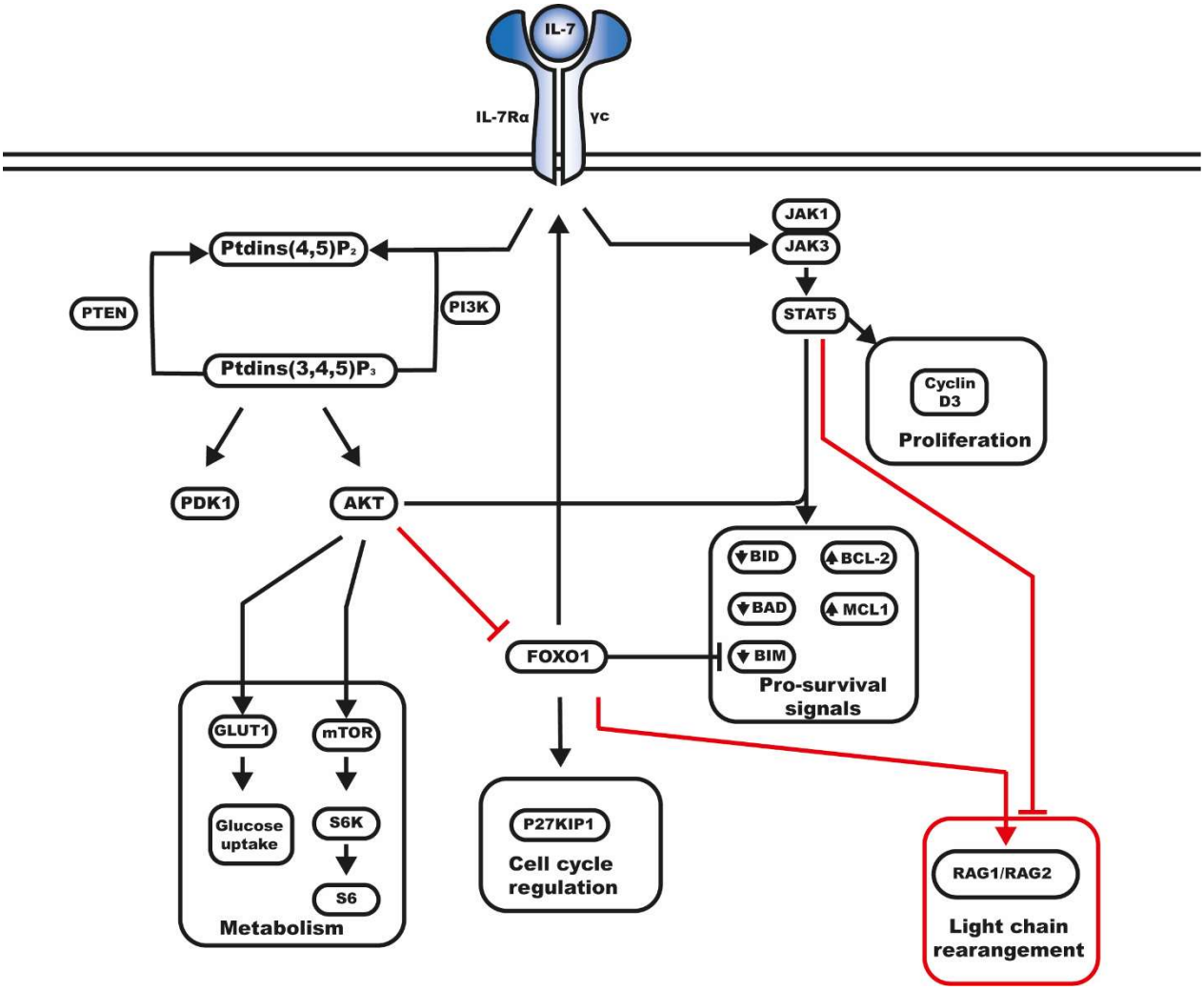


Figure 1.7: Signal transduction pathways activated by IL-7 signalling. IL-7 signalling results in the upregulation of genes involved in metabolism, cell cycle regulation, survival and proliferation. In addition, IL-7 signalling also represses light-chain rearrangement, by downregulation of the *Rag* genes and by decreased light chain accessibility mediated by STAT5.

In addition to regulating B cell specification and proliferation, IL-7 signalling also functions in the developmental stage specific regulation of V(D)J recombination. At the pro-B cell stage, STAT5 binding at the *Igh* locus promotes V_H recombination whereas STAT5 binding to the *Igκ* locus results in the repression of *Igκ* recombination (Bertolino et al., 2005; Johnson et al., 2008). Furthermore, IL-7 signalling through the Phosphoinositide 3-kinase (PI3K) pathway represses the expression of *Rag1* and *Rag2* by the cytoplasmic sequestration of the transcription factor FOXO1 (Amin and Schlissel, 2008), which potentially reduces the aberrant generation of breaks in the rapidly expanding pro-B cell population. *Igh* recombination at the pro-B cell stage is enabled by the oscillation of *Il-7rα* expression. Signalling through the IL-7R results in the downregulation of FOXO1 and as *Il7rα* expression is upregulated by FOXO factors (Clark et al., 2014), IL-7 signalling downregulates the expression of its own receptor, thereby transiently upregulating *Rag* expression to enable heavy chain recombination.

1.9 Pre-BCR signalling

The in-frame rearrangement of the *Igh* locus and its successful expression results in the cell surface expression of a pre-BCR containing Igμ, Igα, Igβ and the surrogate light chain components λ5 and VPREB (Kitamura et al., 1991; Pelanda et al., 2002; Shimizu et al., 2002). The pre-BCR is essential for the continued survival and maturation of B cells and results in the generation of a large pool of precursor B cells that can undergo light-chain rearrangement. Interactions between charged and glycosylated residues on pre-BCR molecules is thought to promote ligand-independent pre-BCR auto-aggregation on the cell surface of pre-B cells (Ohnishi and Melchers, 2003). This aggregation facilitates the phosphorylation of the Igα-Igβ heterodimer by SCR kinases resulting in signal amplification and recruitment of spleen tyrosine kinase (SYK) and ζ-chain-associated protein kinase of 70 kDa (ZAP70) (Rickert, 2013). These tyrosine kinases phosphorylate B cell linker protein (BLINK), which recruits phospholipase Cγ2 (PLCγ2) and Bruton tyrosine kinase (BTK) which promotes the activation of extracellular signal-regulated kinase (ERK) by PLCγ2 binding to RAS (Imamura et al., 2009). This results in the increased proliferation of pre-B cells.

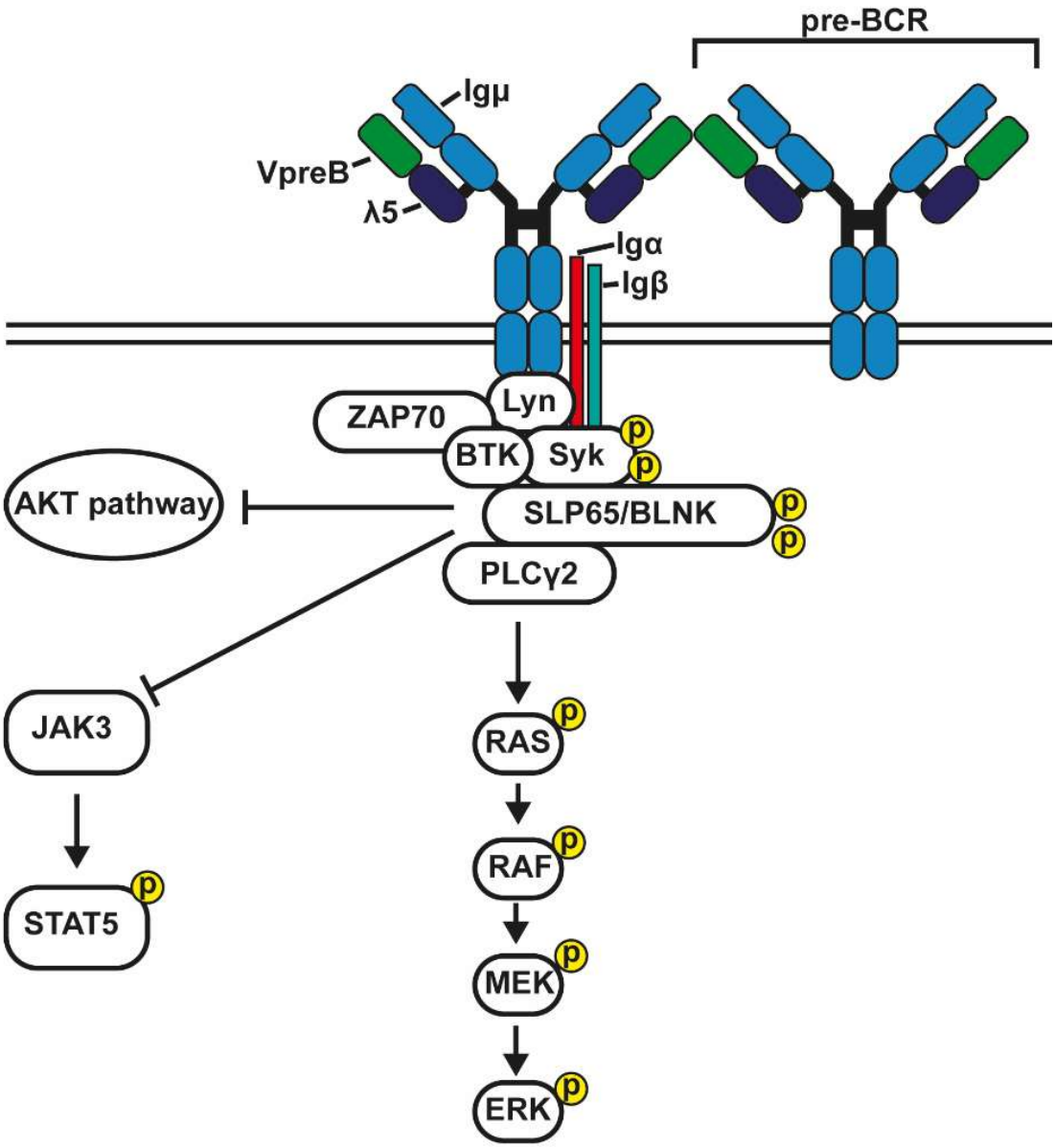


Figure 1.8: Simplified overview of pre-BCR signalling. Pre-BCR aggregation triggers the phosphorylation of the Igα-Igβ heterodimer by SCR kinases resulting in the recruitment of spleen tyrosine kinase (SYK) and ζ-chain-associated protein kinase of 70 kDa (ZAP70) resulting in the phosphorylation and activation of B cell linker protein BLNK and phospholipase Cγ2 (PLCγ2). This results in the downregulation of AKT signalling and inhibition of STAT5 phosphorylation by the inhibition of JAK3. PLCγ2 activates the RAS pathway resulting in the phosphorylation of ERK.

Signalling by the pre-BCR results in the upregulation of pro-survival proteins such as BCL-2, increases cell proliferation by activation of the *Ras-Erk* pathway and interestingly also limits the proliferation of large pre-B cells, due to the actions of BLNK and BTK, to enable further maturation (Rickert, 2013). Notably, pre-B cell clonal expansion *in vivo* is limited to four to five cell divisions (Rolink et al., 2000). The proliferation limit imposed by the pre-BCR is achieved by several mechanisms. Pre-BCR signalling upregulates the transcription factors Ikaros and Aiolos which downregulate components of the surrogate light chain (Ferreirós-Vidal et al., 2013; Heizmann et al., 2013; Ma et al., 2008) resulting in the cessation of proliferative signals. The activation of BLNK by pre-BCR signalling establishes a regulatory network that downregulates *Il7rα* (Clark et al., 2014). Furthermore, the upregulation of C-X-C chemokine receptor type 4 (CXCR4) by pre-BCR signalling leads to the migration of pre-B cells away from IL-7 expressing stromal cells (Johnson et al., 2008). These effects of pre-BCR signalling all contribute to the reduction in proliferative signalling by IL-7.

In addition to mediating the proliferation of pre-B cells, pre-BCR signalling also upregulates numerous transcription factors required for light chain recombination, for example E2A, MEF2C, IRF4 and Ikaros (Heizmann et al., 2013; Herglotz et al., 2015; Johnson et al., 2008; Lazorchak et al., 2006) and removes the repression of light-chain recombination imposed by IL-7 signalling, thus enabling light chain recombination to occur.

D) Regulation of V(D)J recombination

1.10 Regulation of *Rag* expression

As V(D)J recombination involves the production of potentially carcinogenic double stranded breaks, it must be highly regulated to prevent oncogenic events. Regulation takes place at many levels, the first of which is lineage specificity. The recombinase is only expressed in cells of the lymphoid lineage i.e. developing B and T cells as the *Rag2* promoter is activated solely in cells of the lymphoid lineage due to lymphoid specific elements within the promoter region (Kishi et al., 2000; Lauring and Schlissel, 1999). Furthermore, RAG proteins are only expressed at specific stages of lymphocyte development, coinciding with recombination events (Grawunder et al., 1995). Early lymphoid committed progenitors express low levels of RAG1 and RAG2 and exhibit low levels of recombinase activity. Expression of the RAG proteins is increased in pro-B cells, the stage at which immunoglobulin heavy chain assembly begins. Successful heavy chain rearrangement results in *Rag* downregulation. A second wave of *Rag* expression is induced at the non-cycling pre-B stage which is then substantially decreased following successful light chain recombination and the production of a BCR (Grawunder et al., 1995). *Rag* expression can be activated later in development if the BCR recognises a self-antigen, enabling continued rearrangement of the light chains, termed receptor editing (Jankovic et al., 2004).

Control of *Rag* transcription is mainly provided by the *Rag*1/*Rag2* enhancer *Erag*. *Rag* transcription was not detectable in a chromosomally integrated bacterial artificial chromosome (BAC) construct containing the *Rag* locus when *Erag* was absent from the construct (Hsu et al., 2003), implying that *Erag* plays a vital role in activating the *Rag* promoters. The regulation of FOXO1 and FOXP1 binding to *Erag* (Amin and Schlissel, 2008) by IL-7R and pre-BCR signalling appears to play a vital role in restricting *Rag* expression to the correct developmental stages.

In addition to developmental stage specific upregulation and down regulation, the activity of RAG2 is restricted to specific phases of the cell cycle, namely G_0 and G_1. At the G_1 to S phase transition, CDK2/Cyclin E phosphorylates RAG2 at threonine-490 resulting in its degradation via the ubiquitin-proteasome pathway (Li et al., 1996; Zhang et al., 2011). Consequently, as V(D)J recombination relies on the activity of both RAG1 and RAG2, phosphorylation of RAG2 prevents recombination outside the G_0/G_1 phases. When expression of RAG2 is enforced at an inappropriate stage in the cell cycle, aberrant recombination products accumulate (Lee and Desiderio, 1999). Regulation of RAG2 expression therefore ensures that RAG liberated ends are repaired by the NHEJ machinery before re-entry into the cell cycle, preventing the use of other repair pathways e.g. homologous recombination and the introduction of potentially carcinogenic double stranded breaks during periods of rapid proliferation.

Furthermore, several studies have shown that DNA damage can regulate *Rag* expression and activity. DNA damage, signalling through ATM, results in the downregulation of *Rag1* and *Rag2* mRNA (Fisher et al., 2017; Ochodnicka-Mackovicova et al., 2016). Downregulation of *Rag* expression appears to be mediated by the abolishment of FOXO1 binding to the enhancer *Erag,* likely mediated by ATM as treatment with an ATM kinase inhibitor restores FOXO1 binding (Fisher et al., 2017; Ochodnicka-Mackovicova et al., 2016). There is also evidence that phosphorylation of RAG2 at S365 by ATM following RAG mediated cleavage could inhibit the ability of the RAG recombinase to catalyse additional recombination events (Hewitt et al., 2017). Additional *in vivo* studies are required to determine the effect of this modification on the regulation of V(D)J recombination.

1.11 RSS accessibility regulates V(D)J recombination

Although the strict control of RAG expression explains the lymphoid lineage specificity of V(D)J recombination, an additional level of control must exist in order to explain why a specific locus undergoes recombination at a specific developmental stage. The ordered regulation of V(D)J recombination can be best explained by the 'accessibility hypothesis' proposed by Yancopoulos and Alt (1985). This proposes that the gene segments that are not recombining are inaccessible to the recombination machinery and alterations in chromatin structure are required to render the correct gene segments accessible at the appropriate developmental stage. This is evidenced by the fact that expression of RAG1 and RAG2 in fibroblasts is able to recombine extrachromosomal substrates, but not endogenous antigen receptor loci (Schatz et al., 1992) which are inaccessible in fibroblasts.

Genomic DNA is packaged into a highly folded and condensed structure known as chromatin. The base component of chromatin is the nucleosome, which consists of 146 base pairs of DNA stably packed onto the surface of a core histone octamer, containing two of each of the four core histones (H2A, H2B, H3 and H4), by numerous electrostatic interactions and hydrogen bonds (Luger et al., 1997; Figure 1.9). Histone H1 binds DNA at the entry/exit points of the nucleosome, partially protecting and organising an additional 20-50bp of linker DNA (Luger et al., 2012).

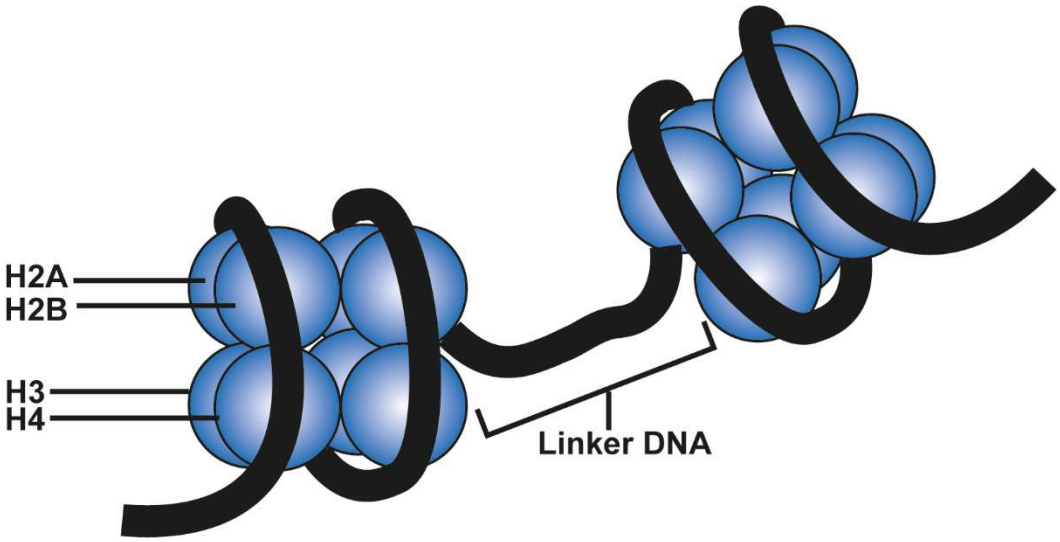


Figure 1.9: Schematic representation of a di-nucleosome. The nucleosome core is composed of a histone octamer two of each of the core histones (H2A, H2B, H3 and H4). The DNA double helix is wrapped around the histone octamer resulting in 146 bp of DNA being tightly associated with the nucleosome leaving 20-60 bp as linker DNA.

The linear polymer of nucleosomes is extensively condensed within interphase chromosomes and is packaged even further in mitotic chromosomes. Nucleosomes separated by linker DNA generate the 10 nm fibre and under physiological salt conditions it has been suggested that the 10 nm fibre forms a helical 30 nm structure that acts as a folding intermediate in the assembly and maintenance of chromosomes. This could potentially be mediated by the interaction of an H4 tail with the H2A/H2B-acidic patch of another nucleosome (Dorigo et al., 2003, 2004; Robinson et al., 2006; Schalch et al., 2005; Song et al., 2014).

Studies of purified or reconstituted chromatin have provided evidence supporting a compacted 30-nm-diameter fibre with a solenoidal (one-start helix; Finch and Klug, 1976; Robinson et al., 2006; Widom and Klug, 1985) or twisted zigzag (two-start helix; Bednar et al., 1998; Dorigo et al., 2004; Horowitz et al., 1994; Schalch et al., 2005; Song et al., 2014) arrangement of nucleosomes. The solenoid model proposes that the linker DNA between nucleosomes is in a bent conformation resulting in a helical path formed with six to eight nucleosomes per turn (Finch and Klug, 1976). In contrast, the zigzag model proposes nucleosomes 'zigzag' back and forth, connected by relatively straight linker DNA.

Studies using cross-linked nucleosome arrays and the low resolution crystallisation of a tetranucleosome have provided support for the zigzag model (Dorigo et al., 2004). In contrast, cyro-EM on *in vitro* assembled chromatin fibres proposed a model whereby each nucleosome contacts the fifth and sixth nucleosome along the nucleosome path, which is consistent with the solenoid model (Robinson et al., 2006). It is also possible that the existence of either model is not mutually exclusive. The analysis of chromatin structure by the combination of cell irradiation with high-throughput sequencing suggested that a solenoid structure is predominant in euchromatin, whereas a zigzag structure is predominant in heterochromatin (Risca et al., 2017).

Several recent studies, however, have failed to identify a defined structure beyond the 10 nm fibre. Super resolution imaging indicated that nucleosome structures are heterogenous and poorly defined *in vivo* (Ricci et al., 2015). Furthermore, the ChromEMT technique has shown that nucleosomes are assembled into disordered 5-24 nm diameter chains with different nucleosome arrangements, densities and conformations (Ou et al., 2017). These studies suggest that the defined 30 nm structure might not be prevalent *in vivo* and that potentially chromosomes are assembled through long range interactions of 10 nm fibres (Bilokapic et al., 2018). Whilst the presence of the 30 nm fibre, *in vivo*, is under some debate, chromatin is believed to be folded further into larger fibres of 100-200 nm to form highly condensed chromatin.

Chromatin architecture has a substantial effect on V(D)J recombination. When packaged in a nucleosome, such that an RSS is occluded by the histone octamer, RAG mediated cleavage is prevented (Golding et al., 1999; Kwon et al., 1998; McBlane and Boyes, 2000). Furthermore, establishment of inactive chromatin suppresses V(D)J recombination (Osipovich et al., 2004). Adding further support to the hypothesis, activating epigenetic modifications such as H3 and H4 acetylation, H3 lysine 4 trimethylation and DNaseI accessibility, directly correlate with V(D)J recombination (Schlissel, 2003).

1.12 Histone acetylation

Histone acetylation is the most well studied histone modification at the antigen receptor loci. Histone acetyltransferases (HATs) and histone deacetylases (HDACs) carry out the acetylation and deacetylation, respectively, of the ε-amino group of specific lysine residues. The presence of the acetyl group neutralises positive charges on the lysine amino group. In the case of nucleosomes, the acetylation of lysine residues reduces the level of nucleosome compaction, which can alter higher level chromatin architecture and lead to functional alterations in gene expression (Eberharter and Becker, 2002).

The importance of histone acetylation in the regulation of V(D)J recombination was first identified by ChIP analysis of the TCRα locus. Acetylation of H3 was found to be tightly linked to V(D)J recombination (McMurry and Krangel, 2000), strongly implying a role for H3 acetylation in the regulation of recombination. Furthermore, the treatment of developing lymphocytes with drugs that globally increase histone acetylation increases the efficiency of V(D)J recombination at endogenous loci (McBlane and Boyes, 2000).

Notably, the repressive effects of the incorporation of an RSS into a mononucleosome can be relieved by histone acetylation and ATP-dependent chromatin remodelling (Kwon et al., 2000; McBlane and Boyes, 2000; Patenge et al., 2004). Furthermore, histone acetylation and the chromatin remodelling factor BRG1 are enriched at recombinationally active antigen receptor loci, compared to inactive antigen receptor loci (Mandal et al., 2015), implying that histone acetylation and chromatin remodelling play a role in regulating RSS accessibility *in vivo*. Histone acetylation appears to facilitate V(D)J recombination by increasing the accessibility of nucleosome remodelling complexes to their targets, providing accessibility to the recombinase (Nightingale et al., 2007). Histone acetylation alone, however, is insufficient for the activation of V(D)J recombination (Bevington and Boyes, 2013) and is unable to overcome nucleosome mediated repression (Bevington and Boyes, 2013; Golding et al., 1999; Hesslein et al., 2003; McBlane and Boyes, 2000).

1.13 Histone methylation

Whilst histone acetylation and chromatin remodelling have roles in the regulation of V(D)J recombination, there are many layers of regulation. When key transcription factors (e.g. Pax5) are deleted, recombination at the *Igh* locus is impaired but histone acetylation is unaffected (Hesslein et al., 2003). This implies that additional regulatory mechanisms must exist. The tail residues of histones can be modified by the addition of methyl groups (one to three depending on the lysine that is modified) conferring active or inactive chromatin states. An important methylation modification is histone 3 lysine 4 trimethylation (H3K4me3) which is associated with transcriptional active chromatin and is critical for efficient V(D)J recombination, *in vivo*.

Studies at the *Igh* and TCRβ loci provided the first evidence of a correlation between H3K4me3 and V(D)J recombination, as peaks of H3K4me3 were identified at the ends of regions actively involved in recombination (Morshead et al., 2003). Further studies at the *Igκ* locus correlated the stage specific initiation of V(D)J recombination with increased levels of H3K4me3, indicating that this modification is likely to be involved in the regulation of V(D)J recombination (Perkins et al., 2004). In addition, the H3K4me3 modification can be bound by the plant homeodomain (PHD) of RAG2 (Liu et al., 2007).

Mutation of critical residues in the PHD finger of RAG2 significantly reduces the level of recombination observed in cell-based assays (Matthews et al., 2007). Notably, patients with mutations in a key tryptophan residue of the RAG2 PHD finger develop primary immunodeficiency (Notarangelo et al., 2016). These studies imply that the binding of RAG2 to regions marked by H3K4me3 is vital for V(D)J recombination. There are two non-exclusive explanations for this observation, firstly, H3K4me3 is present at the 5' end of actively transcribing genes therefore this modification targets the RAG recombinase to accessible regions. In fact, RAG2 binding sites correlate very strongly with sites containing the H3K4me3 modification (Teng et al., 2015), suggesting that this is the major factor regulating the localisation of RAG2. Additionally, RAG2 binding to H3K4me3 through its PHD finger overcomes the autoinhibition of cleavage imposed by RAG2 (Grundy et al., 2010; Shimazaki et al., 2009), thus enhancing recombination.

1.14 Histone ubiquitination

In addition to histone acetylation and methylation, histone tails can be ubiquitinated (Cao and Yan, 2012). The polyubiquitination of a protein marks it for proteasomal degradation, however, monoubiquitination of histones is associated with gene expression and cell signalling processes (Baarends et al., 1999). Histone H2A, H2B H3 and H1 have all been found to be ubiquitinated, but the ubiquitination of H2A and H2B appears to be the most prevalent and most well studied (Belle and Nijnik, 2014). Remarkably a role for H3 ubiquitination in V(D)J recombination has been identified. The N-terminal region of RAG1 contains a RING (Really Interesting New Gene) domain with E3 ubiquitin ligase activity (Yurchenko et al., 2003). Interestingly, not only can this domain catalyse the autoubiquitination of RAG1 but it can also directly ubiquitinate histone H3 (Grazini et al., 2010). It appears that RAG1 is restrained to chromatin by non-ubiquitinated H3 and released by RAG1-mediated H3 ubiquitination, thereby enabling RSS cleavage to occur (Deng et al., 2015). While it is clear that histone ubiquitination is involved in the regulation of V(D)J recombination, further study is required to investigate the exact effects of this modification.

1.15 Non-coding transcription

In 1985, Yancopoulos and Alt observed that non-coding transcripts were upregulated during V(D)J recombination in the *Igh* locus (Yancopoulos and Alt, 1985). These transcripts were later found to correlate with the initiation of recombination in the *Igκ* locus (Schlissel and Baltimore, 1989). In fact, non-coding transcripts, also known as sterile or germline transcripts that initiate at V, D and J segments have been found to developmentally coincide with the activation of V(D)J recombination at each antigen receptor locus (Stubbington and Corcoran, 2013).

The essential role of non-coding transcription in V(D)J recombination was demonstrated by the Krangel group. In the *TCRα* locus, non-coding transcription is initiated by the T early α (TEA) promoter. By the insertion of a transcriptional terminator downstream of the TEA promoter Vα to Jα recombination was suppressed. Interestingly, a decrease in acetylation and trimethylation of H3K4 and H3K36 was also observed, which suggested that non-coding transcription may regulate these modifications (Abarrategui and Krangel, 2006).

Notably, histone acetyltransferases and methyltransferases associate and travel with the elongating form of RNA polymerase II (Li et al., 2007), resulting in the deposition of active histone modifications. This implies that transcription itself could remodel chromatin for V(D)J recombination by both trimethylating H3K4, resulting in the targeting of RAG2 to accessible regions and relieving the autoinhibition of cleavage imposed by RAG2 in the absence of H3K4me3. This is in addition to generating a more open chromatin conformation, allowing the RAG recombinase to access to RSSs.

Whilst the presence of histone acetylation and methylation correlate with the onset of V(D)J recombination, histone acetylation and H4K4me3 deposition alone are insufficient for the initiation of V(D)J recombination (Bevington and Boyes, 2013). In order for recombination to occur, the nucleosomes occluding RSSs must be disrupted. Transcription has been shown to transiently evict H2A/2B dimers (Kireeva et al., 2002) which results in the release of 35-40 bp of DNA (Kulaeva et al., 2009). Transcription mediated H2A/H2B eviction increases the accessibility of RSSs for RAG binding and cleavage, both *in vitro* and *in vivo* (Bevington and Boyes, 2013). Interestingly, the transient nature of this event would permit RSS cleavage whilst preventing wide-spread RAG cleavage, that could be detrimental to genomic stability.

1.16 Allelic exclusion

A fascinating aspect of V(D)J recombination is the phenomenon of allelic exclusion, whereby productive antigen receptor rearrangements are limited to a single allele. The homologous allele either remains in its germline configuration, has an incomplete D-J rearrangement (in the cases of *Igh* or *Tcrβ*) or carries a non-productive rearrangement. This ensures that each lymphoid cell expresses a single antigen receptor with a defined ligand binding specificity (Outters et al., 2015). The mechanistic basis of allelic exclusion is very poorly understood, however, there are several models to explain how allelic exclusion potentially occurs. It has been heavily debated whether allelic exclusion occurs in a stochastic or deterministic manner (Outters et al., 2015). Stochastic models emphasise mechanisms that decrease the efficiency of recombination making it unlikely that recombination can occur on both alleles. In contrast, deterministic models suggest that the chromosomes are somehow marked, generating substrates that are not functionally equivalent.

Stochastic models

One of the earliest models developed was the “pure” stochastic model which implies that the generation of a productive V(D)J exon on both alleles is so infrequent that allelic exclusion is a consequence of this low probability, given that one out of three V(D)J recombination events are in the correct reading frame (Coleclough et al., 1981). As this model would allow ~20% of cells to rearrange both alleles (Wabl and Steinberg, 1992), far more than the ~1% observed *in vivo* (Barreto and Cumano, 2000), this model cannot accurately explain how allelic exclusion occurs.

The probabilistic model suggests that the low efficiency of recombination results in asynchronous recombination of each allele. As approximately 5% of *Igκ* alleles are activated to a high level in pre-B cells, the probability of rearrangement on both alleles would be less than 0.25% which is consistent with *in vivo* observations (Barreto and Cumano, 2000; Liang et al., 2004). It is however possible the extent of *Igκ* activation has been underestimated (Taylor et al., 2009). The degree of *Igκ* locus activation was assessed by Liang et al 2004 using a GFP reporter inserted into the *Igκ* locus. Taylor et al 2009 observed a disparity between GFP protein expression and non-coding transcripts, which can in part be explained by alternative promoter utilisation, resulting in the production of non-functional GFP (Taylor et al., 2009). This study therefore casts doubt upon the findings of Liang et al 2004.

Strong evidence for the probabilistic models has been provided by a study of the *Tcrβ* locus (Schlimgen et al., 2008). Both *Tcrβ* alleles were shown to interact frequently with pericentromic heterochromatin and the nuclear lamina which are known to be repressive compartments (Deniaud and Bickmore, 2009). Following *Tcrβ* recombination, the recombined allele was positioned away from these repressive compartments. Interestingly, the insertion of an ectopic enhancer, Eα, into one Tcrβ allele reduced the association of allele with repressive compartments. Furthermore, this also resulted in an increased number of mature T cells containing two rearranged alleles (51.2% vs 38.5%; Schlimgen et al., 2008). This suggests that allelic exclusion may be achieved by asynchronous recombination of TCRB alleles in a stochastic rather than deterministic manner.

Deterministic models

Deterministic models suggest that the allele to be recombined is marked during development. The asynchronous replication model suggests that asynchrony in allele replication is established during embryonic development. This marks one allele resulting in it being more accessible for recombination (Mostoslavsky et al., 2001).

Asynchronous replication of immunoglobulin alleles has been observed by several studies (Farago et al., 2012; Mostoslavsky et al., 2001; Outters et al., 2015) and in most cases this appears to result in the early-replicating allele being preferentially demethylated (Goldmit et al., 2002; Mostoslavsky et al., 1999, 2001) and recombined, as the later replicating allele is recruited to pericentric heterochromatin before recombination. How asynchrony is established is poorly understood, but it is clear that genomic imprinting is not involved in this process (Gebert et al., 2017). Moreover, the developmental stage at which asynchrony is established has been debated. Farago et al. 2012 showed that asynchrony is established stochastically in the early embryo whereas Khor and Sleckman 2005 suggested that chromosome choice occurs later in lymphocyte development (Farago et al., 2012; Khor and Sleckman, 2005). Additional studies are therefore required to further explore the feasibility of this model.

E) Regulation of the antigen receptor loci

Whilst the control of non-coding transcription provides a mechanism for the regulation of V(D)J recombination, the regulation of non-coding transcription at the antigen receptor loci is highly complex and not fully understood. The *Igh* and *Igκ* loci are the best studied antigen receptor loci and their regulation together with that of the *Igλ* locus will be discussed in the following sections.

1.17 Structure of the immunoglobulin heavy chain locus

The murine *Igh* locus, chr12:113258768-116009954, spans 2.8 Mb and consists of 110 functional V_H gene segments, eight to twelve D_H segments (dependent on the mouse strain), four J_H segments (Ye, 2004) and eight constant regions, (Figure 1.10).

The first regulatory element to be identified in the *Igh* locus was the heavy chain enhancer, Eμ, located in the intergenic region between the J_H gene segments and the constant region exons (Gillies et al., 1983; Figure 1.10). This enhancer is vital for *Igh* recombination as the replacement of Eμ with an oligonucleotide (Serwe and Sablitzky, 1993) or the neomycin resistance gene (Chen et al., 1993) reduced the level of *Igh* non-coding transcription and impaired V_H-DJ_H recombination. The Eμ enhancer also regulates D_H-J_H recombination, but the absence of Eμ is not sufficient to completely prevent this rearrangement event, implying additional regulatory elements are present. Indeed, two additional regulatory elements have been identified at the 3' end of the locus, belonging to the regulatory region known as the 3' regulatory region (3'RR; Figure 1.10). Whilst this region contains several strong enhancers (Matthias and Baltimore, 1993; Pettersson et al., 1990), the 3'RR is not required for V(D)J recombination but appears to be essential for class switch recombination (CSR; Cogné et al., 1994) and somatic hypermutation (SHM; Rouaud et al., 2013).

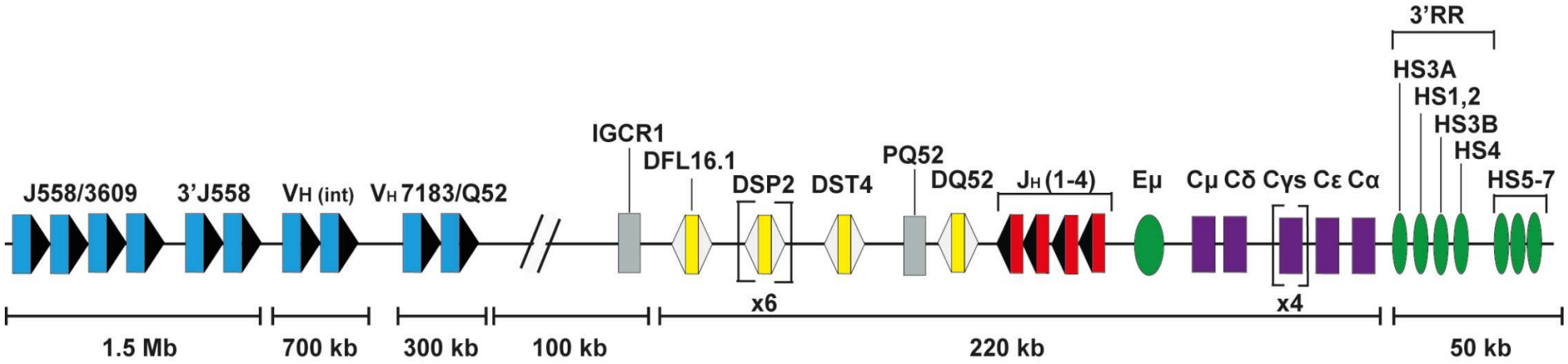


Figure 1.10: The murine immunoglobulin heavy chain locus. A simplified schematic of the murine *Igh* locus. The locus contains 110 functional V_H gene segments (blue), eight to twelve D_H gene segments (yellow), four J_H gene segments (red) and eight constant region exons (purple). Recombination sequences are represented by black (23-RSS) or white (12-RSS) triangles. The CTCF binding elements Intergenic control region 1 (IGCR1) and promoter of non-coding transcription (PQ52) are indicated (grey). Regulatory elements including the Eμ enhancer are shown in green. 3'RR – 3' regulatory region; HS – hypersensitive site.

In addition to enhancer elements, the *Igh* locus also contains several regions with multiple CTCF binding sites, IGCR1 and HS5-7, which demonstrate insulator activity and also appear to be involved in the regulation of long range interactions at the locus. CTCF binding sites play a crucial role in the regulation of *Igh* recombination (Guo et al., 2011a) and analogous CTCF binding sites have been identified in the *TCRβ* and *TCRδ* loci between V and J gene segments. The deletion of these sites alters the frequency of V gene segment utilisation at these loci (Chen et al., 2015; Majumder et al., 2015). Furthermore, the introduction of new CTCF binding sites into the *TCRβ* locus reduces rearrangements by varying degrees depending on its location (Chen et al., 2016; Rawat et al., 2017). These studies imply that CTCF binding to the appropriate location is essential for efficient locus folding at the antigen receptor loci.

1.18 Structure of the immunoglobulin κ locus

The murine *Igκ* locus, chr6:67555636-70726754, spans approximately 3.2 Mb and consists of 160 Vκ gene segments, of which approximately 100 are functional. The locus also contains five Jκ segments and a single constant region exon (Figure 1.11). Notably, unlike the *Igh* locus, Vκ gene segments are in forward and reverse orientations (Zachau, 1993), thus enabling deletional and inversional recombination events to occur.

Recombination of *Igκ* appears to mainly be regulated by two enhancer elements, iEκ and 3'Eκ. The first enhancer element to be identified was iEκ (Queen and Baltimore, 1983) and was found to be required for the promotion of *Igκ* recombination as its deletion reduced the level of *Igκ* non-coding transcription and *Igκ* recombination (Inlay et al., 2002). The replacement of this enhancer with the *Igh* enhancer Eμ resulted in the premature activation of *Igκ* at the pro-B cell stage. Notably however, the Eμ enhancer was unable to activate *Igκ* at the pre-B cell stage (Inlay et al., 2006). This implies that iEκ plays a vital role in restricting *Igκ* recombination to the pre-B cell stage.

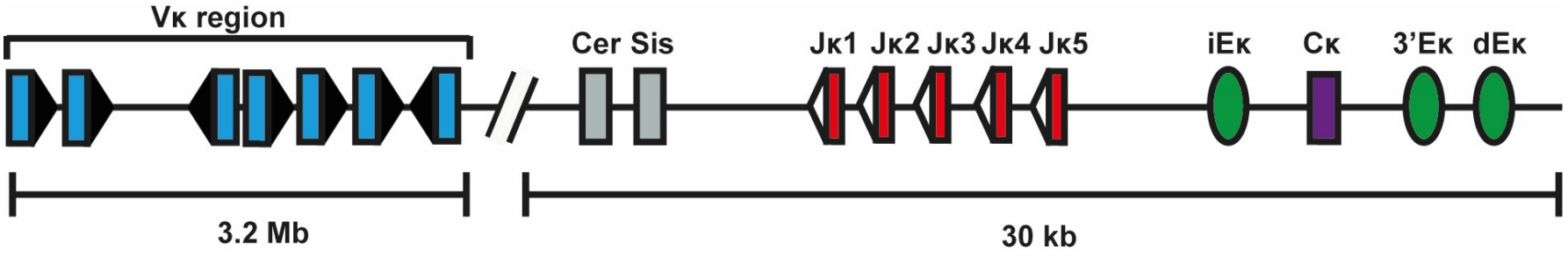


Figure 1.11: The murine immunoglobulin kappa chain locus. A simplified schematic of the murine *Igκ* locus. The locus contains approximately 100 functional V_H gene segments (blue), five J_H gene segments (red) and a single constant region (purple). Recombination sequences are represented by black (23-RSS) or white (12-RSS) triangles. The CTCF binding elements contracting element for recombination (Cer) and silencer in the intervening sequence (Sis) are indicated (grey). Enhancer elements are shown in green.

The second enhancer, 3'Eκ, located approximately 9 kb downstream of the constant exon, was identified in the plasmacytoma cell line S107 (Atchison and Perry, 1987) which is defective in the induction of NF-κB expression (Baumann et al., 1998). Notably this enhancer appears to enhance β-globin transcription approximately seven-fold higher than iEκ (Meyer and Neuberger, 1989) suggesting that 3'Eκ is the stronger enhancer. 3'Eκ knockout mice exhibit a similar phenotype to that of iEκ mutant mice (Gorman et al., 1996). Whilst there is redundancy between the two enhancers, *Igκ* recombination is almost completely repressed when both enhancers are removed (Gorman et al., 1996; Inlay et al., 2002; Xu et al., 1996) implying that both enhancers are vital for efficient V(D)J recombination. Notably, unlike iEκ, the 3'Eκ enhancer is required for the expression of mature *Igκ* transcripts (Gorman et al., 1996; Inlay et al., 2006).

Recent studies have identified a third more distal *Igκ* enhancer, dEκ (Liu et al., 2002). The deletion of dEκ has little impact on *Igκ* recombination (Xiang and Garrard, 2008), however, the removal of both 3'Eκ and dEκ abolishes *Igκ* recombination and furthermore, compensation for the loss of 3'Eκ can be observed from dEκ (Zhou et al., 2010). Whilst this suggests some involvement of dEκ in *Igκ* recombination, it is more likely that the enhancer is predominantly involved in enhancing the transcription of the *Igκ* light chain (Zhou et al., 2010) and somatic hypermutation (Xiang and Garrard, 2008).

In addition to the well characterised three *Igκ* enhancers discussed, several other enhancer elements have been identified, which have been excluded from Figure 1.11. A conserved region of open chromatin (HS10; Zhou et al., 2012) has been observed approximately 2 kb downstream of the *Igκ* flanking gene ribose 5-phosphate isomerase (*Rpia*). Deletion of this element did not impact *Igκ* recombination and appears to mainly function to promote maximal expression of mature *Igκ* transcripts, in addition to a minor role in SHM (Zhou et al., 2012). Unbiased chromatin profiling has also indicated the presence of an additional enhancer like element within the *Vκ* region, κRE1, which displays enhancer activity in a reporter assay (Predeus et al., 2014). *In vivo* evidence regarding the involvement of this enhancer in *Igκ* recombination is, however, lacking.

Two CTCF binding elements, contracting element for recombination (Cer) and silencer in the intervening sequence (Sis) play crucial roles in the regulation of correct *Igκ* recombination by altering the three-dimensional structure of the locus. The mechanism by which these elements regulate *Igκ* recombination will be discussed in Section 1.24.

1.19 Structure of the immunoglobulin λ locus

The murine *Igλ* locus spans 230 kb, chr16:19026858-19260844, and contains only six functional V and J gene segments, three J gene segments and three V gene segments (Figure 1.2). In addition, each J gene segment is associated with its own constant exon (Eisen and Reilly, 1985). The *Igλ* locus in mice appears to have arisen from an evolutionary duplication event (Hayzer, 1990), giving rise to two recombination domains: One domain consists of *Vλ2, Vλx, Jλ4* and *Cλ4* in addition to the $Eλ_{2-4}$ enhancer whilst the other domain contains *Vλ1, Jλ1, Jλ3, Cλ1* and *Cλ3* together with the $Eλ_{3-1}$ enhancer. These two domains appear to be independent as recombination is usually observed between the gene segments within domains (Sanchez et al., 1991). Interestingly, there is a significant bias in recombination frequency between *Vλ1* and Jλ1. *Vλ1-Jλ1* recombination can be observed at approximately 70% of *Igλ* rearranged alleles but the mechanism behind this bias has not been well studied.

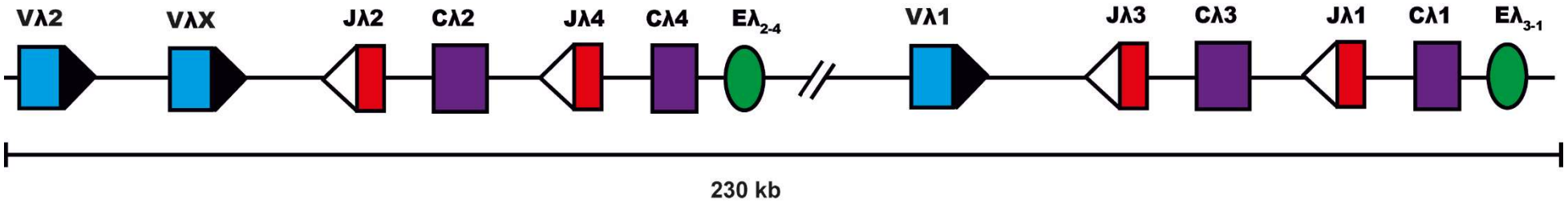


Figure 1.12: A simplified overview of the murine immunoglobulin lambda locus. A simplified schematic of the murine *Igλ* locus. The locus contains three functional V gene segments (blue), four J_H gene segments (red) and a single constant region (purple). Recombination signal sequences are represented by black (23-RSS) or white (12-RSS) triangles. Enhancer elements are shown in green.

Early studies investigating the regulation of the murine *Igλ* locus identified two strong transcriptional enhancers ($E\lambda_{3\text{-}1}$ and $E\lambda_{2\text{-}4}$) which share 90% sequence homology. The $E\lambda_{3\text{-}1}$ enhancer appears to be essential for *Vλ1-Jλ1* recombination, as the deletion of the enhancer within a BAC construct containing the 3' half of the *Igλ* locus dramatically reduced the level of recombination observed in transgenic mice (Haque et al., 2013). DNaseI footprinting identified two key binding domains, λA and λB, within these enhancers which appear to be essential for enhancer activity (Eisenbeis et al., 1993; Rudin and Storb, 1992).

The λA domain contains MEF2C and E2A binding sites (Satyaraj and Storb, 1998). The λB domain plays a key role in the developmental stage specific activation of these enhancers. This domain contains binding sites for the transcription factors IRF4 and PU.1 (Eisenbeis et al., 1993, 1995). Notably, loss of IRF4 prevents *Igλ* and *Igκ* recombination (Lu et al., 2003) and its re-introduction is sufficient to enable *Igλ* recombination (Johnson et al., 2008). Additionally, the level of IRF4 increases from the pro-B to pre-B stage of development (Muljo and Schlissel, 2003) correlating with a 10 fold increase in *Igλ* recombination. Remarkably, transgenic mice (PIP3; Bevington and Boyes, 2013) in which the pro-B specific promoter and locus control region from the λ5 locus (Sabbattini et al., 1999) were used to drive the expression of an IRF4 transgene, resulting in a pre-B cell level of IRF4 at the pro-B cell stage is sufficient to upregulate non-coding transcription and trigger premature *Igλ* recombination (Bevington and Boyes, 2013).

1.20 Subnuclear positioning

Early fluorescence in-situ hybridisation (FISH) studies indicated that the *Igh* and *Igκ* loci are positioned at the nuclear periphery in hematopoietic progenitor cells and pro-T cells (Kosak et al., 2002). The *Igh* locus is anchored at the nuclear periphery by the distal V_H genes with the proximal *Igh* domain orientated towards the centre of the nucleus, facilitating D_H-J_H rearrangements (Fuxa et al., 2004). Both loci are then repositioned to a more central location in pro-B cells coinciding with increased chromatin accessibility and non-coding transcription of the V_H genes in preparation for V_H-DJ_H recombination (Kosak et al., 2002). Examination of the subnuclear localisation of the *Igh* and *Igκ* in *Rag2*$^{-/-}$ pro-B cells suggests that the repositioning of the loci occurs prior to the onset of recombination. The nuclear periphery is a repressive compartment that is important for the propagation of inactive genes (Deniaud and Bickmore, 2009), implying that the localisation of these loci to the nuclear periphery is responsible, at least in part, for preventing V(D)J recombination. Repositioning of the loci is likely to be one of the first stages in the activation of the antigen receptor loci, however, as repositioning of the *Igκ* locus occurs at the pro-B cell stage, before recombination occurs, subnuclear localisation is not sufficient for locus activation.

1.21 Locus contraction

In the *Igh* and *Igκ* loci, the V gene segments are spread over a large genomic region (2.5 – 3.2 Mb). To facilitate V(D)J recombination, the immunoglobulin loci undergo locus compaction (Kosak et al., 2002). Immunoglobulin locus contraction was first observed at the *Igh* locus, when the three-dimensional structure of the locus was examined by 3D-FISH with probes specific for both ends of the locus (Fuxa et al., 2004). These studies revealed that in non-B cells and lymphoid progenitors the *Igh* locus is present in an extended conformation, evidenced by the relatively large separation of the two FISH probes. In contrast, in the nuclei of pro-B cells, the locus is contracted, indicating that contraction occurs at the developmental stage at which rearrangements occur (Fuxa et al., 2004). Analysis of *Igκ* locus by Hi-C and 3C-seq studies have found interactions between iEκ or 3'Eκ and the distal *Vκ* region in pro-B cells implying that the locus is contracted at the pro-B cell stage (Lin et al., 2012; Ribeiro de Almeida et al., 2011; Stadhouders et al., 2014) and that locus contraction occurs before the induction of recombination (Fitzsimmons et al., 2007). Locus contraction appears to be achieved by long-range interactions between proximal and distal regions of the immunoglobulin loci, a process that brings distal V gene segments into close proximity with the D or J regions, enabling efficient V(D)J recombination. The mechanisms by which locus contraction is achieved are discussed in Sections 1.23 and 1.24.

1.22 Transcription factors regulating V(D)J recombination

The driving force behind the complex cascade of events culminating in V(D)J recombination at the antigen receptor loci is a number of developmental stage and tissue-specific transcription factors. These key factors include: EBF1, PAX5, E2A, PU.1, IRF4, Ikaros, STAT5, CTCF and YY1.

EBF1

EBF1 is a pivotal factor determining B-cell lineage fate, together with E2A and PAX5. EBF1-deficient mice display a complete block in B lymphopoiesis at the pre-pro-B cell stage (Lin and Grosschedl, 1995). Notably, EBF1-decificent cells cultured *in vitro* show D_H-J_H but not V_H-D_HJ_H recombination (Pongubala et al., 2008). It is difficult to establish the exact role of EBF1 in V(D)J recombination due to the impairment of lymphopoiesis in *Ebf1*$^{-/-}$ mice and altered expression of factors involved in V(D)J recombination such as RAG1 and RAG2 (Zandi et al., 2008). Expression of EBF1 in non-lymphoid cells, together with RAG1/RAG2 can induce *Igκ* rearrangements (Goebel et al., 2001) and this is most likely mediated by the direct binding of EBF1 to the Vκ region. Interestingly, EBF1-binding does not appear to induce non-coding transcription at the Vκ region but instead appears to act by promoting chromatin remodelling (Agalioti et al., 2000; Cote et al., 1998; Zhao et al., 2003). This suggests that EBF1 is involved in immunoglobulin locus activation but is unlikely to directly regulate this process.

E2A

E2A proteins are comprised of two basic helix-loop-helix transcription factors, E12 and E47, which differ in their DNA-binding domain. Both E12 and E47 are produced through alternative splicing of the *Tcfe2a* gene. Similar to *Ebf1*$^{-/-}$ mice, *Tcfe2a*$^{-/-}$ mice are blocked at the pre-pro-B cell stage. E2A is a key regulator of *Igκ* recombination and was originally identified as a transcription factor binding to iEκ (Murre et al., 1989). As with EBF1, the ectopic expression of E2A, together with RAG1 and RAG2, is sufficient to facilitate *Igh* and *Igκ* recombination in an embryonic kidney cell line (Romanow et al., 2000). Moreover, *Tcfe2a*$^{-/-}$ knockout mice are unable to undergo *Igκ* recombination (Goebel et al., 2001; Lazorchak et al., 2006) and the reintroduction of E47 is sufficient to upregulate *Igκ* non-coding transcription (Lazorchak et al., 2006). Binding of E2A to the *Igh* enhancer Eμ and the three enhancer elements present in *Igκ* has been well established and notably, targeted deletion of the E2A binding sites in iEκ impairs *Igκ* recombination to the same extent as the removal of the entire element (Inlay et al., 2004). Furthermore, E2A proteins have been observed to bind close to Vκ gene segments, with the binding of E2A proteins correlating with the segment utilisation. E2A is likely to mediate the upregulation of non-coding transcription by two main mechanisms: the regulation of chromatin architecture by recruitment of HAT complexes (Beck et al., 2009; Lazorchak et al., 2006; Sakamoto et al., 2012) and by recruitment to 3'Ek mediated by IRF4 (Lazorchak et al., 2006; Nagulapalli and Atchison, 1998).

PAX5

Paired box protein, PAX5, is a vital transcription factor in B cell development (Adams et al., 1992; Barberis et al., 1990; Fuxa and Busslinger, 2007). At the *Igh* locus, PAX5-deficient pro-B cells exhibited an approximately 100-fold decrease in recombination efficiency of the distal V_H558 gene segments, despite normal D-J_H recombination (Nutt et al., 1997). Analysis of the *Igh* locus in *Pax5*$^{-/-}$ pro-B cells by 3D-FISH showed that the locus is in an extended conformation, preventing the distal V_H genes from recombining with the proximal D_HJ_H domain (Fuxa et al., 2004). Fourteen PAX5-activated intergenic repeat regions (PAIR) elements bound by PAX5, E2A, YY1 and CTCF/cohesin have been identified by the analysis of active chromatin marks (Ebert et al., 2011). These elements not only give rise to PAX5 regulated non-coding transcripts but also appear to be involved in the mediation of long-range interactions resulting in the looping of distal V_H gene segments for V_H-DJ_H recombination (Ebert et al., 2011; Medvedovic et al., 2013a). PAX5 has also been implicated in the removal of the repressive H3K9me2 modification (Johnson et al., 2004) and the recruitment of RAG1/RAG2 to V_H genes by deposition of H3K4me3 (Zhang et al., 2006).

Despite its extensive role at the *Igh* locus, the role of PAX5 in the regulation of *Igκ* recombination is not as well understood. Binding sites for PAX5 have been identified in 3'Eκ (Roque et al., 1996) and PAX5 appears to be necessary for the activation of the *Igκ* locus as PAX5-deficient pre-B cells fail to upregulate *Igκ* non-coding transcription (Sato et al., 2004), but this may be due to defects in pre-BCR signalling.

PU.1

PU.1 is a member of the ETS-family and is required for the development of many hematopoietic cells (Scott et al., 1994). Transgenic mice that contain homozygous mutations of the PU.1 DNA binding domain lack mature B cells (McKercher et al., 1996). In B cells, PU.1 is vital for the regulation of both heavy and light chain recombination as binding sites are present at the heavy chain intronic enhancer (Nelsen et al., 1993; Rivera et al., 1993) and at both the light chain enhancers (Eisenbeis et al., 1995; Pongubala et al., 1992). At all loci PU.1 binding results in increased locus accessibility and together with other transcription factors, it leads to the activation of non-coding transcription (Eisenbeis et al., 1995; Nelsen et al., 1993; Nikolajczyk et al., 1999; Pongubala et al., 1992; Rivera et al., 1993).

IRF4

IRF4 is a lymphoid restricted member of the interferon regulatory factor family of transcription factors (Brass et al., 1996; Eisenbeis et al., 1995). IRF4 plays a critical role at the pre-B cell stage of development. Notably, IRF4 has minimal DNA binding due to the presence of an autoinhibitory domain (Eisenbeis et al., 1995) and requires an interaction with other proteins such as PU.1 to form a stable interaction with DNA (Eisenbeis et al., 1995; Pongubala et al., 1993). *Irf4*$^{-/-}$/*Irf8*$^{-/-}$ knockout mice exhibit a block at the pre-B cell stage due to a defect in light chain recombination (Lu et al., 2003) and binding sites for IRF4 are present at the *Igκ* enhancer 3'Eκ and the *Igλ* enhancers Eλ$_{3\text{-}1}$ and Eλ$_{2\text{-}4}$ (Eisenbeis et al., 1995; Pongubala et al., 1992). As *Irf4* transcription increase at the pro-B to pre-B cell transition (Johnson et al., 2008; Muljo and Schlissel, 2003), IRF4 is likely to play a key role in the regulation of light chain recombination. In support of this hypothesis, IRF4 has been observed to stimulate the deposition of activating histone modifications and activate *Igκ* germline transcription (Johnson et al., 2008; Lazorchak et al., 2006; Ma et al., 2006). Furthermore, IRF4 plays a crucial role in regulating V(D)J recombination at the *Igλ* locus (Bevington and Boyes, 2013).

Ikaros

The transcription factor Ikaros is expressed in almost all murine and human hematopoietic cell types (Georgopoulos, 2002), in fact, in the absence of Ikaros, multipotent progenitors cannot differentiate into common lymphoid progenitors. Interestingly, pro-B cells from Ikaros knockout mice have been shown to fail to upregulate *Rag* expression in addition to exhibiting *Igh* decontraction, decreased *Igh* accessibility and reduced heavy chain recombination (Reynaud et al., 2008). In direct contrast, rearrangement of the *Igh* locus was shown to be unaffected in three conditional *Ikaros* knockout mice (Heizmann et al., 2013; Joshi et al., 2014; Schwickert et al., 2014). These studies suggest that Ikaros does not have a direct regulatory role in *Igh* recombination but is essential for other aspects of B-cell development.

Ikaros-null mice as well as Ikaros dominant-negative mutant mice arrest B cell development at the pre-B cell stage before *Igκ* recombination. This appears to be due to defects in pre-BCR signalling, integrin signalling and cellular adhesion (Joshi et al., 2014). However, a massive recruitment of Ikaros to the *Igκ* locus at the pre-B cell stage correlates with the induction of *Ikaros* expression and the onset of Vκ to Jκ recombination.

Interestingly, Ikaros appears to play both activating and repressive roles in *Igκ* recombination. Ikaros associates with the Sis regulatory element and localises the bound allele to pericentric hetermochromatin which is associated with transcriptional repression (Liu et al., 2006). On the other hand, rescue experiments have shown that Ikaros can induce *Igκ* non-coding transcripts within four hours (Heizmann et al., 2013), suggesting a crucial role of Ikaros in *Igκ* recombination.

STAT5

Signal transducer and activator of transcription (STAT)5A and STAT5B, referred to as STAT5, control the survival of pro-B cells (Malin et al., 2010). STAT5 is phosphorylated in pro-B cells upon signalling via IL-7R which results in the formation of STAT5 dimers. Activation of STAT5 in pro-B cells inhibits *Igκ* non-coding transcription and interferes with E2A binding to iEκ (Mandal et al., 2011) by recruiting the polycomb protein enhancer of zeste homolog 2 (EZH2) to this region which results in the deposition of the repressive histone modification H3K27me3. Furthermore, STAT5 is also capable of displacing PU.1 at 3'Eκ (Hodawadekar et al., 2012) resulting in the loss of IRF4 occupancy. STAT5 therefore appears to act as a molecular switch that controls transcription factor occupancy at the *Igκ* enhancers. In contrast, STAT5 binding has not been observed at the *Igλ* locus (Bevington and Boyes, 2013), implying that this mechanism is not conserved between the light chain loci.

CTCF

CTCF was discovered as an essential factor for transcription of the *c-myc* gene (Baniahmad et al., 1990; Lobanenkov et al., 1990). In addition to a role in promoting *c-myc* transcription, CTCF binding was also detected at insulator sequences within the chicken β-globin locus (Bell et al., 1999; Furlan-Magaril et al., 2011) and the imprinted *Igf2/H19* locus (Bell and Felsenfeld, 2000). CTCF is able to function as an insulator, blocking the ability of an enhancer to activate a promoter (Recillas-Targa et al., 2002), however, subsequent work has revealed an essential role for CTCF in the mediation of enhancer-promoter interactions (Guo et al., 2015). Together with cohesin, a protein complex involved in sister chromatid cohesion during mitosis (Parelho et al., 2008; Wendt et al., 2008), CTCF forms long-range chromatin loops which alter the topology of the genome. In V(D)J recombination, loops mediated by CTCF and cohesin have been shown to be vital for efficient recombination at the *Igh*, *Igκ* and *Tcrα* loci (Guo et al., 2011; Ribeiro de Almeida et al., 2011; Seitan et al., 2011).

Notably, over 90% of interacting CTCF binding sites are in a convergent orientation (de Wit et al., 2015) implying that CTCF binding site polarity plays a role in directing CTCF mediated chromatin loop formation. The importance of CTCF binding site orientation has also been highlighted by the CRISPR/Cas9 mediated inversion of CTCF binding sites (Guo et al., 2015). The disruption of CTCF binding site orientation is sufficient to abolish chromatin loops between the mutated binding site (Guo et al., 2015), implying that only CTCF sites in a convergent orientation are likely to interact.

YY1

The transcription factor Yin Yang 1 (YY1) is expressed in all mammalian cells, it was first identified as a key regulator for viral genes (Montalvo et al., 1991; Seto et al., 1991), in addition to being a repressor of the *Igκ* enhancer 3'Eκ and *Igh* intronic enhancer (Park et al., 1991). YY1 can activate or repress genes by the recruitment of a wide variety of coactivators and corepressors which recruit activating and repressive histone modifying complexes (Gordon et al., 2006).

Within the B lineage, YY1 plays several critical roles at the pro-B cell stage. The conditional deletion of YY1 results in a developmental block at the pro-B cell stage due to aberrant *Igh* locus contraction and impaired heavy chain recombination (Liu et al., 2007). YY1 also appears to be essential for light chain recombination as it binds to sites within the *Igκ* locus where it potentially plays a role in regulating Vκ segment utilisation (Pan et al., 2013). In addition to mediating transcription and transcriptional repression, YY1 also mediates long-range interactions via interaction with the cohesin complex (Weintraub et al., 2017).

Analysis of pre-B cells expressing a deletion mutant of YY1 without its REPO domain, which facilitates the PcG function of YY1, shows that this mutant protein enables *Igh* recombination to progress, but, interestingly skews recombination at the *Igκ* locus. YY1ΔREPO expressing cells had a limited V*κ* segment utilisation with over 40% of recombination events limited to two *Vκ* segments (Pan et al., 2013). This strongly implies that YY1 is involved in the regulation of *Igκ* recombination, however, the mechanism by which it performs this function is unclear.

1.23 Regulation of *Igh* recombination

D_H to J_H recombination

The recombination of D_H to J_H gene segments in the *Igh* locus is mediated by long-range interactions that result in the formation of four separate domains, Figure 1.13. Chromosome conformation capture on chip studies using pro-B cells first identified an interaction of the *Igh* enhancer Eμ with sequences 5' of DFL16.1 (PQ52) and with HS5-7 in the 3'RR (Guo et al., 2011). The interaction of Eμ with PQ52 results in the formation of an approximately 5 kb chromatin loop, containing the J_H gene segments, whereas the interaction of Eμ with the DNaseI hypersensitive sites HS5-7 generates an approximately 200 kb chromatin loop that contains the constant region exons. The third domain is created via the interaction of the CTCF binding element IGCR1 with Eμ, which produces an approximately 70 kb domain containing the D_H gene segments (Verma-Gaur et al., 2012) and separates the V_H gene segments into a distinct domain.

The three loop domains (Figure 1.13) have distinct functions in regulating *Igh* regulation. The smallest loop, containing the J_H gene segments, exhibits the highest level of RAG1/RAG2 binding observed in the *Igh* locus (Schatz and Ji, 2011), facilitated by high levels of H3K4me3 (Teng et al., 2015). It is therefore likely that the function of this domain is the generation of a chromatin structure to facilitate RAG binding. The deletion of CTCF binding elements within IGCR1 results in the increased frequency of recombination between proximal V_H gene and DQ52 (Guo et al., 2011; Lin et al., 2015). Furthermore, RAG1/RAG2 tracking along DNA appears to be limited to the chromatin loop between Eμ and IGCR1 (Hu et al., 2015). This implies one of the functions of the domain containing the D_H gene segments is the prevention of V_H to D_H recombination prior to D_H to J_H rearrangement. The function of the large (~ 200 kb) loop is unclear as the deletion of HS5-7 only results in a modest decrease in locus contraction (Volpi et al., 2012). However, other CTCF binding sites could potentially compensate for the loss of this region and additional studies are required to identify the function of this domain. When the correct three-dimensional structure and epigenetic modifications (Schatz and Ji, 2011) are present, the close proximity of the D_H loop to the J_H loop, where RAG binding occurs, results in the capture of a D_H RSS, facilitating D_H to J_H recombination.

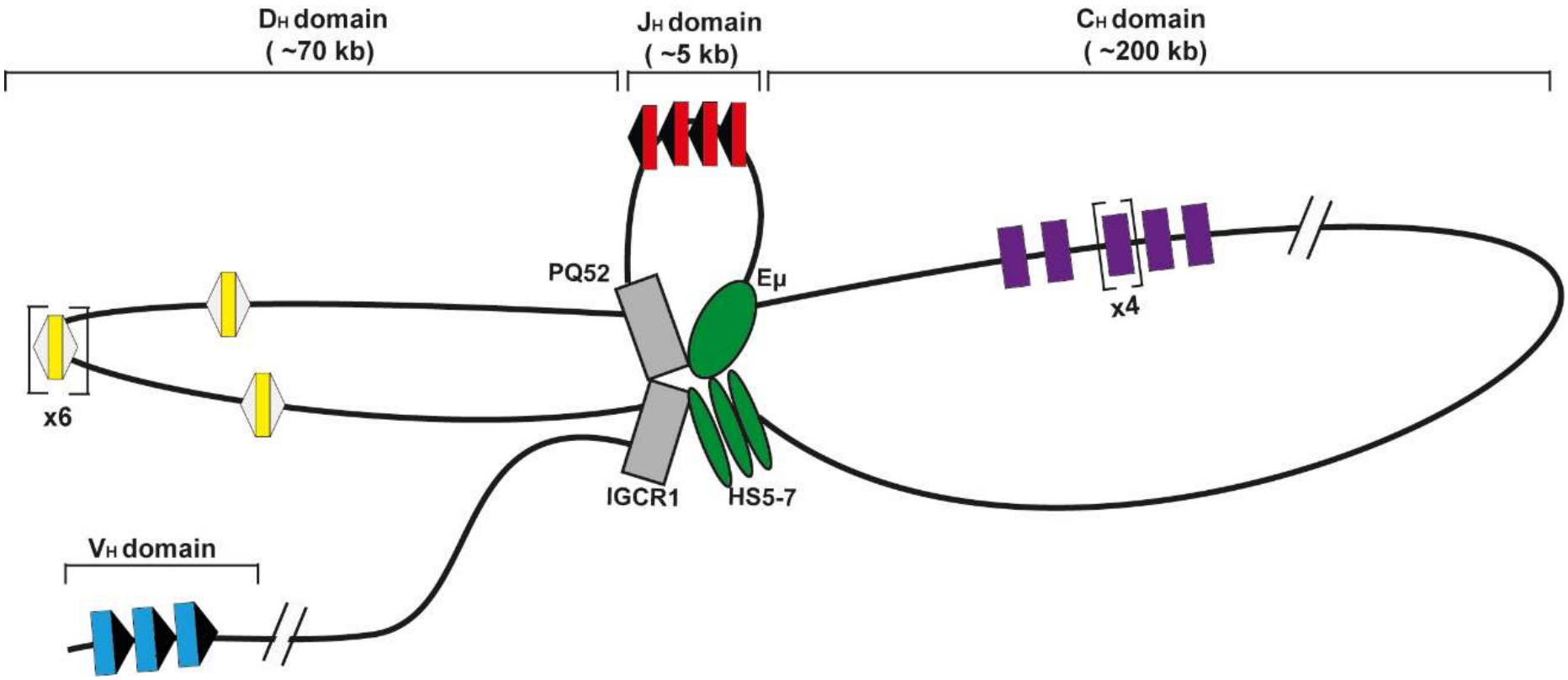


Figure 1.13: Long range interactions mediated by Eμ. Interactions between Eμ and IGCR, PQ52 and HS5-7 result in the formation of three loops domains within the 3' end of the *Igh* locus. The smallest loop (~5 kb) formed by interaction of Eμ with PQ52 contains the J_H gene segments (red). The interaction of Eμ with IGCR1 results in an approximately 70 kb domain containing the D_H gene segments (yellow) and the interaction of Eμ with HS5-7 contains the constant exons (purple). The V_H gene segments are separated from these three domains by Eμ mediated interactions. Adapted from Kumari and Sen 2015.

V_H to DJ_H recombination

Following DJ_H recombination, a series of epigenetic alterations occur at the rearranged DJ_H segment to promote V_H to DJ_H recombination. These include increased histone acetylation and deposition of active histone methylation marks e.g. H3K4me3. The presence of the latter modification is associated with bidirectional non-coding transcription from the rearranged segment, which strongly correlates with the recruitment of the RAG recombinase and appears to be mediated by interaction of the DJ_H promoter with the Eμ enhancer (Puget et al., 2015; Subrahmanyam et al., 2012).

In order for V_H to DJ_H recombination to occur, V_H gene segments must be brought into close proximity to the rearranged DJ_H segment. Furthermore, to ensure a diverse array of antigen receptors each V_H must be given an approximately equal chance of undergoing recombination. The *Igh* locus undergoes large scale compaction to facilitate recombination between distal V_H gene segments and the rearranged DJ_H gene segment. The transcription factor PAX5, together with E2A and CTCF, binds to PAIR elements located within the distal V_H domain. PAX5 binding to PAIR elements is specific to the pro-B cell stage, when locus contraction occurs, and in PAX5 deficient pro-B cells the distal and proximal V_H gene families are spatially dissociated. This strongly implies that PAX5 binding to PAIR elements has a function in the contraction of the V_H domain. In addition to PAX5, CTCF and YY1 have been strongly implicated in mediating *Igh* locus contraction (Guo et al., 2011; Medvedovic et al., 2013).

Interestingly, 4C analysis of the *Igh* locus has indicated the presence of a continuum of chromatin loops across the *Igh* locus, whilst 5C analysis of the *Igh* locus has indicated defined boundaries between the subdomains of the *Igh* locus (Medvedovic et al., 2013; Montefiori et al., 2016). Regardless of the extent of domain definition, there is sufficient evidence to suggest that the V_H domain is folded into several smaller domains. Furthermore, the formation of these domains is likely to be dependent on PAX5 (Montefiori et al., 2016) and not CTCF. A recent analysis of the 121 CTCF binding sites within the V_H region revealed that all bound CTCF sites were orientated towards the D-J-C region (Loguercio et al., 2018). As the majority of interacting CTCF binding sites exist in a convergent orientation (de Wit et al., 2015) this therefore suggests that CTCF/CTCF interactions are unlikely to mediate the contraction of the V_H domain.

Following the contraction of the V_H domain into smaller domains and the 300 kb 3' domain (Figure 1.13), evidence suggests that additional steps are required to fully establish the pre-rearrangement structure of the locus. Alleles where IGCR1 and Eμ are mutated display a marked decrease in distal V_H recombination (Guo et al., 2011). Furthermore, YY1 deficiency results in locus decompaction and preferential recombination of proximal V_H genes, even though the major YY1 binding sites are within the 3' *Igh* domain (Liu et al., 2007; Verma-Gaur et al., 2012). Interactions between V_H3'558 and V_H5'7183 with Eμ, mediated by interactions involving YY1 and CTCF (Benner et al., 2015) with the 3' RR are believed to be responsible for the recruitment of V_H loops to the to the distal end of the locus (Guo et al., 2011; Verma-Gaur et al., 2012). The recruitment of V_H loops to the distal domain results in the activation of the V_H gene segments by interaction with Eμ and also relocates V_H gene segments into closer proximity with the DJ_H segment for recombination.

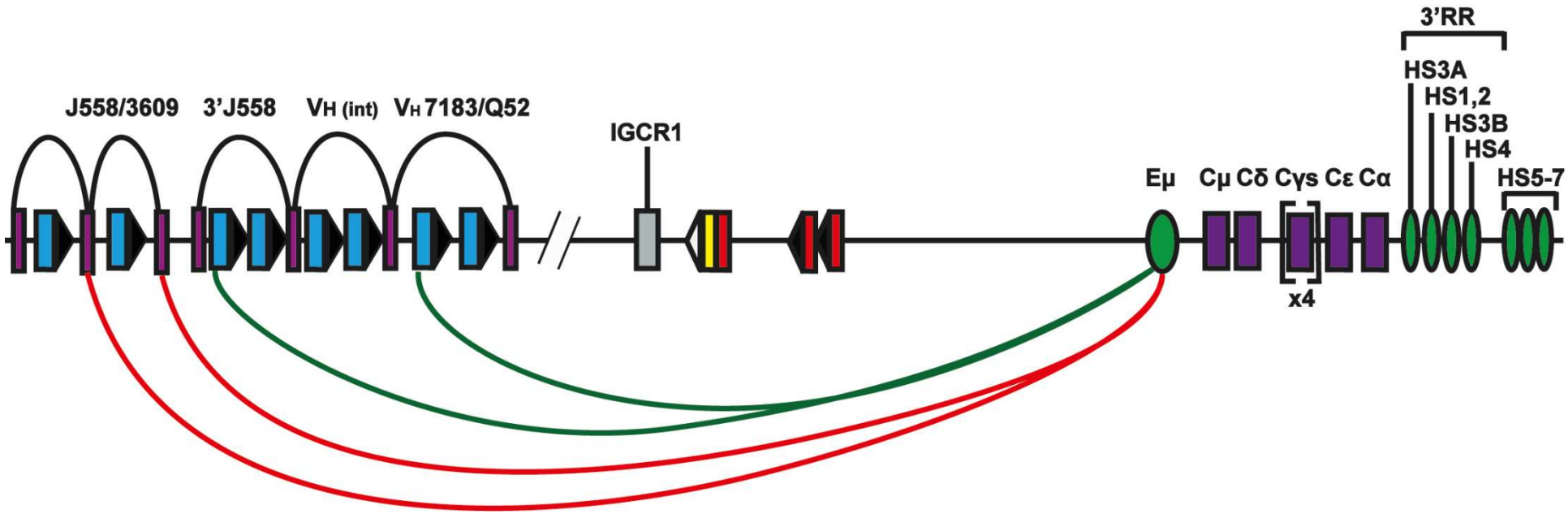


Figure 1.14: Long-range interactions mediating V_H to DJ_H recombination. A simplified schematic of long-range interactions formed at the *Igh* locus. A number of PAX5/YY1 mediated interactions are formed in the V_H region forming a number of dynamic loops (black curves). Locus contraction is mediated by interactions between the V_H region and Eμ. This is facilitated by CTCF/PAX5 mediated interactions between Eμ and PAIR elements (red curves) in addition to YY1 mediated interactions (green curves).

Diversity of V_H gene segment usage is potentially achieved by several mechanisms. CTCF binding appears to be dynamic (Hansen et al., 2017), implying that interactions between CTCF binding site may alter, to enable interactions with alternative V_H domains. Alternatively, different CTCF binding sites could be used in different cells (Levin-Klein et al., 2017). Secondly histone modifications and transcription factor binding may play a role in determining V_H gene segment usage. Interestingly, active histone modifications are enriched at distal V_H gene segments in comparison to proximal V_H genes segments (Matheson et al., 2017). As proximal V_H segments are closer to the DJ_H region these modifications would increase the probability of distal V_H recombination to ensure equal gene segment usage. Furthermore, PAIR elements display promoter activity (Verma-Gaur et al., 2012) dependent on PAX5 binding (Ebert et al., 2011). PAIR elements are activated to different degrees in different cell lines (Verma-Gaur et al., 2012), suggesting that an altered level of transcription originating from PAIR elements could also diversify V_H segment utilisation.

1.24 Regulation of *Igκ* recombination

At the pro-B cell stage, the *Igκ* locus is present in a contracted state (Lin et al., 2012; Stadhouders et al., 2014) likely mediated by CTCF and other factors. YY1 is also likely to play a role in this process as mutation of YY1 has a significant effect on *Vκ* gene segment utilisation (Pan et al., 2013). Moreover, E2A and PU.1 are frequently found at the base of long-range chromatin interaction loops together with CTCF (Lin et al., 2012) implying that multiple factors may play a role in locus contraction.

As with the *Igh* locus, the Vk region contains multiple CTCF binding sites, however, unlike the *Igh* locus these binding sites are not all orientated towards the 3' of the locus. The binding of CTCF strongly correlates with Vk gene segment utilisation (Matheson et al., 2017; Ribeiro de Almeida et al., 2011), suggesting that CTCF plays a role in directing V*κ* gene segment choice. The majority of CTCF sites within the proximal V*κ* region are orientated away from the J gene segments and are therefore not orientated in the correct manner to participate in long range interactions with this region. Conversely, CTCF binding sites within the distal V*κ* region are in the correct orientation to interact with the CTCF binding sites within the Cer and Sis elements (Loguercio et al., 2018) implying that locus contraction is mediated by the interaction of CTCF binding sites within the distal V*κ* region and CTCF binding sites within the Cer and Sis elements. Indeed, the Cer insulator element is crucial for locus contraction (Xiang et al., 2013) and contains two CTCF binding sites, both orientated towards the V*κ* region.

Whilst locus contraction is essential for the activation of the *Igκ* locus, it is not sufficient for the initiation of recombination. An initial increase in *Igκ* locus accessibility is thought to be provided by E2A binding to iE*κ* (Inlay et al., 2004, 2006; Lazorchak et al., 2006). Signalling by the pre-BCR results in the upregulation and increased binding of IRF4 to the 3'Ek enhancer (Johnson et al., 2008). The further recruitment of E2A to both iE*κ* and 3'Ek in addition to V*κ* genes (Lazorchak et al., 2006) allows promotion of *Igκ* non-coding transcription by both enhancers resulting in the initiation of recombination.

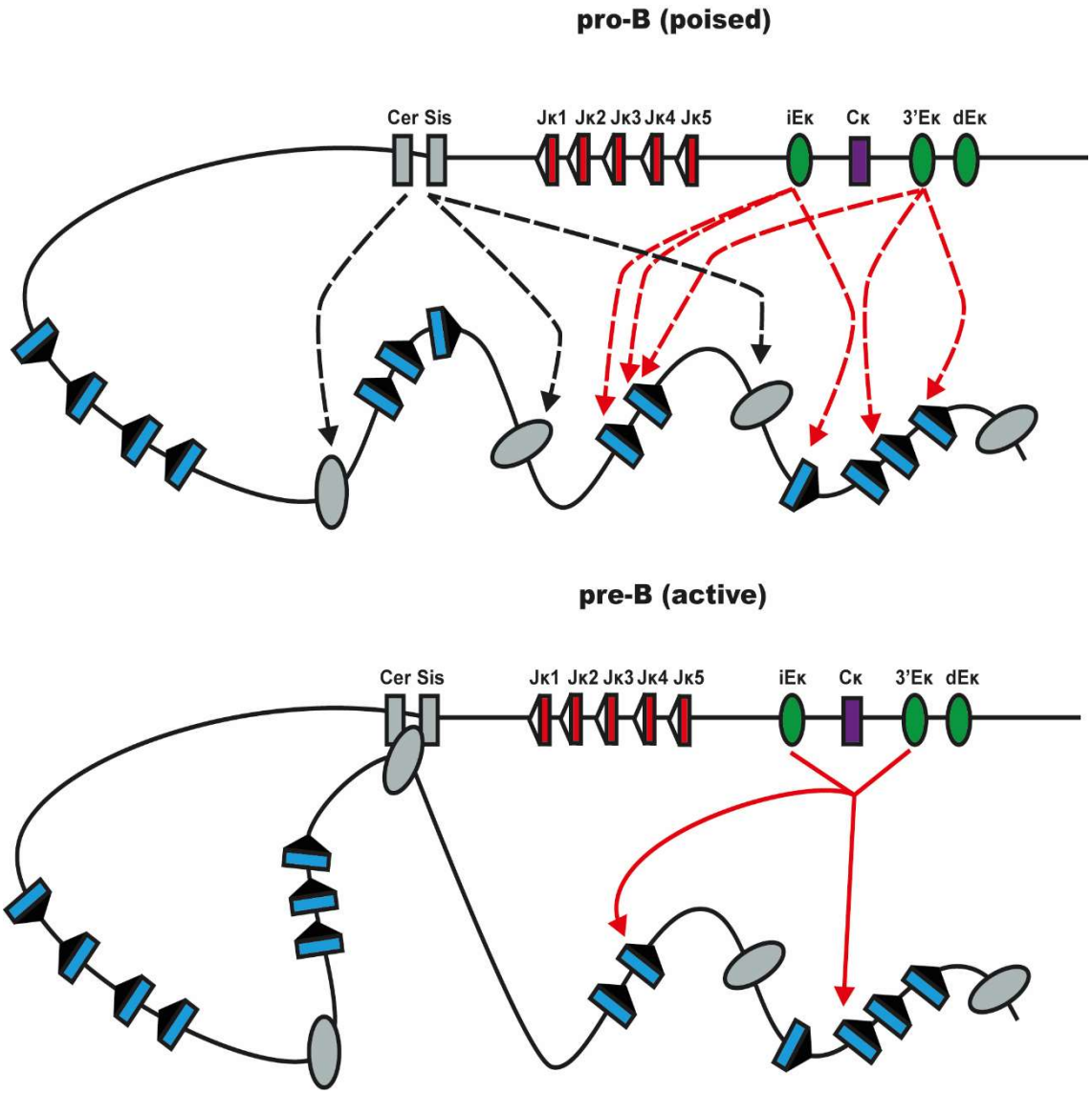


Figure 1.15: Long-range interactions mediating Igκ recombination. A simplified schematic of long-range interactions formed at the Igκ locus. Locus contraction occurs at the pro-B cell stage but interactions between the Vκ regions and the Cer/Sis elements or iEκ/3'Eκ enhancers. Upon differentiation to pre-B cells, pre-BCR signalling triggers an increase in Vκ region interaction with the Cer/Sis elements. Furthermore, the interactions of the iEκ and 3'Eκ enhancers become highly focussed and these elements coordinate to upregulate Vκ and Jκ non-coding transcription and Igκ recombination. Adapted from de Almeida et al. 2015.

F) Generation of additional antigen receptor diversity

Antigen binding to naïve B cells, in coordination with signals from helper T-cells, results in a cascade of signalling events which promote the rapid clonal expansion of the B cell (Berek and Ziegner, 1993). Following clonal expansion, the population undergoes affinity maturation to further increase the specificity of the immunoglobulin-antigen interaction. Affinity maturation consists of two key processes: somatic hypermutation (SHM) and class switch recombination (CSR).

1.25 Somatic hypermutation

During immune responses, in germinal centres, B lymphocytes undergo multiple cycles of immunoglobulin variable region hypermutation followed by clonal expansion and selection based on antigen affinity to select the most specific antibodies (Berek et al., 1991). Somatic hypermutation introduces point mutations into the variable regions of the rearranged immunoglobulin genes in an attempt to increase the specificity of the antibody for its antigen. Both SHM and CSR require the action of Activation-induced cytosine deaminase (AID). Upon stimulation by antigen, B cells upregulate the expression of *Aid* via the action of the transcription factors PAX5 and E2A. As AID can only act on single stranded DNA, transcription through the variable exons is required to generate single stranded ‘transcription bubbles’ to allow SHM to occur (Peled et al., 2008). Furthermore, the transcription mediated deposition of the H3K4me3 modification has been suggested to play a role in the recruitment of AID to SHM hotspots (Begum et al., 2012).

AID introduces point mutations by converting deoxycytosine (dC) to deoxyuracil (dU) by deamination (Muramatsu et al., 2000), resulting in a uracil:guanine (U:G) mismatch. The U:G mismatch is resolved in one of three ways: DNA replication, base excision repair (BER) or mismatch repair (MMR). DNA replication resolves the U:G mismatch by interpreting the uracil as a thymidine, resulting in a C to T or G to A (on the opposite strand) transition. Resolution of the U:G mismatch by BER involves the action of uracil DNA glycosylase (UNG), which excises the uracil and results in the generation of an abasic site. The action of apurinic/apyrimidinic endonuclease 1 (APE1) generates a single stranded DNA break which is repaired by an error prone polymerase e.g. REV1 or DNA pol μ (Peled et al., 2008). Alternatively, the U:G mismatch can be recognised by MutS like homologue 2 (MSH2) and MutS like homologue 6 (MSH6) resulting in the excision of the mismatch and several surrounding nucleotides. Mutations are generated by the resynthesis of DNA by error prone translesion polymerases (Wilson et al., 2005) and the nick is repaired by the action of DNA ligase I (Peled et al., 2008).

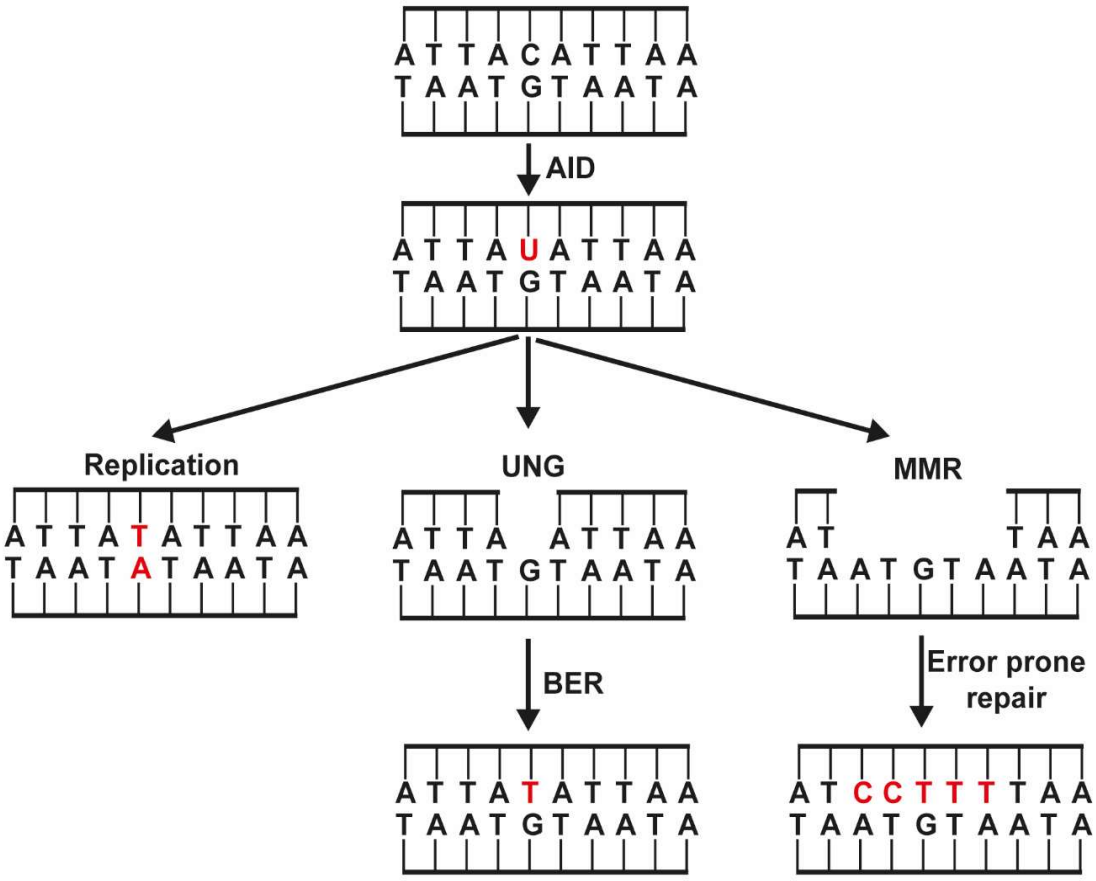


Figure 1.16: Simplified overview of somatic hypermutation. Somatic hypermutation is initiated by the action of activation induced deaminase (AID) which deaminates cytosine to uridine. This is processed in one of three ways: replication, base excision repair (BER) or mismatch repair (MMR). When repaired by replication, the uridine is recognised as a thymidine resulting in a C to T transition. Repair by BER involves the action of uracil DNA glycosylase (UNG) followed by the action of apurinic/apyrimidinic endonuclease 1 (APE1) generating an abasic site which is repaired by an error prone polymerase. In MMR basepairs surrounding the deaminated cytosine are also removed and repaired in an error prone manner.

1.26 Class switch recombination

CSR changes the antibody class, which alters the properties of the antibody, by selecting an alternative heavy chain constant region. Switch regions present upstream of the *Igh* constant regions, with the exception of Cδ, vary in length from 1-10 kb and contain several groups of short (20-80 bp) G-rich tandem repeat sequences. Whilst naïve B cells have the capacity to switch to any isotype, cytokines secreted by T cells are able to direct isotype switching. The major mechanism by which this occurs is the regulation of non-coding transcription through switch regions, as only transcriptionally active switch regions are subject to the activity of AID (Stavnezer et al., 2008).

Non-coding transcription enables AID to act on both strands of a switch region, resulting in the deamination of cytosine residues and their subsequent removal by UNG followed by the formation of two single stranded nicks by the action of APE1/APE2. It is thought that the presence of multiple AID targets in a switch region overwhelms the ability of the BER pathway to repair these nicks (Stavnezer et al., 2008), thus increasing the likelihood of forming two single stranded breaks in close enough proximity to generate a double strand DNA break. The overhangs formed at donor and acceptor switch regions are repaired by the actions of error-prone DNA polymerases (5’ overhangs) or ERCC1-XPF (3’ overhangs) before being joined via the NHEJ machinery in the same manner as V(D)J recombination (Stavnezer et al., 2008). As with V(D)J recombination, the intervening region is excised.

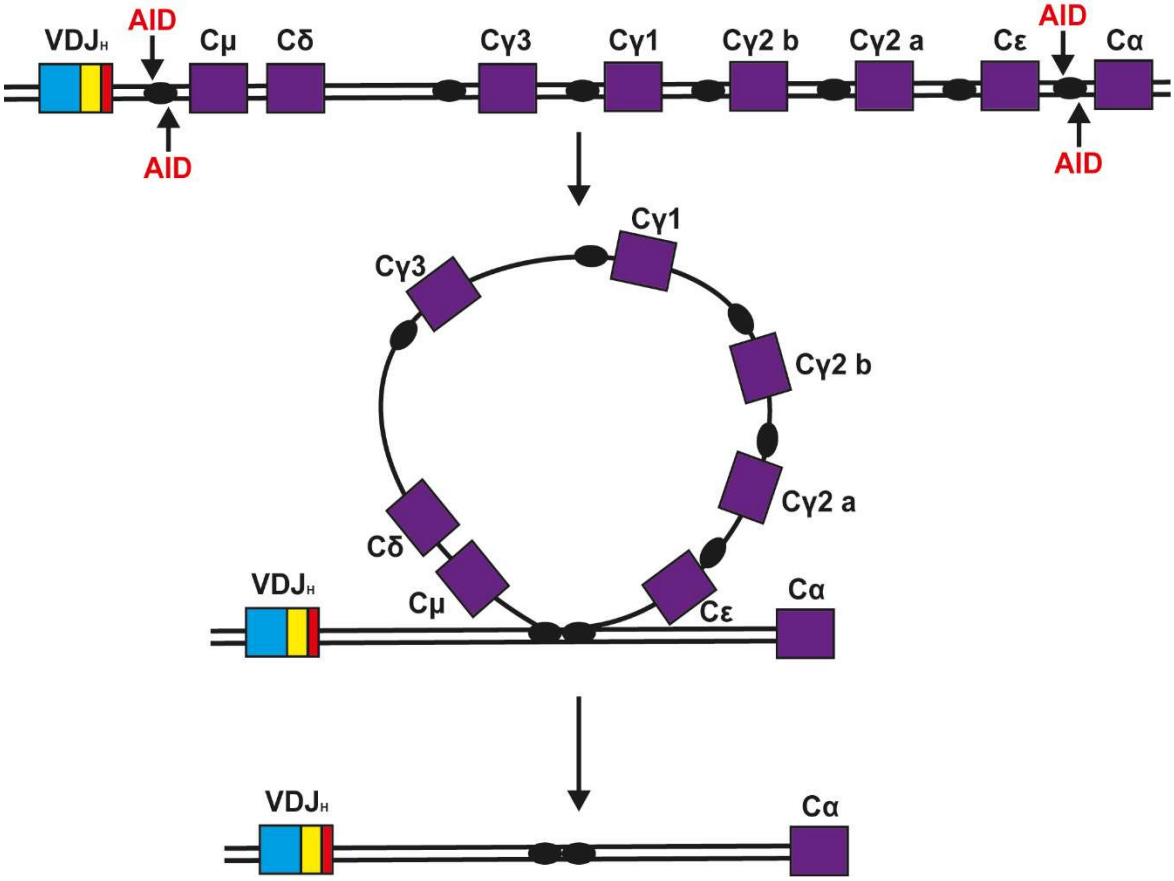


Figure 1.17: Overview of class switch recombination. Class switch recombination is initiated by AID-mediated double strand breaks (two single stranded DNA breaks in close proximity) at switch regions (black ovals). This results in recombination between switch regions leading to an alternative constant region exon being adjacent to the rearranged variable exon.

G) V(D)J recombination and leukaemia

As V(D)J recombination introduces a Double strand break (DSB) into the genome, it poses a risk of insertion/deletion events or even chromosome rearrangements. Chromosome translocations are frequent genetic events in haematological cancers (Jankovic et al., 2007; Tsai and Lieber, 2010), however, translocation events are often insufficient to drive full cancerous transformation due to the requirement for additional genetic or epigenetic changes (Robbiani and Nussenzweig, 2013). Nevertheless, translocations can be key initiating events, promoting oncogenic transformation by providing a proliferative advantage (Tsai and Lieber, 2010).

While some leukaemias are strongly linked to aberrant V(D)J recombination, the mechanisms by which aberrant recombination leads to leukaemia remains poorly understood. Analysis of translocation breakpoints suggested that the incorrect selection of cryptic RSSs (cRSSs), which bear homology to genuine RSSs (Kirsch et al., 1982) could account for a large proportion of translocations (Tsai and Lieber, 2010). A notable example is the t(11;14)(p13;q11) translocation which results in LIM only 2 (LMO2) being inserted into the *Tcrδ* locus (Garcia et al., 1991). As many of these translocation events have heptamer like sequences at the breakpoint in addition to frequently exhibiting signs of N-nucleotide addition (Garcia et al., 1991) it is highly likely that these translocations are mediated by the RAG recombinase.

RAG mediated deletions and translocations have also been identified in subtypes of acute lymphoblastic leukaemia (ALL). Next-generation sequencing identified RSS motifs at deletion breakpoints which implicated aberrant RAG-mediated cleavage in the deletion of the Ikaros Family Zinc Finger 1 (*IKZF1)* and cyclin-dependent kinase inhibitor 2A (*CDKN2A)* genes (Iacobucci et al., 2009; Mullighan et al., 2008). Additional evidence that RAG mediated deletions are key drivers of oncogenesis has been identified in B cell-precursor ALL patients harbouring a *ETV6-RUNX1* translocation. Frequently acquired *in utero* (Greaves and Wiemels, 2003), this translocation slows early B-lineage differentiation, enabling high levels of RAG expression to persist and results in on-going V(D)J rearrangements. Whole genome sequencing studies of acute lymphoblastic leukaemia (ALL) patients carrying the *ETV6/RUNX1* translocation observed both deletions and chromosome rearrangements at RSSs which are thought to play an integral role in disease progression (Papaemmanuil et al., 2014).

Some cRSSs are used efficiently e.g. the LMO2, TAL1 and TAL2 cRSSs, while, the cRSSs at other common translocation breakpoints e.g. BCL2 are used very inefficiently (Marculescu et al., 2002). It is possible that the presence of H3K4me3 may increase the efficiency of cleavage at some cRSSs, as RAG2 is able to bind to this modification (Matthews et al., 2007; Shimazaki et al., 2009). Nevertheless, a notable fraction of translocations do not occur at the cRSS heptamer boundary which is inconsistent with RAG mediated cleavage (Brandt and Roth, 2009). This suggests that many B cell translocations occur via other mechanisms.

The frequency of chromosome translocations to a single RSS implies that an end-donation mechanism, whereby a RAG mediated DSB is joined to a break mediated by another process, could be the predominant mechanism by which these translocations occur. DSBs can be generated in response to ionising radiation, oxidising radicals, the collapse of a replication fork or the inadvertent action of nuclear enzymes (Lieber, 2010). In addition, there is substantial evidence that the RAG proteins can cleave non-B form DNA (Raghavan et al., 2004) or cause DSBs at deaminated methylated CpGs (Tsai et al., 2008), to generate partner ends for translocations. End donation events are thought to occur in 30-40% of follicular and mantle cell lymphomas (Jäger et al., 2000; Nussenzweig and Nussenzweig, 2010). For example, the t(14;18)(q32;q21) translocation (Cleary et al., 1986) juxtaposes BCL-2 to the *Igh* locus, placing the anti-apoptotic gene under the control of strong regulatory elements in the *Igh* locus (Duan et al., 2008). The IgH/BCL-2 translocation is the major translocation observed in follicular lymphoma and displays hallmarks of an end donation event mediated by aberrant V(D)J recombination. Notably, the study by Papaemmanuil et al. 2014 mapped RSSs at both breakpoints and at just one of the breakpoints (Papaemmanuil et al., 2014) in approximately equal proportions, suggesting that both cryptic recombination and end-donation may both be prominent causes of genome instability.

H) Model systems to investigate V(D)J recombination

As V(D)J recombination occurs at specific transient stages of lymphocyte development, determining how errors in recombination occur is a complicated prospect. Lymphoid cells transformed by the Abelson murine leukaemia virus (A-MuLV; Rosenberg 1976) have provided a classic model for the study of early B-cell development and immunoglobulin rearrangement. Unlike their wild-type pre-B counterparts, most Abelson transformants do not rearrange light-chain loci or go on to express surface IgM. These cells also have very low levels of *Igκ* or *Igλ* non-coding transcription and the rearrangement of light-chain loci is prevented by the downregulation of *Rag* expression (Chen et al., 1994). However, cells transformed in this way can undergo inducible recombination, for example, the 103/BCL-2 cell line which has been transformed with a temperature sensitive *v-abl* mutant. Whilst at the permissive temperature (33ºC), 103/BCL-2 adopt a pro-B cell like phenotype and exhibit a very low level of recombination but when shifted to the non-permissive temperature (39ºC), these cells exhibit a substantial increase in RAG expression, light chain loci non-coding transcription and V(D)J recombination (Chen et al., 1994; Xu and Feeney, 2009). Despite the advances made using these transformed pro-B cells, a temperature sensitive *v-abl* mutant is not ideal to investigate immunoglobulin recombination as inactivating the *v-abl* protein results in substantial changes in gene expression (Muljo and Schlissel, 2003) thus decreasing the similarity to recombination in wild-type pre-B cells.

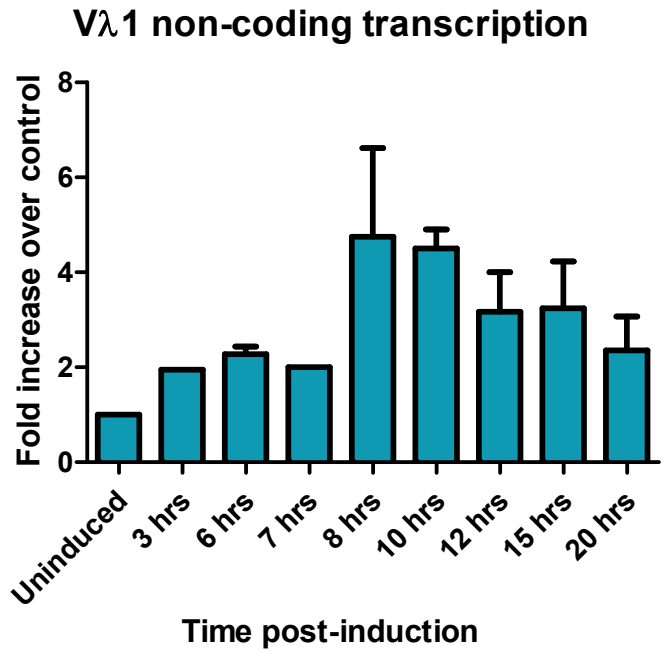
Remarkably, it was shown in the Boyes lab that the expression of a pre-B cell level of IRF4 at the pro-B cell stage is sufficient to sufficient to fully activate *Igλ* recombination and all associated chromatin changes (Bevington, 2009; Bevington and Boyes, 2013). This observation enabled the generation of transgenic mouse line, PIP-ER, in which *Igλ* recombination can be induced. This is achieved by the fusion of the estrogen receptor hormone binding domain to the N-terminus of IRF4. This fusion gene is expressed specifically at the pro-B cell stage under the control of the *Igll1* (λ5) promoter and locus control region (LCR) together with the VpreB exon and intron (Sabbattini et al., 1999), in PIP-ER transgenic mice. Following extraction of PIP-ER bone marrow and expansion in culture PIP-ER pro-B cells can be induced by the estrogen antagonist tamoxifen or its active metabolite 4-hydroxytamoxifen (4-OHT) which activates IRF4-ER and enables the events of *Igλ* activation to be followed temporally. Importantly, following induction, V(D)J recombination and VJ interactions occur at the same level as observed in pre-B cells. Therefore, this model system has the potential to answer many questions regarding the regulation of V(D)J recombination and enables the first temporal analysis of the process in a natural locus.

I) Aims

Whilst the PIP-ER experimental system has several advantages over other model systems, it is reliant on the recovery of primary pro-B cells from the bone marrow of PIP-ER transgenic mice. Pro-B cells are typically isolated via FACS purification using CD19 and CD43 cell surface markers. As pro-B cells exist in relatively low numbers, few cells are obtained from each mouse and additionally pro-B cells have a very limited half-life *ex vivo*. This imposes severe limits on the experiments that can be performed i.e. temporal chromatin immunoprecipitation and dramatically increases experimental variability. Therefore, the generation of a cell-line from PIP-ER pro-B cells would result in an unprecedented number of cells, enabling key experiments to uncover the true temporal order of events leading to the activation of V(D)J recombination.

It is vital that non-coding transcription at antigen receptor loci is tightly regulated to ensure that the correct gene segments are recombined and potentially to avoid aberrant recombination. A crucial observation from the analysis of PIP-ER pro-B cells was that *Vλ1* and Jλ1 non-coding transcription is coordinately upregulated at eight hours post induction (Figure 1.18). Determining how this regulation is achieved would provide one of the first temporal insights into the regulation of V(D)J recombination.

A)



B)

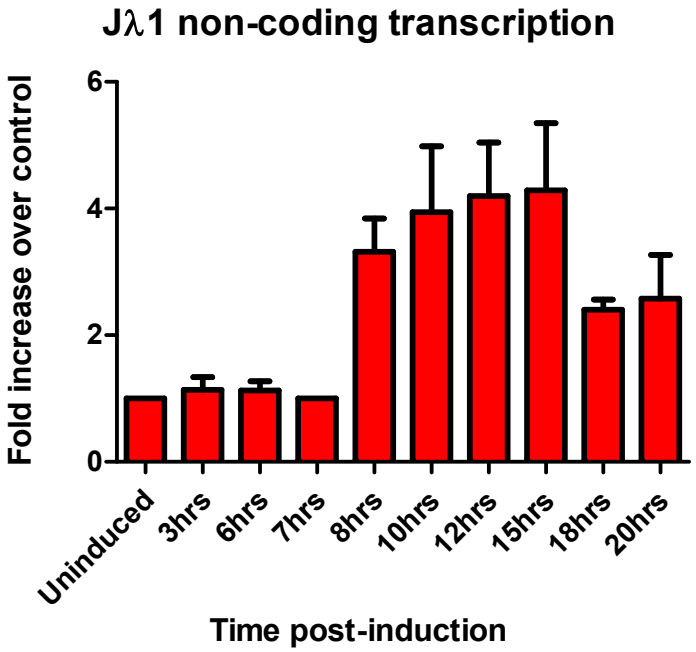


Figure 1.18: Vλ1 and Jλ1 non-coding transcription is coordinately upregulated. RT-qPCR analysis of Vλ1 (A) and Jλ1(B) non-coding transcription in PIP-ER pro-B cells following induction with tamoxifen. The non-coding transcription of both gene segments is upregulated coordinately at eight hours post-induction.

In order to determine how the coordinate activation of *Vλ1* and Jλ1 gene segments is achieved, I therefore sought to:

1. Generate and characterise a pro-B cell line from PIP-ER non-transgenic mice
2. Determine the factors regulating the coordinate activation of *Vλ1* and Jλ1 non-coding transcription
3. Use the generated cell line to determine the temporal order of events preceding the activation of *Vλ1* and Jλ1

These aims are addressed in the following results chapters. In the first results chapter, I describe the generation and characterisation of a pro-B cell line capable of inducing *Vλ1* and Jλ1 non-coding transcription. Furthermore, I also show that as in PIP-ER transgenic mice, *Vλ1* and Jλ1 non-coding transcription is coordinately upregulated. In the next chapter, I investigate the regulatory elements and long-range interactions that mediate the upregulation of *Vλ1* and Jλ1 non-coding transcription. In the final results chapter, I use the generated cell line to determine the temporal order of events that occur prior to *Vλ1* and Jλ1 non-coding transcription and propose a model to describe how this coordinate activation occurs.

Chapter 2 – Materials and Methods

A. Common buffers

Phosphate-buffered saline pH 7.4 (PBS)

137 mM NaCl

2.7 mM KCl

4.3 mM Na_2HPO_4

1.47 mM KH_2PO_4

TE (Tris-EDTA) buffer
10 mM Tris pH 8.0
1 mM EDTA

1 x TAE

40 mM Tris pH 8

20 mM Acetic acid

1 mM EDTA

Sort Buffer for flow cytometry

2.5 mM Hepes pH 7.9

2% Foetal Calf Serum (PAA)

1 mM EDTA

PBS to a final volume of 50 ml

Alkaline lysis buffer I

50 mM Glucose

25 mM Tris-HCl pH 8.0

10 mM EDTA

Alkaline lysis buffer II

0.2 M NaOH

1% (w/v) SDS

Alkaline lysis buffer III

3 M KOAc

5 M Acetic acid

Protein loading buffer (2x)

20% Glycerol

100 mM Tris-HCl pH 6.8

284 mM β-Mercaptoethanol

2% (w/v) SDS

0.025% (w/v) Bromophenol Blue

SDS-PAGE running buffer

25 mM Tris-HCl pH 8.5

192 mM Glycine

0.1% (w/v) SDS

B. Media

Dulbecco's Modified Eagle Medium (DMEM)

DMEM medium (Sigma-Aldrich, D6546) supplemented with:

10% Foetal Calf Serum (PAA)

2 mM L-Glutamine

50 µg/ml Streptomycin

50 U/ml Penicillin

Pro-B cell medium

McCoy's 5A (modified) medium (Life Technologies, 26600-023) supplemented with:

15% Foetal Calf Serum Gold (PAA Laboratories)

100 U/ml Penicillin

100 µg/ml Streptomycin

0.4x Essential amino acids (50x stock, Life Technologies, 11130)

0.4x Non-essential amino acids (100x stock, Life Technologies, 11140)

1 mM Sodium Pyruvate

0.1 % Sodium Bicarbonate

1.6 mM L-Glutamine

0.16 mg/ml L-Asparagine

0.16 mg/ml L-Serine

1x Vitamin mix (Life Technologies, 11120)

50 µM β-Mercaptoethanol (added fresh before use)

RPMI media for A-MuLV infection

RPMI-1640 medium (Sigma-Aldrich) supplemented with 15% Foetal Calf Serum Gold (FCS; PAA Laboratories),

2 mM L-Glutamine

50 µg/ml Streptomycin

50 U/ml Penicillin

50 µM β-mercaptoethanol

RPMI media for A-MuLV immortalised pro-B cells

RPMI-1640 medium (Sigma-Aldrich) supplemented with 10% Foetal Calf Serum Gold (FCS; PAA Laboratories),

2 mM L-Glutamine

50 µg/ml Streptomycin

50 U/ml Penicillin

50 µM β-mercaptoethanol (added fresh before use)

Opti-MEM media

OptiMEM (Gibco) supplemented with 10% FCS (PAA Laboratories)

100 µg/ml Streptomycin

100 U/ml Penicillin

50 µM β-mercaptoethanol

Lysogeny Broth (LB) medium

1% w/v	Bacto-tryptone
0.5% w/v	Yeast Extract
0.5% w/v	NaCl

Super Optimal Broth (SOB) medium

2% w/v	Bacto-tryptone
0.5% w/v	Yeast Extract
0.05% w/v	NaCl
0.019% w/v	KCl

LB-Agar

1.5% w/v	Agar
1% w/v	Bacto-tryptone
0.5% w/v	Yeast Extract
0.5% w/v	NaCl

For selective medium, 50 µg/ml Ampicillin was added.

C. DNA based methods

2.1 Preparation of genomic DNA

Typically, 1 x 10^6 cells were centrifuged at 504 x *g* for 3 minutes, washed in 10 ml PBS and cells were re-pelleted by centrifugation at 504 x g for 3 minutes. Cells were resuspended in 0.5 ml genomic digestion buffer (100 mM NaCl, 10 mM Tris pH 8, 25 mM EDTA, 0.5% SDS, 0.2 mg/ml Proteinase K) and incubated at 56°C overnight with rotation. Following incubation, four phenol-chloroform extractions were performed, followed by ethanol precipitation. The DNA was resuspended in 150 µl TE, the concentration was determined by a DS11+ spectrometer (DeNovix) and diluted to a final concentration of 5 ng/µl in ddH_2O.

2.2 Phenol/chloroform extraction of DNA

An equal volume of phenol-chloroform (1:1) was added to DNA solutions. The phenol denatures any protein present, whilst the chloroform extracts phenol from aqueous solutions. The mixture was vortexed for approximately 10 seconds before centrifugation at 16,000 x *g* for two minutes at room temperature. The aqueous phase containing nucleic acids was transferred to a fresh microcentrifuge tube. This process was repeated until the interface between the organic and aqueous phase was clear.

2.3 Ethanol precipitation of DNA

Precipitation of DNA was performed by the addition of two volumes of 100% ethanol, 10% volume of 3 M sodium acetate (pH 5.2) and 20 µg glycogen to aid visualisation and precipitation of particularly small amounts of DNA. Samples were incubated on dry ice for 10 minutes before centrifugation at 20,000 x *g* for 10 minutes at 4°C. The pellet was washed with 500 µl of 70% ethanol and centrifuged at 16,000 x *g* for 5 minutes. Pellets were air dried until there was no visible trace of ethanol, and re-suspended in a suitable volume of ddH_2O or TE.

2.4 A typical PCR reaction

A standard PCR reaction consisted of: 50 ng template, 1 x ThermoPol buffer (NEB), 200 nM forward and reverse primers, 200 µM dNTPs and 2 U Taq DNA polymerase (NEB).

A typical thermal profile was as follows:

94°C	5 mins	
94°C	30 s	} 30 cycles
T_m°C	20 s	
68°C	1 min per kb	
68°C	7 mins	

Where Tm = melting temperature of the oligonucleotides.

2.5 Agarose Gel Electrophoresis

DNA fragments were separated by size using agarose gel electrophoresis. Agarose gels were made in 1x TAE buffer and ethidium bromide was added at a final concentration of 1.25 ng/ml, to enable DNA visualisation by UV light. The percentage of agarose used was appropriate for the sizes of DNA being separated (between 0.8 - 2%). Agarose gels were electrophoresed in a Bio-Rad Sub-Cell GT tank, submerged in 1x TAE buffer. Prior to loading samples, DNA loading buffer was added, and the gel was then electrophoresed at 75 V for a small gel (6 x 10 cm), or 90 V for a large gel (14 x 15 cm) until sufficient separation of DNA fragments had been achieved.

2.6 Isolation of DNA fragments by gel extraction

The desired fragment was excised from the gel with a clean scalpel blade. DNA was then purified using a QIAquick gel extraction kit (Qiagen). The gel slice was initially dissolved in three volumes, relative to the weight of the gel slice, of a guanidinium thiocyanate containing buffer (QG) at 50°C. Following the addition of one volume of isopropanol, the dissolved gel was then loaded onto a QIAquick spin column which contained a silica membrane that binds DNA in a high salt. Following centrifugation at 16000 x *g* for one minute, the column was washed with 0.7 ml buffer PE and centrifuged for an additional minute. Residual ethanol was removed, and the DNA was eluted in 25-50 µl pre-warmed TE by centrifugation.

2.7 Ligation of DNA

DNA was ligated using T4 DNA Ligase (NEB). Reactions were typically 10 µl volumes, containing 200 units enzyme, 1x T4 Ligase Buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT) and typically 50 ng vector DNA with a 3-fold molar excess of insert DNA.

2.8 Preparation of competent E.coli using Calcium Chloride

Chemically competent DH5α *E. coli* cells were prepared using a variation of the method described by Cohen in 1972 (Sambrook et al., 1989). Firstly, *E. coli* were grown on a LB Agar plate without antibiotics at 37°C overnight. Following this, a single colony was inoculated into 50 ml LB medium in a 250 ml flask. Following incubation at 37°C overnight, 4 ml was transferred to a 2-litre flask containing 400 ml LB. This was then grown at 37°C to an OD600 of 0.375. Following this, eight 50 ml aliquots of cells were prepared, cooled on ice for 5 to 10 minutes, and were harvested by centrifugation at 1600 x *g* for 7 minutes at 4°C. Each cell pellet was gently resuspended in 10 ml of ice-cold $CaCl_2$ solution (60 mM $CaCl_2$, 15% glycerol, 10 mM PIPES, pH 7) before centrifugation at 1100 x *g* for 5 minutes at 4°C.

Following resuspension in 10 ml ice-cold $CaCl_2$ solution, cells were incubated for 30 minutes on ice, before centrifugation at 1100 x *g* for 5 minutes. Each cell pellet was finally resuspended in 2 ml ice-cold $CaCl_2$ and 200 µl aliquots of cells were made and frozen immediately at -80°C, until required.

2.9 Transformation of *E.coli*

For routine cloning, 5 µl of ligation or 2-10 ng of a plasmid was added to 50 µl competent cells and incubated on ice for 30 minutes. The cells were then heat shocked at 42°C for 45 seconds, and 300 µl of LB pre-warmed to room temperature was added. Cells were then incubated at 37°C for 45 minutes, after which 150 µl was spread on each of two LB-agar plates, containing an appropriate dilution of antibiotic. The plates were incubated at 37°C overnight.

2.10 Small scale extraction of plasmid DNA from *E.coli* (Miniprep)

Small-scale plasmid preparations were performed according to the mini-prep protocol described in Sambrook *et al.* (1989). A single *E.coli* colony was inoculated into 2 ml LB containing 50 µg/ml ampicillin and incubated overnight at 37ºC with shaking at 250 rpm. 1.5 ml of the culture was transferred to an Eppendorf tube and centrifuged at 12,000 x *g* for 30 seconds. The supernatant was discarded and the pelleted *E.coli* were resuspended in 100 µl alkaline lysis solution 1 (50 mM glucose, 25 mM Tris-HCl pH 8.0). To the resuspended *E.coli,* 200 µl alkaline lysis solution 2 (0.2 M NaOH, 1% SDS) was added. The tube was mixed by gentle inversion and placed on ice for up to five minutes. 150 µl alkaline lysis solution 3 (3 M KOAc, 11.5% (v/v) glacial acetic acid) was added followed by mixing and incubation on ice for five minutes.

Samples were centrifuged at 12,000 x *g* for five minutes and the supernatant was transferred to a fresh tube. An equal volume of phenol-chloroform was added, the solution was vortexed and centrifuged at 12,000 x *g* for two minutes. The aqueous phase was transferred to a new tube and ethanol precipitated by the addition of two volumes ethanol. The pellet was washed in 70% ethanol and resuspended in 50 µl ddH_2O containing 100 µg/ml RNaseA.

2.11 Large scale extraction of plasmid DNA from *E.coli*

Plasmid DNA was extracted from 150 ml of *E.coli* using a Qiagen Midi prep kit according to manufacturers' instructions. Extracted DNA was resuspended in 100 µl TE and the concentration was determined by a DS11+ spectrometer (DeNovix).

D. RNA methods

2.12 RNA extraction

RNA was extracted from approximately 2 x 10^6 cells using TRIzol (Invitrogen) following the manufacturer's instructions and treated with 2 U DNase I (Worthington) for 1 hr at 37°C in 100 µl of 1 x NEB DNase I buffer (10 mM Tris pH 7.5, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$). Following phenol-chloroform extraction and ethanol precipitation, RNA concentration was determined by a DS11+ spectrometer (DeNovix).

2.13 Generation of cDNA

1-3 µg of RNA was reverse transcribed with M-MuLV reverse transcriptase (Invitrogen). Briefly, 1-3 µg of RNA was added to 0.5 µg oligo dT primer, 500 µM dNTPs and ddH_2O to give a total volume of 12 µl. This was incubated at 65°C for 5 minutes and immediately placed on ice. To the reaction, 4 µl first strand buffer (Invitrogen), 10 mM DTT and 1 µl RNasinPlus (Promega) were then added followed by incubation at 37°C for 2 minutes and addition of 1 µl Moloney-Murine Leukaemia Virus Reverse Transcriptase (Invitrogen). The reaction was incubated at 37°C for 50 minutes before heat inactivation of the enzyme at 70°C for 15 minutes.

2.14 Quantitative PCR (qPCR) using SYBR green

Quantitative PCR was performed using a Corbett Rotor-Gene 6000 machine. A typical qPCR reaction contained 5 µl 2x SYBR Green master mix (NoROX SensiMix, Bioline), 4 pmol of each primer (Table 2.4), 0.5 - 5 ng of template DNA or cDNA at a final dilution of 1:100, in a total volume of 10 µl. All reactions were performed in duplicate and a melt curve, to determine the purity of the amplicon, for each sample was produced by analysis of fluorescence as the temperature was increased from 72°C to 95°C.

A typical cycle consisted of:

95°C	3 min	
95°C	5 s	} 40 cycles
T_m	10 s	
72°C	10 s	

Where Tm = melting temperature of the oligonucleotides.

Reverse transcription-qPCR (RT-qPCR) samples were quantified by determining the relative number of copies from a standard curve (typically composed of 0.4, 1.2 and 3.6 µl of a sample) using the Rotor-Gene Q (v2.3.1.49) software. All samples were then normalised to the calculated copies of an internal control, the housekeeping gene (*Hprt*), to account for discrepancies in sample quantification and the RT procedure.

ChIP-qPCR quantification was performed by determining the fold enrichment of the region of interest over a negative control region (IntIII). This was calculated using the following formula:

$$\text{Fold enrichment} = \frac{2^{(\text{Region of interest input Ct} - \text{Region of interest bound Ct})}}{2^{(\text{Control region input Ct} - \text{Control region bound Ct})}}$$

2.15 Classical 5' Rapid Amplification of Complementary DNA Ends (RACE)

This protocol was performed according to "Rapid amplification of 5' cDNA ends" (Sambrook 2001) with modifications. RNA (1 µg) was reverse transcribed (Section 2.13) the oligo dT primer was removed by the addition of three volumes of buffer QG (Qiagen) and one volume isopropanol before application to a Qiagen quickspin column (Qiagen) and elution according to manufacturer's instructions. To generate A tailed cDNA, the extracted cDNA was then added to 1 x Terminal Transferase buffer (NEB), 250 µM $CoCl_2$ (NEB), 100 µM dATP, 10 U Terminal Transferase (NEB) and ddH_2O to generate a total volume of 50 µl. The tailing reaction was performed at 37°C for 30 minutes after which the enzyme was inactivated by heating at 75°C for 15 minutes. The volume was then increased to 100 µl with ddH_2O to dilute the components of the tailing reaction.

To generate 5' RACE products, the A-tailed cDNA was subjected to PCR, in a reaction that consisted of approximately 5-20 µl diluted A-tailed cDNA (~50 ng assuming 1:1 RNA to cDNA conversion), 1 x Q5 Reaction buffer (NEB), 200 nM dT adaptor primer and *Vλ1* specific primer (Table 2.5), 200 µM dNTPs and 2.5 U Q5 Hot-Start polymerase (NEB). A touchdown protocol was used in order to increase the specificity of the PCR, the thermal profile consisted of 98°C for 3 minutes followed by 15 cycles of 98°C for 10 s, 71°C for 20 s and 72°C for 1.5 minutes, 10 cycles of 98°C for 10 s, 68°C for 20 s and 72°C for 1.5 minutes, 15 cycles of 98°C for 10 s, 65°C for 20 s and 72°C for 1.5 minutes and a final extension at 72°C for 3 minutes.

The highest intensity bands as visualized on a UV transilluminator were excised and gel extracted (Section 2.6). Gel extracted products were eluted in 30 µl ddH_2O and 1 µl was used in a PCR reaction designed to add a Hind III restriction site to the 3' end of the product, to enable cohesive end cloning (a Xho I recognition site was present in the dT Adaptor primer). The PCR reaction consisted of 1x ThermoPol Buffer (NEB), 200 µM dNTPs, 200 nM dT adaptor primer and 200 nM Vλ1-GSP4-2-Hind III and 2 U Taq polymerase (NEB), in a final volume of 50 µl. The thermal profile was: 94^oC for minutes followed by four cycles of 94^oC for 30 s, 58^oC for 20 s, and 68^oC for 2 minutes and 16 cycles of 94^oC for 30 s, 60^oC for 20 s, 68^oC for 2 minutes with a final extension at 68^oC for seven minutes. 5' RACE products were purified by phenol-chloroform extraction and ethanol precipitation, cloned into pBluescript SK- and sequenced.

2.16 Modified 5' Rapid Amplification of Complementary DNA Ends (RACE)

This protocol is as described by Dallmeier and Neyts, 2013 with modifications. PIP3 pro-B cell RNA (2 µg) was reverse transcribed with 1 µM phosphorylated Vλ1-GSP2 (sense transcripts) or Vλ1-GSP2-2 (antisense transcripts) (Table 2.5) using Protoscript II reverse transcriptase (NEB) following the manufacturer's instructions. Following reverse transcription, 10 µg RNaseA was added to the cDNA and incubated at 37^oC for 30 minutes in the presence of 300 mM NaCl. After phenol/chloroform extraction and ethanol precipitation, cDNA was circularised by T4 RNA Ligase 1, in a reaction that consisted of: cDNA in 20 µl ddH_2O, 1x T4 RNA Ligase Buffer (NEB), 50 µM ATP, 10% PEG 8000 and 10 U T4 RNA Ligase 1 in a total volume of 50 µl. The reaction was incubated at 16^oC overnight to maximise circularisation events.

Following circularisation, inverse PCR was performed with the primers Vλ1-Upstream-2 and Vλ1-UpstreamF (sense transcripts) or V1_RACE_+367_F_HindIII and V1_RACE_+325_R_XbaI (antisense) (Table 2.5). The PCR reaction consisted of 2 µl circularised cDNA, 1x ThermoPol Buffer (NEB), 200 µM dNTPs, 200 nM of each primer and 2 U Taq polymerase (NEB), in a final volume of 50 µl. The resulting products were phenol/chloroform extracted and ethanol precipitated before digestion with the appropriate restriction enzymes overnight. Digested products were phenol/chloroform extracted and ethanol precipitated before being cloned into pBluescript SK- and sequencing.

E. Protein extraction and analysis

2.17 Luciferase reporter assay

Pro-B cells (103/BCL-2) were transfected as described in Section 2.26 with 1 µg luciferase vector and 10 ng Renilla luciferase reporter vector. After 48 hours cells were centrifuged at 386 x *g* for three minutes and washed in PBS. Cells were pelleted as described and lysed in passive lysis buffer (Dual glo luciferase assay; Promega) for five minutes at room temperature with gentle agitation. Luciferase and Renilla expression was examined using the Dual glo luciferase assay system (Promega; E2920) according to the manufacturer's instructions.

2.18 Preparation of whole cell and nuclear extracts

For whole cell extracts, pro-B cells were washed with PBS and resuspended at a concentration of 2 x 10^4 cells/µl in a 3:1 mix of RIPA (25 mM Tris pH 8.2, 50 mM NaCl, 0.5 % NP40, 0.5 % sodium deoxycholate, 0.1 % SDS) and lysis buffer (5% SDS, 0.15 M Tris pH 6.7, 30% glycerol) supplemented with protease inhibitors (Complete™, Mini Inhibitor Cocktail Tablets, Roche). The samples were rapidly boiled for 5 minutes, to inhibit protein degradation, and centrifuged at 1600 x *g* for 10 minutes at 4°C.

Nuclear extracts were prepared from pro-B cells by resuspending PBS washed cells at a density of $1x10^6$ cells/ml in 1 ml of lysis buffer (10 mM Tris pH 8, 10 mM NaCl, 0.2% NP-40, 50 µg/ml PMSF, 1 x Complete protease inhibitor cocktail (Roche)) and incubating on ice for 20 minutes. Nuclei were pelleted at 800 x *g* for 2 minutes before resuspension in 100 µl 1 x laemmli loading buffer and boiling for 5 minutes to inhibit protein degradation and to lyse the nuclei.

Both whole cell and nuclear extracts were either used immediately for Western blotting or flash frozen in dry ice and stored at -80°C until required.

2.19 Western blotting

Proteins were separated according to their size via SDS-PAGE. Extracted proteins were mixed with an equal volume of 2 x laemmli loading buffer, boiled for five minutes and separated on a 10% SDS-PAGE gel for 1.5 hours at 180 V in a Bio-Rad Protean III gel tank. The SDS-PAGE gel was then soaked in transfer buffer (20 % methanol, 192 mM glycine, 25 mM Tris pH 8) and proteins were transferred to a polyvinylidene difluoride membrane (Immobilon-P, IPVH00010, Millipore) by the use of an electroblotter. Blotting was performed at 0.68 mA/cm^2 for one hour. The membrane was then blocked with 10 ml of a solution of 5% dry milk powder in TBS-T (50 mM Tris pH 7.6, 150 mM NaCl, 5% milk, 0.05% Tween-20) and incubated with rocking for one hour. Primary hybridisation was performed overnight at 4°C in 2 ml TBS-T-milk with an appropriate dilution of primary antibody (Table 2.10).

Following hybridisation, the membrane was washed with TBS-T for one hour, replacing the buffer every ten minutes. Secondary hybridisation with an appropriate dilution of HRP conjugated secondary antibody (Table 2.10), raised against the species from which the primary antibody was isolated, was performed for one hour in 2 ml TBS-T-milk at room temperature. The membrane was washed as above. Following this 1 ml enhanced chemiluminescence substrate (Thermo Scientific) was added, and the membrane was imaged using a G:BOX Chemi XT4 system (Syngene).

F. Cell culture methods

2.20 Culture of adherent cells

Cos7, NIH-3T3 and 293T cells were maintained in complete DMEM media. Typically, adherent cells were cultured in 20 ml of media in a T75 flask and cells were passaged 1:10 at ~90% confluency. Cells were split by aspirating the medium and washing cells with 10 ml PBS. The cells were then incubated at room temperature for five minutes with 1 ml trypsin/EDTA (Sigma Aldrich T3924). Trypsin was quenched by the addition of 10 ml complete DMEM media, followed by centrifugation at 386 x *g* for three minutes. Pelleted cells were resuspended in 10 ml DMEM media, before 1 ml of cells were added to 19 ml fresh DMEM in a new T75 flask.

2.21 Culture of 103/BCL-2 cells

103/BCL-2 cells were maintained in RPMI media, at a density of 0.5 - 2 $x10^6$ cells/ml. Cells were diluted in fresh medium every 2 - 3 days or when cell density was approaching 2 $x10^6$ cells/ml.

2.22 Culture of 1D1-T215 cells

1D1-T215 cells were cultured in RPMI media at a density of 0.2 – 2 x 10^6 cells/ml. Every two days, or when the cell density approached 2 x 10^6 cells/ml, cells were diluted 1:20 in fresh RPMI media.

2.23 Preparation of IL-7

The stromal cell line Mo-IL-7 (a kind gift from Prof A. Rolink) was used as the source of IL-7. Mo-IL-7 cells were grown in DMEM supplemented with 4 µg/ml mycophenolic acid and 12.5 µg/ml xanthine until three days past confluency at which point IL-7 was harvested from the supernatant. IL-7 containing supernatant was centrifugated at 386 x *g* for three minutes and filtered through a 0.2 µm filter and stored at -20^{o}C. IL-7 concentration was determined by enzyme-linked immunosorbent assay (ELISA) using a Mouse IL-7 Quantikine ELISA kit (R&D Systems). Before addition to cultures, IL-7 aliquots were thawed, re-filtered and aliquoted into 10 ml working stocks. Working stocks were stored at -20^{o}C until use and upon thawing were stored at 4-8^{o}C.

2.24 Culture of primary pro-B cells

Femurs were dissected from 5-7-week-old mice and bone marrow was extracted by flushing with a G-25 needle and syringe into 10 ml pro-B cell medium. IL-7 was added to a final concentration of 5 ng/ml. Cells were expanded in a T25 flask at 33^{o}C in 5% CO_2 with 5 ml of IL-7 supplemented media being added every four days.

2.25 Transfection of 293T cells

293T cells were plated at a density of 3 x 10^5/ml 24 hrs before transfection in 10 ml DMEM media in a 10 cm dish. The media was replaced with antibiotic free DMEM approximately three hours before transfection. Polyethylenimine (PEI) and plasmid DNA were diluted separately in 500 µl serum free/antibiotic free DMEM at a ratio of 3 µg PEI: 1 µg DNA. The PEI solution was added to the DNA solution, immediately vortexed, incubated for 10 minutes at room temperature before being added dropwise to the plated 293T cells.

2.26 Transfection of 103/BCL-2 cells by electroporation

Electroporation was performed using a Nucleofector I device (Lonza) following manufacturers' instructions. Plasmid DNA, 2 µg, was added to 8x10^6 103/BCL-2 cells resuspended in 100 µl of supplemented Mouse B cell Nucleofector solution (Lonza; cat VPA-1010). The suspension was transferred to an electroporation cuvette and placed in a Nucleofector I device set to programme Z-01. RPMI media (500 µl; without antibiotics) was then added to the cuvette and the cells were transferred to three wells of a 12 well plate, resulting in a total volume of 1.5 ml. Cells were incubated for 24 hours prior to temperature shift at 39^{o}C for 16 hours.

2.27 Tamoxifen treatment of cell lines

Tamoxifen was used to activate the IRF4-ER protein in PIP-ER cell lines. The inductions were performed by resuspending 1 - 5 x 10^6 cells at concentration 0.5 x 10^5 cells/ml in RPMI media to which 4-OHT (Insight Biotechnology; Cat HY-16950-2mg), was added to a final concentration of 2 µM. To ensure good mixing, the 10 mM stock solution was diluted in 500 µl of RPMI media prior to addition to the cells. Treated cells were incubated at 37ºC with 5% CO_2 for the number of hours indicated.

G. Viral methods

2.28 Preparation of retrovirus

The Phoenix cell line (Grignani et al., 1998) was used to package the retrovirus. Phoenix cells were plated at a density of 3 x 10^6 cells in a 10 cm dish 24 hrs before transfection. Cells were transfected with 4 µg of MSCV-IRF4-ER-IRES-GFP construct using 12 µg PEI. Media was replaced 24 hrs post-transfection and retrovirus containing supernatant was harvested at 48 and 72 hours post-transfection. The retrovirus was filtered through a 0.22 µm syringe filter and used immediately for infection.

2.29 Preparation of Lentivirus

Lentiviral packaging was performed by 293T cells which were transfected with 4 µg of the appropriate transfer vector, in addition to 4 µg and 2 µg respectively of the viral packaging plasmids pCMVR8.74 (Addgene #22036) and pMD2.G (Addgene #12259) using 30 µg PEI. The pCMVR8.74 plasmid encodes the viral polymerase whereas pMD2.G encodes the VSV-G viral envelope. Lentivirus was harvested at 48 and 72hrs post transfection, passed through a 0.22 µm filter to remove any 293T cells and flash frozen in 5 ml aliquots. Aliquots of lentivirus were stored at -80ºC until use.

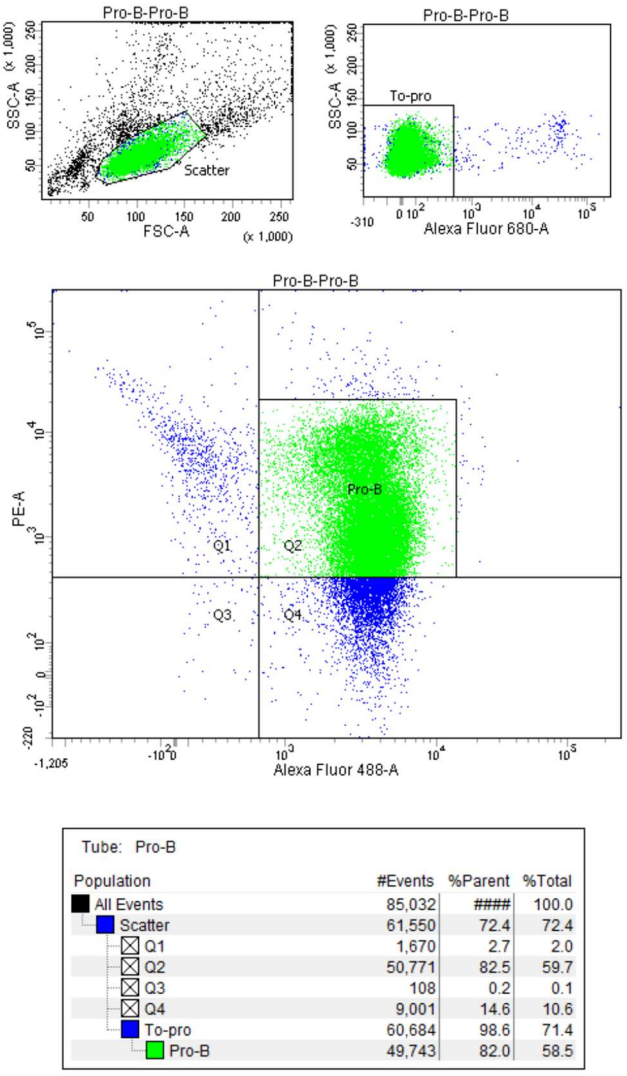
2.30 Retroviral/Lentiviral infection by spin-infection

Approximately 0.5 - 1 x 10^6 cells were centrifuged at 385 x *g* for 3 minutes and resuspended in 500 µl of filtered retrovirus/lentivirus containing 4µg/ml polybrene (Sigma-Aldrich). Cells were centrifuged at 800 x *g* for 90 minutes at 32°C in 12 well plates. Following spin-infection, 1 ml of RPMI media was added to each well and cells were maintained at 37°C for 48 hours before analysis or antibiotic selection.

H. Flow cytometry

2.31 Cell sorting by flow cytometry

Media was removed by centrifugation at 385 x *g* for three minutes. For primary cell cultures, cell pellets were resuspended in 10 ml 168 mM NH_4Cl and incubated at room temperature for 10 minutes to lyse erythrocytes. For pro-B or pre-B cells, cells were pelleted by centrifugation at 385 x *g* for three minutes and resuspended in 1 ml flow cytometry buffer (PBS supplemented with 2.5 mM Hepes pH 7.9, 2% foetal calf serum gold (PAA), 1 mM EDTA). Cells were stained with 8 µl anti-CD19-FITC and 8 µl anti-CD43-PE (Table 2.10) at room temperature in the dark for 10 minutes. The volume was increased to 3 ml with flow cytometry buffer and cells were centrifuged for three minutes at 385 x *g*. Pelleted cells were resuspended in 500 µl flow cytometry buffer and filtered through a 40 µm cell strainer. Figure 2.1 displays the gating strategies for pro-B and pre-B cells by flow cytometry. For GFP expressing cells, pelleted cells were washed in 1 ml PBS, pelleted at 385 x *g* for three minutes and resuspended in 500 µl flow cytometry buffer before filtering through a 40 µm cell strainer. Cell sorting was performed using a Becton Dickinson FACS ARIA 2 or a Becton Dickinson FACS Melody.



Tube: Pro-B

Population	#Events	%Parent	%Total
All Events	85,032	####	100.0
Scatter	61,550	72.4	72.4
Q1	1,670	2.7	2.0
Q2	50,771	82.5	59.7
Q3	108	0.2	0.1
Q4	9,001	14.6	10.6
To-pro	60,684	98.6	71.4
Pro-B	49,743	82.0	58.5

Figure 2.1: Flow cytometry template for purification of pro-B cells. Pro-B cells were isolated from cultured mouse bone marrow. A) Lymphocytes are first gated on forward scatter (FSC) and side scatter (SSC). B) Live cells are gated based on their exclusion of the viability dye TOPRO 3 iodide C) pro-B cells are gated based on staining with staining with anti-CD19-FITC (Alexa Fluor 488-A) and anti-CD43-PE (PE-A) conjugated antibodies.

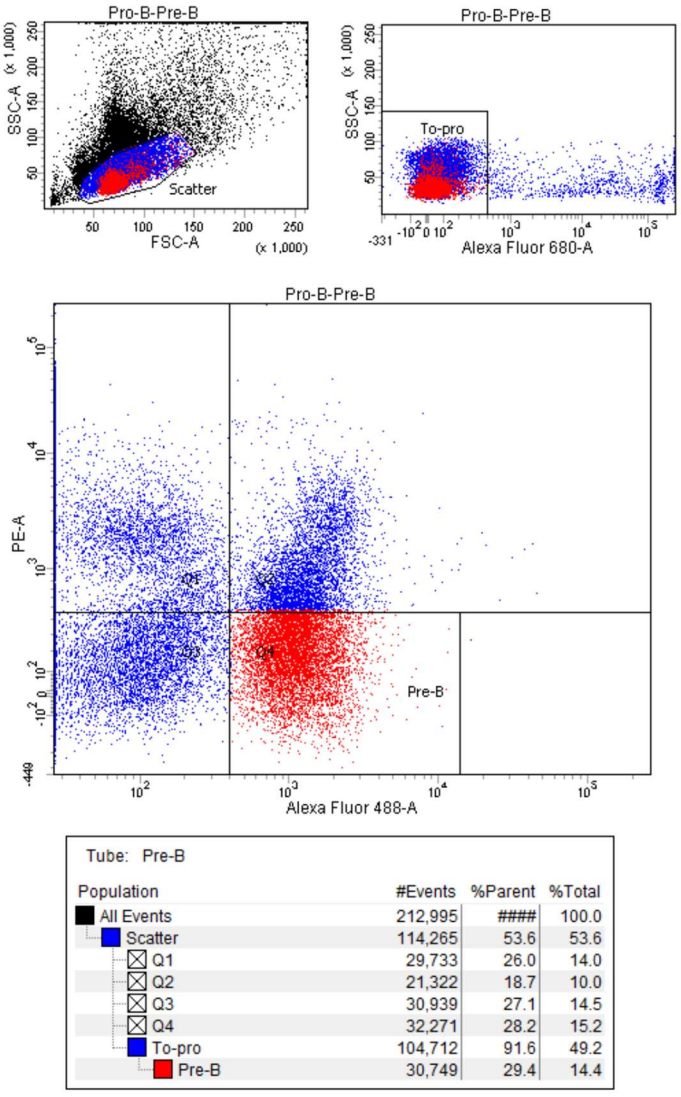


Figure 2.2: Flow cytometry template for purification of pre-B cells. Pre-B cells were isolated directly from mouse bone marrow. A) Lymphocytes are first gated on forward scatter (FSC) and side scatter (SSC) B) Live cells are gated based on their exclusion of the viability dye TOPRO 3 iodide C) pre-B cells are gated based on staining with anti-CD19-FITC (Alexa Fluor 488-A) but not with anti-CD43-PE (PE-A) conjugated antibodies.

2.32 Cell surface analysis by flow cytometry

Media was removed by centrifugation at 504 x g for three minutes. Cells were washed in PBS before centrifugation at 504 x g for three minutes and resuspension in 1 ml flow cytometry buffer (PBS supplemented with 2.5 µM Hepes pH 7.9, 2% foetal calf serum (PAA), 1 nM EDTA). Cells were stained as described in Section 2.31. Stained cells were resuspended in 4-500 µl flow cytometry buffer and analysed on a Becton Dickinson BD-LSR Fortessa. Analysis was performed using BD FACS DIVA software.

I. Generation of a PIP-ER pro-B cell line

2.33 IL-7 dependent immortalization of PIP-ER D pro-B cells

This protocol was adapted from Corfe et al (2007), unless otherwise stated the IL-7 concentration used was 5 ng/ml. Bone marrow cells were extracted from 5 - 7 week old PIP-ER mice and flow cytometry (Section 2.31), was used to isolate CD19 positive cells. The gating strategy is described in Figure 2.3.

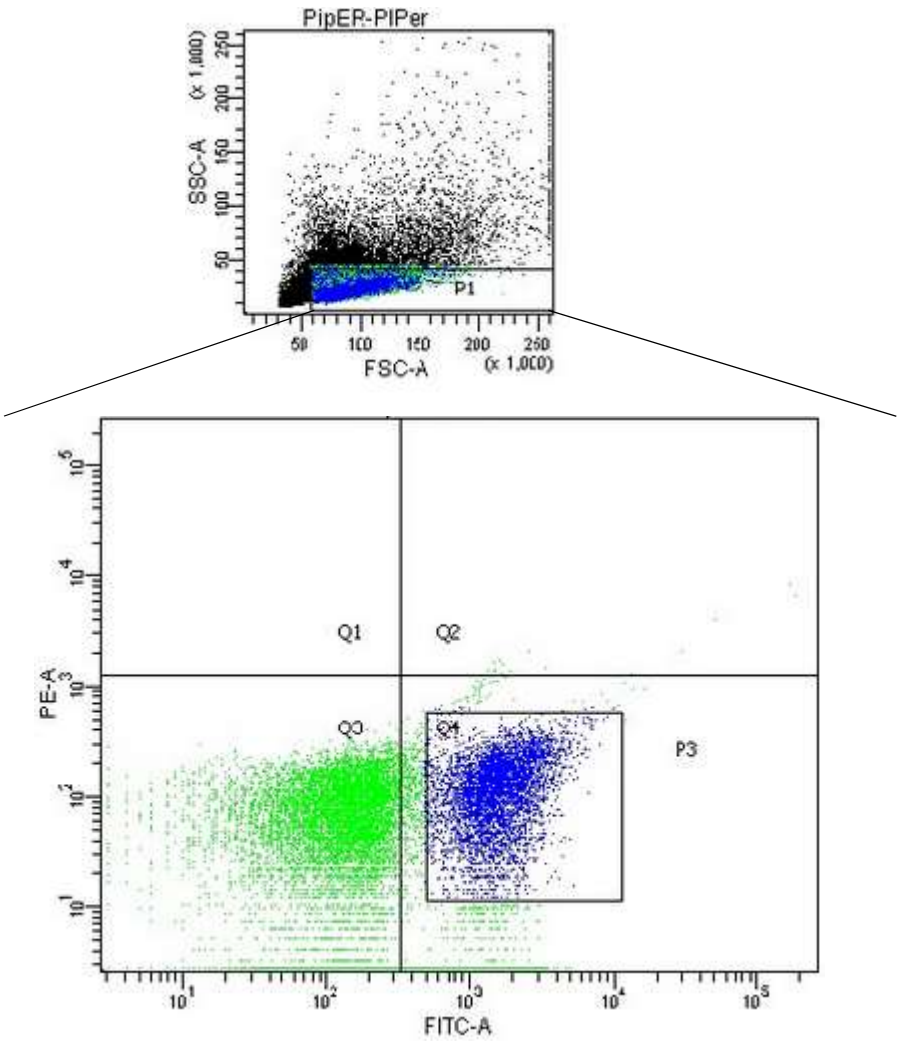


Figure 2.3: Gating strategy to obtain a pure CD19+ population. Flow cytometry template for the purification of CD19+ B-cells. A) Lymphocytes are first gated on forward scatter (FSC) and side scatter (SSC) B) CD19+ cells are gated based on staining with anti-CD19-FITC antibody.

Isolated cells were recovered by centrifuging for 10 minutes at 386 x *g* before resuspension in Opti-MEM medium supplemented with IL-7. Cells were plated at 1 x 10^4 cells per well in 24-well plates and were maintained at 37ºC. 1 ml of IL-7 supplemented media was added every 4-5 days. After day four, every 4-5 days, 1 ml of media was removed from each well, without disturbing the cells, and was replaced with 1 ml of IL-7 supplemented media. Surviving cells were cultured until a density of approximately 5 x 10^5 cells/ml was achieved, at which time 1 ml of cells was added to 2 ml of IL-7 supplemented media and placed into one well of a 6 well plate. Again, surviving cells were cultured until a density of approximately 5 x 10^5 cells/ml. Flow cytometry, following staining with anti-CD19-FITC and anti-CD43-PE antibodies (Table 2.10), was used to verify that the generated cell lines expressed pro-B specific markers. Verified cell lines were transferred to a T25 flask containing 5 ml IL-7 supplemented media.

2.34 Generation of A-MuLV pro-B cell lines

PIP-ER transgenic mice express a fusion protein consisting of the estrogen receptor ligand binding domain fused to the N-terminus of *Irf4* cDNA, under the control of the *λ5* promoter cassette (Sabbattini et al., 1999). Pro-B cells from PIP-ER transgenic mice were immortalised using the following protocol.

Cells of the AB010 cell line (Chen & Rosenberg, 1992), which secretes A-MuLV, were grown for two days past confluency in DMEM. The virus containing supernatant was removed and concentrated. Bone marrow cells were flushed from the femurs of 6-8 week old PIP-ER D mice, with 10 ml PBS. Cells were infected immediately following extraction or after expansion for five to six days in 10 ml pro-B cell media at 33ºC with 5% CO_2 in a humidified atmosphere. Red blood cells were lysed for ten minutes by suspension in 168 mM NH_4Cl.

Infection with A-MuLV was performed by the addition of 1 ml of primary cells at a concentration of 2 x 10^6 cells/ml to 1 ml of concentrated viral supernatant, in the presence of 8 µg Polybrene (Millipore), to increase infection efficiency. Cells were incubated at 37°C for 2.5 hours with agitation every 20 minutes and plated at concentrations of 1 x 10^6 cells/ml in semi-solid agar (RPMI media supplemented with 20% foetal calf serum, 2 mM L-glutamine, 50 µg/ml streptomycin, 50 U/ml penicillin, 50 µM β-mercaptoethanol and 0.3% bacterial agar (Oxoid Ltd)). Following infection, cells were maintained in a humidified atmosphere at 37°C with 5% CO_2 and maintained by adding 1 ml of semi-solid agar every four days.

2.35 Generation of synergistic activation mediator (SAM) cell lines

The synergistic activation mediator (SAM) transcription activator system is based on the expression of three components: a catalytically inert Cas9 protein fused to four copies of the herpes simplex trans-activator VP16, a single guide RNA (sgRNA) modified with MS2 aptamers and an MS2-P65-HSF1 fusion protein. The expression of these components enables the targeting of a strong transcriptional activator complex to a specific location, determined by the sgRNA sequence, resulting in the upregulation of transcription at the targeted gene. As IRF4-ER expression was undetectable in the 1D1 pro-B cell line, I sought to increase the expression of the transgene by targeting the SAM transcriptional activator to the transgene promoter.

The three components were introduced into the 1D1 pro-B cell line using lentiviral transduction. Lentivirus was prepared from lenti-dCas9-VP64 (Addgene plasmid # 61425) and lenti-MS2-p65-HSF1 (Addgene plasmid # 61426) as described in Section 2.28. 1D1 cells (5 x 10^5) were transduced with dCas9-VP64 lentivirus by spin-infection (Section 2.30) and after 48 hours infected cells were selected with 8 µg/ml Blasticidin (Alfa Aesar J61883WDL) for one week. Following selection, 5 x 10^5 dCas9-VP64 expressing cells were spin-infected with MS2-p65-HSF1 lentivirus and selected with 400 µg/ml Hygromycin (Alfa Aesar J60681.MC) after 48 hours.

To generate the sgRNA encoding lentivirus, guide oligonucleotides targeting the *Igll1* (λ5) promoter were phosphorylated by combining 10 µM top strand and bottom strand, 1 x NEB T4 DNA ligase buffer and 10 U T4 PNK (NEB) in a final volume of 10 µl. The mixture was incubated at 37ºC for 30 minutes before heat inactivation and denaturing the oligos by incubation at 95ºC for five minutes. The oligos were annealed by slowly decreasing the temperature by 5ºC per minute, until the oligo duplex was at room temperature. Annealed oligos were then cloned into lenti-sgRNA-Zeo (Addgene plasmid # 61427) by simultaneous digestion with BsmBI (NEB) and ligation. The annealed oligos were diluted 1:250 and 1 µl was combined with, 100 ng lenti-sgRNA-Zeo, 1x NEB Buffer 3.1, 1 µM ATP, 3000 U T7 DNA ligase (NEB) and 10 U BsmBI, in a final volume of 20 µl. The mixture was cycled six times between 37ºC and 21ºC with five minutes at each temperature. As the optimal reaction temperature of BsmBI is 50ºC, a 30-minute incubation step at 50ºC was performed after the cycles of digestion/ligation, to ensure complete digestion of the lenti-sgRNA-Zeo and avoid unnecessary background colonies.

Following bacterial transformation, screening and large-scale plasmid extraction, lentivirus was generated from the lenti-λ5-sgRNA-MS2-Zeo vector and used to transduce dCas9-VP64/MS2-p65-HSF1 expressing cells. Two days after transduction, cells were selected with 100 µg/ml Zeocin (Alfa Aesar J67140) for one week prior to analysis.

2.36 Generation of MSCV-IRF4-ER cell lines

The human oestrogen receptor hormone binding domain, from MyoD-ER, was fused to the N-terminus of *Irf4*, by overlap extension PCR (Sarah Bevington, unpublished) and cloned into pRc/CMV (Invitrogen). The cDNA encoding the IRF4-ER fusion was PCR amplified using IRF4-F and BGHR and cloned into the blunted EcoRI and XhoI sites of MSCV-IRES-GFP (Addgene #20672). Retrovirus was produced as detailed in Section 2.28 from MSCV-IRF4-ER-IRES-GFP (Figure 2.5) and the A-MuLV infected PIP-ER pro-B cell line 1D1 was transduced by spin-infection. Infection of 1D1 cells was examined by GFP expression using flow cytometry. Due to the low infection efficiency observed with this construct, infected cells were enriched by three rounds of FACS sorting based on GFP expression and subsequent culture. To generate monoclonal cell lines, 1 x 10^4 cells expressing the highest level of the GFP reporter (Figure 2.4) were plated in 10 ml of semi-solid agar (Section 2.34) and maintained for approximately 10 days. Macroscopic colonies were transferred to 1 ml of RPMI in 24 well plates and cultured until sufficiently dense for transfer to a T25 flask and analysis.

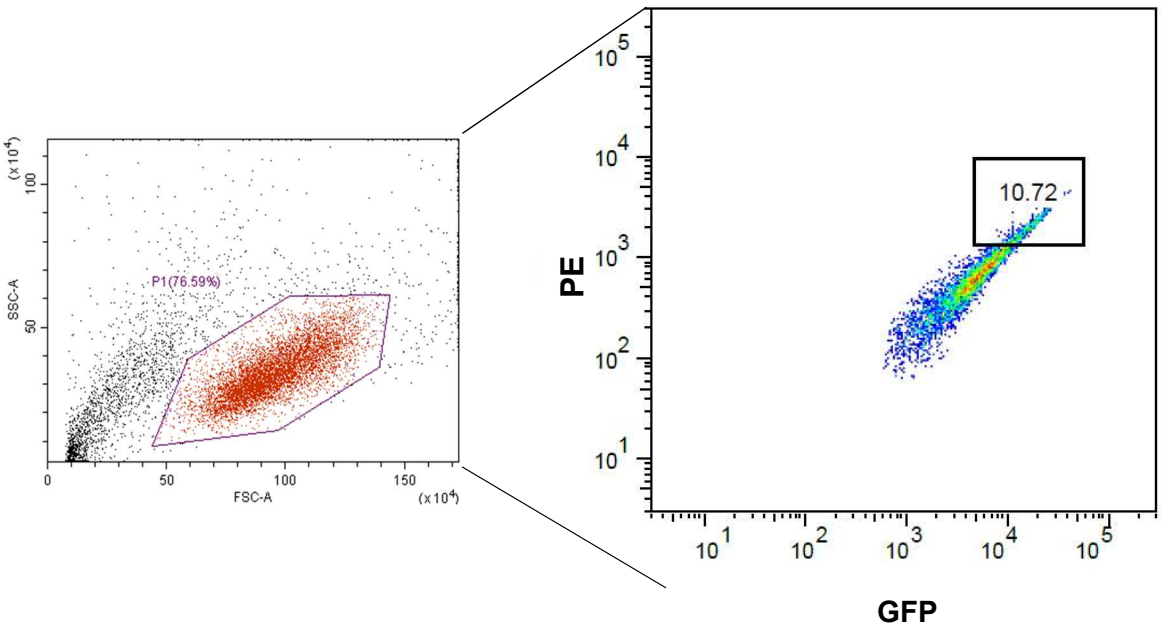


Figure 2.4: Gating strategy for the isolation of the cells with highest transgene expression. Viable lymphocytes were first gated on forward scatter (FSC-A) and Side scatter (SSC-A). The top 10% GFP expressing cells were gated and collected to generate monoclonal cell lines.

2.37 Generation of IRF4-DD cell lines

The E.coli Dihydrofolate reductase (ecDHFR) destabilisation domain (DD) with a C-terminal GGGGS linker and HA tag was ordered as a GeneBlock^{Tm} from Integrated DNA Technologies. pBluescript-IRF4 was linearized to remove the *Irf4* stop codon, and the ecDHFR DD was ligated in frame with the *Irf4* C-terminus. Next, the lenti-CRISPRv2 vector (Addgene #49535) was modified by the removal of the sgRNA scaffold by digestion with Acc651 and EcoRI followed by the removal of overhangs, circularisation and transformation into *E. coli*. The modified vector was then linearized by PCR (Table 2.9) to remove *Cas9,* generating the lenti-P2A-puro vector (Figure 2.6A). Following the generation of lenti-P2A-puro, *Irf4-DD* was amplified and cloned into this vector to generate lenti-IRF4-DD-P2A-puro (Figure 2.6B). Lentivirus was generated from lenti-IRF4-DD-P2A-Puro (Section 2.28) and 1D1 cells were infected as above. Infection was repeated on three subsequent days to increase the expression of the transgene and increase the number of infected cells. Two days after the final infection, 0.2 or 2 µg/ml Puromycin (Alfa Aesar J61278) was added and cells were selected for one week.

2.38 Generation of MSCV-Irf4-ER (T2) cell lines

To avoid activation of IRF4-ER by estrogenic compounds present within culture medium and increase the sensitivity to 4-OH tamoxifen, point mutations M543A/L544A were made within the ER domain of IRF4-ER by Q5® Site-Directed Mutagenesis (NEB; primers are given in Table 2.9). Retrovirus was prepared (Section 2.28) and infections were performed as described in Section 2.31. GFP positive cells were sorted and monoclonal cell lines generated as described in Section 2.36.

2.39 Knockout of the PU.1/IRF4 binding site in the HS6 and Eλ3-1 enhancers in 1D1-T215 cells

Large-scale deletion

Transduction of pro-B cells is highly inefficient therefore to increase the percentage of infected pro-B cells expressing two sgRNAs, I modified the lenti-CRISPRv2 plasmid to enable the co-expression of two sgRNAs. Two sgRNAs flanking Eλ3-1 or HS6 were designed using the MIT CRISPR design tool (http://crispr.mit.edu/) and modified to generate complementary overhangs to those produced by the BsmBI restriction enzyme. The sgRNA specifying oligonucleotides were annealed and cloned into lenti-CRISPRv2 in the same manner as described in Section 2.35 for lenti-sgRNA-MS2-Zeo. The U6 promoter to sgRNA scaffold was PCR amplified from the lenti-CRISPRv2 vector encoding the 3' sgRNA and cloned into the NheI and EcoRI sites of the lenti-CRISPRv2 vector encoding the 5' sgRNA (Figure 2.7). Retrovirus was generated from this vector and 1D1-T215 cells were transduced (Sections 2.28, 2.31). Transduced cells were selected after 48 hours with 0.25 µg/ml Puromycin for one week before the generation of monoclonal cell lines using semi-solid agar.

PU.1/IRF4 binding site mutation

With the exception of the insertion of the second sgRNA cassette, all steps were performed as described for the large-scale deletion method.

PU.1/IRF4 binding site deletion

Two CRISPR sgRNA specifying oligos flanking the PU.1/IRF4 sites in each enhancer element (Eλ3-1 and HS6) were designed as above. Eλ3-1 guide 2 / HS6 guide 1 oligos were annealed and cloned into lenti-CRISPR v2 and Eλ3-1 guide 3 and HS6 guide 3 oligos were cloned into lenti-sgRNA-MS2-Zeo by Golden gate cloning. Lentiviral production was performed as described (Section 2.28). Due to the very low transduction efficiency of PIP-ER T215 cells, infections were performed in a sequential manner. $5x10^5$ PIP-ER T215 cells were spin-infected (Section 2.31) with 500 µl of Eλ3-1 guide 3 or HS6 guide 3 lentiviruses.

After 48hrs, transduced cells were selected with 100 µg/ml Zeocin. Following one week of selection, cells were spin-infected with $E\lambda_{3\text{-}1}$ guide 2 or HS6 guide 1 lentivirus and selected for one week with 0.25 µg/ml puromycin to identify infected cells. Monoclonal cell lines were generated using semi-solid agar and clones were screened for knockouts by PCR using the primers HS6delF2 and HS6del4R (HS6) and $E\lambda_{3\text{-}1}$ del F/R ($E\lambda_{3\text{-}1}$). Monoclonal cell lines appearing homozygous for the deletion had these regions amplified using the above primers and cloned for analysis by Sanger sequencing.

J. ChIP and 3C

2.40 Chromatin Immunoprecipitation (ChIP)

ChIP was performed according to Boyd and Farnham with minor modifications (Boyd & Farnham, 1999). 2 x 10^7 cells were centrifuged at 504 x *g* for three minutes at 4°C and resuspended in 25 ml PBS. DNA was cross-linked by the addition of formaldehyde to a final concentration of 0.8%. Cells were incubated at room temperature for 10 minutes with gentle agitation after which the formaldehyde was neutralised by five minutes incubation with 125 mM glycine, with gentle agitation. Cells were pelleted by centrifuging at 700 x *g* for four minutes at 4°C and washed three times in ice cold PBS. Cells were resuspended in 2 ml lysis buffer 1 (10 mM Tris pH 8, 10 mM NaCl, 0.2% NP-40, 10 mM sodium butyrate, 50 µg/ml PMSF, 1 x Complete protease inhibitor cocktail (Roche)) and incubated for 10 minutes on ice. Cells were centrifuged at 504 *g* for five minutes and resuspended in 1 ml lysis buffer 2 (10 mM Tris pH 8.1, 10 mM EDTA, 1% SDS, 10 mM sodium butyrate, 50 µg/ml PMSF, 1 x Complete protease inhibitor cocktail (Roche)) and incubated for 10 minutes on ice. Chromatin was fragmented by sonication with a Sanyo Soniprep 150 at level 10 for 6 x 20 seconds with one minute on ice between each burst of sonication.

Debris was removed by centrifugation at 1200 *g* for 10 minutes at 4°C. Chromatin was diluted fivefold with dilution buffer (20 mM Tris pH 8, 150 mM NaCl, 2 mM EDTA, 0.01% SDS, 1% Triton, 50 µg/ml PMSF, 10 mM sodium butyrate, 1 x Complete protease inhibitor cocktail (Roche)) and pre-cleared with 30 µl herring sperm/ protein G beads (Sigma) for 30 minutes at 4°C on a rotating wheel. Following pre-clearing, a 150 µl aliquot was taken as input and the supernatant was transferred to a fresh tube. Immunoprecipitation was carried out using 10 µl anti-IRF4 antibody (Proteintech 11247-2-AP) at 4°C overnight with rotation.

To the chromatin, 80 µl herring sperm/protein G beads were added followed by incubation at 4°C for 2 hours. Immunoprecipitated complexes were pelleted by centrifugation at 254 *g* for two minutes and washed twice in wash buffer 1 (20 mM Tris pH 8, 50 mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton x 100), twice in high salt buffer (10 mM Tris pH 8, 500 mM NaCl, 2mM EDTA, 0.01% SDS, 1% Triton x 100), once in wash buffer 2 (10 mM Tris pH 8, 0.25 M LiCl, 1 mM EDTA, 1% NP-40, 1% Sodium deoxycholate) and twice in TE. Bound DNA was eluted by resuspension in 200 µl elution buffer (100 mM $NaHCO_3$, 1% SDS). Crosslinks were reversed, and RNA was removed by the addition of 30µg RNaseA, 300 mM NaCl and incubation at 65°C for five hours. Following the reversal of crosslinks samples were incubated with proteinase K at 45°C overnight. DNA was purified by phenol-chloroform extraction and ethanol precipitation and subjected to qPCR. All samples were normalised to an intergenic control region (IntIII).

2.41 Chromosome conformation capture (3C)

Approximately 1 x 10^7 cells were resuspended in in 10 ml PBS + 10% FCS and fixed by incubation with 2% formaldehyde for 10 minutes. Formaldehyde was quenched by the addition of glycine at a final concentration of 140 mM. Samples were centrifuged at 500 x *g* for 5 minutes, resuspended in 5 ml lysis buffer (10 mM Tris pH 8, 10 mM NaCl, 0.2% NP-40, 50 µg/ml PMSF, 1 x Complete protease inhibitor cocktail (Roche)) and incubated on ice for 45 minutes. Nuclei were centrifuged at 700 x *g* for 5 minutes at 4ºC, resuspended in 1 ml of PBS before centrifugation at 1000 x *g* for 2 minutes. Nuclei were flash frozen in a dry ice/ethanol bath and stored at -80ºC.

Stored nuclei were resuspended in 500 ul of 1.2x NEB DpnII buffer and permeabilised by the addition of SDS to a final concentration of 0.3%. Nuclei were incubated for one hour at 37ºC with rotation. To prevent aggregation nuclei were agitated by pipetting every 20 minutes. Triton X-100 was added to a final concentration of 3% and the nuclei were incubated at 37ºC for one hour. DpnII (100 U; NEB R0543M) was added and nuclei were incubated at 37ºC for four hours before an additional 100 U of DpnII respectively were added and the chromatin was digested overnight. Chromatin digestion was examined, and nuclei were digested for an additional four hours. The restriction enzyme was heat inactivated at 65oC for 20 minutes. Ligation was performed in 7 ml of 1 x ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM MgCl2, 1 mM ATP, 5 mM DTT) with 25 U T4 DNA ligase (Roche) at 16ºC overnight.

The 3C template was extracted by the addition of 7 ml phenol and centrifugation at 1000 x *g* for 15 minutes. To remove the phenol, 7 ml chloroform was added and the 3C template was centrifuged at 1000 x *g* for 15 minutes. The template was precipitated by the addition of 0.7 ml NaOAc, 14 ml ethanol and 100 µg glycogen. Precipitations were incubated on dry ice for a minimum of one hour before centrifugation at 1000 x *g* for 30 minutes. The precipitate was washed with 70% ethanol and centrifuged 1000 x *g* for 15 minutes before residual ethanol was removed and the pellet dried. The 3C template was resuspended in 100 µl TE enhanced by incubation at 37ºC overnight.

2.42 Preparation of BAC template

Bacterial artificial chromosome (BAC) Rp23-24i11, obtained from CHORI (Children's Hospital Oakland Research Institute), containing the 3' half of the Igλ, was used to generate the 3C control template. As DpnII is sensitive to Dam methylation, the isoschizomer Sau3A1 (NEB) was used to digest the BAC before it was ligated at a high concentration to generate all possible 3C ligation products. 20 µg BAC DNA was digested overnight at 37°C with 25 U of Sau3A1 in a total volume of 500 µl. Contaminants were removed from the digested DNA by phenol-chloroform extraction before the DNA was recovered by ethanol precipitation and resuspended in 40 µl. DNA was ligated in a final volume of 60 µl with 2000 cohesive end units/ml of T4 DNA ligase (NEB) at 16^{o}C overnight before purification by phenol/chloroform extraction and ethanol precipitation. The precipitated control template was resuspended in 100 µl TE.

2.43 3C primer design

Due to the high sequence homology between restriction fragments in the *Igλ* locus, fragments of low sequence similarity were identified using a custom python script. This identified all possible restriction fragments and compared the final 200 bp of each fragment using BLAST. All fragments with greater than 90% similarity were removed, and the remaining fragments were used for the design of appropriate primers using PRIMER3 (http://primer3.ut.ee/), with parameters set to qPCR.

2.44 Analysis of 3C templates by qPCR

For each primer pair, qPCR was performed in duplicate in 10 µl volumes with 5 µl of 1:5 diluted 3C template, 400 pM each primer, 100 pM 5' nuclease probe and 5 µl qPCRBIO probe mix (PCRBIO PB20.21-05). Additionally, for each viewpoint a standard curve consisting of 1 µl of a 1 x 10^{-2} - 1 x 10^{-5} dilutions of the control BAC template was used to normalise primer efficiencies. Furthermore, all 3C samples were normalised by analysis of an interaction in the *Ercc3* locus which is expected to be consistent across all cell types (Palstra et al., 2003).

K. Bioinformatic methods

2.45 Motif analysis

The HOMER program homer2 was used to predict transcription factor motifs present within specific regions. The command was as follows: homer2 find -m <Concatemerised motif files in HOMER format> -i <sequence in fasta format>. The motif files used for transcription factor analysis were supplied by HOMER v4.9.

2.46 Analysis of published ChIP-seq data

Read files in fastq format were downloaded from the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena) and sequencing adapters were removed by Trim_Galore. Reads were aligned to the *Mus musculus* (mm10) genome using Bowtie2 using default parameters and multimapping reads in addition to poor quality alignments were removed using samtools. Peaks were called using MACS2, for transcription factors the default parameters were used, however, for histone modifications the '—broad' peak calling mode was used. Visualisation was performed using the Integrated Genome Browser IGV after converting the bedgraph output from MACS2 into a binary 'tiled' format using IGV tools.

2.47 Analysis of published ATAC-seq data

Read files (Table 2.3) were downloaded from the ENA, trimmed and aligned as above with one exception, the Bowtie2 max insert parameter (--X) was set to 2000 to enable the mapping of large inserts that are typical of ATAC-seq. Multimapping reads were removed by samtools before peak calling. ATAC-seq peaks were called by MACS2 with the parameters –nomodel –shift 150 –extsize 300.

2.48 RNA-seq analysis

The adapters from read files downloaded from the ENA were trimmed by trim_galore before alignment to the mm10 genome by HISAT2, using the default parameters. For paired end sequencing data, alignment was performed in paired end mode using the --paired flag. Transcripts were assembled and merged using Stringtie. Visualisation of transcript 5' ends was performed using IGV.

2.49 Hi-C analysis

Read files (fastq) were downloaded and trimmed as above, before being aligned separately to the mm10 genome using Bowtie2. The HOMER program makeTagDirectory was used to process the aligned reads into a tag directory for downstream analysis. Significant interactions occurring in the *Igλ* locus were identified with the HOMER script analyzeHiC. This command was run with the following parameters: -res 20000 -interactions <interaction_file> -pos <region of interest> --center. This script identifies and reports pairs of regions that have a significantly increased number of interactions than would be expected from the background model. The 'center' argument re-centres the regions outputted to the average of the position of the Hi-C reads participating in the interaction. The visualisation of the Hi-C interactions was performed using Circos.

2.50 Igλ locus homology analysis

Analysis of the sequence homology between the 5' and 3' halves of the Igλ locus was performed using a custom python script. The 3' half of the Igλ locus (chr16:19,001,956-19,089,641) was split into 1 kb regions which were then compared to a BLAST nucleic acid database generated from the 5' half of the locus (chr16:19,145,047-19,270,546). Regions with similarity greater than 90% were reported and Circos was used to visualise the results.

L. Software

Table 2.1: Software utilised during the course of this project.

Software	Version	URL
HOMER	4.9	http://homer.ucsd.edu/homer/download.html
Trim_Galore	0.5.0	https://github.com/FelixKrueger/TrimGalore
Bowtie2	2.3.4.2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools	1.9	https://github.com/samtools/
Macs2	2.1.0	https://github.com/taoliu/MACS
IGV	2.4.2	https://software.broadinstitute.org/software/igv/
IGV tools	2.3.98	https://software.broadinstitute.org/software/igv/igvtools
Hisat2	2.0.5	https://ccb.jhu.edu/software/hisat2/index.shtml
Stringtie	1.3.4	http://ccb.jhu.edu/software/stringtie/
Circos	0.69	http://circos.ca/software/download/
BLAST	2.7.1	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/

M. Accession numbers

Table 2.2: Accession numbers of published ChIP-seq data analysed.

Factor	Cell type	Sample Accession	Input Accession	Additional notes
CTCF	pro-B	GSM672401	Not available	$Rag2^{-/-}$
E2A	pro-B	GSM546523	GSM546540	$Rag1^{-/-}$
H3K27Ac	pro-B	GSM1463433	GSM1463439	N/A
H3K4me1	pro-B	GSM1463434	GSM1463439	N/A
IRF4	pro-B	GSM1296534	GSM1296537	$Rag2^{-/-}$
MED1	pro-B	GSM1038263	GSM1038264	v-abl immortalised line 38B9
MEF2C	pro-B	GSM1894135	GSM1894134	N/A
P300	pro-B	GSM1290115	Not available	Haftl derived line c10
PU1	pro-B	GSM1290093	Not available	Haftl derived line c10
RAD21	pro-B	GSM672403	Not available	$Rag2^{-/-}$
YY1	pre-B-like	GSM1897389	GSM1897390	Expresses Igμ

Table 2.3: Accession numbers of published ATAC-seq and RNA-seq data.

Cell type	Sample Accession
pre-B	GSM1545327
pre-B	GSM1545327
pre-B	GSM1849923
pre-B	GSM1849924
pre-B	GSM1849925

N. Oligonucleotides used

The primers used during the course of this project are shown in the following sections. Primers were ordered from either Sigma-Aldrich (Poole, UK) or Integrated DNA Technologies (IDT; EU) and were purified by desalting. Primers were resuspended to 100 µM in ddH_2O before use.

Table 2.4: Oligonucleotides used for RT-qPCR.

Name	Sequence
Cxcr4_F	CCGGTACCTCGCTATTGTCC
Cxcr4_R	CTATCGGGGTAAAGGCGGTC
HPRT_F	GGGGGCTATAAGTTCTTTGC
HPRT_R	TCCAACACTTCGAGAGGTCC
Irf4_F	GGAGCAAAGCAGCTCACTTTG
Irf4_R	CATGGGGTGGCATCATGTAG
Irf4-ERF1	TGAAGTGCAAGAACGTGGTG
Irf4-ERR1	TGCCCGTCTCCAAGTTCATA
J1GT3_F	ACTTGAGAATAAAATGCATGCAAGG
J1GT3_R	TGTGGCCTTGTTAGTCTCGA
Pu.1_F	CTGAGAACCACTTCACAGAGCTGCA
Pu.1_R	TGGGCTGGGGACAAGGTTTGATAAG
V1nctF (Vλ1ts1R)	GTGAATTATGGCCTGGATTTCACT
V1ntcR(Vλ1-GSP2)	GAGCGACAAGTGAGTGTGAC

Table 2.5: Oligonucleotides used for 5' RACE.

Name	Sequence
RACE-dT-adaptor-L	CCGGACTCGAGTCGACATCGATTTTTTTTTTTTTTTT
V1_RACE_+325_R_XbaI	AATCTCTAGACCTGAGAATCTGGCAGGAACAC
V1_RACE_+367_F_HindIII	CTGCAAGCTTCATCACAGGGGCACAGACT
VL1-Upstream-2	TTACCACATGATAATGCCCTACTCACAGC
Vλ1-GSP2-2	AAACAGTCACACTCACTTGTCGCTC
Vλ1-GSP4-2-HindIII	CGGGAAAGCTTCCCTGCAAACAGATGAGAAATCCAGT
Vλ1-Upstream-F	TGCAGTTGAGCCTTTCCAAA

Table 2.6: Oligonucleotides used for ChIP.

Name	Sequence
IntgeneIIIF	CAAGGAAAGGCCAACCAATA
IntgeneIIIR	TAACCCTTTCCCCAGCTCTT
HS6_ChIP_F	AGGCAGCATCAGGCCTTAGGACTA
HS6_ChIP_R	AGCATGACAAACAGAACCAGGTGT
HSCl1_F	AGGAATAGAACATGAGTAAGCTGT
HSCl1_R	TGCATGAGTTGAGTCTAGTCCTGAGT
HS7_ChIP_F	ACCTTCTCTTTGCTCTGCAGGCA
HS7_ChIP_R	ACCCAGAGGCTTTCCTGCAATGT
HSVl1_ChIP_F	ACACTGTAAGGGGCCAATGA
HSVl1_ChIP_R	GCAGCTTGGCAAATAAATGTAGG
LambdaenF ($E\lambda_{3\text{-}1}$)	GACATTACAAGCTCTGTGGAAG
LambdaenR ($E\lambda_{3\text{-}1}$)	GCTAATGGACTTGGTTTCAGTTCC

Table 2.7: sgRNA oligonucleotides. The sequence of the sgRNA is underlined.

Name	Forward	Reverse
Eλ3-1 g2	CACCGTGCCTTGCTGCTAATGGACT	AAACAGTCCATTAGCAGCAAGGCAC
Eλ3-1 g3	CACCGTTTAGCTTGTGGACTCTCAA	AAACTTGAGAGTCCACAAGCTAAAC
HS6 IRF4 g3	CACCGTTGACTGAAACAAAAACTCA	AAACTGAGTTTTTGTTTCAGTCAAC
HS6 3'A	CACCGCCCTCCCAGTTTTCCCAATG	AAACCATTGGGAAAACTGGGAGGGC
HS6 5'A	CACCGCAGAATTTCTGGTCATGAGG	AAACCCTCATGACCAGAAATTCTGC
HS6 IRF4 g1	CACCGCAGAAGTGAAACCAAGCTGA	AAACTCAGCTTGGTTTCACTTCTGC

Table 2.8: Oligonucleotides used for the confirmation of HS6 and Eλ3-1 deletions.

Name	Sequence
HS6_del_F2	AGGCAGCATCAGGCCTTAGGACTA
Eλ3-1_del_F	GACATTACAAGCTCTGTGGAG
Eλ3-1_del_R	CCTTCCAGGATTCAAGTGTT
HS6_del_R4	GGGTAGGGTTGATTGTGGAGGT
HS6_del_R2	TGTGTGGAAGGTCATCAGCA

Table 2.9: Oligonucleotides used for cloning.

Name	Sequence
Cas9_del_F	GCAACAAACTTCTCTCTGC
Cas9_del_R	CTGTGTTCTGGCGGCAAAC
ecDHFR_F	ATGGGAGGAGGGGGCTCAAT
ecDHFR_R	TCAAGTACTGGCGTAATCAGGAACA
HS6_BamHI_F	GTTAGGATCCAGGGTTGATTGTGGAGGTAGG
HS6_SalI_R	TTTTGTCGACCAAACGCACACACACGAAAC
Irf4_F	ATGTCTGCTGGAGACATG
BGHR	TAGAAGGCACAGTCGAGG
IRF4-ER_M543A-L544A-F	GCTGCTGGAGGCGGCTGACGCCCACCGC
IRF4-ER_M543A-L544A-R	AGGTCATAGAGGGGCACC
pBluescript_IRF4_C_term_F	TATTGGTCCATTTCTCAG
pBluescript_IRF4_C_term_R	CTCTTGGATGGAAGAATG
Vl1_promoter_anti_F	TATATTTGTTCTTTTCCTGGCTTG
Vl1_promoter_anti_R	CATAACTACCATCTTCTTAACAGGTGG
Vl1_promoter_sense_R	CTATAAGCTTCAATATTGGTCAGCAGCAGG
Vl1_promoter_sense1_F	CTAGCTCTGCTTTGTTGCTTGAAA
Vl1_promoter_sense2_F	CAGTGCTAGCTCTGCTTTGTTGCTTGAAACTGC

O. Antibodies

Table 2.10: Antibodies together with their application and amount used per experiment. The antibodies' species of origin and supplier are also noted. WB – Western blotting, ChIP – Chromatin Immunoprecipitation, FC – Flow cytometry.

Antibody	Application	Species	Amount/ dilution	Supplier	Catalogue no.
α-CTCF	ChIP	Rabbit	8 µg	Millipore	07-729
α-E2A	ChIP	Rabbit	16 µg	Santa Cruz Biotechnology	sc-763
α-H2A	WB	Mouse	1:1000	Abcam	Ab-18255
α-IRF4	WB, ChIP	Rabbit	1:1000 WB 4 µg ChIP	Proteintech	11247-2-AP
α-mouse-HRP	WB	Goat	1:10000	GE Healthcare	NXA931
α-PU.1	ChIP	Rabbit	16 µg	Santa Cruz Biotechnology	sc-352
α-rabbit-HRP	WB	Goat	1:10000	Sigma-Aldrich	AP156P
α-YY1	ChIP	Rabbit	16 µg	Santa Cruz Biotechnology	sc-1703
α-β-tubulin	WB	Mouse	1:300	Sigma-Aldrich	T4026
α -CD19-FITC	FC	Rat	8 µl/ml	BD Bioscience	553785
A-CD43-PE	FC	Rat	8 µl/ml	BD Bioscience	553271

P. Plasmid maps

Maps for the plasmids generated during the course of this project are shown below. Plasmid maps were generated using snapgene viewer v.2.2.2.

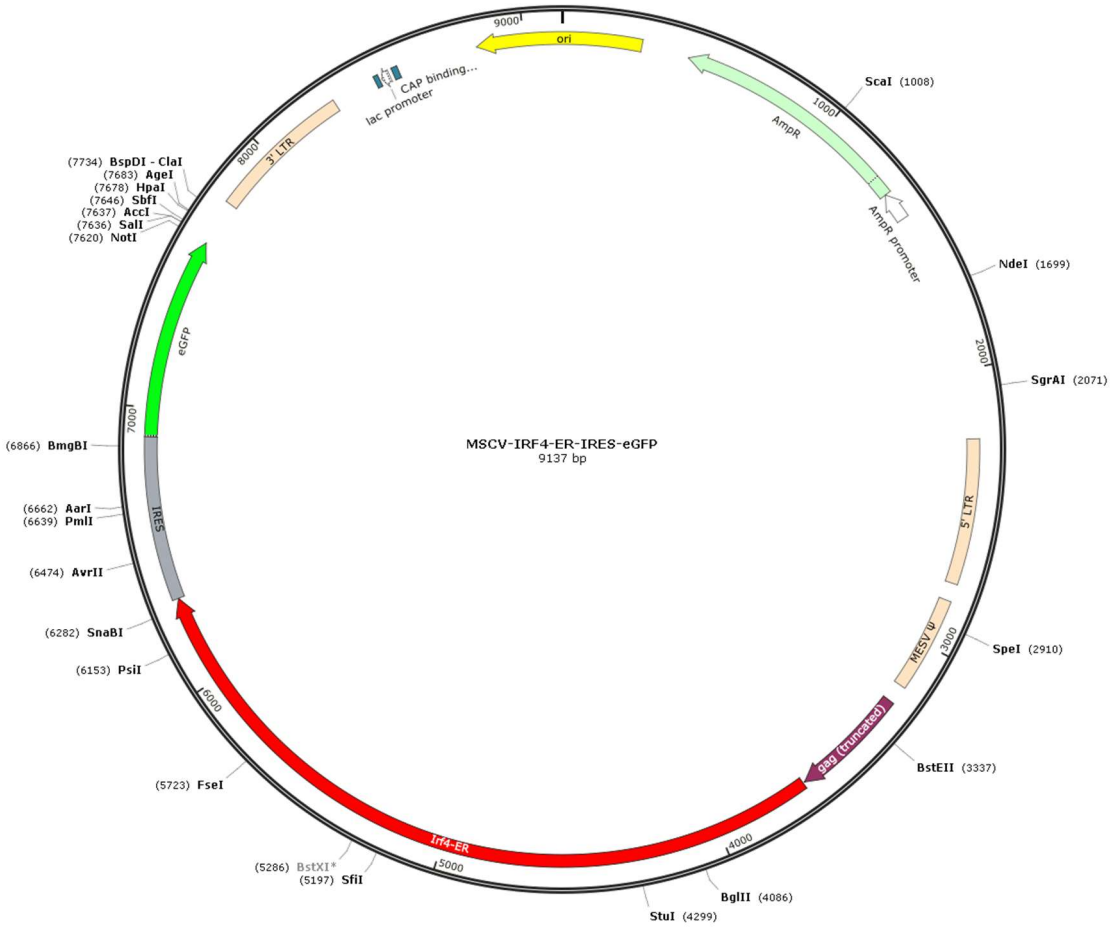
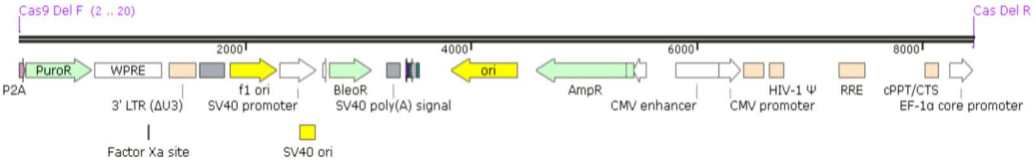


Figure 2.5: Map of MSCV-IRF4-ER-IRES-GFP. IRF4-ER cDNA was cloned into the blunted EcoRI and XhoI sites of MSCV-IRES-eGFP. Unique restriction sites are highlighted.

A)



B)

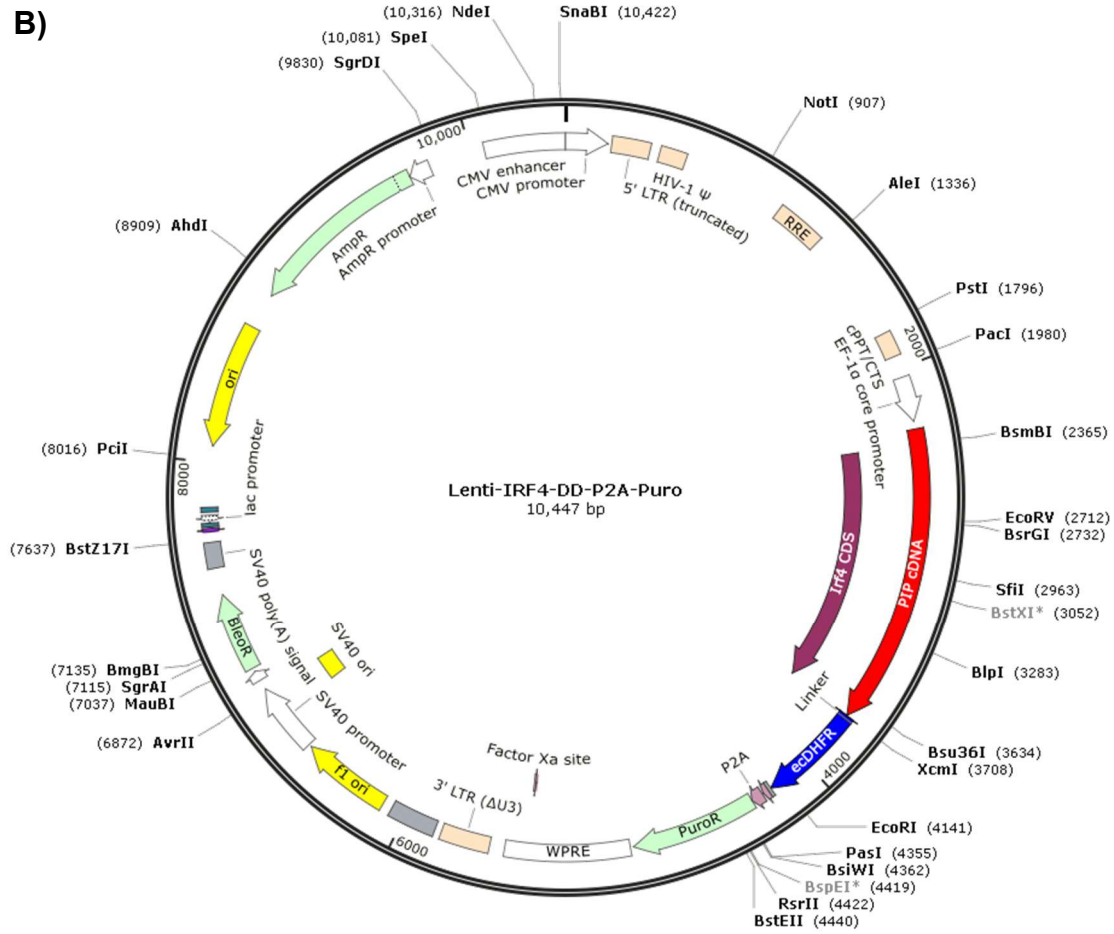


Figure 2.6: Map of lenti-P2A-Puro (A) and lenti-IRF4-DD-P2A-Puro (B). Lenti-P2A-Puro was generated by PCR amplification of a modified lenti-CRISPRv2 vector. IRF4-DD was PCR amplified and ligated to lenti-P2A-Puro to generate lenti-IRF4-DD-P2A-Puro. Unique restriction sites are highlighted.

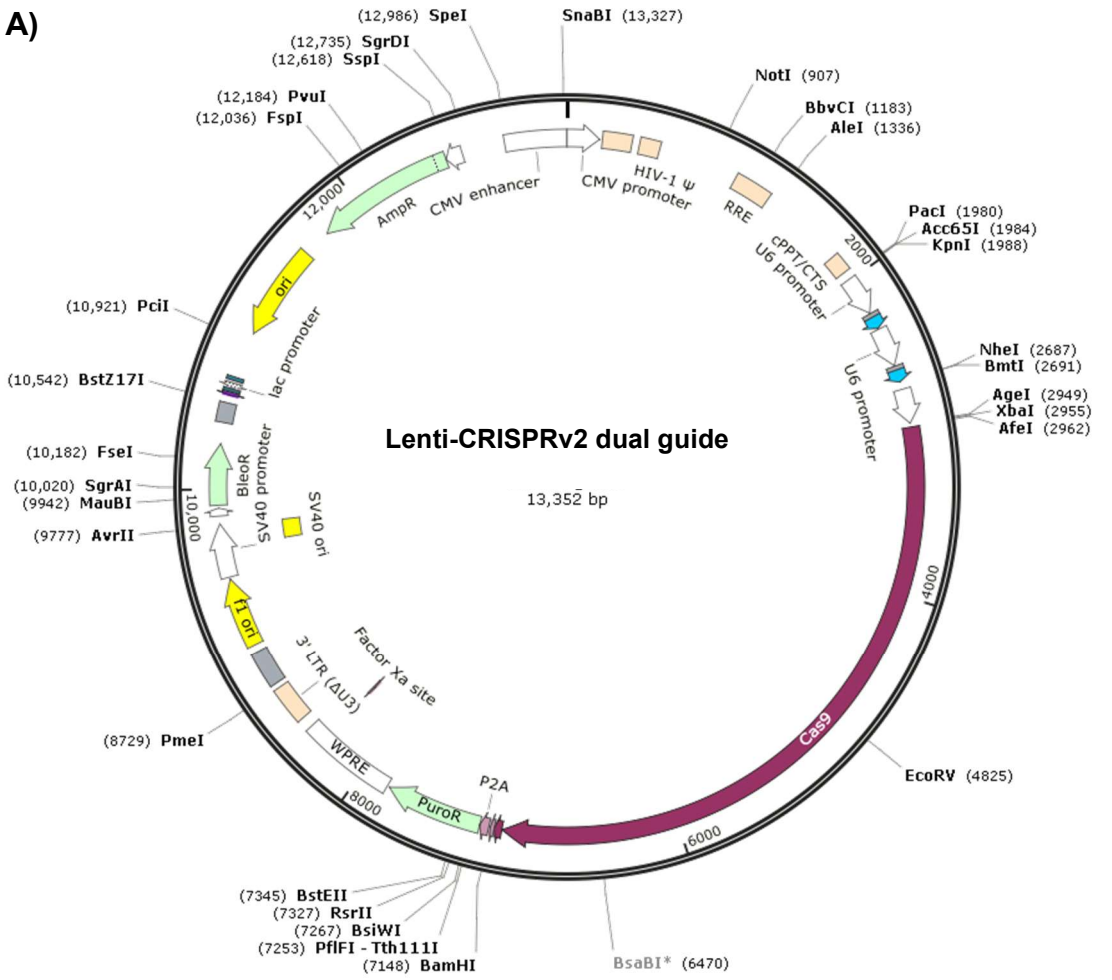


Figure 2.7: Map of a modified lenti-CRISPRv2 vector with two sgRNA cassettes (A). The upstream sgRNA was cloned using BsmBI and the second sgRNA cassette was cloned into the EcoRI and NheI sites of this vector. B) A higher resolution map of the two sgRNA cassettes. Unique restriction sites are highlighted.

Chapter 3 – Generation of a cell line capable of inducing the *Igλ* locus

A. Introduction

V(D)J recombination occurs in a developmental stage specific manner, with recombination at the *Igh* locus occurring at the pro-B cell stage, preceding light-chain recombination (*Igλ* and *Igκ*) that occurs at the pre-B cell stage (Akashi et al., 1999). The murine *Igλ* locus is an ideal model system to investigate gene segment activation as it is the smallest antigen receptor locus, consisting of just six recombining gene segments (Eisen and Reilly, 1985). Furthermore, recombination at the murine *Igλ* locus is unusually biased to events involving *Vλ1* and Jλ1 as approximately 70% of recombination events occur between these gene segments (Boudinot et al., 1994). This enables experimental efforts to be focussed on these gene segments specifically. Most importantly, expressing a pre-B cell level of IRF4 at the pro-B cell stage in transgenic (PIP3) mice is sufficient to prematurely activate *Igλ* recombination and all associated chromatin changes (Bevington and Boyes, 2013).

Subsequently, the Boyes lab developed an inducible V(D)J recombination system by the fusion of the estrogen receptor hormone binding domain (ER-HBD) to the N-terminus of IRF4. This fusion gene is expressed specifically at the pro-B cell stage under the control of the *Igll1* (λ5) promoter and locus control region (LCR; Sabbattini et al., 1999) together with the VpreB exon and intron, in PIP-ER transgenic mice. Following extraction from PIP-ER bone marrow and expansion in culture, PIP-ER pro-B cells can be induced by the estrogen antagonist tamoxifen or its active metabolite 4-hydroxytamoxifen (4-OHT) which activates IRF4-ER and enables the events of *Igλ* activation to be followed temporally.

PIP-ER mice have enabled the first temporal analysis of the regulation of non-coding transcription at a natural locus. However, as pro-B cells exist in relatively low numbers in mouse bone marrow, a limited number cells are obtained per mouse. In addition, pro-B cells have a very limited life span *ex vivo*, and therefore obtaining a sufficient number of cells for experiments such as ChIP, 3C and Capture-C would require the use of a very large number of transgenic mice. Furthermore, additional experimental variability is introduced as this is not a homogenous system and factors such as the age the mouse lead to experimental variability. This chapter describes the generation and analysis of a PIP-ER cell line, that can potentially replace the use of PIP-ER transgenic mice.

Normal mammalian somatic cells proliferate a limited number of times *in vitro,* the maximum number of divisions is referred to as the 'Hayflick limit' (Hayflick, 1965; Hayflick and Moorhead, 1961). Cellular senescence occurs after cell reach this limit and exhibit an array of biochemical and morphological changes (Campisi, 1997). To avoid cellular senescence, several methods have been established to produce immortalised cell lines from primary cells. The most common method is the overexpression of the SV-40 large T-antigen, which represses the tumour suppressors retinoblastoma (Rb) and p53 (Garbe et al., 1999) enabling indefinite proliferation. Alternatively, the overexpression of Telomerase Reverse Transcriptase protein (TERT), results in the lengthening of telomeres thus preventing telomere lengths from reaching that required to cause replicative senescence. Therefore expression of TERT allows cells to avoid replicative senescence and permits these cells, in theory, to divide indefinitely (Blackburn, 2005).

The generation of a pro-B cell line is not compatible with these well-established methods. Without the suppression of heavy and/or light chain recombination, *Igλ* activation cannot be induced. Therefore, to generate a cell line from PIP-ER mice, I required immortalisation procedures that halt the progression of pro-B cells by inhibiting V(D)J recombination. For the generation of pro-B cell lines, there are two well-established methods: Infection of bone marrow with the Abelson Murine Leukaemia Virus (A-MuLV) (Rosenberg et al., 1975) and long-term growth in the presence of IL-7 (Corfe et al., 2007). These methods leverage the signalling pathways regulating pro-B cell development that are orchestrated by IL-7 signalling (Figure 1.7).

Infection of pro-B cells with A-MuLV is the best established method for the generation of pro-B cell lines (Rosenberg, 1976; Rosenberg et al., 1975). A-MuLV infection results in the malignant transformation of early pro-B cells due to the expression of the viral oncogene *v-abl*, a constitutively active ABL tyrosine kinase generated by the fusion of Moloney murine leukemia virus *gag* with endogenous murine *c-abl* (Goff et al., 1980; Reddy et al., 1983). V-abl expression renders pro-B cells factor-independent due to the constitutive activation of JAK1/3 which in turn mediate the phosphorylation of STAT-1,-3 and -5 that are activated by the cytokines IL-3, IL-4 and IL-7 (Danial et al., 1995). This ultimately results in pro-B cells that can proliferate in the absence of IL-7.

As IL-7 signalling activates multiple genes involved in proliferation (e.g. *Ccnd3*), metabolism (e.g. *Glut1*), cell survival (e.g. *Bcl-2*) and represses pro-apoptotic genes (e.g. *Bim*), v-abl expressing pro-B cells can rapidly proliferate and avoid apoptosis, resulting in their indefinite proliferation (Clark et al., 2014). In addition, to maintain pro-B cells in a constantly proliferating state, IL-7 signalling also represses the *Rag* genes via the *PI3K-AKT* pathway and decreases *Igκ* accessibility, resulting in inhibited heavy and light chain recombination (Clark et al., 2014). Analysis of v-abl immortalised cell lines by microarray studies has confirmed that *Rag* expression is significantly downregulated by v-abl (Muljo and Schlissel, 2003), implying these cells are unable to undergo V(D)J recombination. Furthermore, the treatment of *v-abl* expressing pro-B cell lines with the ABL tyrosine kinase inhibitor STI-571 or inactivation of a temperature sensitive *v-abl* mutant is sufficient to enable light chain recombination in these cell lines (Chen et al., 1994; Muljo and Schlissel, 2003), demonstrating that *v-abl* immortalised cell lines are capable of undergoing V(D)J recombination.

A more recently established method to generate pro-B cell lines is the growth of pro-B cells in high concentrations of IL-7 (Corfe et al., 2007). Multiple studies have described the creation of IL-7 dependent cell lines in the presence of stroma to support cell growth (Ishihara et al., 1991; Miyamoto et al., 1998; Park et al., 1990). Corfe et al (2007) demonstrated that pro-B cell lines can be generated in the absence of stromal cells by selecting for CD19+ cells with long term growth potential in the presence of the growth factor IL-7 (Corfe, Gray and Paige, 2007). As IL-7 signalling is inhibited by B cells at developmental stages later than pro-B i.e. from the pre-B cell stage (Clark et al., 2014), IL-7 signalling selectively promotes the expansion and survival of pro-B cells, thus providing a mechanism of specifically generating a pro-B cell line.

In this chapter, I describe the generation and characterisation of a pro-B cell line capable of inducing *Igλ* non-coding transcription.

B. Results

3.1 Immortalisation by long term growth in IL-7

As preliminary data with a v-abl-immortalised cell line (103/BCL-2) suggested that the expression of *v-abl* could potentially reduce the expression of the IRF4-ER transgene, I initially attempted to generate a cell line using the method devised by Corfe et al (Corfe et al., 2007). Bone marrow cells were flushed from the femurs of PIP-ER transgenic mice, and CD19+ cells were isolated by FACS. CD19+ cells were plated at 1 x 10^4 cells/ml in Opti-MEM media containing 5 ng/ml IL-7. Following a two-day period post-sorting where no growth was observed, rapid proliferation occurred for approximately 14 days. From day 12, cell death was observed in all wells, yet substantial variability in the percentage of surviving cells was also observed.

Two distinct cell morphologies were observed by bright field microscopy, with the larger of the two cell types consistently surviving for a longer period of time. Total cell death in approximately 90% of wells was observed by day 21 whereas the surviving cells continued to proliferate. At various points after day 21, total cell death was observed in the majority of the surviving wells. The wells which continued to proliferate were expanded and cell surface markers were analysed, as was the capability for inducible recombination. Six IL-7 dependent pro-B cell lines were generated by this method with a success rate of approximately 1.5% (6/400 wells), which is significantly lower than the 49% of wells that gave rise to cell lines when the procedure was performed by Corfe et al (Corfe et al., 2007).

The IL-7 dependent lines were CD19/CD43 double positive implying that they were indeed pro-B cells (Figure 3.1). Despite evidence that these lines expressed *Irf4-ER*, all IL-7 dependent pro-B cell lines failed to display any evidence of an increase in *Vλ1* or *Jλ1* non-coding transcription following induction by 4-OHT. Additionally, no evidence of *Vλ1-Jλ1* recombination could be detected, implying that *Igλ* activation was not impaired because the locus had already undergone recombination. The most likely explanation for the absence of *Igλ* induction in the IL-7 dependent PIP-ER pro-B cell lines is the complete absence of IRF4-ER. Whilst IRF4 could be readily detected in these cells, IRF4-ER expression was not detectable by Western blotting (Data not shown).

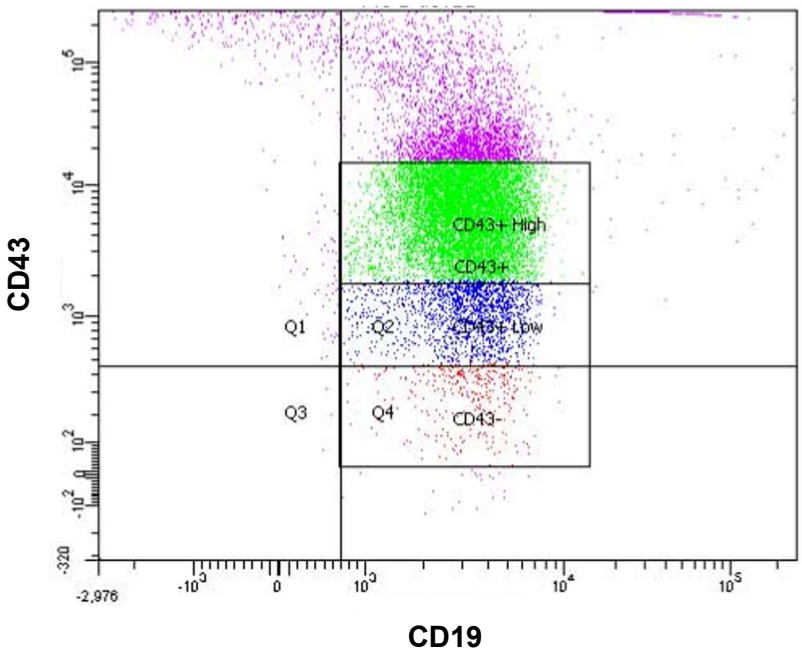


Figure 3.1: IL-7 dependent cell lines have a pro-B cell phenotype. Analysis of CD19 and CD43 expression of a representative IL-7 dependent cell line (d61B2) by flow cytometry. This line expresses CD19 and CD43 which are characteristic markers of pro-B cells.

3.2 Immortalisation by infection with A-MuLV

It was unclear if the lack of IRF4-ER expression in the IL-7 dependent pro-B cell lines was due to the immortalisation procedure used. The λ5/VpreB cassette which drives the expression of IRF4-ER has not been tested in IL-7 dependent pro-B lines, however, A-MuLV immortalised pro-B cells have been demonstrated to retain expression of transgenes under the control of the λ5/VpreB cassette (Sabbattini et al., 1999). I therefore sought to generate pro-B cell lines by infection with A-MuLV.

PIP-ER bone marrow, extracted from six week old mice, was infected with A-MuLV as described (Rosenberg, 1976). Since the length of time in culture can affect the ability of pro-B cells to undergo V(D)J recombination (Sarah Bevington, unpublished observations), PIP-ER bone marrow was infected either immediately following extraction or on the fifth or sixth day of culture and semi-solid agar was used to isolate individual colonies, which were removed from agar at day 11 post infection (Table 3.1). Following removal from agar, substantial cell death was observed for a period of approximately 14 days. After this time, only a minimal level of cell death was detected and all surviving cell lines (Table 3.1) were cryopreserved and analysed.

Table 3.1: The number of A-MuLV infected cell lines generated immediately following extraction or after five/six days in culture. The total number of infections performed at each time point following extraction from the bone marrow and the number of colonies removed from semi-solid agar at 11 days post-infection are displayed together with the number of cell lines generated.

Days pre-infection	Infections	Colonies extracted	No. cell lines
0	4	149	21
5	4	22	1
6	5	10	0

To test if the generated cell lines were at the correct developmental stage, the cell surface markers were screened by flow cytometry by staining with anti-CD19 and anti-CD43 antibodies and analysed. Predominately, the cell lines consisted of a homogenous population of CD19 and CD43 positive cells (Figure 3.2). From the analysis of CD19 (expressed from the pro-B cell stage throughout B-cell development) and CD43 (expressed from CLP to large-pre-B cells) and BP-1, (expressed late-pro-B to small pre-B; not shown), the data strongly indicate that the generated cell lines represent the early pro-B cell developmental stage, previously characterised as Hardy Fraction B (Hardy et al., 1991).

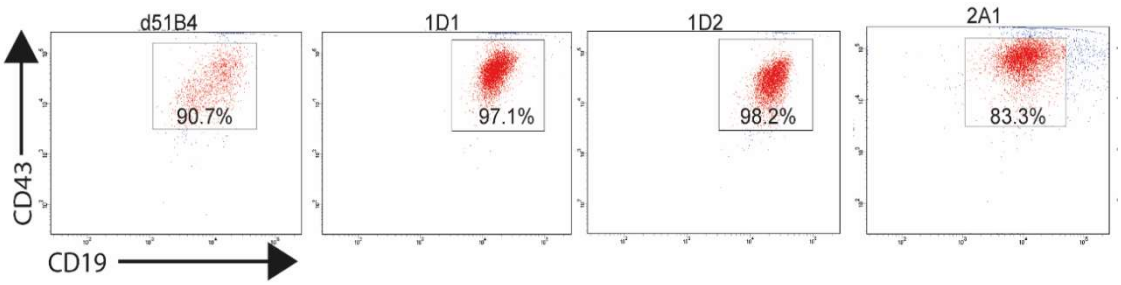


Figure 3.2: The generated cell lines have pro-B cell markers. Cell surface marker analysis by flow cytometry. The black box indicates CD19 and CD43 positive cells and the percentage of the population of each cell line expressing both CD19 and CD43 is given. All generated cell lines express CD19 and CD43, which are characteristic of pro-B cells.

The correct level of IRF4 expression is crucial for the induction of the *Igλ* locus (Bevington and Boyes, 2013). Furthermore, whilst IRF4 can bind weakly to DNA in the absence of a partner (Brass et al., 1996), IRF4 DNA binding is substantially increased by cooperative binding of PU.1 which prevents the auto-inhibition of IRF4 (Eisenbeis et al., 1995). Therefore, A-MuLV immortalised cell lines were first screened by examining the expression of *Irf4* and *Pu1*. Whilst a few cell lines exhibited a minor alteration in *Irf4* and *Pu1* expression compared to primary PIP-ER pro-B cells most pro-B lines had expression levels similar to that of PIP-ER pro-B cells (Figure 3.3).

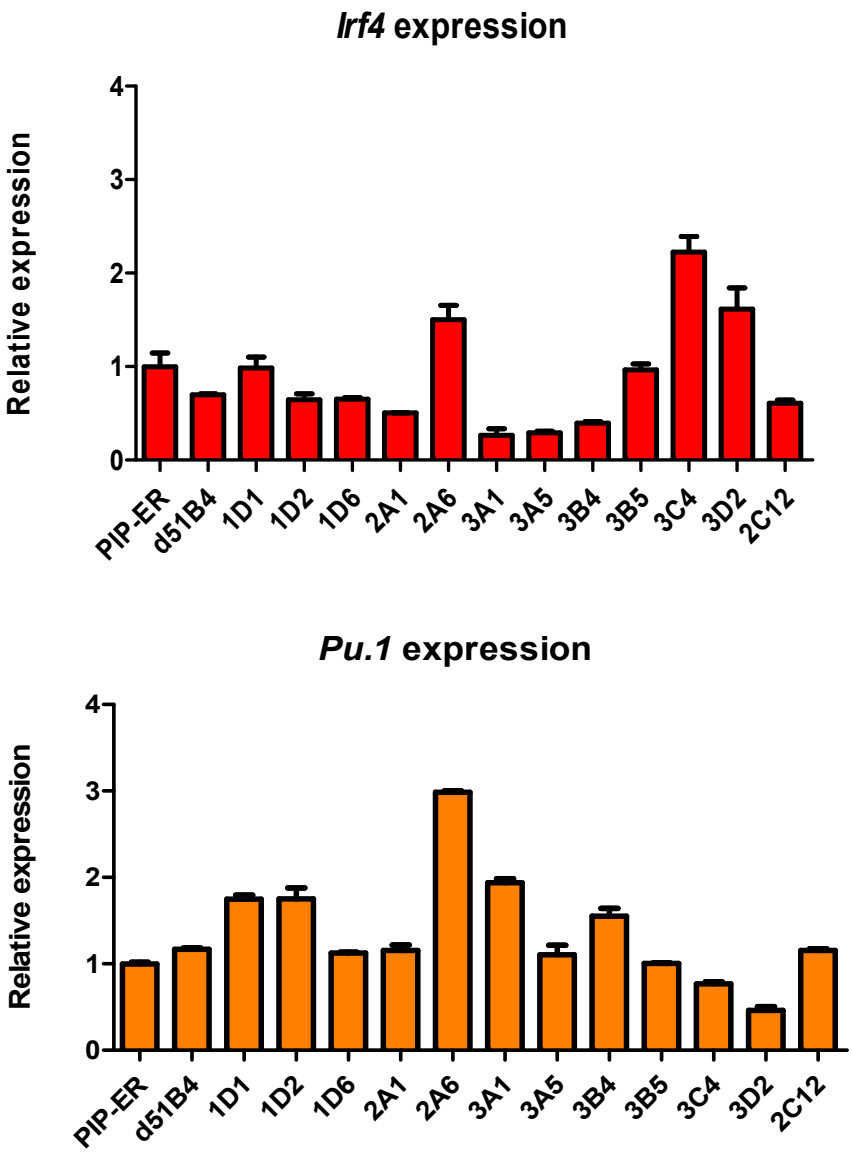


Figure 3.3: *Irf4* and *Pu.1* expression in v-abl immortalised pro-B cells. The expression of A) *Irf4* and B) *Pu.1* in v-abl immortalised cell lines is similar to that of primary pro-B cells for the majority of cell lines. Expression is normalised to *Hrpt* and compared to expression in PIP-ER pro-B cells. Error bars represent standard error of the mean (SEM; n = 3).

To determine if these cell lines were capable of activating *Igλ* non-coding transcription, the upregulation of *Vλ1* non-coding transcription following eight hours of induction with tamoxifen was examined in the highest *Irf4* and *Pu.1* expressing cell lines (1D1, 2A6, 3B5, 3C4 and 2D2). Surprisingly, *Igλ* activation could not be identified in any of the examined cell lines. As it was demonstrated in primary pro-B cells from PIP3 transgenic mice that the level of IRF4 is crucial for *Igλ* activation (Bevington and Boyes, 2013), I therefore examined the expression of IRF4-ER to determine if it was expressed at the correct level in these cells. Whole cell protein extracts were prepared from the cell lines 1D1, 1D2, 3D2 and 2C12 which were selected to maximise the range of IRF4-ER expression. Surprisingly, despite repeated attempts to detect IRF4-ER by Western blotting, the presence of IRF4-ER could not be detected in any *v-abl* immortalised cell line (Figure 3.4).

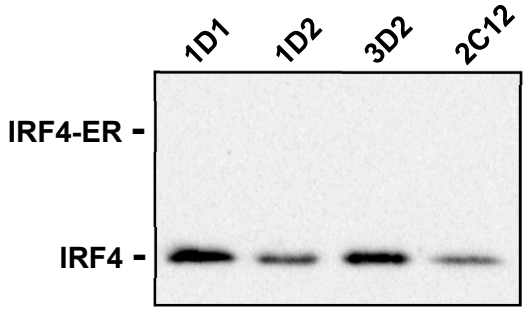


Figure 3.4: Analysis of IRF4-ER expression in *v-abl* immortalised cell lines by Western blotting. IRF4-ER expression could not be detected in any of the cell lines examined, whereas IRF4 is readily detectable. The locations of IRF4 (51 kDa) and IRF4-ER (87 kDa) are indicated.

A possible explanation for the lack of IRF4-ER was that immortalisation by *v-abl* expression silences the λ5 promoter cassette. However, Abelson immortalised cell lines have previously been generated from transgenic mouse lines expressing transgenes from this cassette (Sabbattini et al., 1999). An alternative explanation is that the *IRF4-ER* transgene had been silenced in PIP-ER mice and interestingly, analysis of IRF4-ER expression in primary PIP-ER pro-B cells by James Scott failed to detect the presence of IRF4-ER in these cells, suggesting silencing of the transgene.

As PIP-ER primary pro-B cells displayed no evidence of IRF4-ER protein, I sought to modify an immortalised pro-B cell line to generate the required inducible system. The *v-abl* immortalised pro-B line selected for this purpose was 1D1 as *Vλ1* non-coding transcripts in untreated 1D1 cells were two to four-fold higher than other tested lines, suggesting that this line was the least repressed by v-abl. Two main strategies were used in pursuit of this goal: upregulation of the IRF4-ER transgene by a CRISPR/Cas9 based activator (Konermann et al., 2015) and re-introduction of IRF4-ER by infection with a Murine Stem Cell Retrovirus (MSCV; Hawley, 1994).

3.3 Upregulation of IRF4-ER by CRISPR activator

Since it appeared that IRF4-ER was no longer being expressed at the correct level in the immortalised pro-B lines, I sought to upregulate the expression of IRF4-ER, the Cas9 Synergistic Activation Mediators (SAM) system from the Zhang lab appeared to be the most suitable for this purpose. The system involves introducing three components: a nucleolytically inactive Cas9-VP64 fusion, a sgRNA incorporating two MS2 RNA aptamers and a MS2-P65-HSF1 fusion protein (Konermann et al., 2015). This system enables the targeting of strong transcriptional activators to a specific genomic location i.e. the promoter of the gene of interest (Figure 3.5) and has been has been demonstrated to significantly upregulate target genes, without the need for multiple sgRNAs to improve the upregulation, due to the delivery of several strong transcriptional activators (VP64, p65 and HSF1) to a single location. Therefore, targeting this complex to the λ5 promoter in a PIP-ER pro-B cell line should result in increased expression of IRF4-ER.

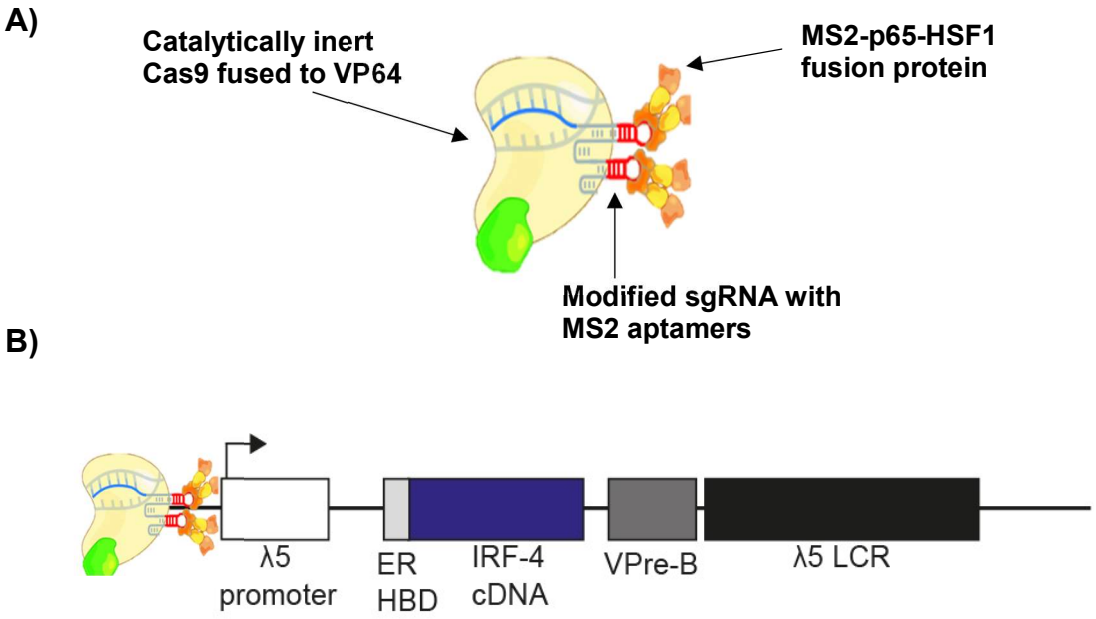


Figure 3.5: Schematic of the Synergistic Activation Mediator (SAM) system targeted to the λ5 cassette. A) Overview of the SAM system consisting of a dCAS9 fused to VP64 which complexes with a modified sgRNA containing MS2-aptermers. These MS2 stem-loops are bound by a MS2-p65-HSF1 fusion protein. B) Cartoon of the SAM system targeted to the λ5 promoter.

Guide RNAs that target the λ5 promoter were cloned into lenti-sgRNA-MS2-Zeo and lentivirus was produced for the introduction of all activator components. The pro-B cell line 1D1 was infected sequentially with lentiviruses containing dCas9-VP64, MS2-P65-HSF1 and one of the λ5 promoter guides. Infection of each component was confirmed by selection with the relevant antibiotic for one week. Selected cells were then expanded, induced with 4OHT and examined for upregulation of *Vλ1* non-coding transcription.

Despite repeated analysis, 1D1 cells infected with λ5 guide 1 and λ5 guide 2 failed to demonstrate any increase in *Vλ1* non-coding transcription, but a small increase (~2 fold) was detectable in cells expressing λ5 guide 3. As the increase in *Vλ1* non-coding transcription was minimal, it seemed highly unlikely that this method would result in a cell line that would be suitable for further analysis.

3.4 Re-introduction of IRF4-ER by MSCV-IRF4-ER

As upregulation of the endogenous *IRF4-ER* by the CRISPR based activator failed to generate a cell line that could reliably activate the *Igλ* locus, I next looked to introduce additional copies of IRF4-ER, by viral infection, to improve the expression.

IRF4-ER was cloned upstream of the IRES element in MSCV-IRES-GFP and this vector was transfected into Phoenix cells (Grignani et al., 1998) in order to produce retrovirus. The pro-B cell line 1D1 was infected with MSCV-IRF4-ER-IRES-GFP retrovirus and the GFP reporter was used to analyse the level of infection 48 hours later. Unfortunately, pro-B cells are very resilient to viral infection and despite repeated attempts to improve infection efficiency, a 0.2% was achieved in all cases. To enrich the infected cells, two rounds of FACS sorting followed by expansion in culture was performed. Following enrichment, infected pro-B cells represented > 95% of the total population. *Irf4* expression was 0.699 ± 0.025-fold that of pre-B cells in the population of infected cells, suggesting that the viral transgene was expressed but at a lower level than required. Moreover, transgene expression was lost rapidly from the infected population and substantial variation in transgene expression was observed within this population precluding its continued analysis. Therefore, monoclonal cell lines were generated from the highest (10%) GFP expressing pro-B cells following the purification of this population by flow cytometry.

Twenty-three monoclonal pro-B lines were recovered from semi-solid agar, expanded and screened by flow cytometry, using the GFP reporter gene to identify high IRF4-ER expressing lines (Figure 3.6A). Strikingly a substantial amount of variation in transgene expression was observed among the generated cell lines, ranging from lines barely expressing the transgene (Figure 3.6B) to cells that expressed a high level of GFP. Additionally, transgene stability within each monoclonal line was very variable with some lines exhibiting no alteration in transgene expression for several months and others exhibiting signs of transgene silencing within one to two weeks (Figure 3.6C). This suggested that the selection of a cell line with an appropriate site or sites of viral integration was vital.

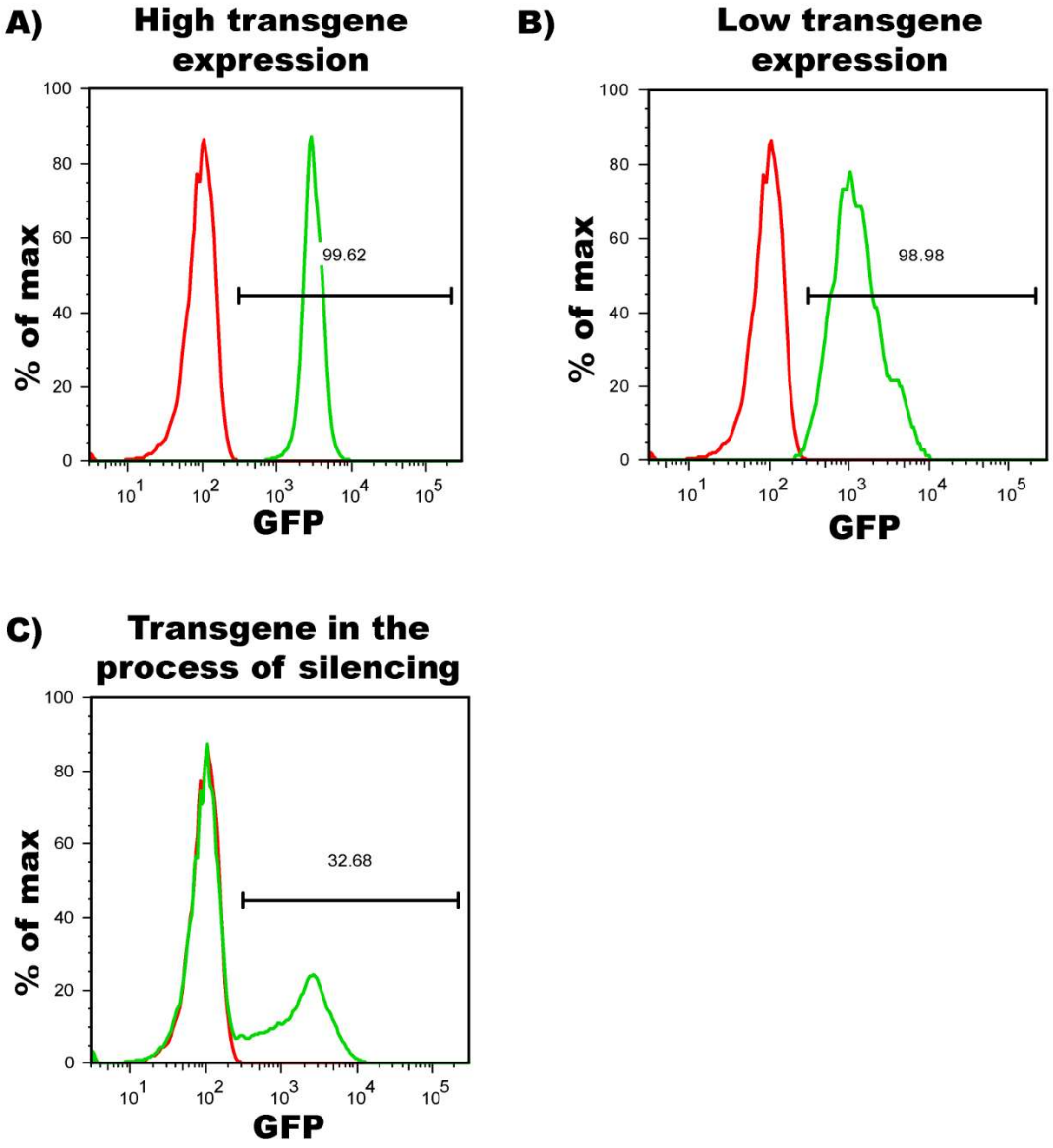


Figure 3.6: Analysis of transgene expression in MSCV-IRF4-ER-IRES-GFP infected 1D1 cells. Flow cytometry analysis of GFP expression in infected 1D1 pro-B cells. As the IRF4-ER transgene is polycistronic GFP expression was used as a reporter of IRF4-ER expression. The generated cell lines exhibited substantial variability in transgene expression and stability within the population. The black bar indicates GFP positive cells and the percentage of the population expressing GFP is shown above the bar. A) An example of a cell line with high transgene expression, B) A cell line with low transgene expression, C) A cell line exhibiting low transgene stability as transgene expression is in the process of silencing.

The MSCV-IRF4-ER infected pro-B cell lines with the highest transgene expression and stability, as determined by re-analysis of GFP expression after one week in culture, were subsequently analysed for the ability to induce *Vλ1* non-coding transcription following an eight-hour induction with 4-OHT. Four out of five pro-B lines examined exhibited increased *Vλ1* non-coding transcription post induction (Figure 3.7), with the 1D1-A4 line displaying the greatest induction in *Vλ1* transcription. Consequently, this line was selected for further analysis.

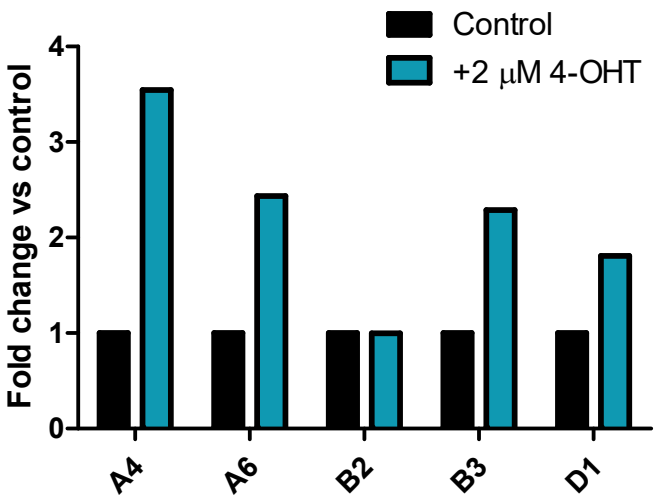


Figure 3.7: Alteration in *Vλ1* non-coding transcription in the MSCV-IRF4-ER infected pro-B cell lines after induction. RT-qPCR analysis of *Vλ1* non-coding transcription eight hours after the addition of 2 μM 4-OHT. Expression was normalised by *Hprt* and compared to uninduced cells of the same cell line (n = 1). Except for the B2 cell line, all MSCV-IRF4-ER-IRES-GFP infected monoclonal cell lines exhibited increased *Vλ1* non-coding transcription eight hours post-induction.

To ensure that the induction of *Vλ1* transcription observed was due to the expression of IRF4-ER, I first confirmed that IRF4-ER was expressed in the 1D1-A4 cell line by Western blotting (Figure 3.8A). As IRF4-ER was clearly expressed, I next sought to determine if *Vλ1* non-coding transcription in the generated cell line mirrored that of primary PIP-ER pro-B cells. Interestingly, I observed that *Vλ1* non-coding transcription in the generated cell line was very similar to that of PIP-ER pro-B cells, with a large increase in *Vλ1* transcription at eight hours post-induction (Figure 3.8B), as observed in the primary cells. The only substantial difference between the A4 cell line and PIP-ER primary pro-B cell was a slightly earlier initial increase in *Vλ1* non-coding transcription at seven hours post-induction in the cell line.

Whilst the 1D1-A4 cell line appeared to behave in a similar manner post-induction to primary PIP-ER pro-B cells, it became evident that there was a substantial problem with this cell line. After examining the timing of *Vλ1*, I sought to ascertain the time post-induction that IRF4-ER enters the nucleus to determine if part of the observed delay in induction was due to nuclear translocation. Strikingly, I consistently observed the presence of IRF4-ER protein in nuclear extract from untreated samples (Figure 3.9). The estrogen receptor ligand binding domain prevents nuclear localisation of any fusion protein, by interaction with HSP90, maintaining the fusion protein in the cytoplasm in the absence of ligand (Eng et al., 1997; Knoblauch and Garabedian, 1999). Furthermore, I also observed a dramatic decrease (~20 fold) in *Vλ1* non-coding transcription following the removal of the weak estrogen mimic phenol red (Moreno-Cuevas and Sirbasku, 2000) from 1D1-A4 culture media and altering the batch of foetal calf serum used in media preparation. These observations suggested that presence of estrogen in 1D1-A4 media resulted in the premature induction of IRF4-ER activity.

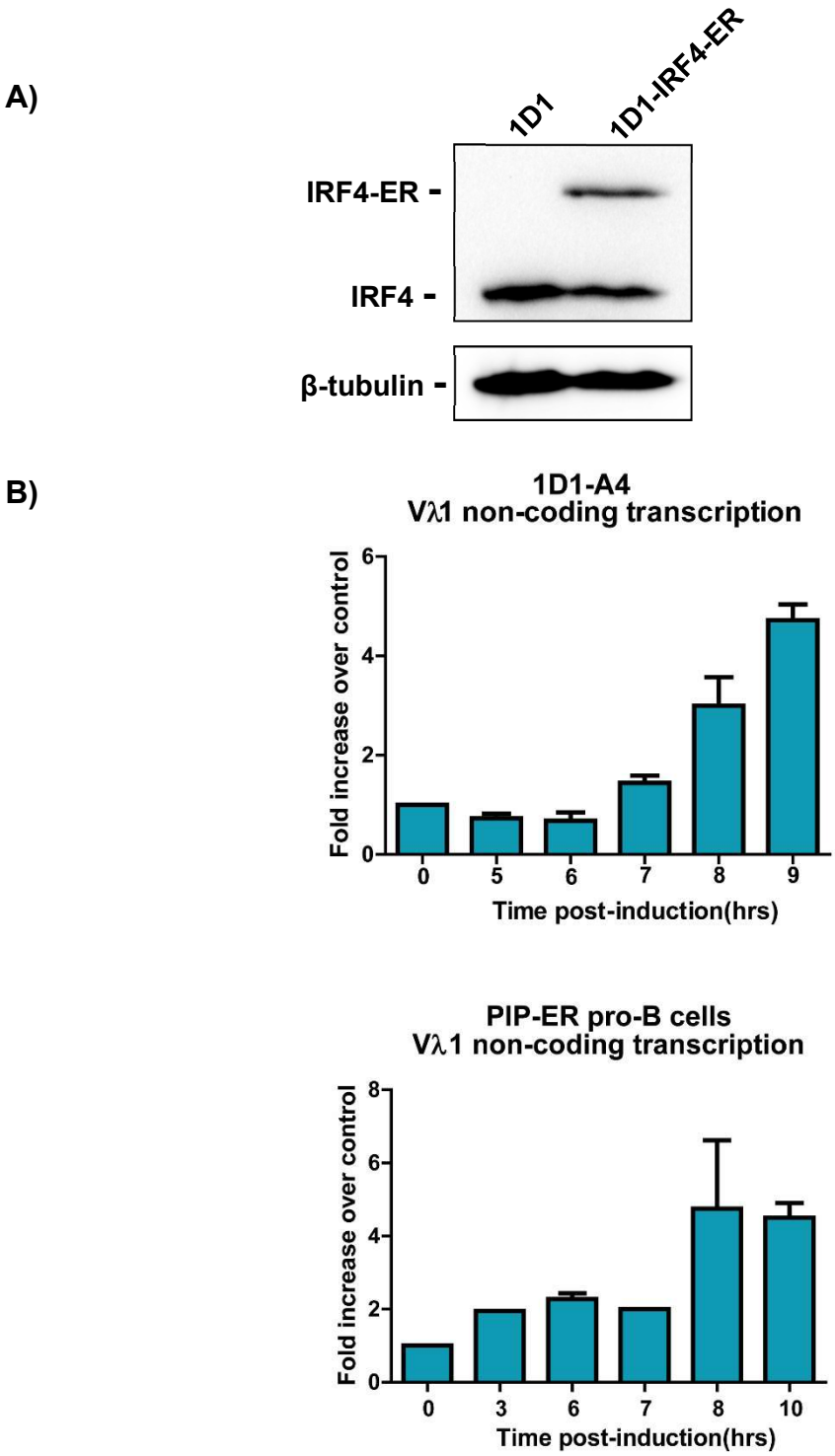


Figure 3.8: IRF4-ER expression is readily detectable in the 1D1-A4 cell line and Vλ1 non-coding transcription in this line mirrors PIP-ER primary pro-B cells. A) IRF4-ER expression analysis by Western blotting in the parental *v-abl* immortalised line 1D1 and the MSCV-IRF4-ER infected cell line. The location of bands corresponding to IRF4 (51 kDa) and IRF4-ER (87 kDa) are indicated. The blot was stripped and re-probed for β-tubulin to test if equivalent amounts of extract were loaded. B) Comparison of *Vλ1* non-coding transcription assessed by RT-qPCR in the 1D1-A4 cell line (left) and primary PIP-ER pro-B cells (right) following induction by 2 μM 4-OHT. Expression was normalised by *Hprt* and compared to that of an uninduced control. Error bars represent SEM (n=4).

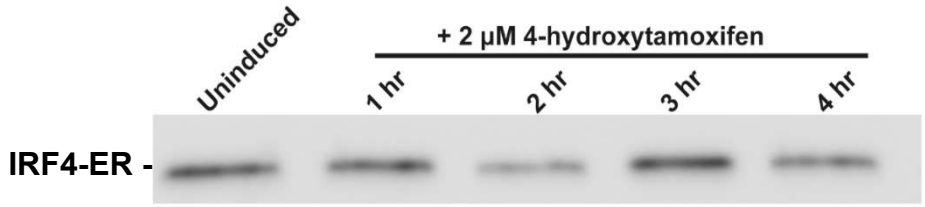


Figure 3.9: IRF4-ER is present in the nucleus of untreated 1D1-A4 cells. Analysis of the presence of IRF4-ER in nuclear extracts from 1D1-A4 cells at one, two, three and four hours post-induction by Western Blotting. IRF4-ER is clearly detectable in uninduced samples.

The IRF4-ER transgene was cloned by fusion of the ER domain from MyoD-ER (Kimura et al., 2008) to the N-terminus of *Irf4*. As the presence of estrogens in mice is sufficient for activation of the fusion protein at picomolar concentrations, hormone binding domains which have dramatically reduced sensitivity to β-estradiol have been developed. The G525R mutation, known as ER_{Tam}, reduces estradiol sensitivity but does not affect the ability of the hormone binding domain to bind the estrogen receptor antagonist 4-hydroxytamoxifen. However this mutation does reduce the activity of the fusion protein (Danielian et al., 1993). An improvement on ER_{Tam} is the ER_{T2} mutant which contains three point mutations: G400V/M543A/L544A which further decrease the sensitivity of the ER_{T2} mutant to β-estradiol compared to the original ER_{Tam} mutant, increasing its suitability for *in vivo* studies (Feil et al., 1997). Therefore, as the 1D1-A4 line displayed signs of being responsive to estrogen, I sequenced the ER domain present in MSCV-IRF4-ER-IRES-GFP and in the genome of PIP-ER mice (Figure 3.10). Whilst the G400V point mutation was present in the ER domains of MSCV-IRF4-ER-IRES-GFP and the PIP-ER transgene, the additional point mutations required to generate the ER_{Tam} or ER_{T2} variants were not present.

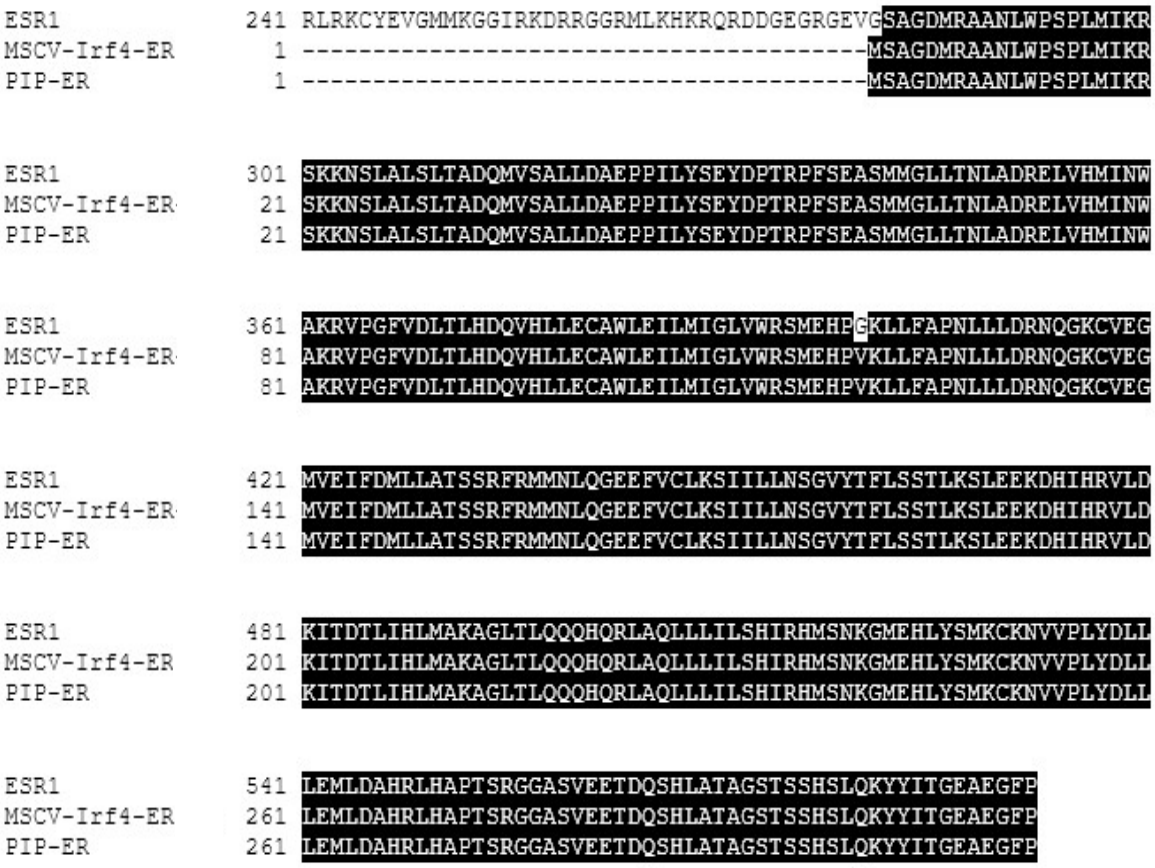


Figure 3.10: The estrogen receptor ligand binding domain of IRF4-ER is the G400V variant. Sequence alignment of hESR1 (human Estrogen receptor α) and the ER domains of IRF4-ER present within MSCV-IRF4-ER-IRES-GFP (MSCV-IRF4-ER) and the PIP-ER transgene (PIP-ER). A single point mutation (G400V) is present within the ER domains of MSCV-IRF4-ER-IRES-GFP and the IRF4-ER transgene in PIP-ER transgenic mice.

The G400V point mutation destabilises the ligand binding domain, which reduces its sensitivity to β-estradiol by approximately 100-fold (Tora et al., 1989), however, this effect can be overcome by increasing the concentration of the ligand (Eng et al., 1997). Approximately 100 pM 17β-estradiol is sufficient for partial activation of the G400V mutant, and complete saturation can be observed between 5-10 nM 17-β-estradiol (Nichols et al., 1997).

The combination of estrogen like compounds within the foetal calf serum and presence of phenol red at an effective concentration of 15 nM (considering the low affinity of phenol red for the estrogen receptor and its concentration in tissue culture media) was likely sufficient for the partial activation of IRF4-ER. In addition to the premature activation of the fusion protein, there is a wealth of evidence suggesting that the binding of β-estradiol to the ER ligand binding domain can result in its proteasome mediated degradation (Alarid et al., 1999; Nawaz et al., 1999; Nirmala and Thampan, 1995). Therefore, the level of IRF4-ER could potentially differ depending on the amount of contaminating estrogen like compounds, resulting in a highly variable experimental system. Therefore, the 1D1-A4 cell line was deemed unsuitable for downstream analysis.

3.5 Generation of an estrogen insensitive IRF4-ER cell line by the use of a destabilisation domain.

As the presence of estrogen prevented additional analysis of the IRF4-ER 1D1-A4 cell line, I sought to produce a cell line incapable of responding to low levels of estrogen present in culture media. Whilst efforts were made to remove contaminating estrogens from the culture media, using dextran/charcoal stripped FCS and phenol red free media, this proved not to be a viable strategy as the cell line grew poorly under these conditions. Therefore, there were two plausible methods to achieve this goal, either to alter the regulatory system to avoid activation by any component of the culture media or to mutate the ER domain present in MSCV-IRF4-ER-IRES-GFP to generate the ER_{T2} variant.

The fusion of an estrogen receptor ligand binding domain is one of the best-established methods for the control of transcription factors (Whitfield et al., 2015), but relatively recently, promising alternatives have been developed. Instead of preventing nuclear localisation and activity, the stability of a protein can be regulated by the fusion of domains that render the fusion protein unstable and exceptionally susceptible to proteasomal degradation, termed destabilisation domains (DD; Banaszynski et al., 2006; Iwamoto et al., 2010). Regulation is achieved by the addition of a small molecule that stabilises the fusion protein and prevents its degradation. There have been multiple DD described in the literature, but the DD derived from Escherichia coli dihydrofolate reductase (DHFR; Iwamoto et al., 2010) was the most suitable for this system due to its inexpensive stabiliser trimethoprim (TMP).

To generate an IRF4-DD expressing pro-B cell line, the ecDFHR DD was fused to the C-terminus of *Irf4* and the fusion protein was cloned into lenti-P2A-Puro. Lentivirus was harvested from 293T cells and 1D1 cells were transduced. Transduced cells were selected using two different concentrations of puromycin (0.2 µg/ml and 2 µg/ml), to select for cells with different levels of IRF4-DD expression. Puromycin selection was performed for one week before IRF4-DD expression was examined 16 hours post-induction with 10 µM TMP (Figure 3.11). Whilst a substantial increase in IRF4-DD stability was observed, induction of *Vλ1* non-coding transcription could not be observed. This was likely due to insufficient degradation of IRF4-DD as the protein is present in uninduced samples.

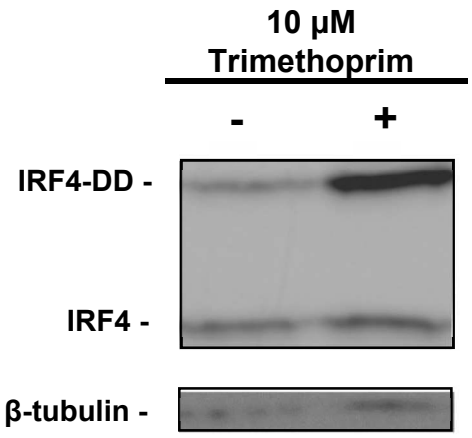


Figure 3.11: IRF4-DD is stabilised in the presence of Trimethoprim (TMP). Western blot analysis of IRF4-DD expression in the presence of 10 μM for 16 hours. IRF4-DD expression is increased approximately two-fold in the presence of TMP.

3.6 Generation of an estrogen insensitive IRF4-ER cell line by mutagenesis of the ER domain in IRF4-ER.

As the ecDFHR DD failed to completely destabilise the fusion protein I next sought to mutate IRF4-ER in MSCV-IRF4-ER to generate the ER_{T2} mutant (Feil et al., 1997). Point mutations, M543A and L544A were generated in MSCV-IRF4-ER-IRES-GFP by Q5® Site-Directed Mutagenesis and retrovirus was produced by transfection of Phoenix packaging cells. The MSCV-IRF4-ER_{T2}-IRES-GFP retrovirus was harvested and cells of the v-abl immortalised PIP-ER pro-B cell line 1D1 were transduced.

To avoid repeated cycles of infected cell enrichment, the highest GFP expressing cells were purified by FACS and monoclonal cell lines were generated immediately from the sorted population to avoid transgene silencing. Twenty cell lines were recovered from semi-solid agar and examined by flow cytometry. Again, substantial variability in transgene expression was observed between cell lines (data not shown), likely due to copy number and insertion site. The cell lines expressing the highest level of GFP were initially screened by analysing the induction of *Vλ1* non-coding transcription following an eight-hour induction with 4-OHT (Figure 3.12). Except for a single cell line, all lines displayed increased *Vλ1* non-coding transcription post induction. The 1D1-IRF4-ER$_{T2}$-Clone15 (referred to as 1D1-T215) displayed the most substantial induction of *Vλ1* non-coding transcription and therefore was selected for further analysis.

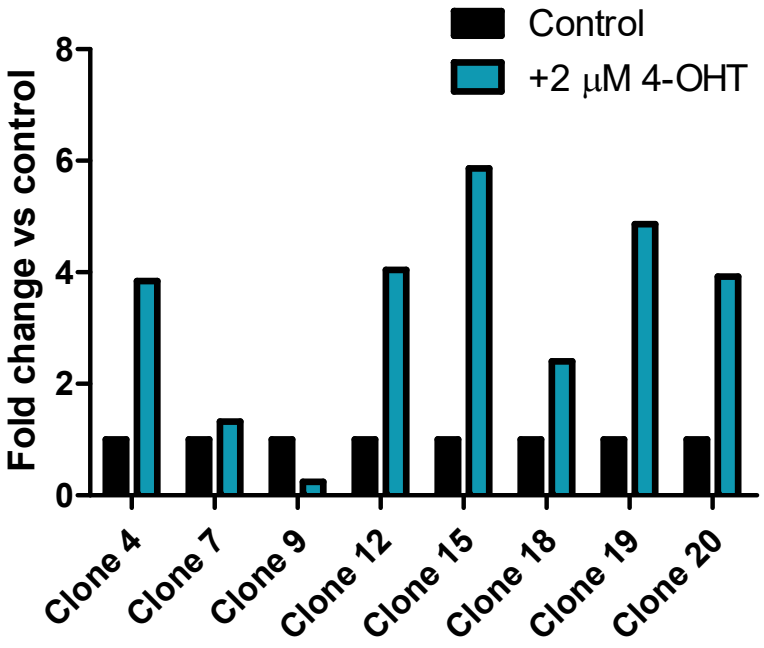


Figure 3.12: Alteration in *Vλ1* non-coding transcription in the MSCV-IRF4-ER$_{T2}$ infected pro-B cell lines after induction. RT-qPCR analysis of *Vλ1* non-coding transcription eight hours after the addition of 2 μM 4-OHT (n=1). Expression was normalised to *Hprt* and compared to an uninduced control from the same cell line. Except for Clone 9, all cell lines exhibited increased *Vλ1* non-coding transcription eight hours post-induction. Clone 15 exhibited the highest fold increase in *Vλ1* non-coding transcription.

To verify that the induction of *Vλ1* non-coding transcription observed in the T215 pro-B cell line was due to the expression of *Irf4*, RT-qPCR was performed. Total *Irf4* expression (endogenous together with *Irf4-ER$_{T2}$*) in T215 cells was found to be very similar to that of pre-B cells, suggesting that the IRF4-ER$_{T2}$ was expressed at the correct level (Figure 3.13A). Furthermore, to confirm that the IRF4-ER$_{T2}$ transgene expressed by T215 cells was not activated by low levels of estrogen presence in culture media, nuclear protein extracts were prepared from untreated samples and at one to four hours post-induction with 4-OHT. Clear nuclear localisation of IRF4-ER$_{T2}$ was observed post-induction, with an 89-fold enrichment of IRF4-ER$_{T2}$ observed after one hour (Figure 3.13B). Moreover, a barely detectable level of IRF4-ER$_{T2}$ protein was observed in the uninduced sample, strongly implying that uninduced IRF4-ER$_{T2}$ remains cytoplasmic, and therefore insensitive to the presence of estrogen or estrogen-like compounds.

I next sought to confirm that this cell line was a suitable model to investigate the coordinate activation of gene segments. Initially, I examined *Vλ1/Jλ1* recombination by qPCR which failed to detect any recombination between the two gene segments (data not shown), suggesting that recombination had not occurred during the generation of the cell line and the subsequent time in culture. To be a reliable model system, IRF4-ER$_{T2}$ expression must remain unaltered for approximately eight weeks to allow the planned genome editing experiments to occur and to obtain consistent replicates. Furthermore, alterations in ploidy can substantially increase the difficulty of obtaining homozygous knockout cell lines. Therefore, I analysed the stability of the IRF4-ER$_{T2}$ transgene after five months in culture by flow cytometry, using the GFP reporter as a measure of IRF4-ER $_{T2}$ expression and found that expression remains stable (Figure 3.14A). In addition, I compared the total DNA content of 1D1-T215 cells with that of primary pro-B using DAPI staining to determine the ploidy status of the cell line (Figure 3.14B). The DNA content of 1D1-T215 cell is identical to that of primary pro-B cell, suggesting no gross alterations in chromosome number had occurred.

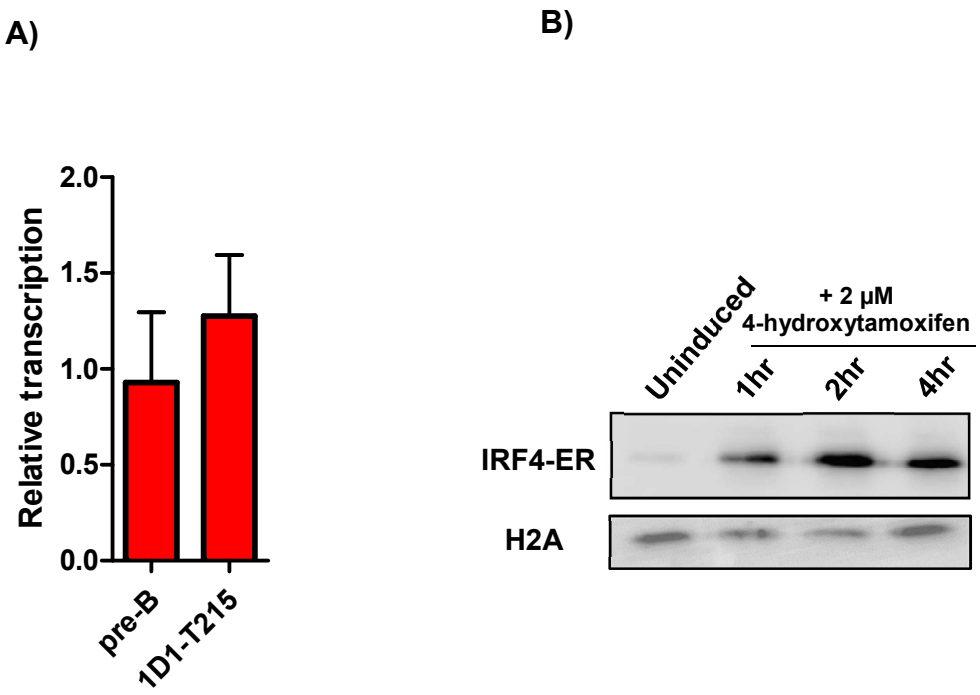


Figure 3.13: *Irf4* is expressed at similar levels to pre-B cells in the 1D1-T215 cell line and IRF4-ER$_{T2}$ nuclear localisation is dependent on 4-OHT. A) *Irf4* expression in pre-B cells (n=3) and the 1D1-T215 cell line (n=3) analysed by RT-qPCR. Expression was normalised to *Hprt*. 1D1-T251 cells express a pre-B cell level of *Irf4*. Error bars represent SEM B) Analysis of IRF4-ER$_{T2}$ in nuclear extracts of 1D1-T215 following induction with 4-OHT. The presence of IRF4-ER$_{T2}$ is barely detectable in the absence of ligand and increases 89-fold one hour after the addition of 4-OHT.

As the IRF4-ER$_{T2}$ transgene in the 1D1-T215 cell line appeared to translocate to the nucleus following induction, as expected, I next sought to confirm that *Vλ1* and *Jλ1* non-coding transcription increases coordinately in this cell line. 1D1-T215 cells were induced, harvested at the relevant time point before RNA was extracted, reverse transcribed and the relative level of *Vλ1* and *Jλ1* non-coding transcription was analysed by qPCR (Figure 3.15). Remarkably, I found that both *Vλ1* and *Jλ1* non-coding transcription is clearly increased four hours post induction confirming that *Vλ1* and *Jλ1* non-coding transcription is coordinately upregulated in this cell line.

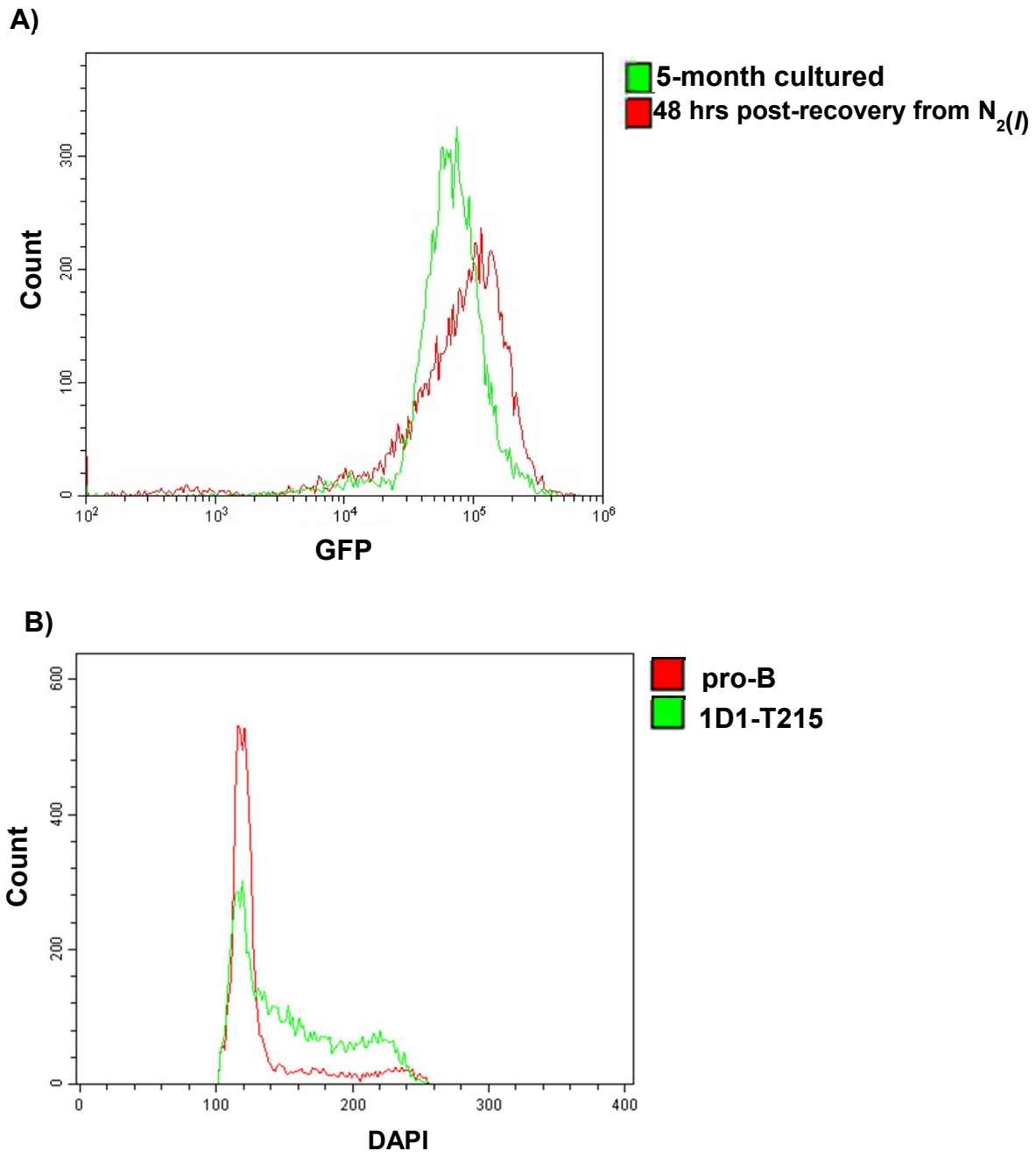


Figure 3.14: IRF4-ER$_{T2}$ is stably expressed in the 1D1-T215 cell line and the 1D1-T215 has no gross copy number alterations. A) Analysis of GFP expression in the 1D1-T215 cell line after five months in culture by flow cytometry. As the transgene is polycistronic IRF4-ER$_{T2}$ expression should mirror GFP expression. This suggests that IRF4-ER is expression is at approximately the same level after five months in culture. B) DNA content analysis of the 1D1-T215 cell line and primary PIP-ER pro-B cells by DAPI-staining followed by analysis using flow cytometry. No substantial change in DNA content can be observed between primary pro-B cells and the cell line.

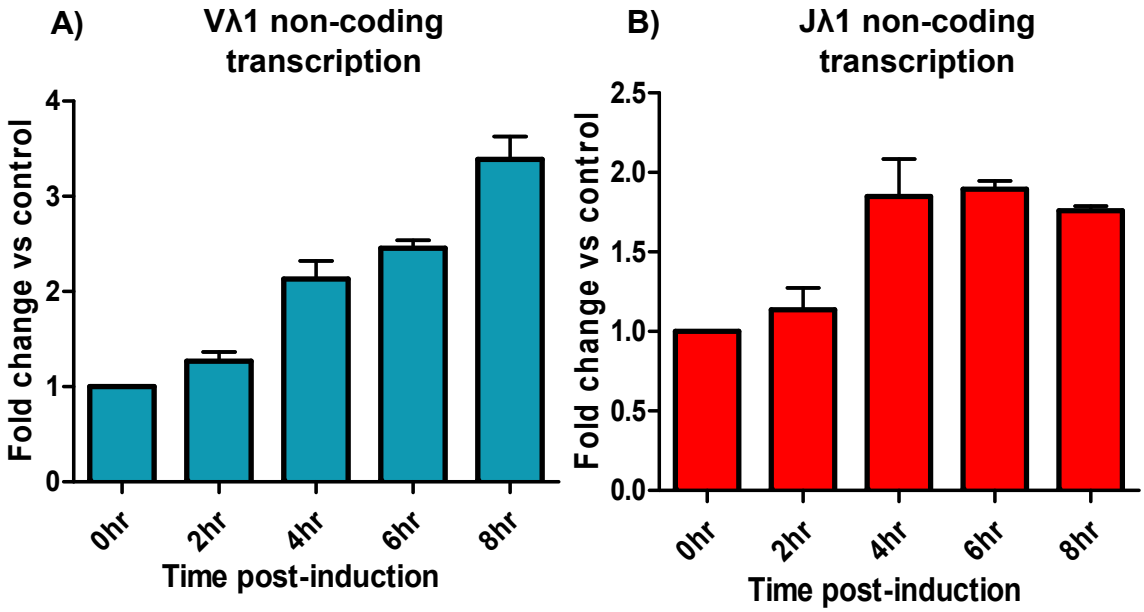


Figure 3.15: *Vλ1* and *Jλ1* non-coding transcription are upregulated by four hours post-induction in the 1D1-T215 cell line. RT-qPCR analysis of *Vλ1* (A) and *Jλ1* (B) non-coding transcription in in the 1D1-T215 following induction. Expression was normalised to *Hprt* and compared to an uninduced control. Non-coding transcription of both gene segments is upregulated coordinately by four-hours post-induction. Error bars represent SEM (n=4).

C. Discussion

3.7 Immortalisation of pro-B cells from PIP-ER transgenic mice

In this chapter, I describe the immortalisation of PIP-ER pro-B cells and the generation of a novel cell line expressing IRF4-ER$_{T2}$ that is capable of inducing *Vλ1* and *Jλ1* non-coding transcription coordinately. My initial aim was to immortalise PIP-ER pro-B cells to generate a cell line capable of replacing transgenic mice. To this end I attempted to generate pro-B cell lines using both IL-7 based immortalisation and A-MuLV infection.

The success rate of the IL-7 based immortalisation method was substantially lower than anticipated, resulting in just six cell lines, despite attempting the protocol five times. Corfe et al 2007 reported that approximately 49% of all seeded wells gave rise to pro-B cell lines which were then cloned and analysed (Corfe et al., 2007). The most likely explanation for the low immortalisation frequency is reduced IL-7 activity compared to that used by Corfe et al. As the IL-7 ELISA assay quantifies the total amount of IL-7 protein present, the activity of the cytokine cannot be measured using this method. Proliferation assays using an IL-7 responsive cell line represents a possible method to assess the activity of the IL-7 used. It is therefore possible that a lower effective concentration of IL-7 could have resulted in this marked decrease in cell line generation.

Unfortunately, all IL-7 dependent cell lines failed to demonstrate any induction of the *Igλ* locus (i.e. upregulated *Vλ1* non-coding transcription or increased *Vλ1-Jλ1* recombination). The most likely explanation for this was the absence of IRF4-ER in these cell lines but the reason for this was unclear. As I deemed it plausible that the immortalisation method could have inadvertently decreased IRF4-ER expression, potentially by the repression of the *λ5/VpreB* promoter, I next generated cell lines by A-MuLV infection. This method generated cell lines with a high efficiency. Furthermore, my data suggested that not only did these cell lines express *Irf4* and *Pu.1* at the primary pro-B cell level but they also expressed the *Irf4-ER* transgene (data not shown). However, these lines failed to activate the *Igλ* locus after treatment with tamoxifen.

The absence of IRF4-ER in these pro-B cell lines explains their failure to induce *Igλ* non-coding transcription. It is possible that the immortalisation of these pro-B cells resulted in the repression of IRF4-ER expression, however, another lab member, James Scott, has also demonstrated that IRF4-ER was not detectable in cell extracts from primary PIP-ER pro-B cell cultures.

3.8 IRF4-ER is not expressed at the correct level in PIP-ER pro-B cells

It is unclear why IRF4-ER was undetectable in PIP-ER pro-B cells. Transgene silencing is a well-known phenomenon (Garrick et al., 1996, 1998; Henikoff, 1998) that is most prevalent when transgenes integrate as multi-copy arrays. This silencing, which may be accompanied by DNA methylation, results from the formation of a repressive chromatin environment (Calero-Nieto et al., 2010). In these cases, silencing appears to be the result of convergent transcription, due to the presence of tail-tail transgene copies (Calero-Nieto et al., 2010). Convergent transcription is believed to result in the inability of the transgene promoter to recruit RNA polymerase II via a mechanism that is likely to involve transcriptional interference (Calero-Nieto et al., 2010).

An alternative explanation for the lack of IRF4-ER expression in the generated cell lines is the proteasome mediated degradation of the fusion protein. Many nuclear receptors, including estrogen receptor α (ERα; Alarid et al., 1999; Nawaz et al., 1999; Nirmala and Thampan, 1995), progesterone receptor (Shen et al., 2001), thyroid hormone receptor (Lee et al., 1995) and androgen receptor (Chymkowitch et al., 2011), are degraded in a proteasome-dependent manner upon addition of their ligand. Ligand mediated degradation of the ER is dependent upon coactivator-binding residues located in the activation function-2 (AF-2) domain of the receptor, indicating that ERα degradation is integrally connected to the receptor activation process (Lonard and Smith, 2002; Lonard et al., 2004). It therefore stands to reason that ER fusion proteins are also degraded by the 26S proteasome pathway following the activation of the ER ligand binding domain. In the case of a myb-ER fusion protein, continuous treatment with β-estrodiol has been shown to result in a barely detectable level of myb-ER, whereas ligand withdrawal resulted in substantially increased myb-ER levels after eight hours (Engelke et al., 1997). This strongly implies that ligand mediated degredation of ER fusion proteins occurs *in vivo*.

As *Irf4-ER* expression in recently isolated primary PIP-ER pro-B cells was equivalent to that of primary PIP-ER pro-B in which induction had previously been detected, it seems unlikely that the absence of IRF4-ER is mediated by an RNA silencing based mechanism. Therefore, the presence of estrogen and estrogen mimics in the culture media is the most likely explanation for the reduction in IRF4-ER expression. This would result in the premature activation and subsequent degradation of IRF4-ER protein in both primary PIP-ER pro-B cells and the generated cell lines. In support of this hypothesis, IRF4-ER was detectable in PIP-ER pro-B cells cultured in the presence of the proteasome inhibitor MG132 (James Scott PhD Thesis, University of Leeds 2016) implying that ligand mediated proteasomal degradation of IRF4-ER was the cause of IRF4-ER absence from PIP-ER pro-B cells.

3.9 Generation of an IRF4-ER expressing pro-B cell line

Numerous strategies were attempted in order to generate a pro-B cell line in which the activity of IRF4 can be induced, these are summarised in Table 3.2. Before identifying that the ER domain of IRF4-ER was responsive to estrogen, I initially attempted to increase the expression of IRF4-ER using a dCas9 based activator and retroviral infection. The use of the dCas9 based transcriptional activator failed to increase *Irf4-ER* expression after infecting the PIP-ER cell line 1D1 with all three components of the activator despite the cells being resistant to all three antibiotics (Blasticidin, Hygromycin and Zeocin) which strongly indicated that all three components of the dCas9 SAM system were expressed in these cells. The location of the sgRNA respective to the transcription start site is vital for the upregulation of the target gene and adjusting the location of the guide sequence by 100 bp can result in more than 1000 fold increase in gene activation or a complete lack of activation (Konermann et al., 2015). Therefore, establishing the correct location for the guide sequence is essential. The ideal placement is variable for each target gene, a general rule, is to design guide sequences ~200 bp upstream of the transcription start site (Zhang et al., 2015). As the location of the *λ5/Vpre-B* promoter transcription start site was difficult to establish, I used RNA-seq from pre-B cells (Table 2.2) to predict the *λ5* transcription start site to design my sgRNAs. The lack of *Irf4-ER* transcriptional upgregulation could therefore be attributed to the failure of selecting an appropriate location for the guide RNA.

Table 3.2: Summary of the strategies used to generate a pro-B cell line capable of inducing the activation of IRF4.

Strategy	Objective	Outcome
CRISPR/Cas9 based transcriptional activator	Upregulate the expression of the endogenous IRF4-ER transgene in the PIP-ER pro-B cell line 1D1.	Failed to upregulate the IRF4-ER transgene.
Expression of IRF4-ER	Overexpress IRF4-ER in the PIP-ER pro-B cell line 1D1 using an MSCV based retrovirus	Failed likely due to the premature activation and degradation of the IRF4-ER fusion protein by estrogen/ estrogen like compounds present in culture media.
Expression of IRF4-DD	Express a destabilised IRF4 in the 1D1 cell line and control IRF4 activity by stabilising the protein.	Failed as the destabilisation domain did not completely destabilise the fusion protein.
Mutation of the IRF4-ER domain and expression of IRF4-ER$_{T2}$	Expression of an estrogen insensitive IRF4-ER fusion protein in 1D1 cells.	This approach resulted in a cell line (1D1-T215) capable of inducing the activity of IRF4.

In addition to the use of the CRISPR/Cas9 activator to upregulate IRF4-ER, I also used an MSCV based retrovirus to introduce additional copies of IRF4-ER. Despite the poor transduction efficiency, after multiple rounds of enrichment and the generation of monoclonal cell lines, the expression of Irf4-ER in the 1D1-A4 cell line was equivalent to *Irf4* expression in pre-B cells. Furthermore, the presence of IRF4-ER could also be confirmed by Western blotting. Using the 1D1-A4 cell line, the upregulation of both *Vλ1* and *Jλ1* non-coding transcripts could be reliably detected following the addition of 4-OHT. Unfortunately, it soon became apparent that there were several problems with this cell line. Initially, I failed to detect an increased level of IRF4 binding at the *Igλ* enhancer $E\lambda_{3\text{-}1}$ by ChIP, despite repeated attempts. Furthermore, whilst attempting to perform a temporal analysis of IRF4-ER nuclear localisation, I repeatedly observed IRF4-ER in nuclear extract from untreated samples.

The sequence analysis of the IRF4-ER estrogen receptor ligand binding domain provided an explanation for these issues as it indicated that this domain only contained the G400V point mutation and lacked further mutations to decrease estrogen sensitivity. Whilst this mutation results in an ER ligand binding domain with a reduced affinity for estradiol (Tora et al., 1989), the relative level of estradiol present in the culture media, accounting for estrogen contamination within the foetal calf serum and the estrogen mimic phenol red, is estimated to be approximately 16 μM. This effective concentration is sufficient for activation of the domain (Eng et al., 1997) which would result in a poor level of non-coding transcript induction and the subsequent degradation of the fusion protein which could explain the failure to detect IRF4-ER binding. The removal of all potential estrogen contaminants from the cell culture media by replacing the FCS with charcoal-stripped FCS and using phenol red free media was not a viable strategy as both primary pro-B cells and pro-B cell lines exhibited markedly decreased proliferation and increased cell death under these conditions, making them unsuitable for further analysis.

3.10 1D1-T215 cells represent an ideal system to investigate gene segment activation

To generate a cell line capable of inducing IRF4, I used two different strategies: expression of an IRF4-DD fusion protein and the mutation of the ER domain fused to IRF4-ER. The analysis of IRF4-DD by Western blot after 16 hrs of treatment with TMP demonstrated a substantial increase in the level of IRF4-DD protein. However, induction of *Vλ1* non-coding transcripts was not detected. A substantial level of IRF4-DD is detectable in control samples suggesting that IRF4-DD was not destabilised to a sufficient level by the addition of the DD. This has been previously observed when DD have been fused to other proteins such as SpCas9 and can be overcome by the fusion of the DD to both the N and C terminus of the protein, which results in a more robust destabilisation of the fusion protein (Maji et al., 2017). However, this was not performed as the mutation of the ER ligand binding domain resulted in a viable cell line.

To avoid activation of an ER ligand binding domain (LBD) fusion protein by estrogen and estrogen like compounds, three ER LBD variants have been generated. These variants, ER_{Tam} and ER_{T2}, are insensitive to low levels of estrogen (< 1 μM) and respond only to the estrogen antagonist tamoxifen or its active metabolite 4-OHT (Feil et al., 1997). Mutation of IRF4-ER to IRF4-ER_{T2} prevented the premature activation of IRF4-ER observed in the 1D1-A4 cell line, as evidenced by the minimal presence of IRF4-ER_{T2} in the nucleus of untreated samples. Most importantly, *Vλ1* and *Jλ1* non-coding transcripts were consistently found to be coordinately upregulated by four hours post-induction in this cell line, suggesting a robust and consistent induction of IRF4-ER_{T2}. Crucially, this finding mirrors the observation of coordinately upregulated non-coding transcription in primary PIP-ER pro-B cells. In PIP-ER pro-B cells and the 1D1-A4 cell line, *Vλ1* and *Jλ1* transcription increases coordinately at seven to eight hours post-induction. Interestingly, the parallel increase in transcription can be observed in the T215 cell line at four hours. As both the 1D1-A4 and T215 cell lines express a near pre-B cell level of *Irf4*, the earlier induction is unlikely to be due to the increased expression of *IRF4-ER_{T2}*.

The most likely explanation for the altered time of upregulation is the altered stability of IRF4-ER$_{T2}$ vs IRF4-ER. β-estradiol promotes the proteasome mediated degradation of ER fusion proteins and interestingly estrogen antagonists can stabilise ER fusion proteins (Kiang et al., 1989; Wijayaratne et al., 1999) by blocking the co-activator interaction with the ER AF2 domain, thus preventing proteasomal degradation (Lonard and Smith, 2002; Lonard et al., 2004). I propose that the delay in the upregulation of non-coding transcription observed in 1D1-A4 and PIP-ER pro-B cells is partially due to the time taken for tamoxifen to stabilise the expression of IRF4-ER. As IRF4-ER$_{T2}$ is insensitive to ligand mediated degradation an increased level of IRF4-ER$_{T2}$, compared to IRF4-ER, is likely to be present, resulting in a rapid induction of non-coding transcription.

Whilst this cell line appears to be a viable model for the investigation of *Vλ1* and *Jλ1* non-coding transcripts, there are two caveats. Firstly, v-abl represses *Rag1* and *Rag2* expression (Chen and Rosenberg, 1992; Muljo and Schlissel, 2003) and therefore without the upregulation of *Rag1/Rag2* expression, analysis of V(D)J recombination in this cell line within a timescale mirroring that of primary PIP-ER pro-B cells cannot be performed. Moreover, *v-abl* also appears to repress *Igλ* non-coding transcription by approximately 1000-fold. The mechanism of this repression is unclear, STAT5 is known to reduce the accessibility of the *Igκ* enhancer Eκi by binding as a tetramer and recruiting Polycomb repressive complex 2 (PRC2; Mandal et al., 2011). Therefore, it is possible that STAT5 binding may also repress *Igλ* accessibility by a similar mechanism. Without the appropriate ChIP-seq data, however, it was not possible to test this hypothesis further.

With the exception of the caveats discussed above, the generated IRF4-ER cell line appears to have all the features required for the investigation into the coordinate activation of *Vλ1* and *Jλ1* non-coding transcription. The IRF4-ER$_{T2}$-IRES-GFP transgene is stably expressed for at least five months, suggesting that the viral integration site is not prone to silencing. Additionally, the tight control of IRF4-ER$_{T2}$ localisation suggests that IRF4-ER$_{T2}$ is unlikely to undergo ligand mediated degradation or exhibit premature activation. Furthermore, the 1D1-T215 cell line appears to have no gross alterations in chromosome number, as assessed by DNA content. These analyses therefore suggest that the 1D1-T215 cell line is highly suitable for gene editing experiments. Together, these data suggest that I have generated a cell line that can be used to examine how the coordinate activation of *Vλ1* and *Jλ1* is mediated.

Chapter 4 – What are the factors regulating the activation of *Vλ1* and *Jλ1* gene segments?

A. Introduction

A series of genetic and biochemical studies have indicated that a single transcription factor, IRF4, appears to be a master regulator of light chain activation and recombination (Lu, 2008). *Irf4*$^{-/-}$/*Irf8*$^{-/-}$ knockout mice are incapable of light chain recombination (Lu et al., 2003) and the re-introduction of IRF4 is sufficient to activate light chain recombination (Johnson et al., 2008). Furthermore, the forced exit from the cell cycle by IL-7 withdrawal from *Irf4*$^{-/-}$ /*Irf8*$^{-/-}$ pro-B cells fails to evoke *Igκ* recombination (Johnson et al., 2008; Ma et al., 2006) indicating that IRF4/IRF8 are directly involved in the regulation of *Igκ* recombination. Interestingly, there is a degree of functional redundancy between IRF4 and IRF8 as the expression of either transcription factor in IRF4,8 deficient pro-B cells is sufficient to induce *Igκ* non-coding transcription and rearrangement (Johnson et al., 2008; Ma et al., 2006). Despite a clear correlation between IRF4/IRF8 and light chain recombination, the mechanism by which these factors regulate recombination is still unclear.

It is conceivable that IRF4/IRF8 regulate pre-B cell maturation and light chain recombination by the upregulation of Ikaros/Aiolos. The phenotype observed in *Ikaros*$^{-/-}$ transgenic mice resembles that of *Irf4*$^{-/-}$/*Irf8*$^{-/-}$ transgenic mice and both IRF4 and IRF8 can induce Ikaros/Aiolos expression (Ma et al., 2008). Furthermore, both IRF4 and Ikaros are capable of inducing *Igκ* non-coding transcription and recombination (Heizmann et al., 2013; Ma et al., 2006) in addition to regulating expression of the chemokine receptor CXCR4 which promotes the migration of pre-B cells away from IL-7 producing stroma (Clark et al., 2014; Johnson et al., 2008). However, in contrast to IRF4, which promotes monoallelic positioning of the *Igκ* locus away from pericentromeric heterochromatin, Ikaros recruits *Igκ* alleles to pericentromeric heterochromatin via binding to the Sis element (Johnson et al., 2008), suggesting negative regulation of *Igκ* by Ikaros.

The level of IRF4 increases from the pro-B to pre-B stage of development correlating with a 10 fold increase in *Igλ* recombination (Muljo and Schlissel, 2003). IRF4 regulates non-coding transcription at the *Igκ* locus by binding cooperatively with E2A to the enhancer 3'Eκ. The full activation of 3'Eκ, by IRF4 and E2A, combined with the activation of iEκ, by increased E2A binding, results in increased *Vκ* and *Jκ* non-coding transcription and recombination. The mechanism by which IRF4 activates *Igλ* recombination is very poorly understood. Interestingly, the increased level of IRF4 at the pro-B cell stage appears to be sufficient for the activation of the *Igλ* locus. This has been demonstrated using both primary pro-B cells from PIP3 transgenic mice (Bevington and Boyes, 2013) and the PIP-ER cell line 1D1-T215. IRF4 has been observed to bind, together with PU.1, to the duplicated *Igλ* enhancers $E\lambda_{2\text{-}4}$ and $E\lambda_{3\text{-}1}$. Furthermore, IRF4 binding to the enhancer $E\lambda_{3\text{-}1}$ increases three-fold from the pro-B to pre-B cell stage, correlating with increased non-coding transcription and *Igλ* recombination in primary pro-B cells. Curiously, the induction of IRF4-ER_{T2} by 4-hydroxytamoxifen upregulates the non-coding transcription of *Vλ1* and *Jλ1* gene segments coordinately, in both primary pro-B cells from PIP-ER transgenic mice and the generated cell line (1D1-T215; Figure 3.15). These gene segments are separated by over 23 kb of intervening DNA, which raises the intriguing question of how this coordinate activation is regulated.

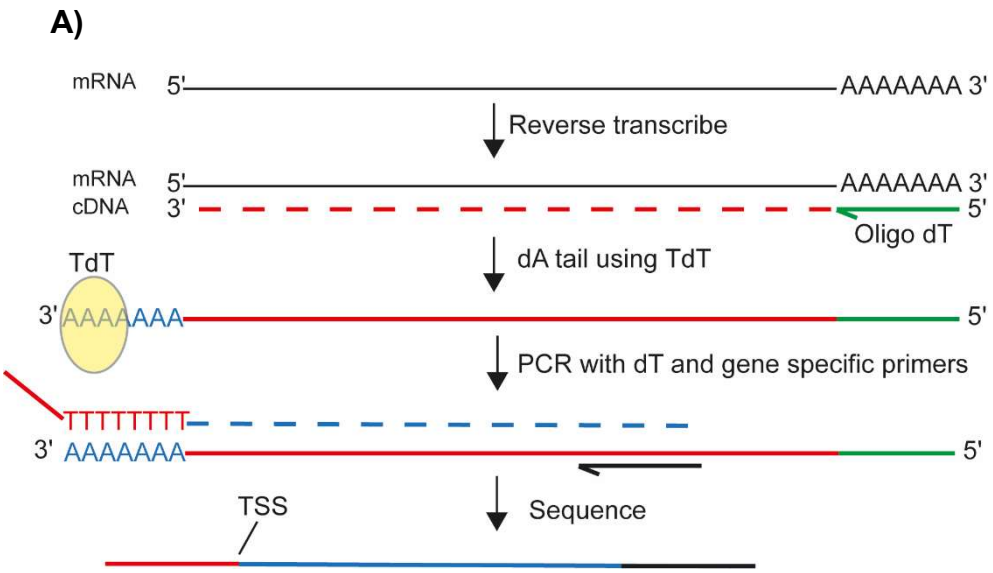
In this chapter I examine the binding of factors involved in the activation of the *Igλ* locus, explore *Igλ* locus for unidentified regulatory elements and investigate the presence of long range interactions that are likely to mediate gene segment activation.

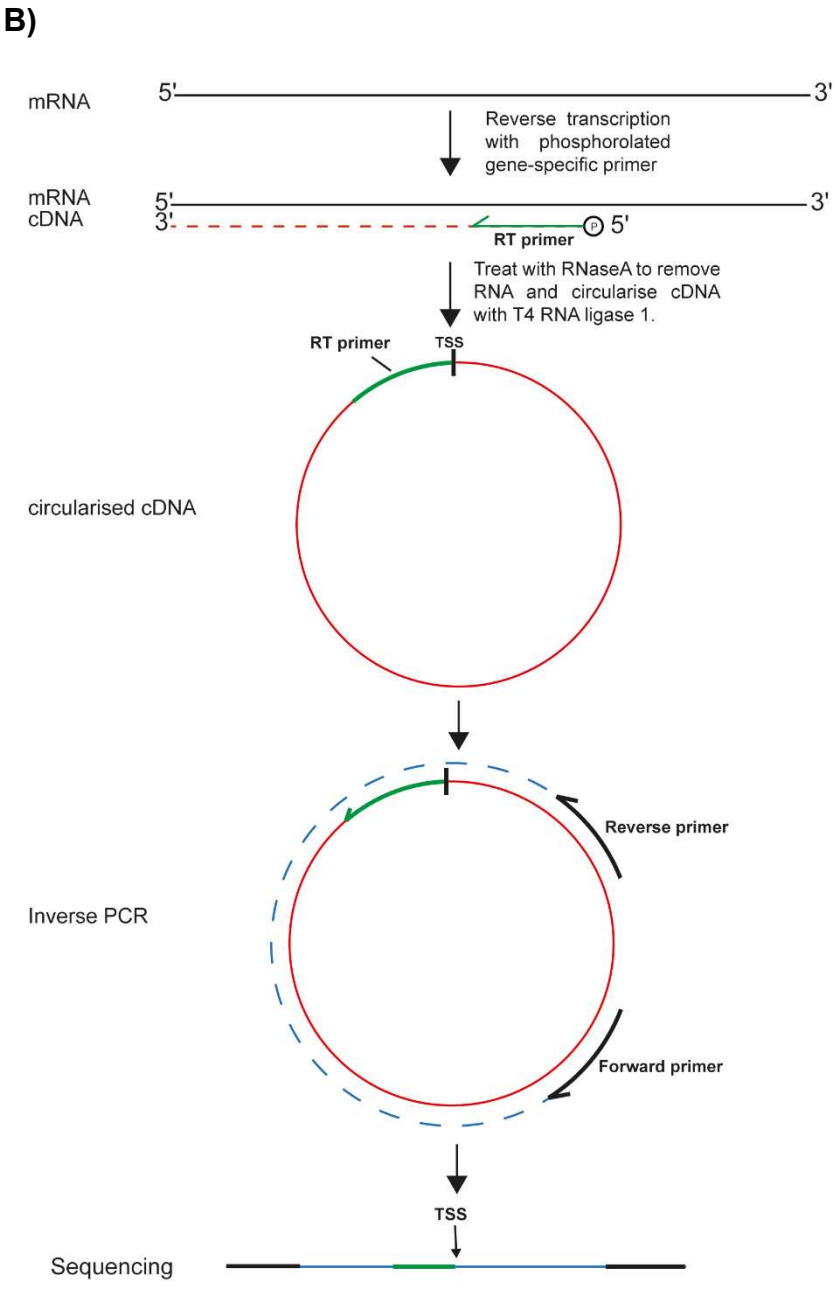
B. Results

4.1 Mapping of the *Vλ1* promoter

To determine how activation of *Vλ1* and *Jλ1* non-coding transcription is coordinately regulated, I first tested the hypothesis that this is controlled by the binding of a single transcription factor or transcription factor complex to the promoters of *Vλ1* and *Jλ1* non-coding transcription. The simultaneous binding of this factor or complex would result in the coordinated increase in non-coding transcription observed. Therefore, I sought to determine which transcription factors bind to the promoters of both the *Vλ1* and *Jλ1* gene segments.

The promoter region of *Jλ1* non-coding transcription had previously been identified (Engel et al., 2001). *Jλ1* non-coding transcripts initiate between 212-293 bp upstream of the *Jλ1* protein coding sequence, implying that the *Jλ1* promoter region is approximately 150-900 bp upstream of the *Jλ1* coding sequence. As the exact location of the *Vλ1* promoter had not previously been determined, I performed 5' RACE using RNA from PIP3 pro-B cells to identify the start sites of *Vλ1* non-coding transcripts. A transcription start site (TSS) 39 bp upstream of the *Vλ1* coding sequence (n=6; Figure 4.1) was identified using the classical 5'RACE technique. This approximately corresponds with the start sites of assembled transcripts from RNA-seq of pre-B cells, suggesting that this is a true transcription start site (Figure 4.2). No additional TSSs were identified using the classic 5'RACE method. To further verify these data, I used a modified 5' RACE method (Dallmeier and Neyts, 2013) that identified a TSS 394 bp upstream of the *Vλ1* coding sequence (n=4, Figure 4.1).


A)
mRNA
5'
AAAAAAA 3'
Reverse transcribe
mRNA
cDNA
5'
3'
AAAAAAA 3'
5'
Oligo dT
TdT
dA tail using TdT
3'
AAAAAAA
5'
PCR with dT and gene specific primers
TTTTTTTT
3'
AAAAAAA
5'
TSS
Sequence


B)
mRNA
5'
3'
Reverse transcription with phosphorolated gene-specific primer
mRNA
cDNA
5'
3'
RT primer
P
5'
Treat with RNaseA to remove RNA and circularise cDNA with T4 RNA ligase 1.
RT primer
TSS
circularised cDNA
Reverse primer
Inverse PCR
Forward primer
TSS
Sequencing

C)

Figure 4.1: Location of *Vλ1* non-coding start sites identified by 5' RACE. A) Schematic of the classical 5' RACE method. Messenger RNA is reverse transcribed using a deoxythymidine oligonucleotide (green) and terminal deoxy transferase (TdT; yellow oval) is used to add multiple adenosine nucleotides to the 3' end of the cDNA. PCR with a thymidine rich primer (green) is used together with a gene specific primer (black) to amplify the 3' end of the cDNA (complementary to the 5' end of the mRNA). This is then sequenced to identify the transcription start site (TSS). B) Schematic of the modified 5' RACE method. Messenger RNA is reverse transcribed with a phosphorylated gene specific primer (green) and the cDNA generated is circularised using T4 RNA ligase I, following treatment with RNaseA. Inverse PCR is performed followed by sequencing to identity the TSS which is the first nucleotide upstream of the reverse transcription primer. C) The location of transcription start sites are indicated by the black arrows. Three transcription start sites were identified at -394, -39 and +450 bp relative to the Vλ1 protein coding sequence (blue). The start site identified at +450 bp is an antisense transcription start site.

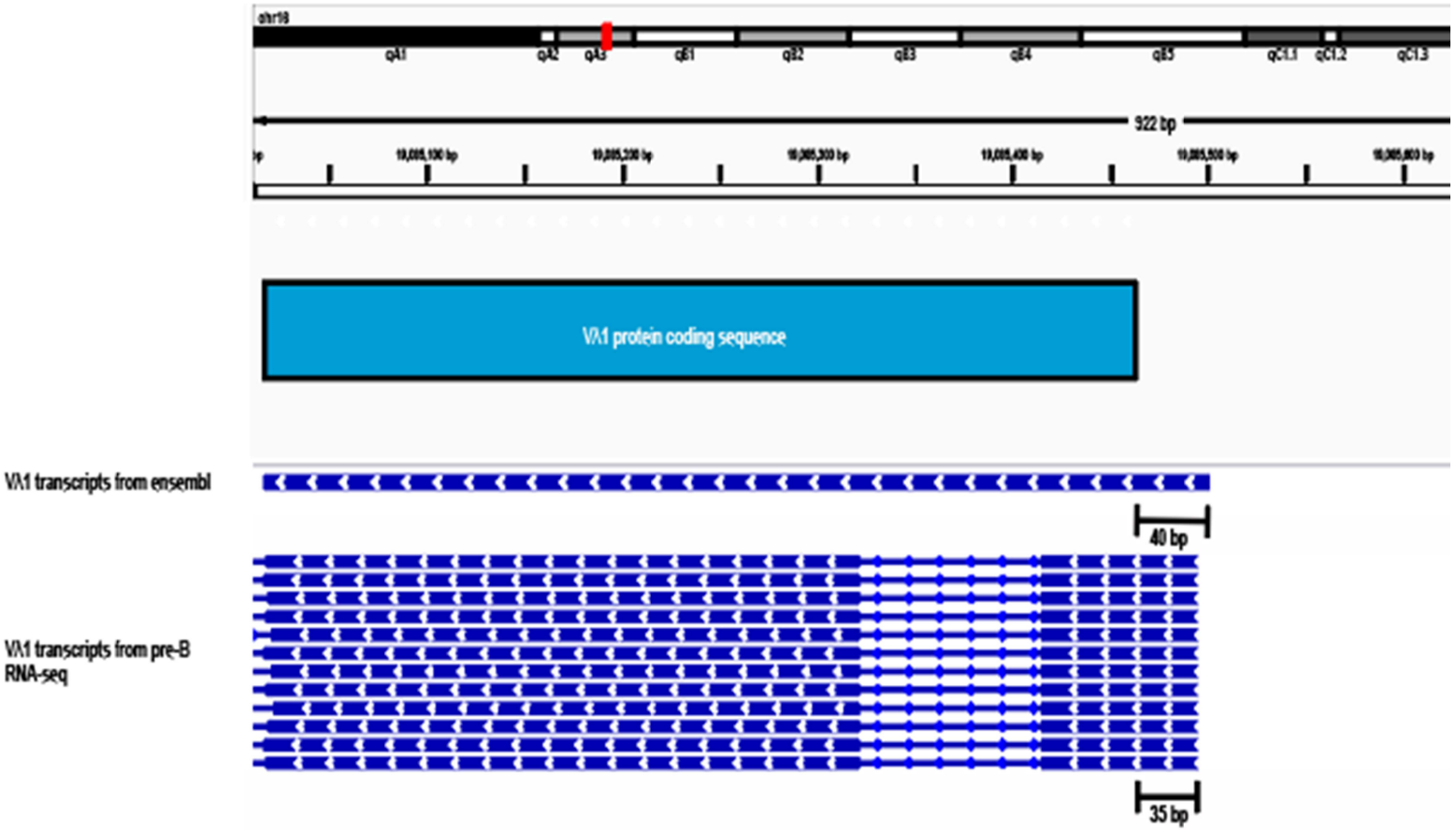


Figure 4.2: Assembled transcripts from pre-B RNA-seq data indicate a TSS of 35-40 bp upstream of the Vλ1 gene segment. Aligned and assembled transcripts from strand specific RNA-seq analysis of pre-B cells, indicating the position of the Vλ1 TSS. The Ensembl reference sequence indicated a start site at 40 bp upstream of Vλ1 and assembled pre-B transcripts suggested a TSS of 35 bp upstream of Vλ1.

In addition to non-coding transcripts in the sense direction, antisense non-coding transcription of V_H, D_H and Vκ gene segments has previously been observed (Bolland et al., 2004, 2007; Stubbington and Corcoran, 2013; Verma-Gaur et al., 2012a). As antisense non-coding transcripts could also regulate RSS accessibility, the presence of these transcripts was assessed. RNA from PIP3 pro-B cells was reverse transcribed with primers specific for sense (V1-GSP2) or anti-sense (*Vλ1*ts1R) transcripts (Table 2.5). When both cDNA samples were amplified by qPCR, amplicons of identical size were generated, indicating that *Vλ1* is bi-directionally transcribed (data not shown). To affect RSS accessibility, antisense non-coding transcripts must originate downstream of the *Vλ1* RSS. A transcription start site 450 bp downstream of the *Vλ1* coding sequence (n = 4, Figure 4.1) was identified by 5' RACE, using an antisense specific primer for the reverse transcription step (*Vλ1*ts1R).

Analysis of *Vλ1* transcription start sites thus implied the presence of three potential transcription start sites: -394 bp (Sense 1), -39 bp (Sense 2) and +450 bp (Antisense 1), relative to the *Vλ1* protein coding sequence. To assess if all the identified potential promoter regions were functional in pre-B cells, 50 bp downstream and 600 bp upstream of each TSS were cloned into a luciferase reporter vector and transfected into 103/BCL-2 cells. The 103/BCL-2 cell line is an Abelson immortalised pro-B cell line, transformed with a temperature sensitive v-abl mutant (Chen et al., 1994). Whilst at the permissive temperature (33^{o}C), 103/BCL-2 adopt a pro-B cell like phenotype and exhibit a very low level of recombination. When shifted to the non-permissive temperature (39^{o}C), these cells exhibit a pre-B phenotype with a substantial increase in RAG expression, light chain loci non-coding transcription and V(D)J recombination (Chen et al., 1994; Xu and Feeney, 2009). Consequently, temperature shifting of 103/BCL-2 cells results in the upregulation of the pre-B cell factors required for *Vλ1* and *Jλ1* non-coding transcription.

After 48 hours, following 103/BCL-2 transfection with the luciferase reporter construct, cells were lysed, and lysates were subjected to luciferase reporter assay. In the presence of the Sense 1 and Antisense 1 putative promoter regions, no significant increase in luciferase expression was observed (data not shown), indicating that these regions are not functional promoters. The presence of the Sense 2 promoter region, however, resulted in a significant increase in luciferase expression compared to the vector control and furthermore luciferase expression was also significantly increased when the Sense 2 promoter region was paired with a putative *Igλ* enhancer HS6 (Figure 4.8) implying that this region is a functional promoter in pre-B cells.

4.2 Identification of consensus transcription factor binding sites

To determine which transcription factors bind to the *Vλ1* promoter region, I initially attempted to use published ChIP-seq data from pro-B cells, focussing on transcription factors involved in *Igλ* activation (IRF4, PU.1, E12/E47, Ikaros and MEF2C/D). For all factors examined, significant enrichment at the *Vλ1* and *Jλ1* promoter regions could not be identified. Therefore, the binding motifs of factors known to be involved in *Igλ* recombination were examined for both the *Vλ1* and *Jλ1* promoters using HOMER (Hypergeometric Optimization of Motif EnRichment; v4.9).

HOMER is a collection of programs designed to perform a variety of functions such as ChIP-seq analysis, Hi-C analysis and motif identification. HOMER identifies transcription factor motifs using a differential motif discover algorithm to identify elements that are specifically enriched in experimental samples compared to a control set, using zero or one occurrence per sequence (ZOOPS; Bailey and Elkan, 1995) together with hypergeometric enrichment calculations (Heinz et al., 2010). The software also includes a substantial number of verified binding motifs, obtained from ChIP-seq experiments, to determine transcription factor binding sites with high accuracy.

The HOMER program findMotifs.pl was used to identify binding sites for factors known to be involved in pre-B cell development: CTCF, E2A, IRF4, IKAROS, MEF2C and PU.1. As IRF4 binds very weakly to DNA in the absence of a partner such as PU.1 (Eisenbeis et al., 1995), the presence of IRF4 binding sites was not assessed. Instead, I examined the incidence of IRF4/PU.1 composite motifs. With the exception of E2A, analysis of the transcription factor binding motifs in the *Vλ1* and *Jλ1* promoter regions (Figure 4.3) revealed no shared transcription factors. The level of active E2A does increase from the pro-B to pre-B cell stage, due to the downregulation of the dominant negative repressor ID3 (Kee et al., 2001), which in turn results in the upregulation of the active E12/E47 heterodimer. There is, however, no evidence that increased expression of IRF4 results in the upregulation of E2A or the downregulation of ID3 and since an increased level of IRF4 appears sufficient for *Igλ* locus activation it was deemed unlikely that E2A was a likely to regulate the coordinate activation of *Vλ1* and *Jλ1*.

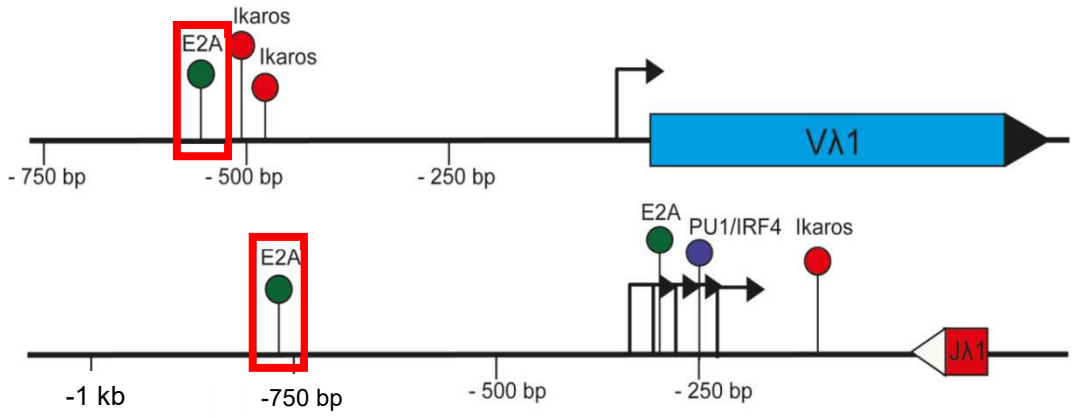


Figure 4.3: Identification of consensus transcription factor binding sites at the *Vλ1* and *Jλ1* promoters by motif prediction. The locations of the predicted transcription factor binding sites are indicated. The only shared transcription factor between the promoters of *Vλ1* and *Jλ1* is E2A, which is indicated by the red rectangle.

4.3 Identification of additional regulatory elements in the *Igλ* locus

Due to the absence of a shared transcription factor between the promoters of *Vλ1* and *Jλ1*, I next tested the idea that a shared enhancer might be involved in gene segment regulation. Chromosome Conformation Capture (3C) experiments undertaken by Sarah Bevington show that the enhancer Eλ$_{3\text{-}1}$ directly contacts the promoters of the *Jλ1* and Jλ3 gene segments in pre-B cells, (Appendix 1). Crucially, these 3C experiments indicated Eλ$_{3\text{-}1}$ does not contact the *Vλ1* promoter. Therefore, I sought to determine if there were other elements present within the *Igλ* locus that could potentially upregulate *Vλ1* non-coding transcription.

Active regulatory element are defined by their chromatin accessibility and therefore Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq; Buenrostro et al., 2013) data from pre-B cells was used to determine regions of open chromatin within the *Igλ* locus. ATAC-seq experiments determine the degree of chromatin accessibility by the use of a hyperactive Tn5 transposase loaded with Illumina sequencing adapters. As this transposon preferentially integrates into regions of open chromatin the read-count for a specific region corresponds to the degree of chromatin accessibility. The data analysed were from ATAC-seq experiments using small pre-B cells (Mandal et al., 2015) and biological replicates were pooled to increase the sensitivity of the assay.

Whilst identifying open regions of chromatin can identify functional elements alone, I sought to also determine if any identified element also displayed enhancer like characteristics. Histone H3 lysine 4 monomethylation (H3K4me1) is a hallmark of all enhancers, whereas the presence of H3 lysine 27 acetylation (H3K27Ac) further defines an active enhancer (Creyghton et al., 2010). In addition, p300 is commonly located at active enhancers as well as at promoters and gene bodies. Therefore, I analysed published ATAC-seq, H3K4me1, H3K4me3, H3K27Ac and p300 ChIP-seq datasets from pre-B cells to locate potential enhancer elements within the *Igλ* locus (Figure 4.4).

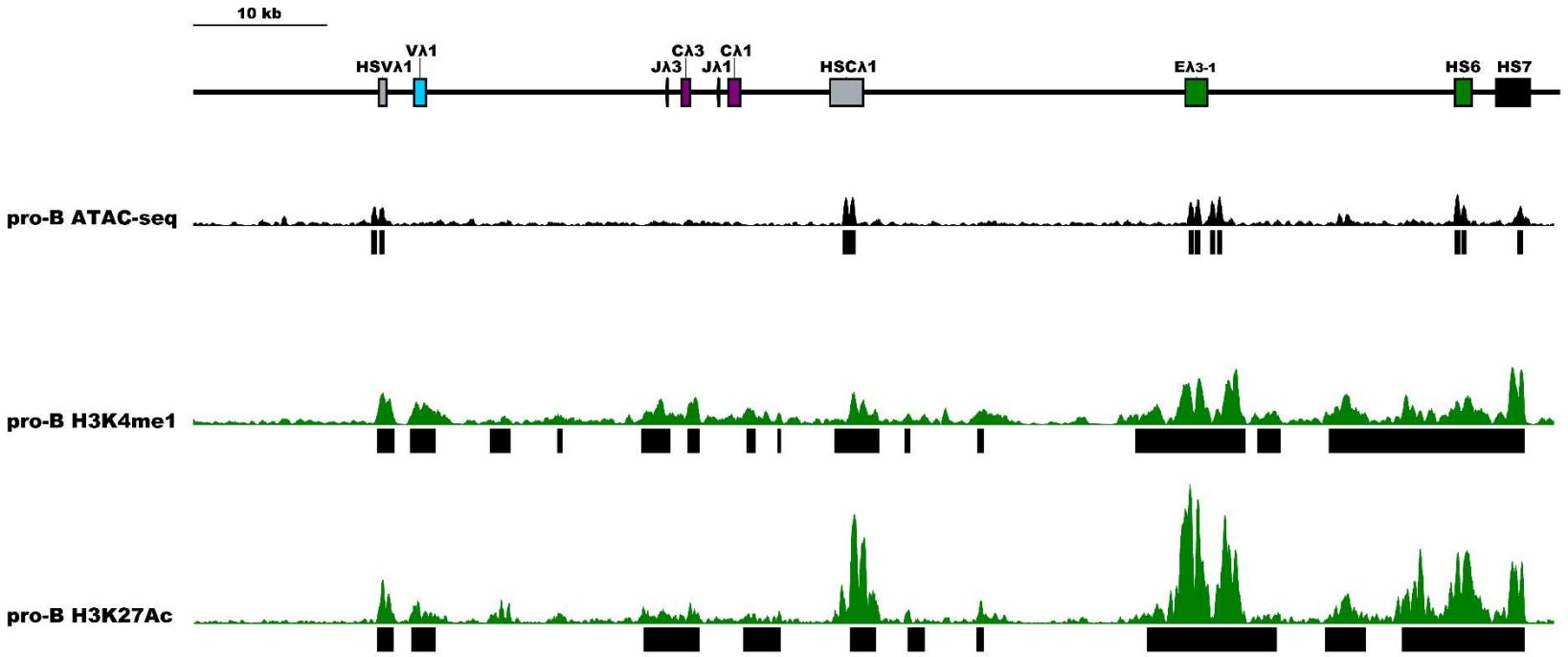


Figure 4.4: Analysis of pro-B ChIP-seq data in the 3' half of the *Igλ* locus to identify elements with enhancer-like characteristics. ATAC-seq data (GSM1545327) together with p300 (GSM1290115), H3K4me1 (GSM1463434) and H3K27Ac (GSM1463433) ChIP-seq data from pro-B cells were mapped onto the 3' half of the *Igλ* locus. Regions with a significant enrichment are annotated beneath each data trace. HSVλ1, HS6 and HS7 have also been defined by measurement of DNaseI hypersensitivity using a published pro-B DNaseI-seq dataset (GSM932968).

4.4 HS6 binds the same transcription factors as Eλ3-1

The data shown in Figure 4.4 highlighted additional hypersensitive sites with unknown function. Analysis of published ChIP-seq and ATAC-seq data revealed a region of open chromatin downstream of the Eλ3-1 enhancer termed HS6. This region displayed all hallmarks of an active enhancer, namely the presence of H3K4me1, H3K27Ac and p300. Several additional accessible regions of unknown function were identified: upstream of *Vλ1* (HSVλ1), downstream of Cλ1 (HSCλ1) and finally downstream of Eλ3-1 (HS7). The potential role of these elements in regulating *Igλ* activation is described later in this chapter.

As HS6 displayed the typical enhancer chromatin profile, I next assessed the transcription factors binding to this putative enhancer. As the Eλ3-1 enhancer binds the transcription factors IRF4, PU.1, E2A and MEF2C (Eisenbeis et al., 1993, 1995; Rudin and Storb, 1992), I examined the binding of these factors, with the exception of MEF2C, to HS6 using published ChIP-seq data from pro-B cells. Whilst ChIP-seq data from pre-B cells would provide a more accurate indication regarding the binding status of these transcription factors, pre-B cell datasets were not available. A low level of light chain recombination can be detected in primary pro-B cells and furthermore the same transcription factors appear to be present at the pro-B and pre-B cell stages, but the level of these factors changes (Muljo and Schlissel, 2003). This implies that ChIP-seq data from pro-B cells can provide preliminary evidence regarding the transcription factors that potentially bind in pre-B cells.

Interestingly, analysis of publicly available ChIP-seq data indicated that HS6 is also bound by the same transcription factors as Eλ3-1 (Figure 4.5). The binding of these factors was confirmed by ChIP-qPCR experiments performed using pre-B cells (Figure 4.6). IRF4 binding was enriched 9.34 ± 2.2 and 12.44 ± 2.2 fold over the intergenic negative control *IntIII* at Eλ3-1 and HS6, respectively. The binding of PU.1 to Eλ3-1 and HS6 was enriched 90.19 ± 4.2 fold and 286.5 ± 88.4 fold, respectively. E12/E47 also appear to bind strongly to both enhancers with binding enriched 4.75 ± 0.05 fold at Eλ3-1 and 8.69 ± 0.83 fold at HS6.

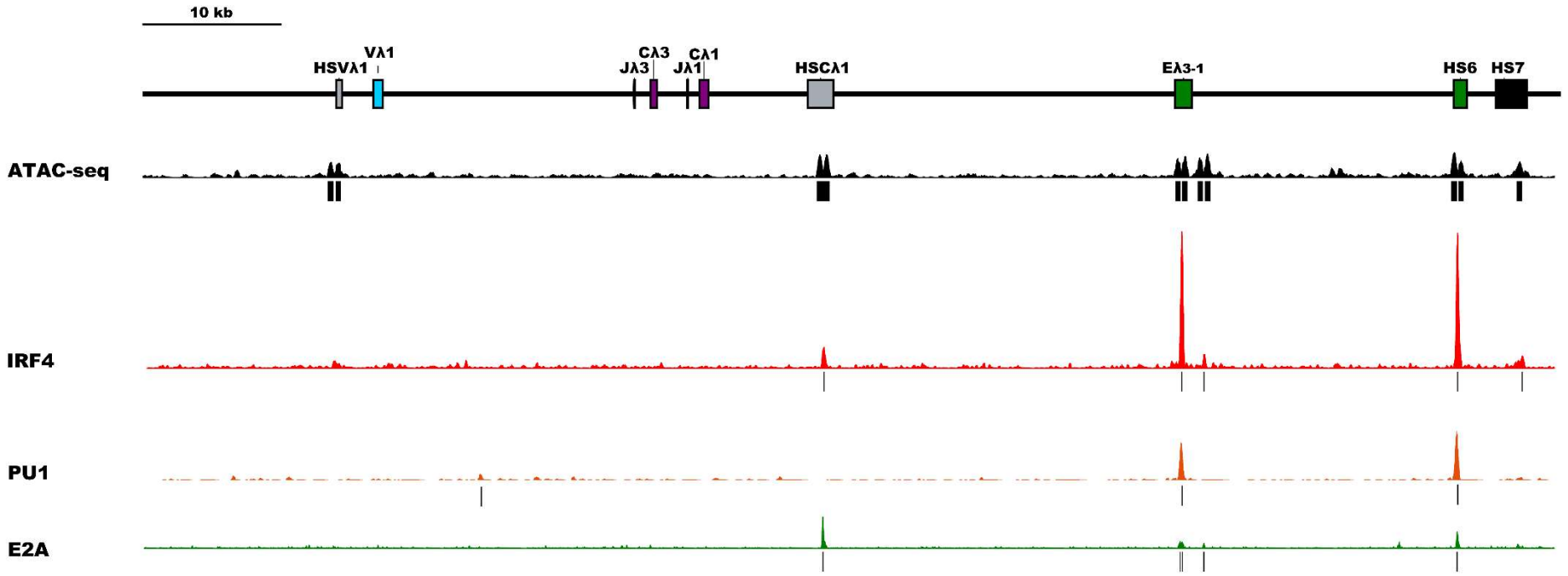


Figure 4.5: HS6 binds the same transcription factors as Eλ$_{3\text{-}1}$. ChIP-seq analysis of transcription factors known to bind to Eλ$_{3\text{-}1}$, IRF4 (GSM1296534), PU1 (GSM1290093) and E2A (GSM546523). Significant enrichment of the transcription factor binding is indicated by black rectangles. IRF4, PU.1 and E2A appear to bind to HS6 at similar levels to Eλ$_{3\text{-}1}$. All tracks shown have been adjusted to the same scale.

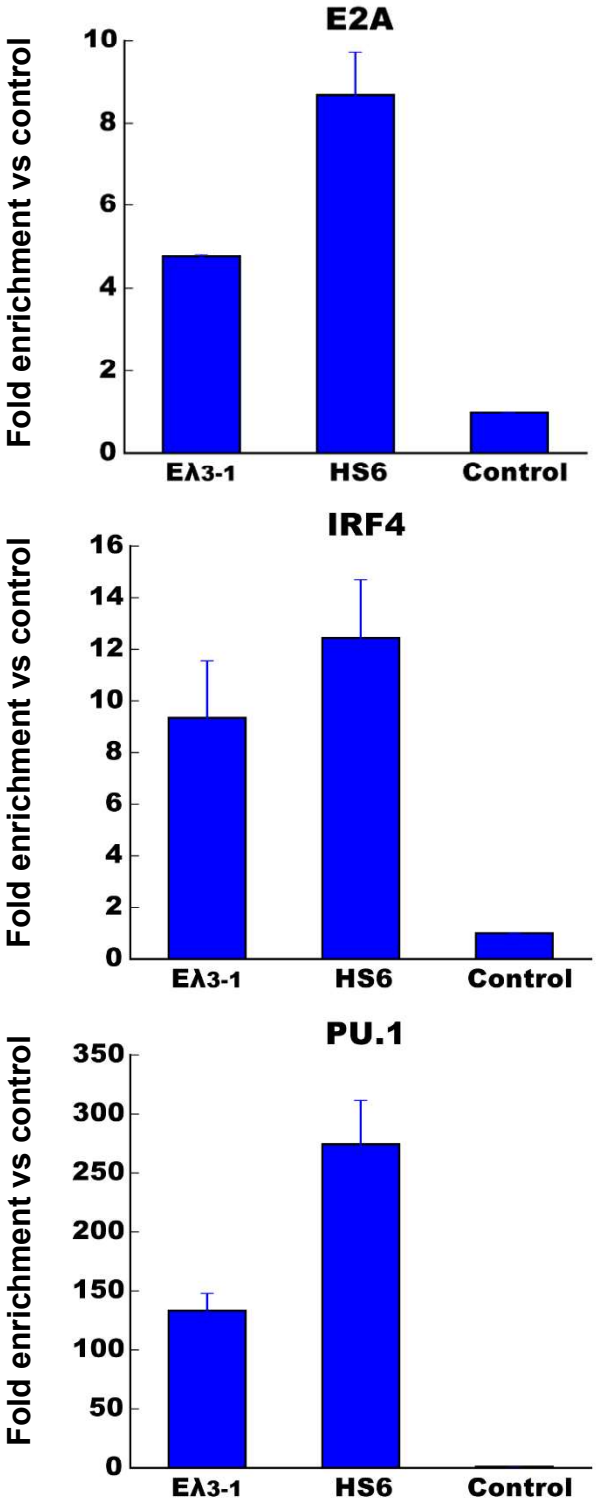


Figure 4.6: ChIP-qPCR analysis of E2A, IRF4 and PU.1 binding to Eλ$_{3\text{-}1}$ and HS6 in pre-B cells. The three factors are all bound at approximately similar levels at Eλ$_{3\text{-}1}$ and HS6. Error bars represent SEM (n = 3). These experiments were performed by James Scott and are included here for completeness.

4.5 HS6 is not a recently diverged duplication of Eλ3-1

As Eλ3-1 and HS6 were found to bind the same core transcription factors, I next analysed the sequence similarity between the two elements. Unexpectedly, Eλ3-1 and HS6 do not exhibit a high degree of sequence similarity (42.83%) implying that HS6 has not recently diverged from Eλ3-1 and is likely to be an unrelated element. Analysis of the transcription factor binding sites within both elements also supported this conclusion, as the relative locations of the transcription factor binding sites are distinctly different between the enhancers (Figure 4.7). Transcription factor binding sites at both Eλ3-1 and HS6 were identified by HOMER. ATAC-seq based bioinformatic footprinting, using the Wellington algorithm implemented in the pyDNase program, was performed on all identified transcription factor binding motifs to obtain additional confidence in the identification of bound motifs.

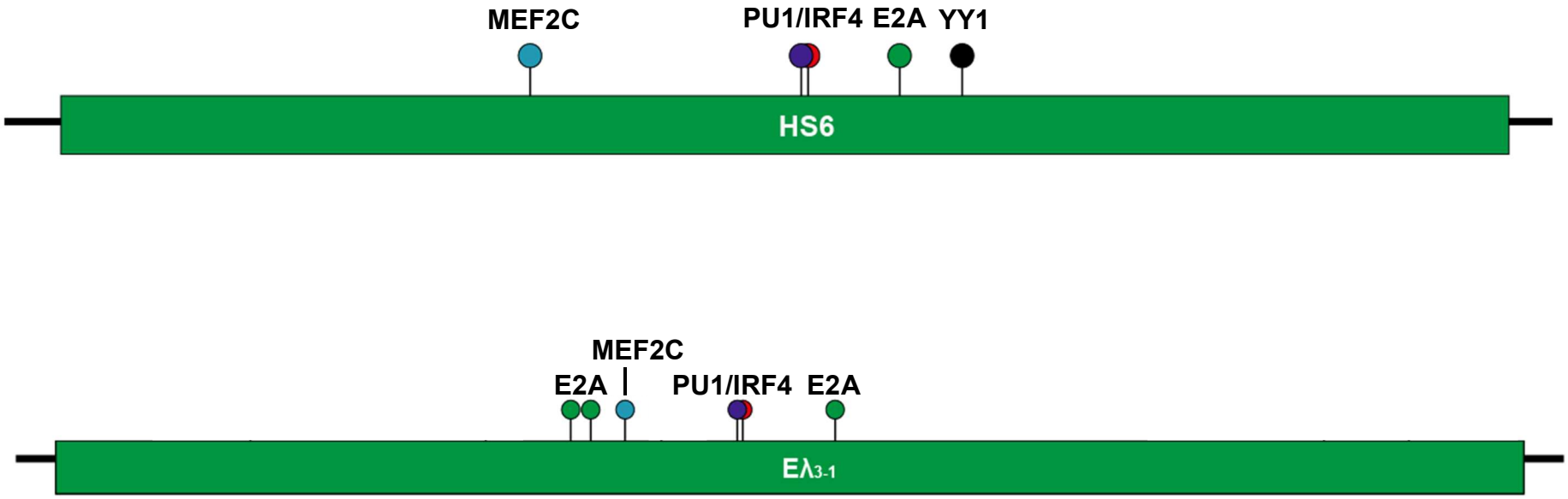


Figure 4.7: Analysis of transcription factor binding sites at HS6 (Top) and Eλ$_{3-1}$ (Bottom). Predicted transcription factor binding motifs are displayed for both enhancer regions. The same four transcription factors (IRF4, PU1, E12/E47 and MEF2C) are predicted to bind to both enhancer regions. The transcription factor binding sites shown have been filtered using ATAC-seq footprinting to identify binding sites with a high likelihood of being bound *in vivo*.

4.6 HS6 is an enhancer of *Vλ1* non-coding transcription in a transient transfection assay.

Due to HS6 exhibiting the typical chromatin profile of an enhancer and binding the same transcription factors as Eλ$_{3\text{-}1}$, I hypothesised that this element could be responsible for the upregulation of *Vλ1* non-coding transcription. To confirm that HS6 behaves as an enhancer of *Vλ1* non-coding transcription, HS6 was cloned into a luciferase expression vector together with the identified functional *Vλ1* promoter. This construct in addition to a *Vλ1* promoter only construct were separately electroporated into 103/BCL-2 cells. After 48 hours, the transfected cells were temperature shifted at 39°C for 16 hours to upregulate pre-B cell factors. Luciferase expression increased significantly (12.8 ± 0.92 fold, Students T-test $p < 0.001$) between the construct containing the *Vλ1* promoter alone compared to when HS6 was present (Figure 4.8). These data confirm that HS6 is an enhancer and importantly, it can enhance *Vλ1* non-coding transcription.

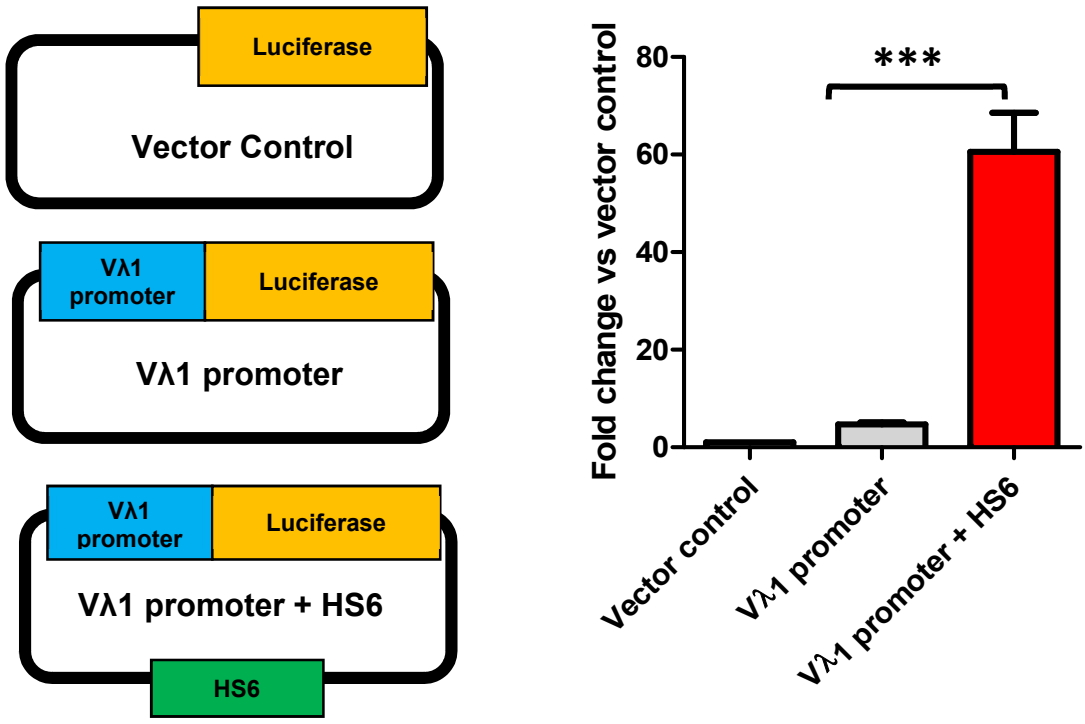


Figure 4.8: HS6 is an enhancer of the *Vλ1* promoter. A) Schematic of the luciferase assay plasmids used in this assay. B) *Vλ1* promoter activity both with and without the presence of HS6 was examined by luciferase reporter assay. A significant increase in luminescence was observed in the presence of HS6.

4.7 HS6 contacts *Vλ1 in vivo*

Whilst HS6 was found to be able to enhance the *Vλ1* promoter *in vitro*, I wanted to investigate if HS6 directly contacts the *Vλ1* promoter *in vivo*. Techniques such as 3C (Dekker et al., 2002) enable the identification of interacting regions of chromatin. Similar to ChIP, in 3C, cells are first cross-linked with formaldehyde to fix interacting regions of the genome. Following cell lysis, chromatin is digested with a restriction enzyme that is designed to fragment the locus of interest. The chromatin is then ligated under dilute conditions to promote intra-molecular ligation of cross-linked fragments, which are reversed, and interactions detected by PCR across the ligation junction (Figure 4.9).

More recent developments of 3C technology include 4C, 5C, Capture-C, Hi-C and Capture-Hi-C (Denker and de Laat, 2016). These methods use next generation sequencing to increase the sensitivity and the number of detected interactions. Whereas 4C, 5C and Capture-C detect interactions formed from a pre-selected viewpoint fragment, Hi-C interrogates all interactions in the genome in an unbiased fashion. To investigate the possibility that HS6 interacts with the promoter of *Vλ1* and to examine any other interactions that exist within the *Igλ* locus, I analysed Hi-C data from a published experiment performed using pro-B cells, as pre-B cell data was not available.

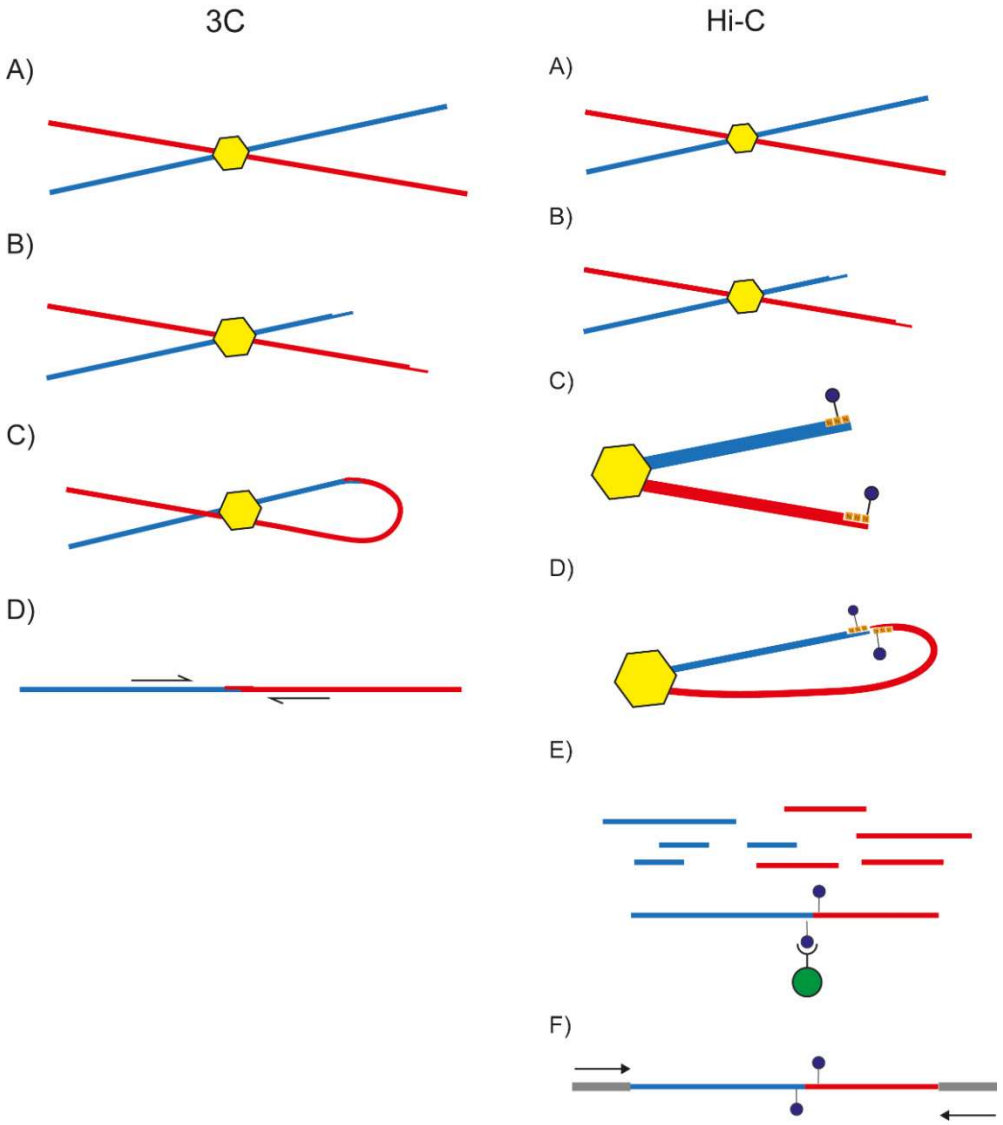


Figure 4.9: Schematic of Chromosome Conformation Capture (left) and Hi-C (right). 3C –A) Interacting regions (red and blue) are fixed by formaldehyde (yellow hexagon) forming a complex. B) The genome is digested by a restriction enzyme. C) Restriction fragments are ligated under very dilute conditions promoting intramolecular ligation. D) Crosslinks are reversed and the interactions are assessed by PCR across the ligation junction.

Hi-C – A) Interacting regions (red and blue) are fixed by formaldehyde (yellow hexagon) before digestion by restriction enzyme (B). C) Restriction fragment ends are repaired with biotinylated (purple) nucleotides before ligation under dilute conditions (D). The genome is then sonicated and interacting fragments are isolated by streptavidin (green) pulldown (E). Sequencing adapters are then ligated before paired end sequencing (F).

The unbiased nature of Hi-C presents substantial difficulties for the data analysis, due to the high coverage and low sequencing depth. To ameliorate this issue, many software packages bin Hi-C reads into genomic intervals of a fixed size, thus increasing the number of reads per region. Combining replicates can also increase the number of interacting reads to enable the significance of the interaction to be determined. Additional complications also arise from the very nature of 'C' based techniques, as fragments have a high tendency to randomly interact due to random collisions during the formaldehyde crosslinking. False positive interactions are accounted for in 3C by analysing the interaction profile across the region of interest. Real interactions are identified by a local peak in interaction frequency, compared to the background interaction frequency. For Hi-C data analysis, the algorithm used by software such as HOMER, involves the generation of a background model from the mapped sequencing data which is used to calculate the expected number of reads for each possible ligation product, under the hypothesis that no interaction exists. For two potentially interacting loci, the software models the randomly expected read count using the cumulative binomial distribution. Regions with only one or two interacting reads between them will have higher p-values, regardless of expected interaction frequency. Significant interactions are determined by examining if the number of interacting reads is significantly higher than the number of reads predicted by the background model.

The pro-B dataset was analysed at a resolution of 20 kb, combining the raw reads from two biological replicates to increase the signal. Hi-C analysis identified 36 significant interactions within the 3' half of the *Igλ* locus, (Appendix 2). Crucially, the Hi-C analysis indicated that HS6 does interact with *Vλ1* in pro-B cells. Furthermore, this analysis also confirmed that Eλ$_{3\text{-}1}$ contacts the *Jλ1* promoter. Interestingly, Hi-C data analysis also suggests that HS6 could also contact the *Jλ1* promoter (Figure 4.10). Notably, interactions were also observed between the hypersensitive site of unknown function HSCλ1 and *Vλ1*, *Jλ1*, HS6 and Eλ$_{3\text{-}1}$. The implications of these potential interactions are discussed later in the chapter.

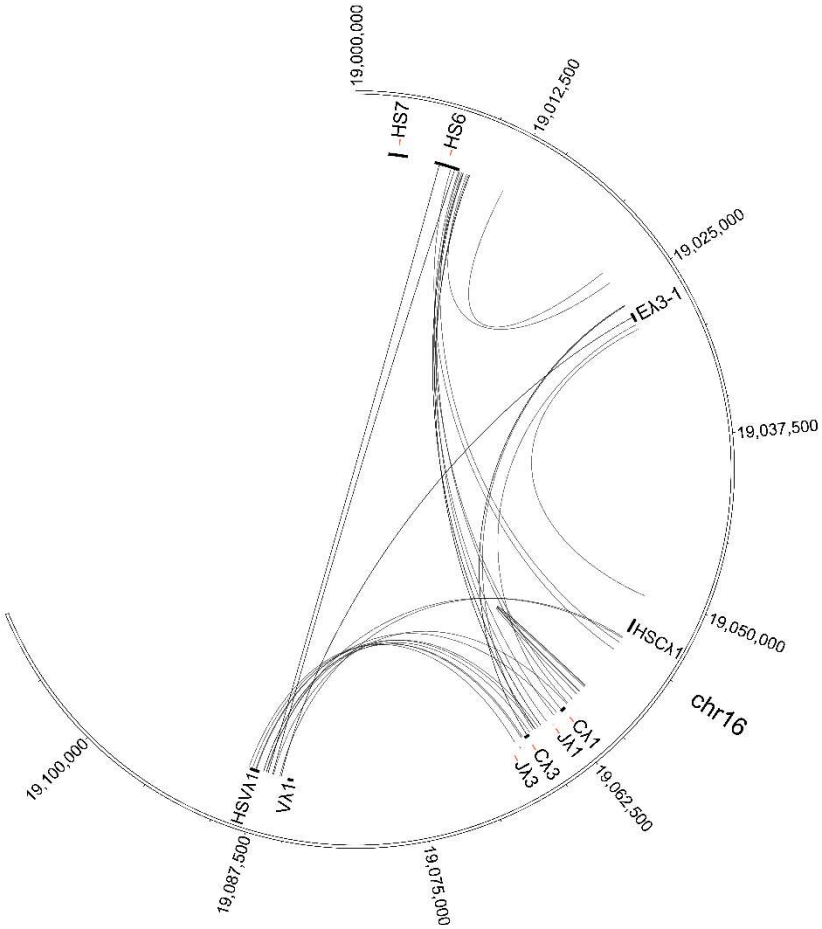


Figure 4.10: Analysis of pro-B Hi-C data performed using the restriction enzyme DpnII at a resolution of 20 kb. Significant interactions between regions in the 3' half of the *Igλ* locus are highlighted with black links. Importantly, clear interactions can be observed between HS6 and *VΛ1* and between Eλ$_{3\text{-}1}$ and *JΛ1*. Interactions can also be observed between HSCλ1 and *VΛ1*, *JΛ1*, HS6 and Eλ$_{3\text{-}1}$.

Analysis of Hi-C data using HOMER outputs a low number of false positive interactions (Forcato et al., 2017), but to confirm HS6 contacts *Vλ1*, I performed 3C on pre-B cells. I initially attempted to use conditions optimised by Sarah Bevington, which use the restriction enzyme BamHI to digest the chromatin. However, BamHI digests crosslinked chromatin very inefficiently in comparison to other enzymes frequently used for 3C for example EcoRI (data not shown). Whilst EcoRI digested crosslinked chromatin efficiently and liberated fragments that separated the promoters and enhancers of interest, this enzyme was unsuitable to analyse the *Igλ* locus. When I examined the final 200 bp of all restriction fragments present in the *Igλ* locus by a custom python script, nearly all fragments had >90% sequence similarity with another restriction fragment. The high degree of sequence homology between the 5' and 3' halves of the *Igλ* locus therefore prevented appropriate primers from being designed using EcoRI digested fragments. Consequently, I used the '4-bp cutter' DpnII, as this restriction enzyme digests much more frequently and generated fragments with a sufficiently unique sequence for the design of PCR primers.

The interaction frequency of HS6 with other elements and control regions was calculated relative to a standard curve. The latter was generated from a bacterial artificial chromosome (BAC) of the 3' half of the *Igλ* locus, Rp23-24i11. DpnII is unable to digest DNA with methylated adenosine and as the transformation of the BAC into a *Dam*- strain of *E.coli* was not an option due to the large size of the vector, the control BAC was digested with the DpnII isoschizomer SauA3. This BAC was then ligated to provide all possible 3C products in an equimolar ratio, preventing differences in primer efficiency from confounding the observed interaction frequency. Furthermore, 3C samples were normalised by the detection of an interaction in the *Ercc3* locus. This region encodes a subunit of TFIIH and is constitutively active. Hence it is assumed that the expression level of the gene and spatial conformation of the locus is the same in all cell types. Interactions are identified by visual inspection of the interaction frequency profile, with interacting regions appearing as local peaks in the profile. Furthermore, it is also of note that interaction frequency declines over linear distance, resulting in interactions far away from the viewpoint fragment appearing less pronounced.

Using HS6 as a viewpoint for the 3C analysis of 3' half of the *Igλ* locus in pre-B cells, I observed an increased interaction frequency of HS6 with *Vλ1*, compared to the control regions flanking *Vλ1*, especially the control region downstream of *Vλ1* (Figure 4.11). This implies that HS6 contacts *Vλ1* and together with the Hi-C analysis, strongly suggests this interaction occurs *in vivo*. In addition, the 3C analysis also supported the interactions of HS6 with *Jλ1* and HSCλ1, which were observed in the Hi-C dataset.

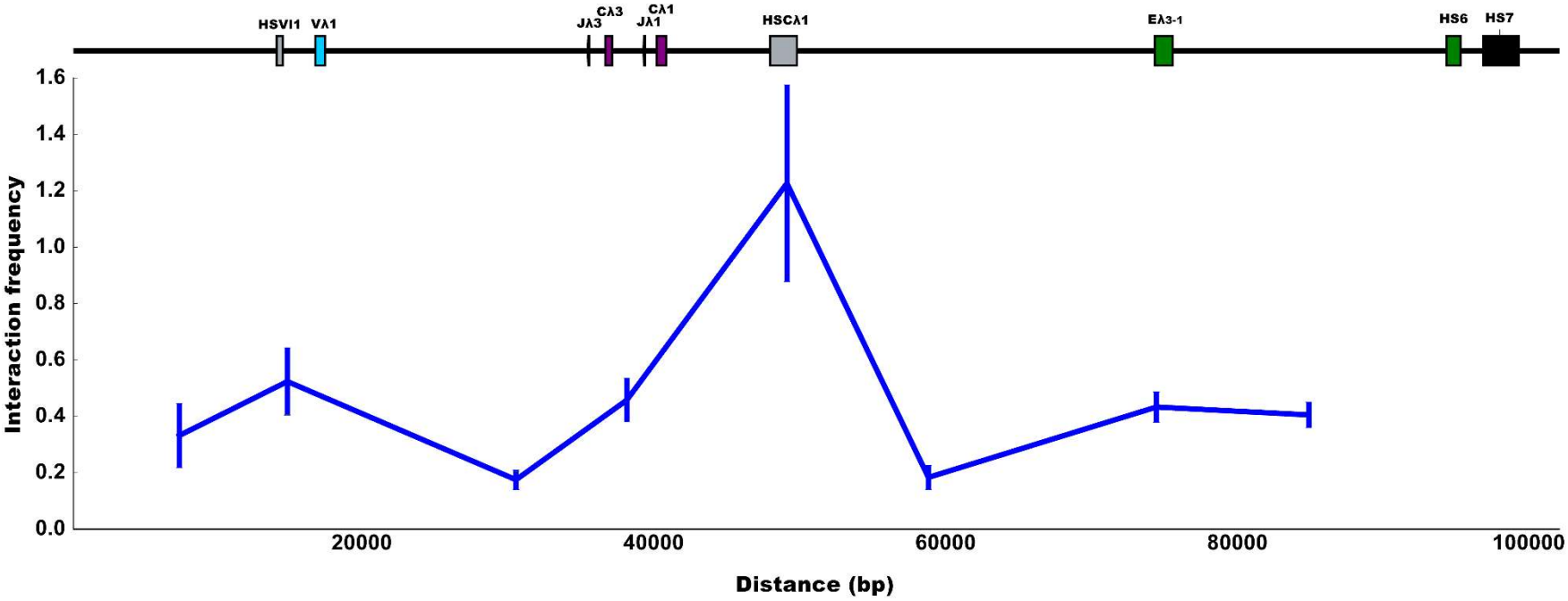


Figure 4.11: 3C analysis of the 3' half of the *Igλ* locus in pre-B cells using HS6 as the viewpoint. Relative interaction frequency of DpnII restriction fragments using HS6 as the viewpoint fragment. The data comprises of three pre-B replicates with error bars representing standard error of the mean (SEM).

4.8 CTCF binding sites mediate the HS6/Vλ1 interaction

After confirming that HS6 interacts with *Vλ1*, I sought to determine how this interaction is mediated. As HS6 and *Vλ1* are 85 kb away from each other, in linear space, I investigated the presence of factors known to be involved in the formation of long range interactions: CTCF, YY1 and the Mediator complex (Figure 4.12). To this end, ChIP-seq data from pro-B cells was examined to determine if these factors are bound within the *Igλ* locus. The presence of the Mediator complex was assessed via the enrichment of a core subunit, MED1. In addition to assessing the presence of these architectural proteins, I also investigated if the cohesin complex was present at any of the regions bound by these proteins by assessing the enrichment of the cohesin component RAD21.

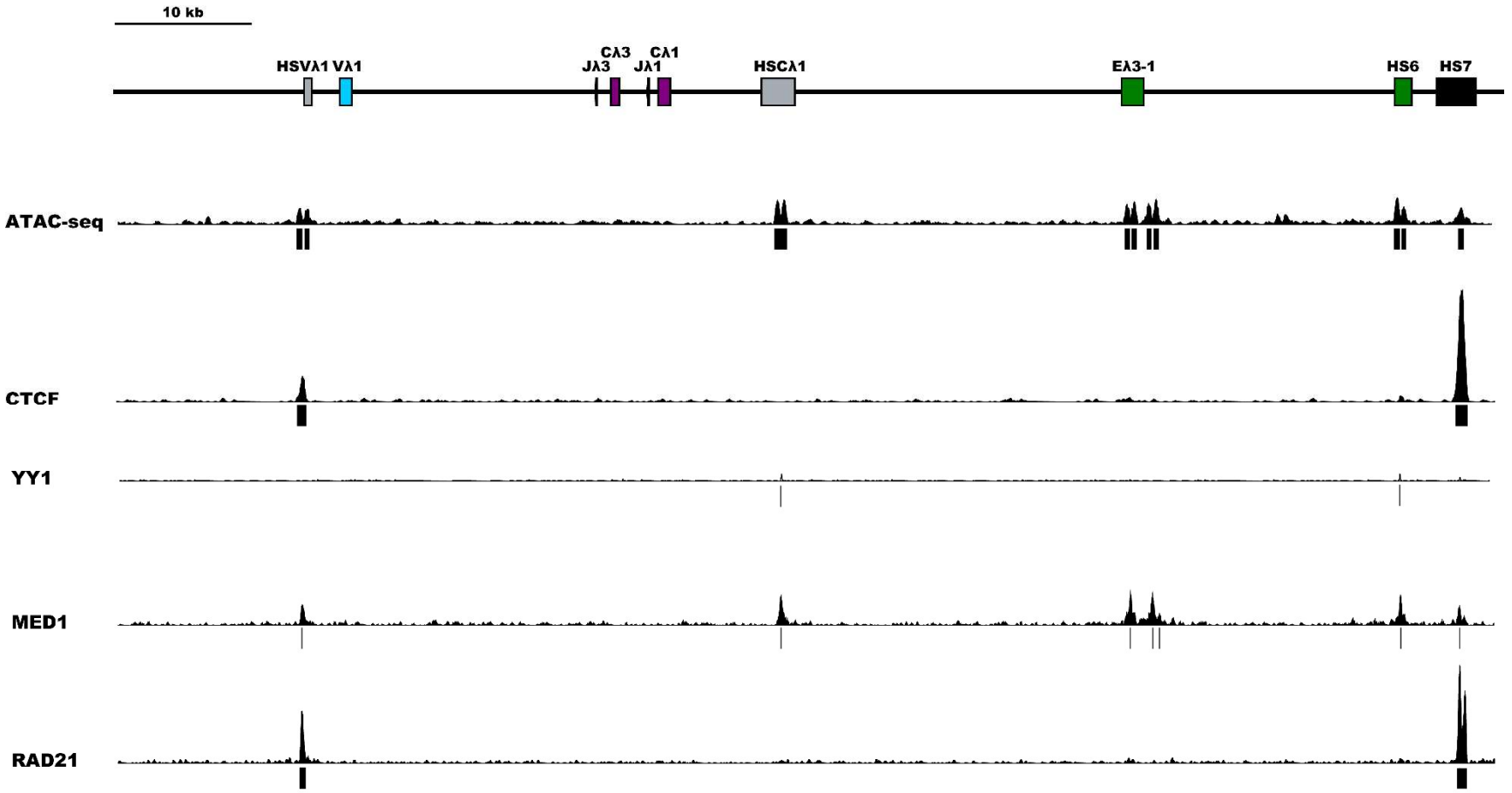


Figure 4.12: ChIP-seq data analysis of factors involved in long range chromatin interactions. ATAC-seq and ChIP-seq analysis of CTCF (GSM672401), YY1 (GSM1897389), a core component of the Mediator complex (MED1; GSM1038263) and a component of the cohesin complex (RAD21; GSM672403) enrichment. Peaks of significant factor enrichment are highlighted beneath the appropriate track. All tracks have been adjusted to the same scale.

Analysis of CTCF ChIP-seq data indicated enrichment of CTCF binding at the hypersensitive site referred to as HS7 and interestingly at the hypersensitive site approximately 3 kb upstream of *Vλ1*, HSVλ1 (Figure 4.12). ChIP-qPCR analysis of CTCF binding to HSVλ1 and HS7 using pre-B cells (Figure 4.13) revealed that CTCF binding is indeed enriched at HS7 and HSVλ1. Notably, the CTCF binding motifs present within these regions of open chromatin are in a convergent orientation (Figure 4.14). As interactions between sites bound by CTCF occur primarily between CTCF sites in a convergent orientation (de Wit et al., 2015), this implied that HS7 and HSVλ1 are likely to interact via CTCF/cohesin. Enrichment of RAD21 was detected at both HSVλ1 and HS7 (Figure 4.12) which again suggested that these regions interact via CTCF/cohesin. Modelling this interaction suggests the formation of an 85 kb domain at the 3' half of the locus (Figure 4.15) which would separate the 3' and 5' halves of the locus.

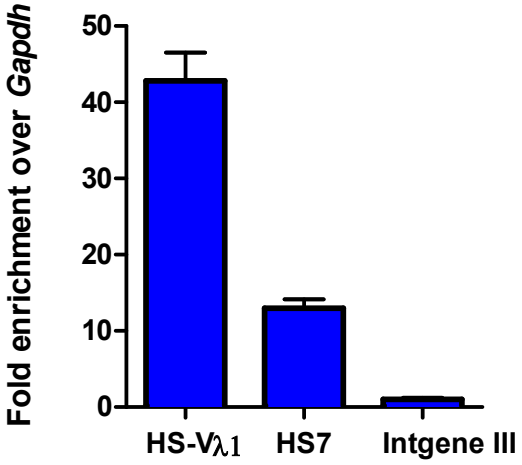


Figure 4.13: ChIP-qPCR analysis of CTCF binding to the regions of accessible chromatin upstream of *Vλ1* (HSVλ1) and downstream of HS6 (HS7) in pre-B cells. CTCF is strongly bound at both regions in comparison to the negative control region (IntIII). All values are normalised to the binding at *Gapdh*. Error bars display the SEM of three independent replicates. Pre-B samples were made in collaboration, but qPCR analysis performed by James Scott.

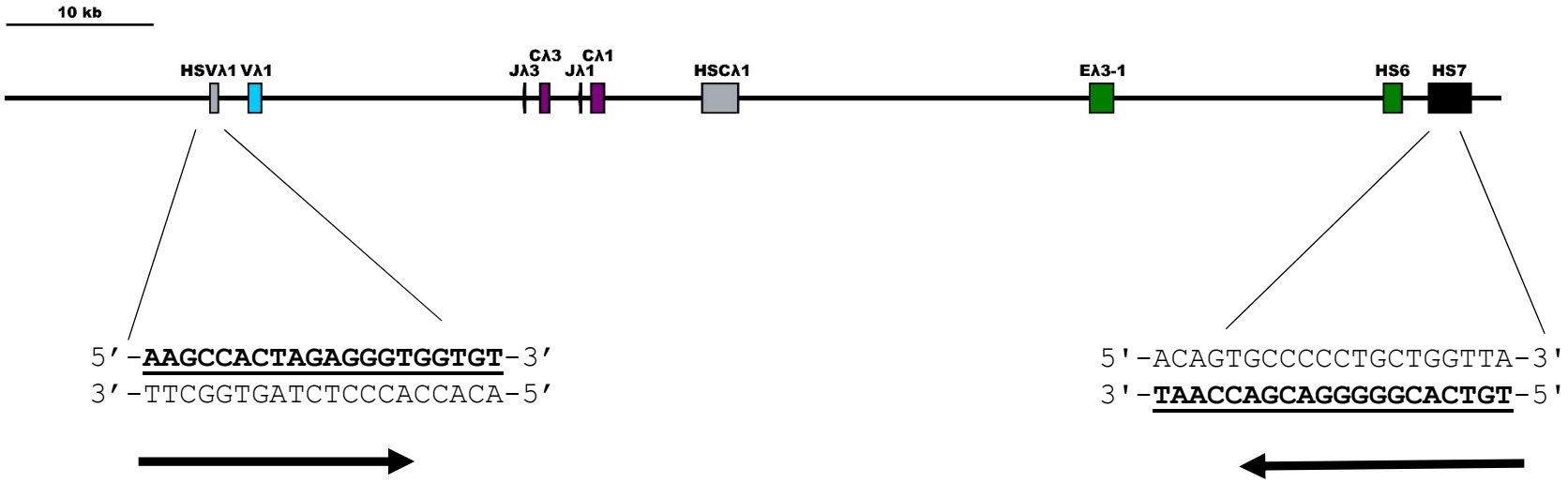


Figure 4.14: CTCF binding sites at HSVλ1 and HS7 are in a convergent orientation. Analysis of the CTCF binding site orientation at HSVλ1 and HS7 by HOMER, binding sites are underlined. The binding sites at these two regions are in a convergent orientation implying interaction between these regions.

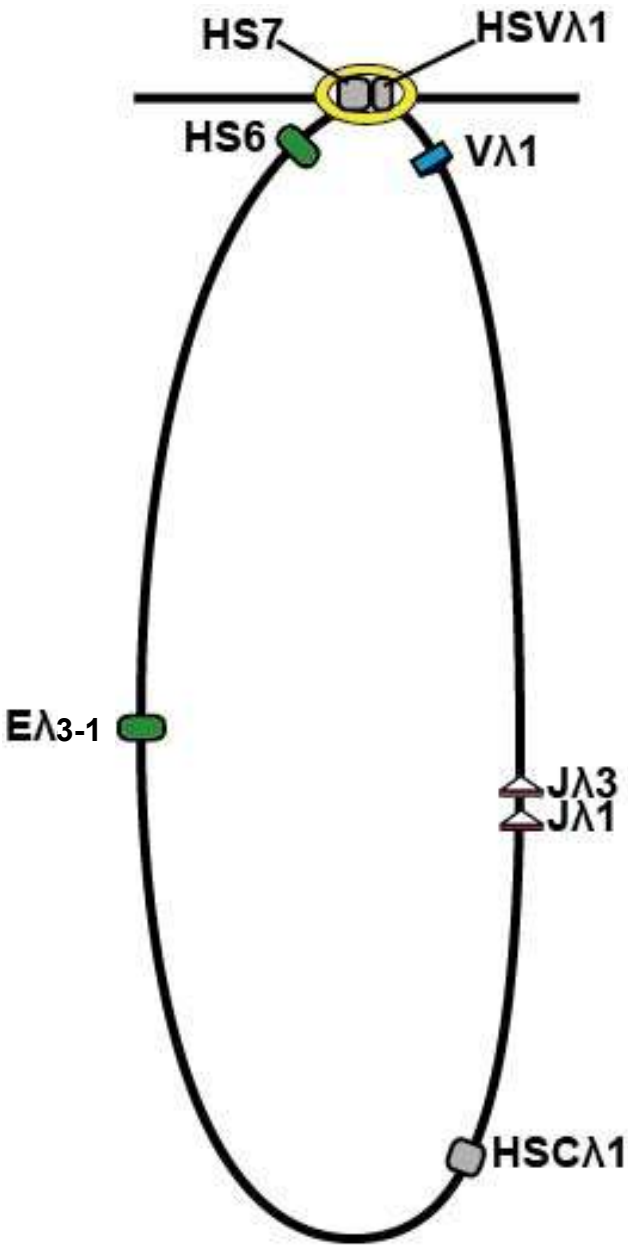


Figure 4.15: Diagram indicating the proposed structure of the 3' half of the *Igλ* locus if HS7 interacts with HSVλ1. Interaction of HS7 with HSVλ1 would result in the formation of a chromatin loop separating the two locus halves. This would also bring *Vλ1* into closer proximity with HS6. Cohesin is indicated in yellow.

4.9 The *Igλ* locus is separated into domains by CTCF/cohesin

Analysis of CTCF binding sites across the entire *Igλ* locus indicated the presence of CTCF binding elements downstream of the enhancer Eλ2-4 (HS7-1), upstream of Vλ2 (HSVλ2) and upstream of *Vλx* (HSVλx) (Figure 4.16). As RAD21 binding colocalised with these CTCF binding sites and furthermore these binding sites are in a convergent orientation (the binding motifs upstream of *Vλ2* and *Vλx* were orientated 5'-3' and the binding site downstream of Eλ2-4 is orientated in the 3' - 5' direction) implying that the CTCF binding sites upstream of upstream of *Vλ2* and *Vλx* interact with the binding site downstream of Eλ2-4, thus forming a domain separating the 5' and 3' halves of the locus.

As the data from the 3' half of the locus strongly implied the presence of a CTCF/Cohesin mediated interaction, I examined the idea that HSVλ2 and HSVλx can interact with HS7-1. To this end, I re-analysed the Hi-C data from pro-B cells to investigate the interactions occurring at the 5' half of the *Igλ* locus. This analysis suggested an interaction between HSVλ2 and HS7-1 in addition to HSVλx and HS7-1 (Figure 4.17). Notably, this analysis revealed that the same interactions, with the exception of those occurring with HSCλ1, are present in the 5' and 3' half of the locus. These data imply that similar long-range interactions govern the 5' as the 3' half of the locus. Figure 4.18 shows the proposed long-range interactions that form the 5' and 3' domains within the *Igλ* locus.

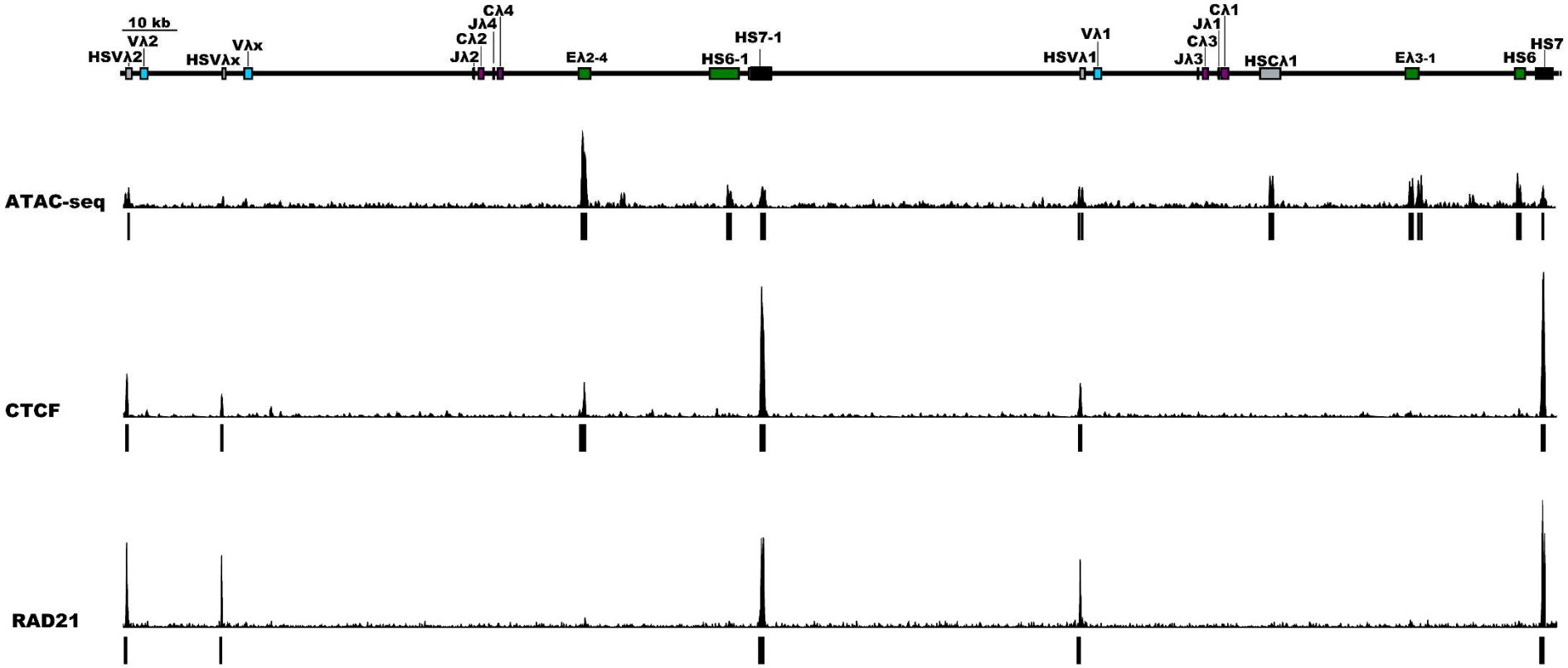


Figure 4.16: The CTCF binding pattern is mirrored in the 5' half of the locus. ChIP-seq analysis of the *Igλ* locus. Regions upstream of *Vλ2* and *Vλx* display CTCF and RAD21 binding. Furthermore, the element termed HS7-1, which exhibits high sequence homology with HS7 also binds CTCF and RAD21. Significant enrichment of the factor within a region is indicated by black rectangles.

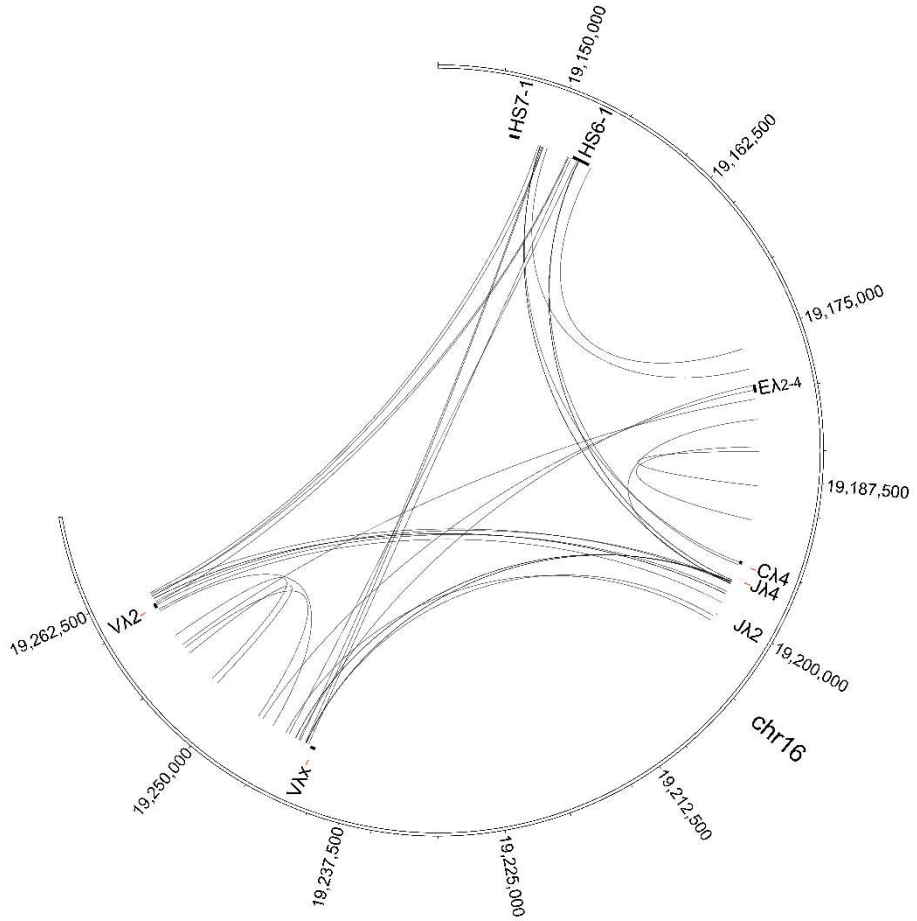


Figure 4.17: Analysis of pro-B Hi-C data performed using the restriction enzyme DpnII at a resolution of 20 kb. Significant interactions between regions in the 5' half of the *Igλ* locus are highlighted with black links. Importantly, clear interactions can be observed between HS6-1/HS7-1 and *Vλ2/VλX.*

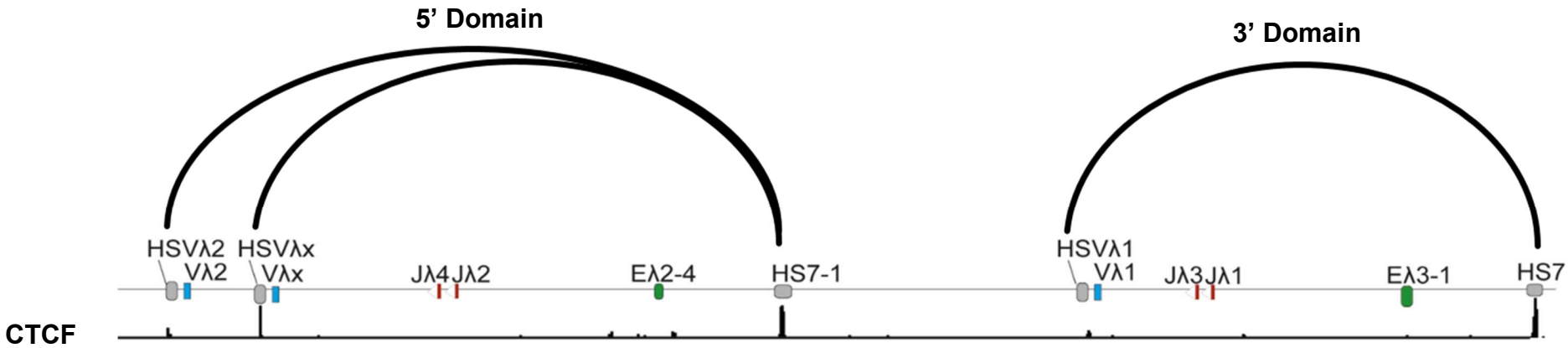


Figure 4.18: Proposed long-range interactions in the Igl locus mediated by CTCF/cohesin. CTCF binding observed at accessible regions upstream of V gene segments interact with CTCF bound regions downstream of the *Igλ* enhancers, mediating the formation of a 5’ domain containing either *Vλ2* or *VλX* to HS7-1 and a 3’ domain containing *Vλ1* to HS7.

As the data suggested that the same long-range interactions regulate the 5’ half of the *Igλ*, I therefore wanted to investigate if these regulatory elements exist due to the duplication event that gave rise to the *Igλ* locus and if there were any conserved regulatory elements that had not been identified. To achieve this, the *Igλ* locus was split into 1 kb windows which were then mapped back to the *Igλ* locus by BLAST (Basic Local Alignment Search Tool), excluding any complete matches. Figure 4.19 shows the regions displaying homology with the 3’ half of the locus. Interestingly, two of the CTCF binding sites observed in the 5’ half of the locus (HSVλ2 and HS7-1) exhibit a high degree of homology (89-94%) with their counterpart in the 3’ half of the locus (HSVλ1 and HS7). Furthermore, an enhancer like element with a 92% sequence identity to HS6, referred to as HS6-1, was identified by this search. This further suggests that the newly identified enhancer element HS6 is very likely to be vital for locus function due to its high conservation. Interestingly, no counterpart for the HSCλ1 element could be observed in the 5’ half of the *Igλ* locus and notably this appeared to be only element without a counterpart.

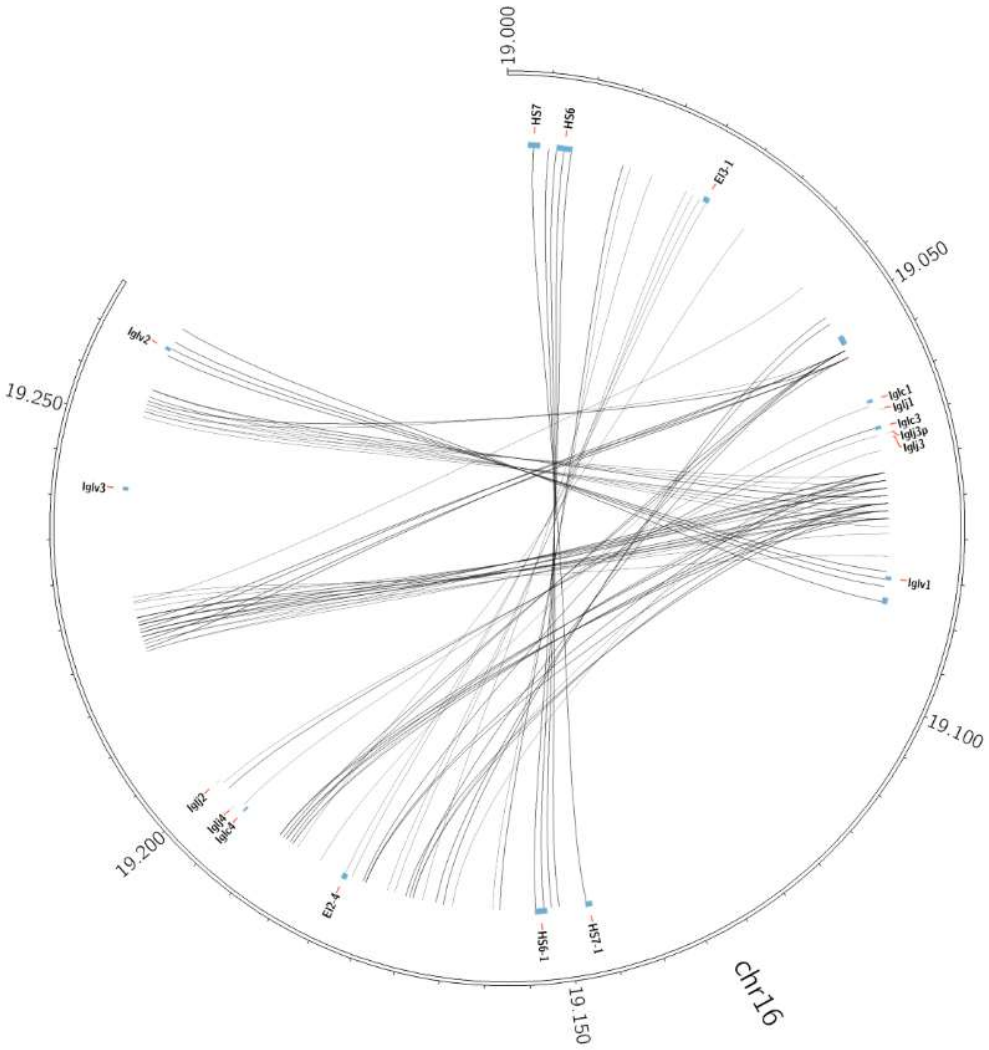


Figure 4.19: Analysis of homologous regions between the 5' and 3' half of the *Igλ* locus. The 3' half of the locus was binned into 1 kb windows and homologous regions were identified by BLAST. Homologous regions are identified by black links. Both HS6 and HS7 have highly homologous counterparts in the 5’ half of the locus.

4.10 How is the enhancer Eλ3-1 involved in *Igλ* locus activation?

Modelling the interaction between HS7 and HS*Vλ1* suggests that this interaction brings HS6 and *Vλ1* into close proximity in three-dimensional space (Figure 4.15). This interaction is likely to mediate the activation of *Vλ1* non-coding transcription by HS6. The enhancer Eλ3-1 has been shown to contact *Jλ1*, however, how this is mediated was not well understood. A surprising finding of the 3C assay was the identification of a potential interaction between HS6 and Eλ3-1. This contact can also be observed using pro-B Hi-C data, and strongly implies that HS6 and Eλ3-1 must be brought into close proximity for the activation of the 3' half of the *Igλ* locus. Interestingly, a YY1 binding site was identified within the HS6 enhancer and furthermore analysis of pro-B ChIP-seq data indicated that this site appears to be occupied (Figure 4.12). However, ChIP-seq and sequence analysis failed to reveal any shared factors between these sister enhancers that could mediate their interaction.

Interestingly, both the analysis of pro-B Hi-C data and 3C analysis using HS6 as the viewpoint identified an interaction between HS6 and HSCλ1. Moreover, this was one of the stronger interactions identified by 3C. As YY1 binding sites can be observed at both HS6 and HSCλ1 and YY1 ChIP-seq using pro-B cells suggests that YY1 is bound to both of these regions, it seemed likely that this interaction is mediated by YY1. Surprisingly, Hi-C data from pro-B cells suggested an interaction between Eλ3-1 and HSCλ1 (Figure 4.10). This interaction was also implied by 3C analysis of pre-B cells using Eλ3-1 as the viewpoint as the interaction frequency does not decline as rapidly as would be expected according to the distance between Eλ3-1 and HSCλ1 (Figure 4.20). It therefore appeared plausible that HSCλ1 could mediate the interaction of the two sister enhancers, implying that a transcriptional hub centred on HSCλ1 is formed.

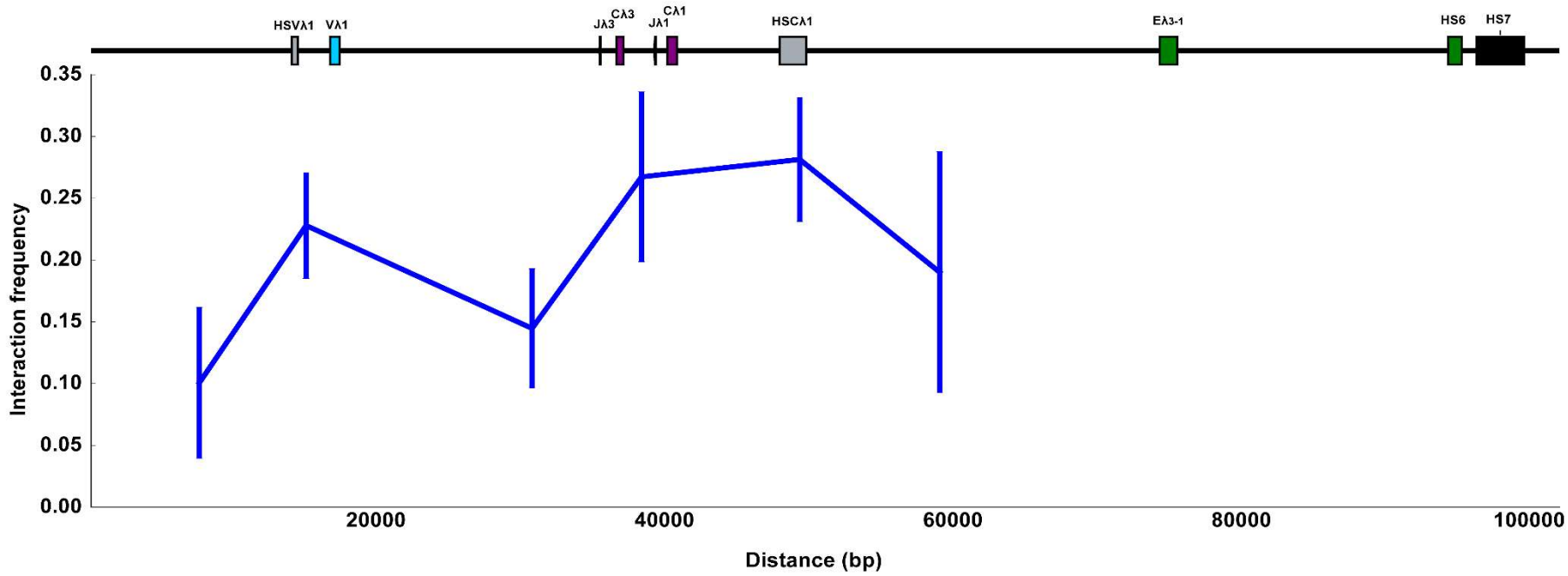


Figure 4.20: 3C analysis of the 3' half of the *Igλ* locus in pre-B cells using Eλ$_{3\text{-}1}$ as the viewpoint. Relative interaction frequency of DpnII restriction fragments using Eλ$_{3\text{-}1}$ as the viewpoint fragment. The data comprises of three pre-B replicates with error bars representing SEM.

4.11 Analysis of transcription factors binding to both enhancer elements

The luciferase and 3C assays indicated that not only does HS6 contact *Vλ1*, it is also able to enhance *Vλ1* transcription. Data from 3C performed by Sarah Bevington together with the pro-B Hi-C data analysis and 3C analysis performed using pre-B cells confirm that Eλ$_{3\text{-}1}$ contacts *Jλ1*. Furthermore, ChIP-seq data and ChIP analysis of pre-B cells indicated that these sister enhancers are bound by the same transcription factors: IRF4, PU.1 and E12/E47.

Non-coding transcription of *Vλ1* and *Jλ1* is increased approximately three-fold from the pro-B to pre-B cell stage and I therefore sought to ascertain how these sister enhancers are activated at this stage. Remarkably, the only observable difference between transcription factor binding at Eλ$_{3\text{-}1}$ and HS6 in between pro-B and pre-B cells was an approximate three-fold increase in IRF4 binding to both enhancers (Figure 4.21), suggesting that IRF4 binding may regulate the activity of these enhancers. Moreover, the activation of a single transcription factor, IRF4, in the pro-B cell line 1D1-T215 is sufficient to upregulate *Igλ* locus non-coding transcription within four hours. This also implies that IRF4 binding is sufficient for enhancer activation and furthermore it was highly intriguing that IRF4 binding to both enhancers was increased at the pre-B cell stage. This possibly suggests that IRF4 binding is increased simultaneously at these sister enhancers, which would result in the parallel activation of both enhancers. This could potentially explain the coordinate increase in *Vλ1* and *Jλ1* non-coding transcription.

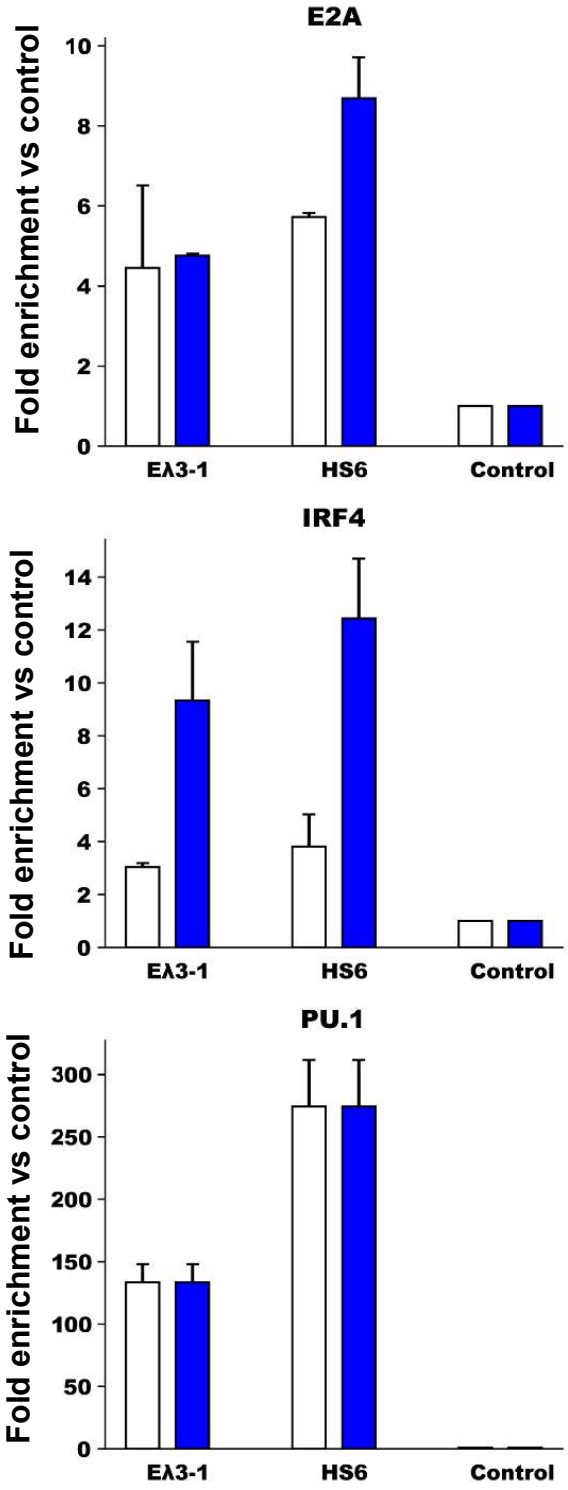


Figure 4.21: ChIP-qPCR analysis of E2A, IRF4 and PU.1 binding to Eλ$_{3\text{-}1}$ and HS6 in pro-B and pre-B cells. Fold enrichment over the negative control *Gapdh* is plotted. Samples were analysed from non-transgenic pro-B (white) and pre-B cells (blue). Error bars represent the SEM from three replicate experiments. The pre-B data shown are also represented in Figure 4.6.

C. Discussion

The precise regulation of gene segment activation is vital for the selection of the appropriate gene segments for V(D)J recombination. Aberrant gene segment activation could generate substrates for oncogenic chromosome translocations and it is therefore vital to better understand how this process is regulated. An ideal model system to investigate the regulation of gene segment activation is the murine *Igλ* locus. The *Igλ* locus is the smallest of the antigen receptor loci and exhibits a substantial bias in recombination frequency between two gene segments, *Vλ1* and *Jλ1.* This bias enables all efforts to be focused on the gene segments that recombine the most frequently. Analysis of *Vλ1* and *Jλ1* non-coding transcription in pro-B cells from PIP-ER transgenic mice and in the cell line 1D1-T215 showed that the non-coding transcription of these gene segments is coordinately upregulated. How this coordinate activation is achieved is a vital question to address.

4.12 Identification of the *Vλ1* promoter

Mapping of the *Vλ1* promoter by 5’ RACE suggested that there are three possible promoter locations. Notably, these three regions all correlated with regions of open chromatin identified using ATAC-seq data. However, analysis of RNA-seq data from pre-B cells only identified one start site at -39 bp, suggesting this start site was the most frequently utilised. Analysis of promoter activity by the luciferase assay indicated that only the 20-600 bp region upstream of the *Vλ1* protein coding sequence exhibited promoter activity. Furthermore, this was also the only potential promoter region to upregulate Luciferase expression when paired with an enhancer, suggesting that this is the only true promoter.

The identification of two regions without any substantial promoter activity could be attributed to the 5' RACE technique used. The 5' RACE protocol used cannot distinguish between partially reverse transcribed transcripts or truncated RNAs and mature mRNAs. The identification of full length mRNA could be substantially improved by 'SMARTER-RACE' methods that include an incubation step of the RNA with alkaline phosphatase to remove terminal phosphates, followed by Tobacco Acid Pyrophoshatase treatment to remove cap structures and generate phosphates for the introduction of tags. Whilst the protocol used to determine the transcription start sites of the *Vλ1* promoter was inefficient, it did identify a promoter region that appears to be functional. A more efficient technique would be the use of capped analysis of gene expression with next generation sequencing (CAGE-seq). In addition to identifying all start sites with base pair resolution, this technique would also enable the quantification of these transcripts to determine the most frequently used site.

4.13 Analysis of transcription factor binding to the *Vλ1* and *Jλ1* promoters

The identification of the *Vλ1* promoter enabled the analysis of shared regulatory elements between the two promoter regions. Initially, I attempted to analyse IRF4, PU.1 and E12/E47 ChIP-seq data performed pro-B cells, however, the peak calling algorithm used by MACS2 failed to identify any significant enrichment in the binding of any of these factors. As the *Igλ* locus only recombines in 5-10% of pre-B cells (Woloschak and Krco, 1987), it is possible that the signal for the binding of these factors at the promoters was below the limit of detection. Therefore, the analysis of transcription factor binding was performed by examining transcription factor binding motifs. Whilst the identification of a transcription factor binding motif strongly indicates that a given factor can bind to a region there are several caveats to this method as the threshold for motif detection is arbitrary, which can prevent the identification of potential transcription factor binding events and also a motif occurrence does not directly correlate to a true binding event. Whilst assessing identified motifs by ATAC-seq footprinting provides additional confidence in the identification of bound motifs, transcription factor binding should still be confirmed *in vivo.*

Analysis of transcription factor binding motifs present in the *Vλ1* and *Jλ1* promoters, by HOMER, revealed that the only factor common to both promoters was E2A. Both E12 and E47 play a critical role in pre-B cells to promote both non-coding transcription and recombination of the *Igλ* locus (Beck et al., 2009). The expression of E47 is vital for B-cell development as E47 knockout mice exhibit a complete block at the pre-pro-B stage (Hardy fraction B; Goebel et al., 2001; Lazorchak et al., 2006). Furthermore, E12 knockout mice exhibit a significant decrease in the fraction of *Igλ*-expressing cells in the bone marrow and the spleen (Beck et al., 2009). Notably, the pre-B cells of E12 knockout mice exhibit a reduction in both *Igλ* non-coding transcription and active histone marks (H3K4me3 and H3Ac; Beck et al., 2009). Whilst E12 and E47 are crucial for correct B-cell development, there is no evidence that IRF4 directly or indirectly mediates the expression of either protein. IRF4 can interact with E12/E47, promoting its binding to DNA and therefore can have a synergistic effect (Nagulapalli and Atchison, 1998; Nagulapalli et al., 2002). However, the addition of a single base pair between IRF4 and E12/E47 binding sites completely abolishes this effect (Nagulapalli et al., 2002). Since overlapping IRF4/E2A motifs were not identified, this synergistic effect is unlikely to mediate the coordinate activation of *Vλ1* and *Jλ1*.

4.14 Identification of an enhancer for the *Vλ1* promoter

Previously, only two enhancers were characterised in the *Igλ* locus, Eλ$_{3\text{-}1}$ and Eλ$_{2\text{-}4}$. These enhancers share greater than 90% homology and are present in the 3' half and 5' half of the locus respectively. Previous 3C analysis by a former lab member, Sarah Bevington, implied that Eλ$_{3\text{-}1}$ contacts the *Jλ1* promoter, resulting in the upregulation of *Jλ1* non-coding transcription. This analysis, however, suggested that Eλ$_{3\text{-}1}$ contacted the promoter of *Vλ1* only infrequently. Analysis of ATAC-seq and histone modifications (H3K4me1, H3K27Ac) identified the presence of a previously uncharacterised enhancer-like element downstream of Eλ$_{3\text{-}1}$, referred to as HS6.

Interestingly the sequence of HS6 is dissimilar to that of Eλ$_{3-1}$, strongly implying that HS6 is not a duplicate of the Eλ$_{3-1}$ enhancer. Furthermore, when the *Igλ* was examined for regions with a high degree of homology, a counterpart to HS6 in the 5' half of the locus was identified (HS6-1). HS6 and HS6-1 exhibit a high degree of sequence conservation (92.03%), suggesting that these regions are involved in the regulation of their respective halves of the locus. I therefore propose that Eλ$_{3-1}$ and HS6 are sister enhancers and that both enhancers are involved in the regulation of gene segment activation.

As HS6 exhibited the chromatin characteristics of an enhancer: high accessibility and enrichment of both H3K4me1 and H3K27Ac, I hypothesised that HS6 could be the unidentified enhancer of the *Vλ1* promoter. The increase in *Vλ1* transcription when paired with HS6 very strongly indicated that HS6 can enhance the expression of *Vλ1* non-coding transcripts. However, a similar effect can also be observed in the presence of other enhancer elements and therefore, whilst HS6 may act as an enhancer *in vitro* it was unclear if the chromatin structure would be permissive for HS6 to act as an enhancer for *Vλ1 in vivo*.

In order to confirm that HS6 could interact with *Vλ1 in vivo,* I utilised both Hi-C data and 3C to investigate the interactions formed by HS6. The analysis of pro-B Hi-C data revealed a significant interaction of the region containing HS6 with the window containing *Vλ1*, suggesting that HS6 contacts *Vλ1 in vivo.* Whilst the algorithm used by HOMER has low rate of error (Forcato et al., 2017), confirmation of these interactions was obtained by the analysis of interactions formed by HS6 in pre-B cells using 3C. The increase in interaction frequency between *Vλ1* and the adjacent fragments strongly implied an interaction between these two regions. Together with the Hi-C data analysis, these data add weight to the hypothesis that HS6 is an enhancer for the *Vλ1* promoter.

4.15 Which interactions mediate the coordinate activation of *Vλ1* and *Jλ1*?

The enhancer HS6 and *Vλ1* are separated by over 85 kb of intervening DNA, and yet appear to interact. Long-range interactions mediated by CTCF, YY1 and cohesin have been shown to be vital for the efficient recombination of the *IgH*, *Igκ* and T cell receptor loci (Benner et al., 2015; Guo et al., 2011; Lin et al., 2015; Medvedovic et al., 2013; Ribeiro de Almeida et al., 2011; Volpi et al., 2012). These interactions are responsible for bringing domains of recombining gene segments into closer proximity. No such interactions have previously been described for the *Igλ* locus, with the prevailing hypothesis being that locus contraction is unnecessary due to the vastly reduced size of the *Igλ* locus compared to the other loci. In contrast, the analysis of the binding of looping factors (CTCF and cohesin) and 3C/Hi-C data strongly imply that the *Igλ* does undergo contraction prior to recombination.

ATAC-seq data analysis indicated the presence of previously uncharacterised regions of open-chromatin, HS*Vλ1*, HSCλ1 and HS7. A hypersensitive site approximately 3 kb upstream of *Vλ1* has previously been identified (Hagman et al., 1990), however, its function was not assessed. CTCF binding analysis by ChIP-seq indicated that these hypersensitive sites both bind CTCF and furthermore, CTCF binding can also be observed at the newly identified site HS7. This binding was not due to sequence artefact as CTCF binding was also detected at these sites by ChIP-qPCR. CTCF forms long range interactions by interacting with cohesin, and a critical subunit of this complex is RAD21 (Wendt et al., 2008). ChIP-seq data analysis for RAD21 in pro-B cells detected the presence of this protein at HSVλ1 and HS7, potentially implying an interaction between these elements. Furthermore, the CTCF binding motifs at these regions are in a convergent orientation. Together the presence of cohesin and the convergent orientation of these CTCF motifs suggest that the 3' half of the locus forms a loop mediated by CTCF and cohesin binding at HS*Vλ1* and HS7. Analysis of Hi-C data and preliminary 3C experiments performed by James Scott also strongly imply the presence of an interaction between these two regions.

Moreover, the formation of a loop between accessible regions upstream of V gene segments appears to be a general regulatory mechanism in the murine *Igλ* locus. The accessible region upstream of *Vλ1*, HSVλ1, has a counterpart upstream of *Vλ2*. These regions share over 90% sequence homology as do *Vλ1* and *Vλ2* suggesting this region has arisen due to duplication. Interestingly a region of high chromatin accessibility is also present upstream of *Vλx*, referred to as HSVλx, which shares little sequence similarity with HS*Vλ1* but appears to bind CTCF. CTCF is bound at the equivalent of HS7 (HS7-1) and is also bound at the hypersensitive sites upstream of *Vλ2* and *Vλx*. Together with the pro-B Hi-C data analysis, these data imply that two possible loops could form in the 5' half of the locus: one between HS7-1 and HS*Vλ2* and another between HS7-1 and HSVλX thereby separating the locus into two domains. I propose that these CTCF mediated interactions between HS7/HS7-1 and HSVλ1/HS*Vλ2* or HSVλX are responsible the condensation of the *Igλ* locus and for bringing HS6 or its duplicate HS6-1 into close proximity with *Vλ1* or its counterpart *Vλ2*/VλX.

The long-range interactions formed in the 5' half of the locus present an interesting conundrum as both interactions, HS7-1 to HSVλ2 or HS7-1 to HSVλx, are unlikely to occur within the same cell. The recombination frequencies of *Vλ2* and *Vλx* are approximately equal (Sanchez et al., 1991), implying that both loops are formed with an approximately equal frequency. Therefore, there are two interesting possibilities, the 5' half of the locus could alternate between a domain formed between HS7-1 and HSVλ2 or HS7-1 and HSVλx. On the other hand, it is possible that one interaction is formed and maintained in each cell and the two potential interactions are observed the population of cells. Separating these possibilities is a potentially difficult task, as 'C' based techniques rely on cell lysis and rare ligation events it would be impossible to ascertain if in the same cell different domains are formed at different timepoints. One alternative is the use of a technique such as single cell DamID (Kind et al., 2013), which involves targeting a Dam methylase to a region of interest, e.g. HS7-1, potentially performed by a Cas9/Dam methylase fusion protein. Regions of the genome interacting with the targeted region have their adenosine methylated by the Dam methylase and these regions can be identified by digestion with DpnI and next-generation sequencing. As this technique allows the identification of multiple interacting

regions, alternating domains would be identified by methylation of both regions and stabilised domains would have a single methylated region.

As HS6 and *Vλ1* are close in linear space to their respective CTCF binding sites, HS7 and HS*Vλ1* respectively, the interaction of these two regions would relocate both HS6 and *Vλ1* into very close proximity (Figure 4.16). This contact, therefore, would facilitate the interaction of HS6 with the *Vλ1* promoter to enable the upregulation of *Vλ1* non-coding transcription and moreover could potentially reduce the three-dimensional distance of Eλ3-1 and *Jλ1*, possibly increasing the likelihood of interaction. Analysis of the interactions formed by Eλ3-1 by 3C using pre-B cells strongly indicated that Eλ3-1 does indeed contact the promoter of *Jλ1.* Remarkably, both enhancers appear to interact, as indicated by both pro-B Hi-C data and 3C using pre-B cells. This interaction appears to be mediated by HSCλ1 as both enhancers display a strong interaction with this element.

The interaction between HS6 and HSCλ1 appears to be mediated by YY1 as this factor binds to both regions. However, the mechanism by which Eλ3-1 interacts with HSCλ1 is unclear as YY1 binding cannot be detected. The Mediator complex has previously been implicated in regulating long-range enhancer-promoter interactions via cohesin (Kagey et al., 2010), therefore one possible explanation for the contact between Eλ3-1 and HSCλ1 is an interaction between the Mediator complex, which can be observed at Eλ3-1, and cohesin present at HSCλ1.

4.16 How might the coordinate activation of *Vλ1* and *Jλ1* be regulated?

Together my data suggest a model in which *Vλ1* and *Jλ1* are activated by the enhancers HS6 and Eλ3-1. The ability of these enhancers to upregulate the non-coding transcription of the gene segments is reliant on long-range interactions mediated by CTCF and YY1. As CTCF and RAD21 binding to HSVλ1 and HS7 can be observed in ChIP-seq data from pro-B cells, I propose that the first interaction to occur is that between HSVλ1 and HS7. This interaction relocates HS6 and *Vλ1* into close proximity but fails to upregulate non-coding transcription due to the insufficient level of IRF4.

IRF4 expression increases approximately three-fold from the pro-B to pre-B cell stage (Muljo and Schlissel, 2003), due to pre-BCR signalling, correlating with a 10 fold increase in *Igλ* recombination. Analysis of transcription factor binding in pro-B and pre-B cells revealed that IRF4 was the only factor to display a change in binding from pro-B to pre-B cells. IRF4 binding increased approximately three-fold at both HS6 and Eλ$_{3\text{-}1}$. As equipping pro-B cells with the pre-B level of IRF4 is sufficient to upregulate non-coding transcription and trigger premature *Igλ* recombination these data together suggest that the increased level of IRF4 binding to the enhancers Eλ$_{3\text{-}1}$ and HS6 is sufficient to activate *Vλ1* and *Jλ1* non-coding transcription.

The temporal events preceding gene segment activation were difficult to determine using pro-B and pre-B cell data. The lack of high resolution temporal data made it impossible to determine if the interaction between HSCλ1 and HS6/Eλ$_{3\text{-}1}$ occurs prior to *Vλ1* and *Jλ1* activation or as a result of gene segment synapsis. The data presented in this chapter, however, suggest that the parallel activation of sister enhancers would enable HS6 to upregulate *Vλ1* non-coding transcription and Eλ$_{3\text{-}1}$ to upregulate *Jλ1* non-coding transcription. It is likely that the upregulation of non-coding transcription is facilitated by the interaction of HSCλ1 with HS6 and Eλ$_{3\text{-}1}$. The parallel activation of both sister enhancers by IRF4 would result in the observed coordinate increase in *Vλ1* and *Jλ1* non-coding transcription.

In summary, the data presented here provides the initial evidence for a model of how the 3' half of the *Igλ* is regulated to facilitate the activation of *Vλ1* and *Jλ1* gene segments. These data support a model in which gene segment activation is facilitated by the formation of chromatin loops and the parallel activation of sister enhancers by IRF4.

Chapter 5 – Temporal analysis of the coordinate activation of *Vλ1* and *Jλ1*

A. Introduction

The temporal dynamics of immunoglobulin locus activation are poorly understood, primarily due to the absence of a temporal system to investigate the changes that occur. Locus contraction of the *Igh* and *Igκ* loci have been extensively studied by mutagenesis, FISH and chromosome conformation capture based techniques e.g. 4C, 5C and Hi-C. For the *Igκ* locus, these studies have proposed a step-wise model for *Igκ* locus activation (Stadhouders et al., 2014).

The current model for *Igκ* locus activation proposes that in pro-B cells the *Igκ* locus is in a contracted state, mediated by the E2A bound to iEκ and CTCF binding to the Sis and Cer elements forming an interaction between these elements and sites bound by E2A and CTCF in the *Vκ* region. Following differentiation to the pre-B cell stage and the subsequent activation of pre-BCR signalling, the 3'Eκ enhancer is activated resulting in a focused contact with the *Vκ* region (Stadhouders et al., 2014).

A substantial problem with these studies is the inability to explore locus contraction and activation in finer detail. Whilst analysis of pro-B and pre-B cells enables the identification of interactions and allows predictions regarding the temporal order of events, these studies cannot truly determine the temporal order of locus contraction in any detail. A significant problem for the temporal analysis of immunoglobulin locus contraction is the lack of a homogenous population of cells in which locus activation can be induced. In Chapter 3, I described the generation and characterisation of an IRF4-ER_{T2} expressing cell line. Using this cell line, I have demonstrated that *Vλ1* and *Jλ1* are consistently and coordinately upregulated at four hours post induction. This cell line enables, for the first time, the analysis of the transcription factors and alterations in three-dimensional chromatin structure regulating the coordinate increase in non-coding transcription.

In Chapter 4, I described the characterisation of an additional enhancer, HS6, in the murine *Igλ* locus. I have shown that this enhancer binds the same transcription factors as the well characterised 3' *Igλ* enhancer Eλ$_{3\text{-}1.}$ Interestingly, analysis of pre-B cells by 3C indicated that this enhancer exhibits a stronger interaction with *Vλ1* whereas the enhancer Eλ$_{3\text{-}1}$ exhibits a stronger interaction with *Jλ1* (Appendix 1; Figure 4.20). HS6 has been shown to be an enhancer of *Vλ1* non-coding transcription *in vitro* and contacts *Vλ1 in vivo* (Figures 4.8, 4.10 and 4.11). The initial studies that investigated the regulation of the *Igλ* locus identified four hypersensitive sites 35 kb downstream of Cλ1 (Eccles et al., 1990), HS1-HS4. The enhancer, Eλ$_{3\text{-}1,}$ is the most 5' hypersensitive site termed HS1. The removal of each of these elements from a BAC construct, followed by the generation of transgenic mice with this construct showed that Eλ$_{3\text{-}1}$ (HS1) was essential for recombination (Haque et al., 2013). Interestingly, however, the BAC construct did not contain the enhancer HS6. On the other hand, the BAC construct exhibited a very low frequency of recombination when integrated into the mouse genome (Haque et al., 2013). It was therefore unclear if the recently characterised enhancer, HS6, is essential for *Vλ1* activation.

In addition, the analysis of pre-B cells indicated the presence of long-range interactions with potentially vital roles in the coordinate regulation of *Vλ1* and *Jλ1* non-coding transcription, however, it was not possible to determine in which order these interactions occur. Notably, a substantial interaction was observed between both enhancers (HS6 and Eλ$_{3\text{-}1}$) and an uncharacterised element referred to as HSCλ1, potentially mediating the formation of a transcriptional hub involving both enhancers. This hub could then possibly control the activation of *Vλ1* and *Jλ1* or alternatively it could be involved in regulating the synapsis between these gene segments for recombination.

In this chapter, I further explore the model proposed for the coordinate activation of *Vλ1* and *Jλ1* described in Chapter 4 in addition to investigating the essential role of IRF4 in *Igλ* locus activation and contraction.

B. Results

5.1 Deletion of the HS6 and Eλ$_{3\text{-}1}$ enhancers

To determine if HS6 and Eλ$_{3\text{-}1}$ are absolutely required for *Igλ* locus activation, I sought to remove these regions from the *Igλ* locus using CRISPR/Cas9. The first strategy utilised to delete the enhancers was use of two sgRNAs upstream and downstream of each enhancer (Figure 5.1A). In order to generate a deletion of the entire enhancer, Cas9 must generate DSBs at both target sites simultaneously, or within a very short period of time. Due to the very low pro-B cell transfection efficiency, lentiviral transduction using the transfer vector lenti-CRISPRv2 was used to introduce the Cas9 and sgRNAs into 1D1-T215 cells. As the lenti-CRISPRv2 vector was only able to accommodate a single guide sequence, this vector was modified to accept an additional guide and sgRNA scaffold (Section 2.39). Lentivirus was generated from these constructs and used to transduce 1D1-T215 cells. Transduced cells were selected using puromycin and monoclonal lines were generated using semi-solid agar, as the cells failed to grow when diluted to single cells in 96 well plates. Clones were screened by PCR for the presence of the required deletion (Figure 5.2) and six clones in which the enhancer was deleted on one allele (as assessed by PCR) were identified. Whilst heterozygote clones exhibiting the required deletion could be detected, homozygote clones could not be identified, suggesting that the efficiency of enhancer deletion was low. This could be due to low expression of the sgRNAs/ Cas9 in pro-B cells, due to the poor transduction efficiency.

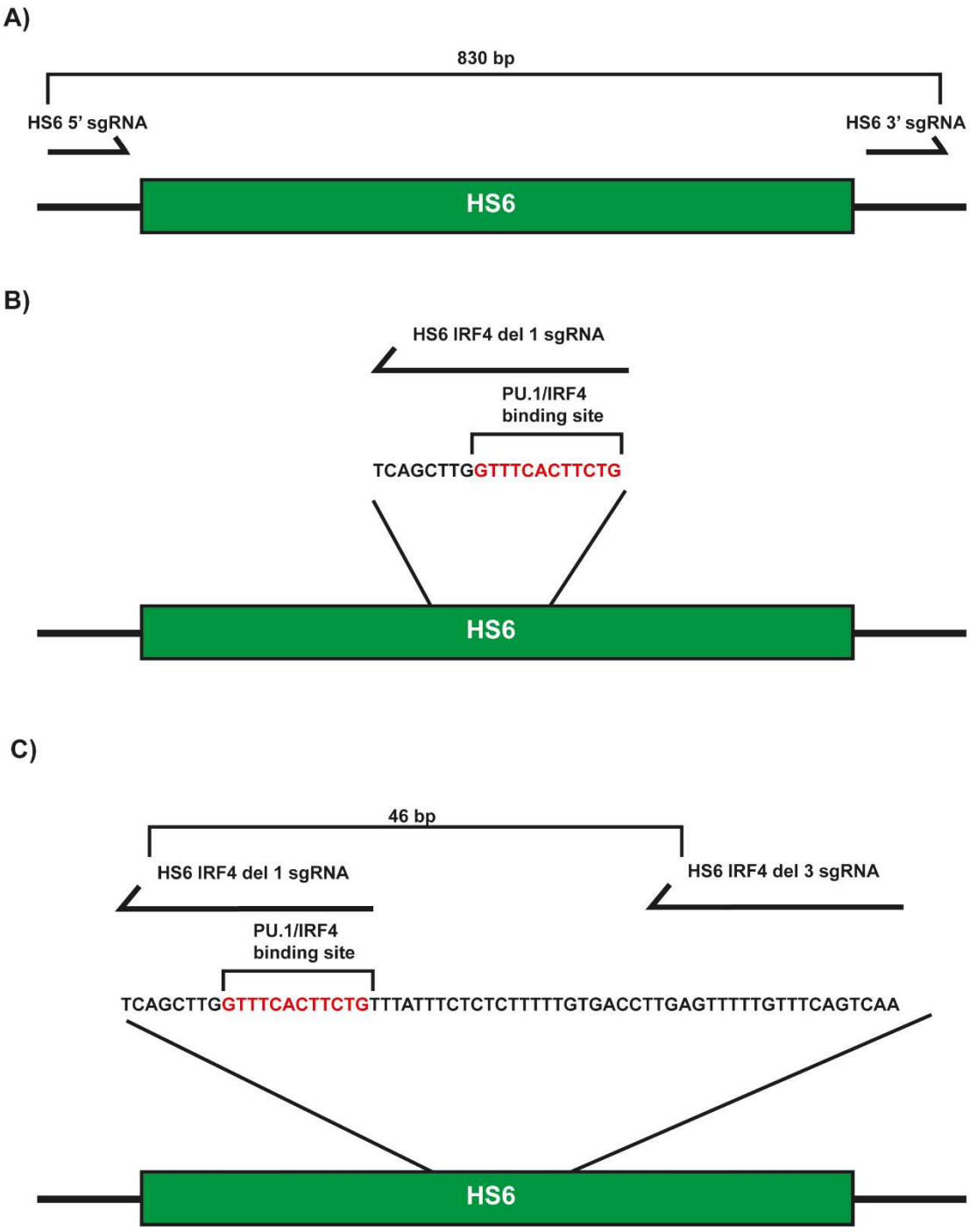


Figure 5.1: Strategies used to mutate the enhancer HS6. A) Large scale deletion mediated by the co-expression of sgRNA 5' and 3' of the enhancer. B) Mutation of the PU.1/IRF4 binding site using a single sgRNA overlapping the binding site. C) Small scale deletion of the PU.1/IRF4 binding site by the co-expression of two sgRNAs flanking the binding site.

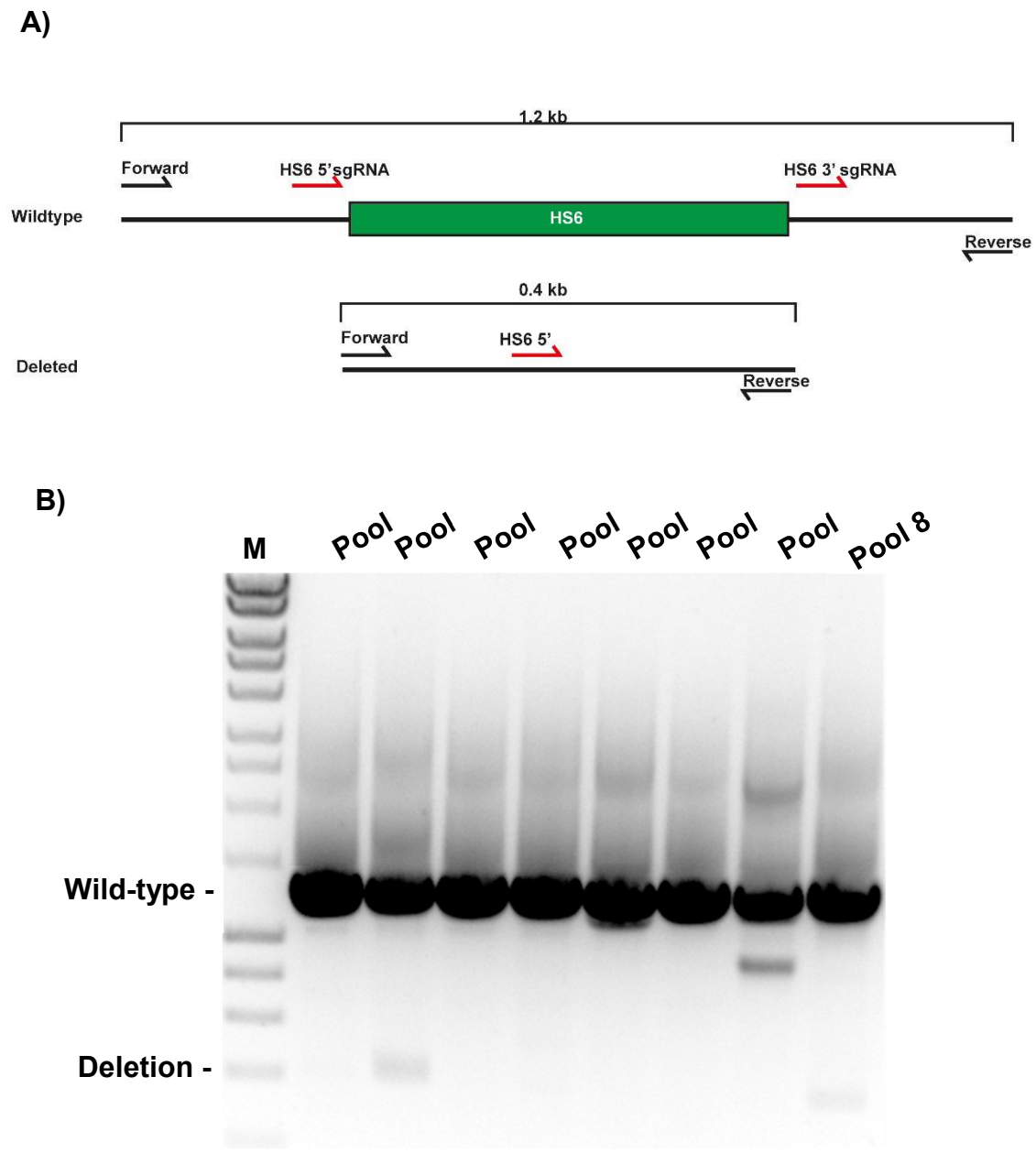


Figure 5.2: Identification of cell line pools containing the required deletion of HS6. A) PCR strategy to identify deleted alleles. Primers are indicated by black half-arrows and sgRNA are indicated by red half-arrows. B) PCR analysis of monoclonal cell lines from a 24 well plate with each row combined together to generate a single pooled sample (Pool 1-8) to identify deletion events. The location of bands corresponding to the wildtype (1.2 kb) and the expected size for deletion mutants (0.4 kb) is indicated. M – 1 kb Hyperladder IV (Bioline)

The use of a single guide can be sufficient to inactivate an enhancer (Kim and Kim, 2017). As IRF4 binding appeared be heavily involved in the activation of Eλ3-1 and HS6, I sought to increase the efficiency of enhancer mutation by targeting a sgRNA to the PU.1/IRF4 binding sites within each enhancer. Guides designed to target the PU1/IRF4 binding sites of each enhancer (Figure 5.1) were cloned into lenti-CRISPRv2 and monoclonal cell lines were generated as described in Section 5.1. The cell lines were screened initially by PCR amplification of the targeted region, followed by Sanger sequencing to identify the presence of a mutation by alterations in the sequencing trace. In the cases where mutations were observed, HS6 and Eλ3-1 were then amplified from their respective mutated cell lines and cloned for sequence analysis. Whilst the mutation of HS6/Eλ3-1 was observed in all clones examined (n = 48), only small deletions 3-12 bp were detected in the majority of cases. Figure 5.3 is an example of a typical mutated cell line. Numerous heterozygote PU.1/IRF4 binding site mutations were identified but no homozygous PU.1/IRF4 binding site mutant cell lines were obtained.

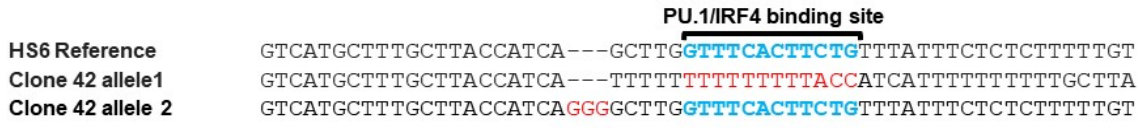


Figure 5.3: Example of a heterozygous HS6 PU.1/IRF4 binding site mutant. Sequence analysis of the HS6 enhancer region in a monoclonal cell line (# 42) in which the HS6_IRF4_del_1 sgRNA and Cas9 had been expressed. One allele has undergone mutation (red) at the PU.1/IRF4 binding site (blue) whilst the other allele contains an insertion mutation (red) away from the binding site.

5.3 Mutation of the PU.1/IRF4 binding sites at HS6 and Eλ$_{3-1}$ via small-scale deletions

As the previous approaches failed to mutate HS6 or Eλ$_{3-1}$ sufficiently to allow analysis of the effects of enhancer deletion, I next designed sgRNAs to make small deletions (50-90 bp) to remove the PU.1/IRF4 binding site at both enhancers. Small deletions have been shown to be more efficient than larger deletions (Shin et al., 2017). Furthermore, I improved the deletion efficiency by the sequential transduction of the guide sequences and subsequent selection of infected cells. For each enhancer, one guide was cloned into lenti-sgRNA-MS2-Zeo, which carries a zeocin resistance gene and the other into lenti-CRISPRv2, which carries a puromycin resistance gene. To ensure both deletion guides were expressed simultaneously, I first transduced 1D1-T215 cells with lentivirus generated from lenti-sgRNA-MS2-Zeo and selected for zeocin resistance. As the lenti-sgRNA-MS2-Zeo did not contain Cas9, this enabled cell lines to be established that expressed one sgRNA but could not cleave the target site. The deletion of the PU.1/IRF4 binding site was achieved by transduction with a lentivirus that expressed Cas9 and the second sgRNA. Cells expressing Cas9 and both sgRNAs were selected by treatment with zeocin and puromycin, before the generation of monoclonal cell lines (n = 48) using semi-solid agar. Monoclonal cell lines were then pooled and screened for the appropriate deletion by PCR (Figure 5.4).

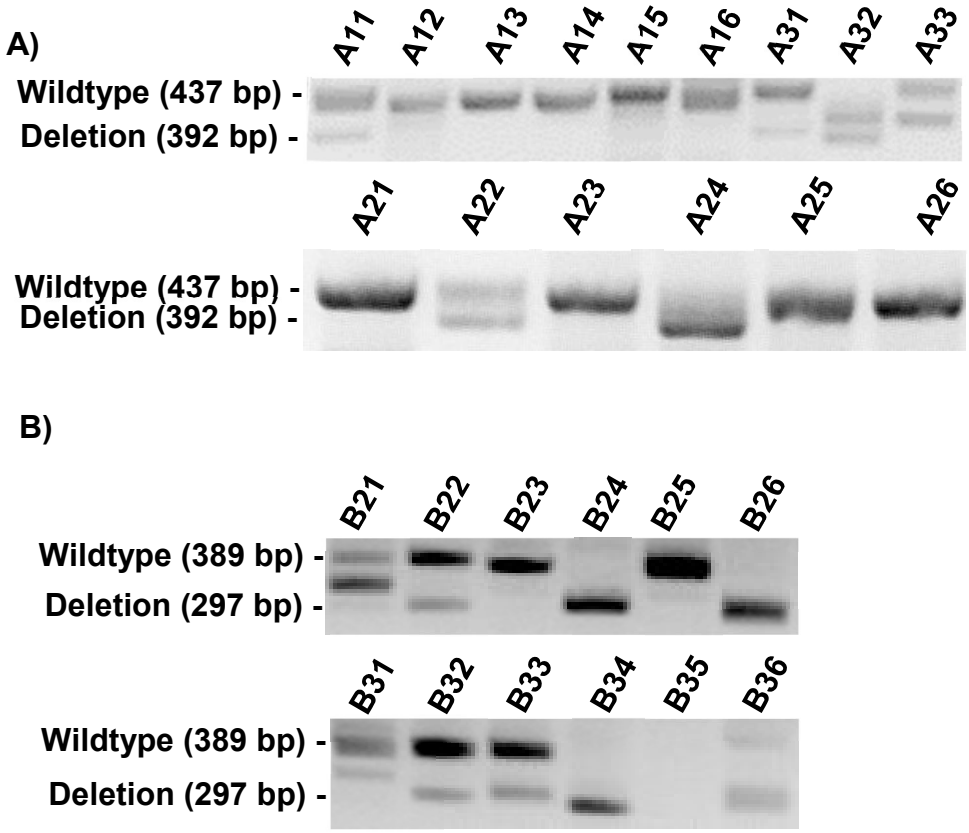


Figure 5.4: Analysis of monoclonal cell lines to identify PU.1/IRF4 binding site deletions at HS6 or Eλ3-1. PCR analysis of the status of HS6 (A) or Eλ3-1 (B) to identify clones with homozygous deletions. Bands corresponding to the wildtype and deletion mutant are indicated.

Screening of monoclonal cell lines (Figure 5.4) identified two cell lines with deletions at HS6 (HS6-A24 and HS6-A32) and six lines with deletions at Eλ3-1, however, only two Eλ3-1 mutant cell lines (Eλ3-1-B24 and Eλ3-1-B26) were subjected to further analysis. Sequence analysis of the enhancer regions confirmed the presence of a homozygous deletion of the expected size for three of the four clones (A24, B24 and B26; Figure 5.5A and B and Figure 5.5D) but interestingly the HS6 mutant A32 is comprised of a perfect PU.1/IRF4 site deletion and a deletion upstream of the PU.1/IRF4 binding site (Figure 5.5B), which removes the E2A binding site. As E2A is essential for *Igλ* locus activation (Beck et al., 2009) this clone was still considered in downstream analysis as an HS6 knockout.

A) HS6 PU.1/IRF4 site deletion clone A24

```
WT          246 CTTCTCGAGGGCCACACCTGGTTCTGTTTGTCATGCTTTGCTTACCATCA
HS6_A24_a1   79 CTTCTCGAGCGCCACACCTGGTTCTGTTTGTCATGCTTTGCTTACCATCA
HS6_A24_a2   73 CTTCTCGAGGGCCACACCTGGTTCTGTTTGTCATGCTTTGCTTACCATCA

WT          297 GCTTGGTTTCACTTCTGTTTATTTCTCTCTTTTTGTGACCTTGAGTTTTT
HS6_A24_a1  129 -------------------------------------------CTCTTTT
HS6_A24_a2  123 -------------------------------------------GAGTTTT
```

B) HS6 PU.1/IRF4 site deletion clone A32

```
WT          195 ATAGTCCTACCTATGAGAGACACATTCAACTACAATCAAAATGGCCTTTT
HS6_A32_a1  130 ATAGTCCTACCTATGAGAGACACATTCAA--------------------
HS6_A32_a2  116 --------------GAGAGACACATTCAACTACAATCAAAATGGCCTTTT

WT          246 CCTTCTCGAGGGCCACACCTGGTTCTGTTTGTCATGCTTTGCTTACCATC
HS6_A32_a1  131 -----------------------------------------------------
HS6_A32_a2  152 CCTTCTCGAGGGCCACACCTGGTTCTGTTTGTCATGCTTTGCTTACC---

WT          266 AGCTTGGTTTCACTTCTGTTTATTTCTCTCTTTTTGTGACCTTGAGTTTT
HS6_A32_a1  131 --CTTGGTTTCACTTCTGTTTATTTCTCTCTTTTTGTGACCTTGAGTTTT
HS6_A32_a2  200 -------------------------------------------GAGTTTT
```

C) Eλ3-1 PU.1/IRF4 site deletion clone B24

```
WT           264 TTACTGAGCTCGCTATGACTGCAGCTGCCACCCTTGAGAGTCCACAAGCT
Eλ3-1_B24_a1  25 TTACTGAGCTCGCTATGACTGCAGCTGCCACCCTTGA-------------
Eλ3-1_B24_a2 123 TTACTGAGCTCGCTATGACTGCAGCTGCCACCCTTGA-------------

WT           315 AAAATTAAATCTGTGATAACTGAAACAAAAACTCATGGTCACAGAAAAAG
Eλ3-1_B24_a1  63 -----------------------------------------------------
Eλ3-1_B24_a2 161 -----------------------------------------------------

WT           366 AGAAATAATAGGAACTGAAACCAAGTCCATTAGCAGCAAGGCATGGCAAG
Eλ3-1_B24_a1  63 -------------------------TCCATTAGCAGCAAGGCATGGCAAG
Eλ3-1_B24_a2 161 -------------------------TCCATTAGCAGCAAGGCATGGCAAG
```

D) Eλ3-1 PU.1/IRF4 site deletion clone B26

```
WT           264 TTACTGAGCTCGCTATGACTGCAGCTGCCACCCTTGAGAGTCCACAAGCT
Eλ3-1_B26_a1  37 TTACTGAGCTCGCTATGACTGCAGCTGCCACCCTTGA-------------
Eλ3-1_B26_a2  41 TTACTGAGCTCGCTATGACTGCAGCTGCCACCCTTGA-------------

WT           315 AAAATTAAATCTGTGATAACTGAAACAAAAACTCATGGTCACAGAAAAAG
Eλ3-1_B26_a1  75 -----------------------------------------------------
Eλ3-1_B26_a2  79 -----------------------------------------------------

WT           366 AGAAATAATAGGAACTGAAACCAAGTCCATTAGCAGCAAGGCATGGCAAG
Eλ3-1_B26_a1  75 -------------------------TCCATTAGCAGCAAGGCATGGCAAG
Eλ3-1_B26_a2  79 -------------------------TCCATTAGCAGCAAGGCATGGCAAG
```

Figure 5.5: Sequence alignments of clones bearing deletions of the PU.1/IRF4 binding site at either HS6 or Eλ3-1. Analysis of both alleles of the HS6 (A and B) or Eλ3-1 (C and D) mutant cell lines. With the exception of HS6 A32 the clones are homozygous for a deletion which remove the PU.1/IRF4 binding site (red) at the appropriate enhancer. WT – wildtype, a1/a2 – allele 1/ allele 2

5.4 HS6 and Eλ3-1 are both required for the activation of *Vλ1* and *Jλ1* non-coding transcription

Following the generation of cell lines that had homozygous mutations of the PU.1/IRF4 binding sites in either HS6 or Eλ3-1, I examined effect of the deletion on the regulation of *Vλ1* and *Jλ1* non-coding transcription. Mutant cell lines were induced for 16 hours, to increase *Vλ1* and *Jλ1* non-coding transcription to a maximal level and compared to a wildtype control (Figure 5.6). The IRF4-ER$_{T2}$ transgene expression has been shown to be stable for approximately five months in culture (Figure 3.14A), however, analysis of GFP expression indicated a slight decrease in IRF4-ER$_{T2}$ expression in two of the mutant cell lines. The potentially confounding effects of the reduced level of IRF4-ER$_{T2}$ were addressed by the normalisation of *Vλ1* and *Jλ1* non-coding transcript levels to the expression of *Cxcr4*, a gene known to be directly upregulated by IRF4 (Johnson et al., 2008).

Deletion of the PU.1/IRF4 binding sites, or the E2A binding site in the case of the HS6 del A32 cell line resulted in a substantial decrease in *Vλ1* and *Jλ1* non-coding transcription (Figure 5.6). The loss of PU.1/IRF4 or E2A binding at HS6 appeared to have the greatest impact on *Vλ1* non-coding transcription, but *Jλ1* non-coding transcription was also markedly decreased. In addition, the mutation of the PU.1/IRF4 binding site at Eλ3-1 dramatically reduced *Vλ1* and *Jλ1* non-coding transcripts to a barely detectable level. Interestingly, these data suggest that there is no redundancy between the two enhancers as the loss of either enhancer is sufficient to prevent the induction of *Vλ1* and *Jλ1* transcription. Moreover, these data suggest that HS6 does not act solely on *Vλ1* and Eλ3-1 does not only enhance the expression of *Jλ1* as the loss of either enhancer dramatically reduces the transcription of both gene segments. This implies that these enhancers cooperate to regulate the 3' half of the locus.

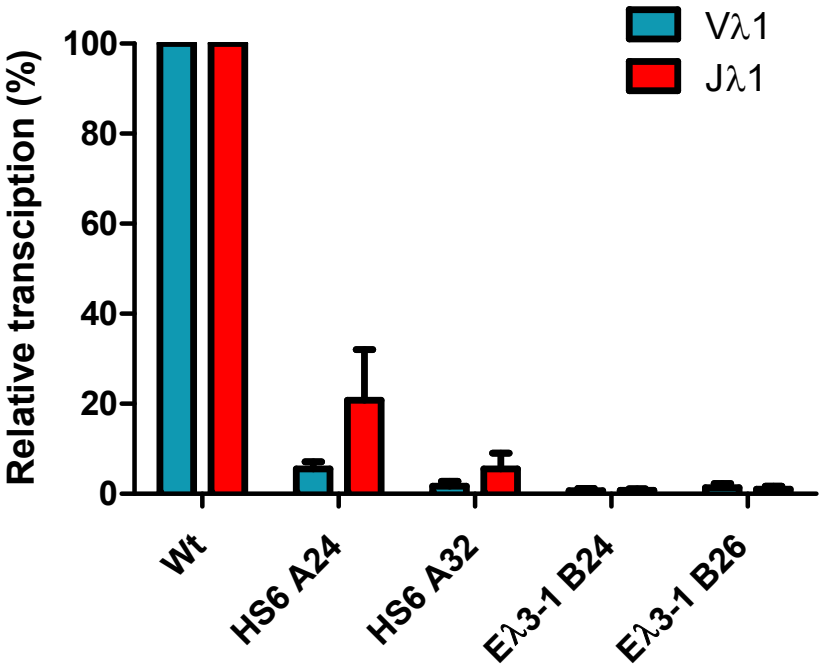


Figure 5.6: *Vλ1* and *Jλ1* non-coding transcription is substantially decreased in the HS6 and Eλ$_{3\text{-}1}$ PU.1/IRF4 binding site mutants. Analysis of *Vλ1* and *Jλ1* non-coding transcription after 16 hours of induction with 2 μM 4-OHT in unmodified 1D1-T215 cells and in the cell lines with the mutated PU.1/IRF4 binding sites. Non-coding transcription was normalised to *Hprt* expression and by *Cxcr4* expression to normalise for any alteration in IRF4-ER$_{T2}$ transgene expression during culture. Error bars represent SEM.

5.5 Eλ$_{3\text{-}1}$ and HS6 are simultaneously bound by IRF4

Analysis of the transcription factors binding to Eλ$_{3\text{-}1}$ and HS6 in pro-B and pre-B cells indicated that IRF4 is the only factor to show substantially increased binding at the pre-B cell stage. The removal of the IRF4 binding sites at either Eλ$_{3\text{-}1}$ or HS6 resulted in a dramatic decrease in *Vλ1* and *Jλ1* non-coding transcription. This strongly implies that increased levels of IRF4 are essential for the activation of both enhancers.

In the previous chapter, I proposed a model in which the simultaneous binding of IRF4 to both enhancers results in their synchronized activation. The IRF4-ER$_{T2}$ expressing cell line, 1D1-T215, that I have generated provides a unique opportunity to explore the temporal events mediated by IRF4 which precede locus activation. Furthermore, this cell line enables temporal analysis at a resolution that was previously unobtainable due to the low number of pro-B cells that can be obtained from transgenic mice. To firstly investigate if IRF4 is simultaneously binds to both Eλ$_{3\text{-}1}$ and HS6, I performed a temporal analysis of IRF4 binding after induction by 4-OHT in the cell line 1D1-T215 by ChIP-qPCR (Figure 5.7).

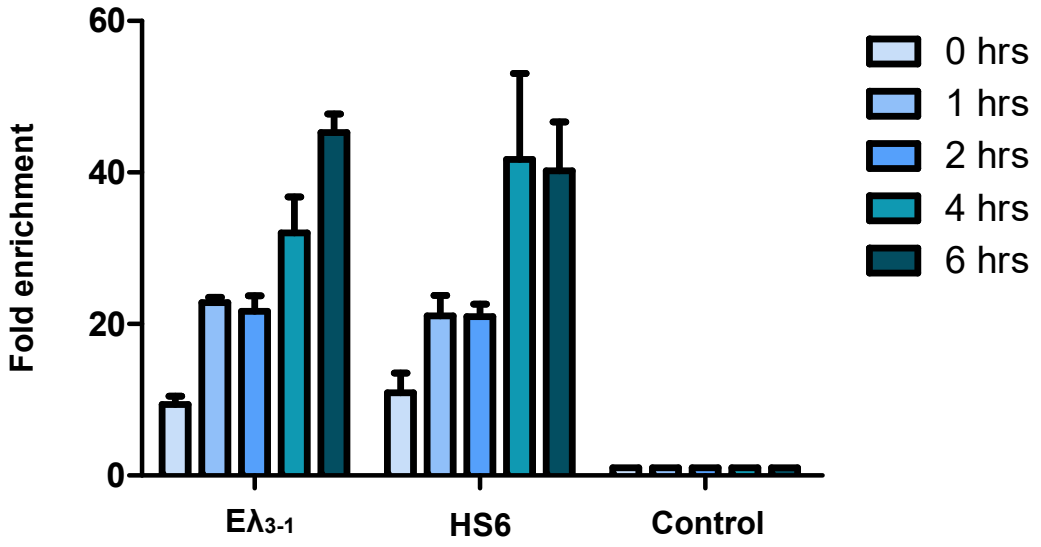


Figure 5.7: IRF4 binding to the enhancers Eλ3-1 and HS6 is simultaneously increased upon induction. Analysis of IRF4 binding to HS6 and Eλ3-1 by ChIP-qPCR at one, two, four and six hours post-induction. A very similar increase in IRF4 binding can be observed at both enhancers. Error bars represent SEM (n=3).

Whilst IRF4-ER$_{T2}$ is elevated to a maximal level in the nucleus by one-hour post-induction (Figure 3.13B), binding to the enhancers appears to increase sequentially. An increased level of IRF4 binding was detected after one hour of induction, this level remains consistent between one and two hours before increasing further at four to six hours. The reason for this delay in IRF4 binding is unclear but it is possible that additional events such as increased histone acetylation are required before IRF4 can bind to its maximal level. Importantly, the temporal analysis of IRF4 binding by ChIP showed a simultaneous increase in IRF4 binding to both HS6 and Eλ$_{3-1}$ enhancer elements, implying that these sister enhancers are activated in parallel by IRF4 binding as suggested by the analysis of pro-B and pre-B cells.

5.6 CTCF/cohesin mediated interactions are present at the pro-B cell stage

These data suggested a model by which *Vλ1* and *Jλ1* coding transcription is coordinately regulated by the simultaneous binding of IRF4 to the Eλ$_{3\text{-}1}$ and HS6 enhancers. The upregulation of non-coding transcription by these enhancers is likely to be aided by interactions between HS*Vλ1* and HS7 which is mediated by CTCF/Cohesin. Furthermore, analysis of interactions in pre-B cells by 3C and in pro-B cells by Hi-C strongly implied that there are interactions between the two sister enhancers and between both enhancers and HSCλ1. This suggests that HSCλ1 mediates the localisation of both enhancers and possibly the gene segments to form a transcriptional hub for *Vλ1* and *Jλ1* activation. However, these data failed to identify the true temporal order of interactions preceding *Igλ* locus activation and I therefore sought to investigate which interactions are altered by IRF4 binding.

Analysis of published CTCF ChIP-seq data using *Rag1*$^{-/-}$ pro-B cells indicated that CTCF is already bound at HS*Vλ1* and HS7 in pro-B cells. I confirmed this via CTCF ChIP-qPCR using pro-B cells (Figure 5.8). Moreover, RAD21 binding was also detectable by analysis of ChIP-seq data from pro-B cells (Figure 4.12), further supporting the idea that the interaction between HS*Vλ1* and HS7 is formed prior to *Igλ* locus activation.

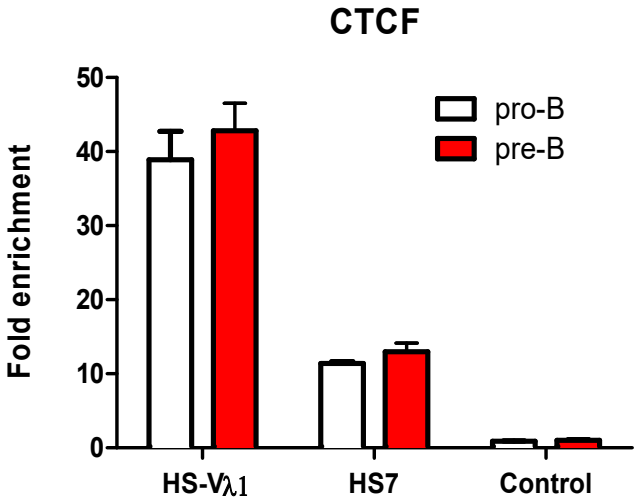


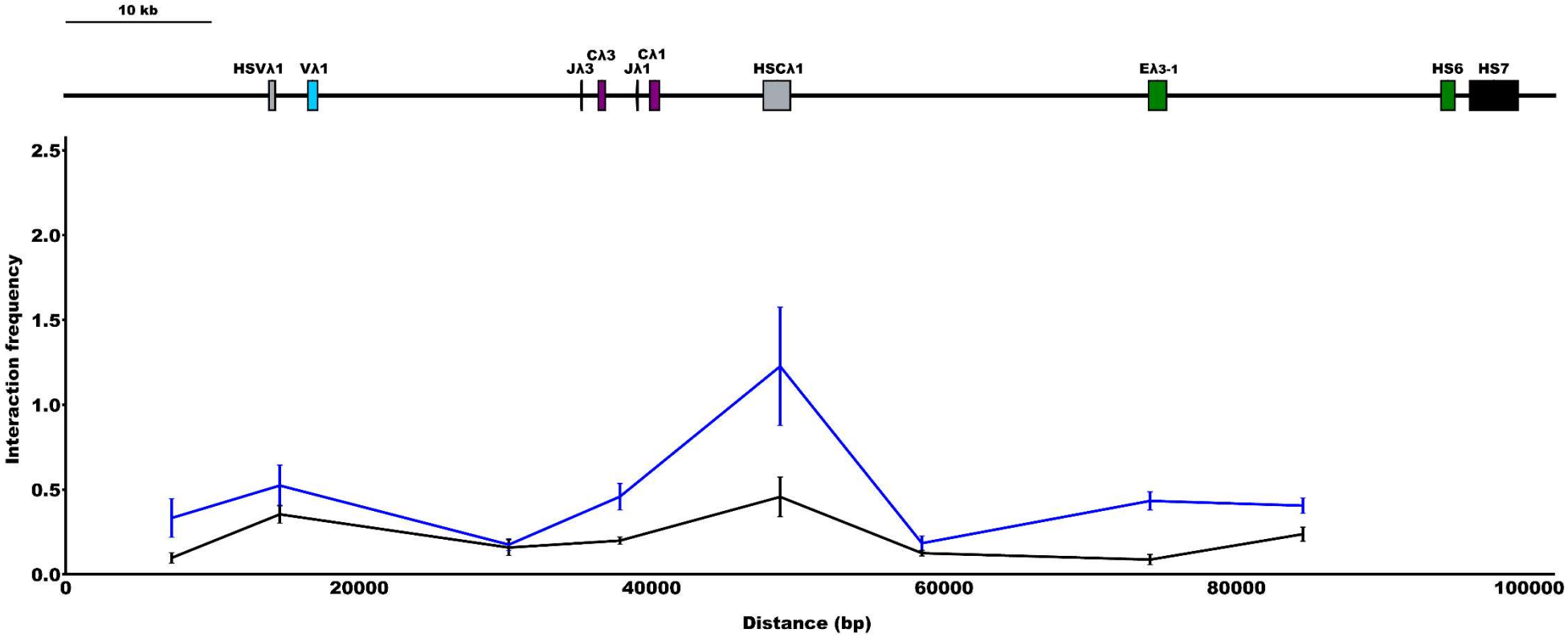
Figure 5.8: CTCF is bound at its maximal level at the pro-B cell stage. CTCF binding determined by ChIP-qPCR is unaltered from the pro-B to pre-B cell stage. Error bars represent SEM (n = 3).

In order to determine the long-range interactions which are likely to be involved in the regulation of *Vλ1* and *Jλ1* non-coding transcription, I performed 3C on pro-B cells and pre-B cells, from both the HS6 (Figure 5.9A) and Eλ3-1 (Figure 5.9B) viewpoints. The analysis of pro-B and pre-B cells by 3C revealed six interactions that exhibit a substantial increase in interaction frequency in pre-B cells (Table 5.1).

Table 5.1: Interactions formed during *Igλ* locus activation. Summary of interactions identified by 3C analysis of pro-B and pre-B cells indicating the viewpoint used, the element contained within the interacting fragment and the mean alteration in interaction frequency observed in pre-B cells compared to pro-B cells.

Viewpoint	Interacting region	Fold change interaction frequency
HS6	Eλ3-1	4.89
Eλ3-1	*Jλ1*	4.29
HS6	HSCλ1	2.69
HS6	*Jλ1*	2.30
Eλ3-1	HSCλ1	2.06
HS6	*Vλ1*	1.48

A)


10 kb
HSVλ1
Vλ1
Cλ3
Jλ3
Jλ1
Cλ1
HSCλ1
Eλ3-1
HS6
HS7
Interaction frequency
2.5
2.0
1.5
1.0
0.5
0.0
0
20000
40000
60000
80000
100000
Distance (bp)

B)

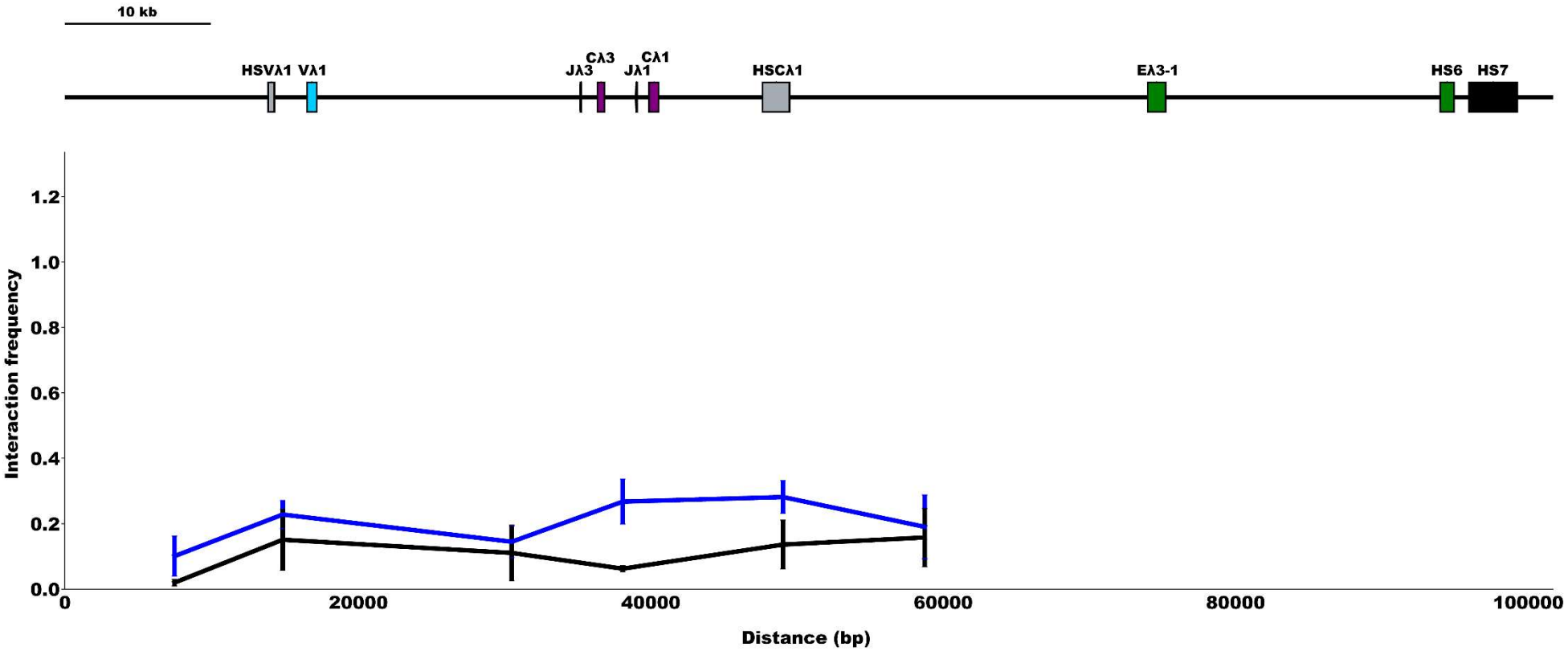


Figure 5.9: Analysis of the interactions formed by the HS6 and Eλ$_{3\text{-}1}$ enhancers in pro-B and pre-B cells. Relative interaction frequency of DpnII fragments from the HS6 (A) or Eλ$_{3\text{-}1}$ viewpoint (B) using pro-B cells (black) and pre-B cells (blue), assessed by qPCR. The data shown are comprised of two biological replicates from pro-B cells and three replicates from pre-B cells. Error bars represent SEM.

In the previous chapter I hypothesised that the two enhancers interact to form a hub centred on HSCλ1, but this hypothesis was based on the interaction observed between HS6 or Eλ3-1 and HSCλ1, in addition to interactions determined by Hi-C data analysis. The analysis of pro-B and pre-B cells by 3C provides the first direct evidence of interaction between these two enhancers. Moreover, the analysis of *Vλ1* and *Jλ1* non-coding transcription in the HS6 and Eλ3-1 mutant cell lines (Figure 5.6) suggests that enhancer cooperation is vital for the upregulation of *Vλ1* and *Jλ1* non-coding transcription. Therefore, the increase in interaction frequency between pro-B cells and pre-B cells supports the idea that enhancer interaction is a crucial feature of *Igλ* locus activation.

Notably, the 3C analysis also strongly indicated that both enhancers interact with both *Vλ1* and *Jλ1* gene segments, supporting the conclusion of the HS6/Eλ3-1 mutation analysis. As would be expected, due to the CTCF/cohesin mediated interaction between HS*Vλ1* and HS7, interaction of HS6 and *Vλ1* was clearly detectable at both the pro-B and pre-B cell stages (Figure 5.9). The interaction between HS6 and *Jλ1* was difficult to discern using 3C data from pre-B cells (Figure 4.11), but when compared to pro-B cells, a substantial increase in interaction frequency between these regions can be observed (Figure 5.9). Furthermore, the interaction between Eλ3-1 and *Jλ1* is markedly increased in pre-B cells, which agrees with 3C experiments performed by Sarah Bevington (Appendix 1), but crucially my data strongly suggest that Eλ3-1 also contacts *Vλ1*.

The 3C analysis of pre-B cells implied that HSCλ1 plays a major role in *Igλ* recombination, potentially by regulating the interaction between the enhancers HS6 and Eλ3-1. The substantial increase in interaction frequency between HS6/Eλ3-1 and HSCλ1 from pro-B cells to pre-B cells supports this idea. As YY1 binding was observed at HSCλ1 and HS6 (Figure 4.12), I reasoned that YY1 could be responsible for directing the HS6 interaction with HSCλ1. The presence of YY1 at HS6 and HSCλ1 was examined in pro-B and pre-B cells by ChIP-qPCR. Interestingly this analysis revealed that not only does YY1 bind to these regions, its binding also increases in pre-B cells (Figure 5.10), which is consistent with the idea that increased YY1 binding may be essential for the increased interaction frequency observed between HS6 and HSCλ1 in pre-B cells. Notably, only a slight enrichment of YY1 binding (~1.5 fold vs control) could be detected at Eλ3-1 by ChIP-qPCR and significant enrichment of the factor could not be determined using YY1 ChIP-seq data from pro-B cells. It is unlikely that YY1 binds to Eλ3-1 although, it is possible that the poor sensitivity of the ChIP-qPCR and ChIP-seq experiments have failed to detect the binding of YY1 to this enhancer.

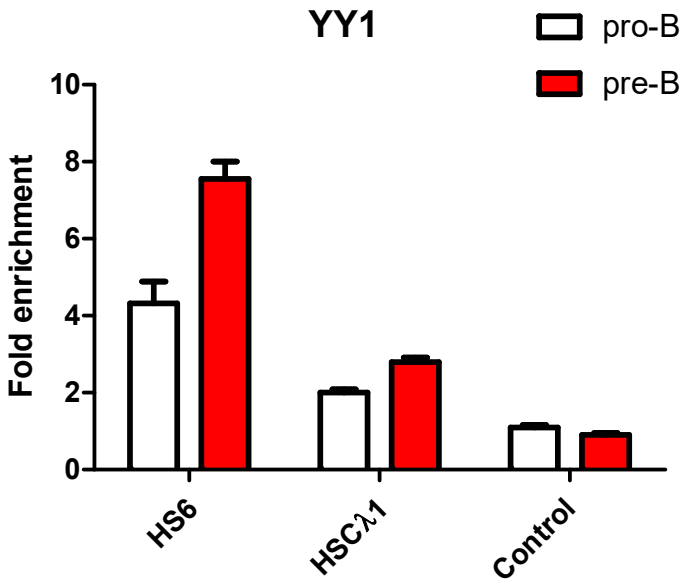
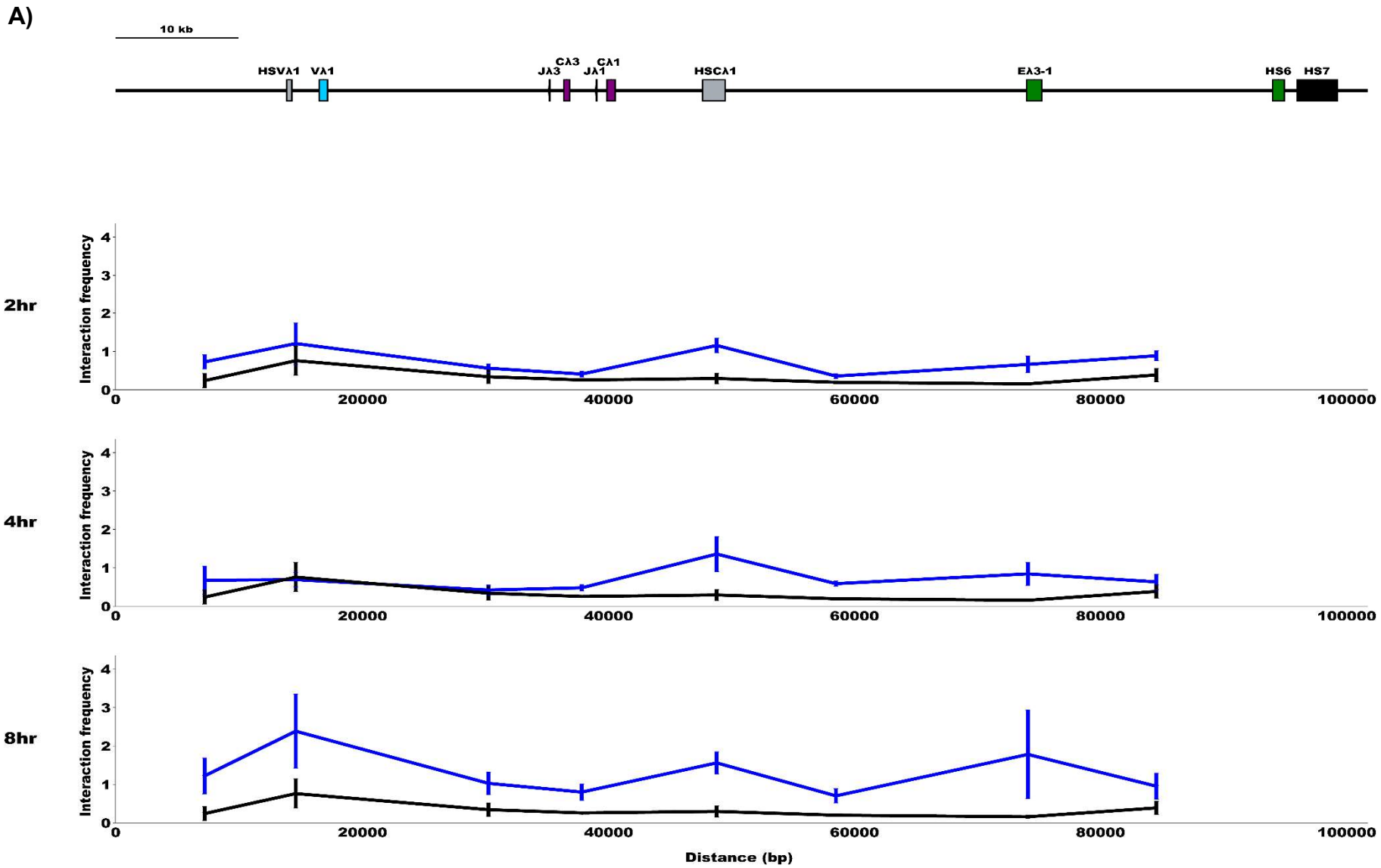


Figure 5.10: The level of YY1 bound at HS6 and HSCλ1 increases during *Igλ* activation. ChIP-qPCR analysis of YY1 binding at HS6 and HSCλ1 in pro-B and pre-B cells. A 1.75 and 1.39-fold increase in YY1 binding at HS6 and HSCλ1, respectively, is observed in pre-B cells. Error bars represent SEM (n = 3).

5.7 Temporal analysis of interactions formed in the *Igλ* locus

Analysis of pro-B and pre-B cells by 3C indicated that multiple interactions occur during locus activation (Table 5.1). Whilst the interaction frequency of these contacts increases during locus activation as observed in pre-B cells, the true temporal order of events was unclear from these data. *Vλ1* and *Jλ1* non-coding transcription in the PIP-ER pro-B cell line 1D1-T215 is upregulated four hours post-induction. Therefore, to determine in which order the interactions that precede *Vλ1* and *Jλ1* non-coding transcription occur, I performed 3C at two, four and eight hours post-induction in the 1D1-T215 cell line using both HS6 and $E\lambda_{3\text{-}1}$ as viewpoints (Figure 5.11).


A)
10 kb
HSVλ1
Vλ1
Cλ3
Jλ3
Jλ1
Cλ1
HSCλ1
Eλ3-1
HS6
HS7
2hr
4hr
8hr
Interaction frequency
0
1
2
3
4
0
20000
40000
60000
80000
100000
Distance (bp)

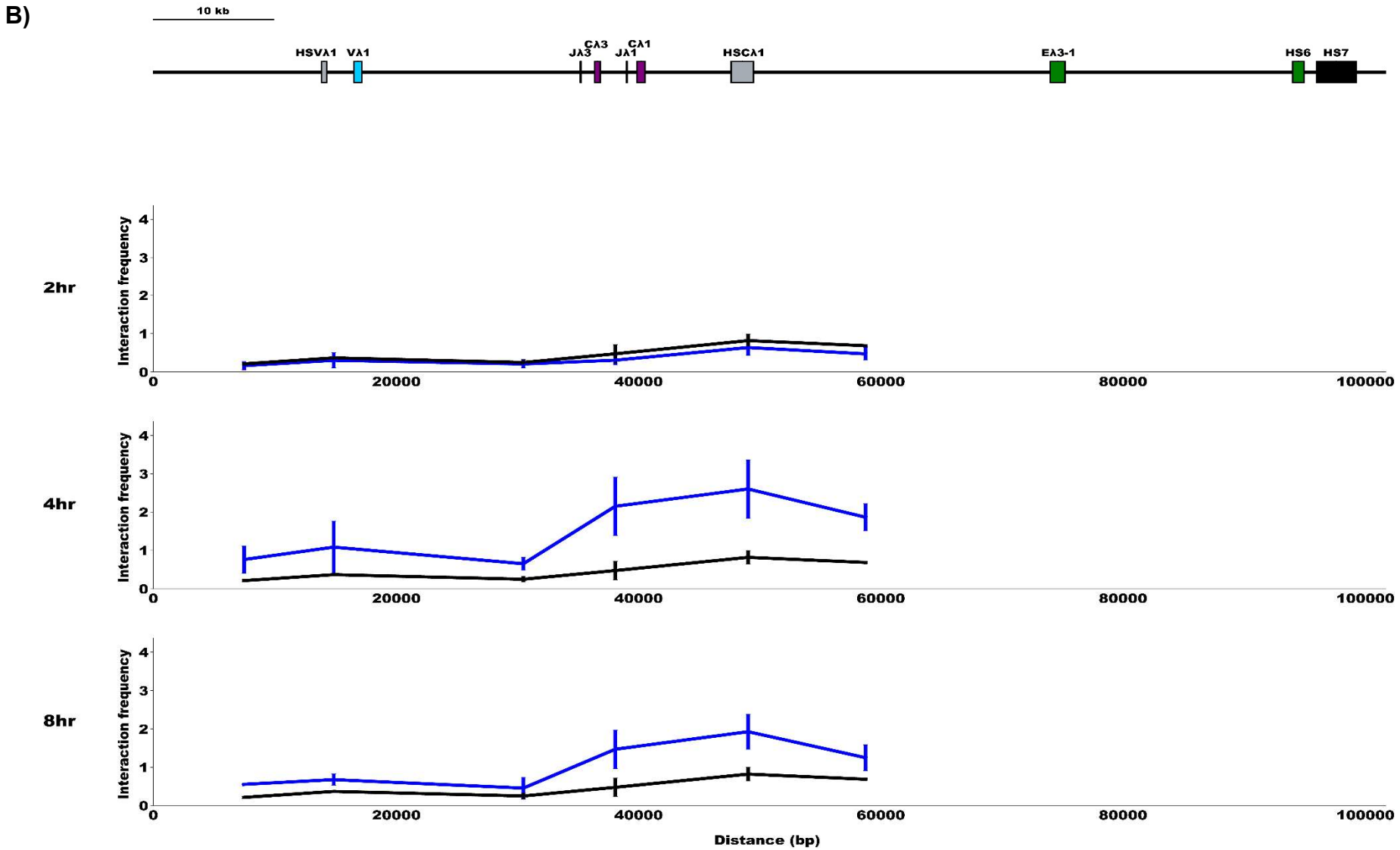


Figure 5.11: Temporal analysis of interactions formed by the 3' *Igλ* enhancers. Analysis of the interaction frequency of DpnII fragments from the HS6 (A) and Eλ$_{3\text{-}1}$(B) viewpoint regions at 2, 4 and 8 hours post-induction (blue) by 4-OHT in the 1D1-T215 cell line compared to an uninduced control (black). Error bars represent SEM and the data consist of three biological replicates of each timepoint.

Interestingly, the first observable alteration in interaction frequency is between HS6 and HSCλ1 at two hours post-induction. The interaction frequency between these two regions increases further at the subsequent timepoints. This suggests that the first interaction involving HS6 after IRF4 binding is to HSCλ1. Notably, increased interaction frequency between Eλ$_{3\text{-}1}$ and HSCλ1 was not observed until four to eight hours post-induction, suggesting a delay in Eλ$_{3\text{-}1}$-HSCλ1 interactions compared to HS6.

Remarkably, the interaction frequency between HS6 and Eλ$_{3\text{-}1}$ appears to correlate well with the degree of contact between these enhancers and HSCλ1. A substantial increase in interaction frequency between the two enhancers was observed at four hours post induction. As the interaction between the enhancer and HSCλ1 has the highest interaction frequency from both viewpoints, it is likely that the interaction with HSCλ1 mediates the interaction between the two enhancers.

Increased interaction frequency was also observed between HS6 and *Vλ1* between four to eight hours post-induction, correlating with the observed increase in *Vλ1* non-coding transcription. This suggests that whilst HS6 and *Vλ1* do interact prior to locus activation, likely due to the CTCF/Cohesin mediated interaction between HS7 and HSVλ1, the stability of this interaction is increased by IRF4 binding.

Analysis from the Eλ$_{3\text{-}1}$ viewpoint indicated that this enhancer exhibits only minimal contact with *Jλ1* before locus activation. This confirms my observations in pro-B and pre-B cells and implies that IRF4 binding alone is essential for initiating the Eλ$_{3\text{-}1}$-*Jλ1* interaction, which appears to occur between two to four hours post induction. Remarkably, the interaction of Eλ$_{3\text{-}1}$ with *Jλ1* also coincides with the increased interaction frequency between the enhancer and HSCλ1, which may suggest that HSCλ1 recruits *Jλ1*.

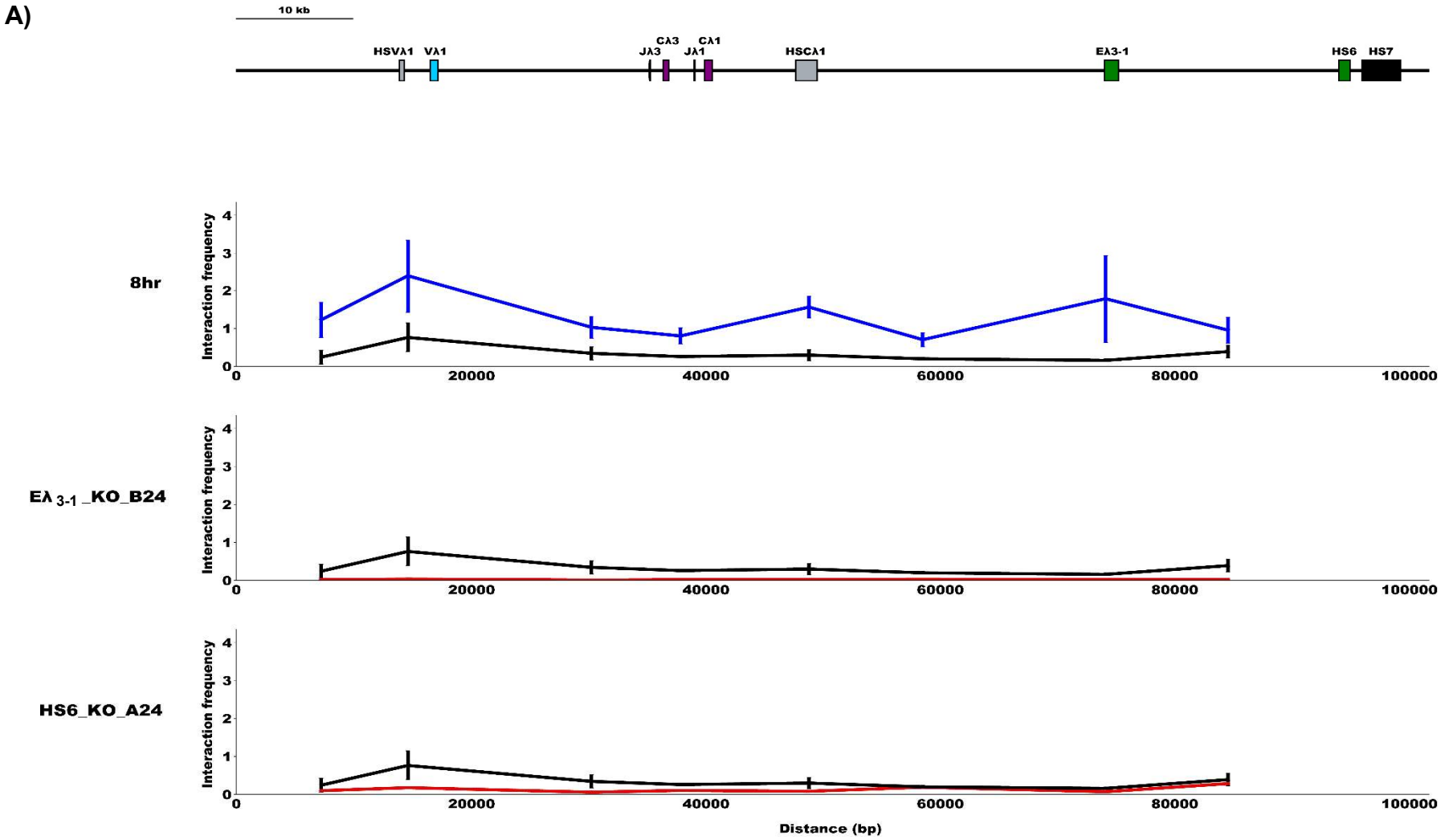
Together, these data suggest a model in which HS6 interacts with HSCλ1, mediated by YY1/Cohesin, at two hours post-induction. My data then suggest that Eλ$_{3\text{-}1}$ then contacts HSCλ1 by four hours post-induction, correlating with the increased interaction frequency between the two enhancers and the onset of *Vλ1* and *Jλ1* non-coding transcription.

5.8 IRF4 binding to HS6 and Eλ3-1 is essential for correct *Igλ* locus contraction

Temporal analysis of the *Igλ* locus by 3C using the 1D1-T215 cell line implied that IRF4 binding is responsible for the altered locus conformation observed in pre-B cells. An alternative explanation is that a downstream target of IRF4, for example Ikaros, is responsible for mediating the changes in locus conformation. To further tests the hypothesis that IRF4 is responsible for locus folding, I examined the effect of the removal of the PU.1/IRF4 binding site from either HS6 or Eλ3-1 on locus contraction. These were the only IRF4 binding sites with a significant enrichment in IRF4 binding (Figure 4.5) identified within the *Igλ* locus. To this end, I performed 3C on the HS6 and Eλ3-1 PU.1/IRF4 binding site mutant cell lines HS6 A24 and Eλ3-1 B24.

To examine the full effect on locus contraction, the HS6 and Eλ3-1 PU.1/IRF4 binding site mutant cell lines were induced for eight hours with 4-OHT, for the maximal level of locus activation. Following induction these cell lines (HS6 A24 and Eλ3-1 B24) were subjected to 3C and interaction profiles were generated from both the HS6 (Figure 5.12A) and Eλ3-1 viewpoints (Figure 5.12B). Surprisingly, the removal of the PU.1/IRF4 binding site at either enhancer dramatically reduced the interaction frequency of HS6 or Eλ3-1 with any region. For the vast majority of regions, the interaction frequency observed was substantially lower than the uninduced control, implying the *Igλ* is incapable of undergoing contraction without IRF4 binding to both enhancers. These data correlate well with the observed decrease in both *Vλ1* and *Jλ1* non-coding transcription in these mutant cell lines and imply that the inability of these lines to induce *Igλ* locus non-coding transcription is due to an aberrant three-dimensional structure. Furthermore, these data strongly imply that IRF4 binding is essential for both locus activation as well as locus contraction that the binding of additional factors is not essential for *Igλ* locus contraction.

A)


10 kb
HSVλ1
Vλ1
Cλ3
Jλ3
Jλ1
Cλ1
HSCλ1
Eλ3-1
HS6
HS7
8hr
Eλ 3-1 _KO_B24
HS6_KO_A24
Interaction frequency
0
1
2
3
4
0
20000
40000
60000
80000
100000
Distance (bp)

B)

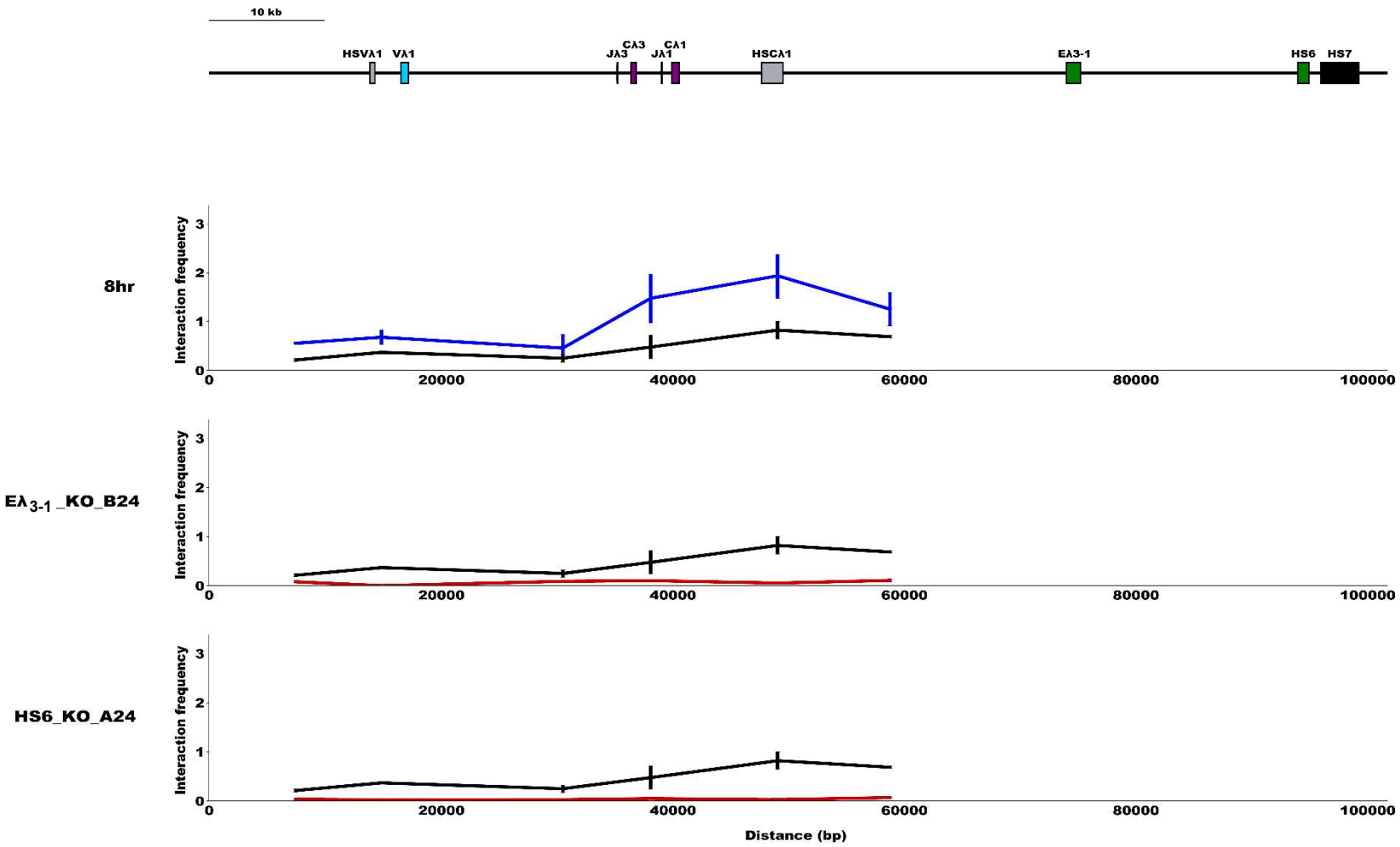


Figure 5.12: Analysis of the interactions formed by HS6 and Eλ3-1 in the HS6/Eλ$_{3-1}$ PU.1/IRF4 binding site mutant cell lines. Relative interaction frequency of DpnII restriction fragments from the HS6 (A) and Eλ3-1 (B) viewpoints using uninduced wildtype 1D1-T215 cells (black, n = 3) and eight hour induced cells (blue, n = 3) compared to the HS6 (A24, n = 2) and Eλ$_{3-1}$ (B24, n = 2) PU.1/IRF4 binding site mutant cell lines, induced for eight hours with 4-OHT (red). Error bars represent SEM.

C. Discussion

5.9 IRF4 binding to the *Igλ* enhancers is crucial for *Igλ* activation

In this chapter, I investigated the role that IRF4 plays in the activation of *Igλ* non-coding transcription and *Igλ* locus contraction. The data show that IRF4 binding to the enhancers HS6 and Eλ$_{3\text{-}1}$ is essential not only for the upregulation of *Vλ1* and *Jλ1* non-coding transcription, but also for *Igλ* locus contraction. Furthermore, IRF4 binding to HS6 and Eλ$_{3\text{-}1}$ increases in parallel and this strongly correlates with the upregulation of non-coding transcription observed. Moreover, I find that five interactions are altered during locus activation and the timing of these interactions is highly correlated with IRF4 binding.

The role of IRF4 in light-chain recombination is poorly understood. *Irf4*$^{-/-}$*Irf8*$^{-/-}$ mice are blocked at the pro-B cell stage, exhibit significantly decreased *Igκ* non-coding transcription and are unable to recombine *Igκ* upon IL-7 withdrawal (Johnson et al., 2008). Interestingly, the reintroduction of IRF4 significantly increases *Igκ* non-coding transcription, and this also results in the repositioning of the *Igκ* locus away from pericentric heterochromatin by an unknown mechanism (Johnson et al., 2008; Ma et al., 2006). This effect on non-coding transcription is likely to be mediated by the binding of IRF4 to the three core *Igκ* enhancers which can be observed at the pro-B cell stage. As significantly decreased histone acetylation at the enhancer 3'Eκ is observed in *Irf4*$^{-/-}$*Irf8*$^{-/-}$ pro-B cells, it is highly likely that IRF4/8 regulates the activity of this enhancer. Furthermore, this effect is likely due to the co-recruitment of E2A by IRF4 followed by the cooperative binding of these factors to 3'Eκ.

Whilst these studies imply that IRF4 binding is involved in the upregulation of non-coding transcription, the effects on locus activation and locus contraction have not previously been explored and furthermore the role that IRF4 plays in *Igλ* activation is very poorly understood. My data show that the binding of IRF4 to the *Igλ* enhancers Eλ$_{3\text{-}1}$ and HS6 is essential for the non-coding transcription of *Vλ1* and *Jλ1* as the removal of the region containing the IRF4 binding site at either enhancer is sufficient to completely inhibit the induction of non-coding transcription.

Notably, studies at the *Igκ* locus have indicated that both IRF4 and Ikaros, a direct target of IRF4, are capable of inducing *Igκ* non-coding transcripts and rearrangements (Heizmann et al., 2013; Ma et al., 2008). Ikaros binding is unlikely to be responsible for *Igλ* induction due to the lack of *Vλ1* and *Jλ1* non-coding transcription in the absence of IRF4. Instead, my data imply that IRF4 is the critical factor involved in the regulation of *Igλ* non-coding transcription as remarkably, the removal of the IRF4 binding site from either 3' *Igλ* enhancer is sufficient to inhibit both *Vλ1* and *Jλ1* non-coding transcription. It is, however, vital to note that the removal of the PU.1/IRF4 binding site at $E\lambda_{3\text{-}1}$ also resulted in the removal of a MEF2C binding site.

The MEF2C binding site at $E\lambda_{3\text{-}1}$ was identified by DNaseI footprinting of the enhancer and was found to bind within the λA motifs that are highly conserved between $E\lambda_{3\text{-}1}$ and $E\lambda_{2\text{-}4}$ (Satyaraj and Storb, 1998). The role of the transcription factor at these enhancers remains poorly understood. MEF2C is a member of the myocyte enhancer factor 2 family of transcription factors and has been implicated as a regulator of B-cell homeostasis and antigen activation (Herglotz et al., 2015). The factor is highly expressed throughout B-cell development and its binding appears to co-localise with IRF4 binding sites. Interestingly, *Irf4* expression is downregulated in *Mef2c*$^{-/-}$/*Mef2d*$^{-/-}$ pre-B cells and furthermore IRF4 can co-immunoprecipitate with MEF2C suggesting that the factor upregulates *Irf4* and could collaborate with IRF4 to enhance transcription (Herglotz et al., 2015).

It is possible that MEF2C plays a role in increasing enhancer accessibility or RNA pol II recruitment. On the other hand, the MEF2C binding site is intact at HS6 in the HS6 PU.1/IRF4 binding site mutant cell lines, and these cells exhibit a very similar phenotype to that of the Eλ$_{3\text{-}1}$ mutant cell lines, suggesting that the effect on non-coding transcription is unlikely to be due to loss of MEF2C. That said, the HS6 mutant cell lines do exhibit slightly increased *Vλ1* and *Jλ1* non-coding transcription compared to the Eλ$_{3\text{-}1}$ mutant cell lines, which could indicate that MEF2C binding plays a role in the regulation of enhancer activity. The relatively weak effect, however, suggests that it is not an essential regulator of non-coding transcription. Unfortunately, it was not possible to design sgRNAs that generate a smaller, more targeted deletion as *Staphylococcus pyogenes* (Sp) Cas9 target sequences must be adjacent to the correct protospacer adjacent motif (PAM; NGG) resulting in the unavoidable removal of the MEF2C binding site when using SpCas9. In order to examine the effect of the removal of only the PU.1/IRF4 binding site the use of a CRISPR/Cas system from another species such as *Neisseria meningitides* (Lee et al., 2016) would be necessary.

Whilst MEF2C may play a role in the regulation of *Igλ* locus non-coding transcription, the binding of IRF4 to HS6 and Eλ$_{3\text{-}1}$ strongly corelates with the formation of the enhancer hub at HSCλ1 and the subsequent upregulation of *Vλ1* and *Jλ1* non-coding transcription. The vital role of IRF4 in *Igλ* locus recombination is also supported by previous studies from the Boyes group which generated transgenic mice with a variable IRF4 transgene copy number (Bevington and Boyes, 2013). The pro-B cells from PIP3 transgenic mice which express a pre-B cell level of IRF4 (3-fold over the pro-B cell level) exhibit a pre-B cell level of *Igλ* locus non-coding transcription and recombination (Bevington and Boyes, 2013). In comparison, the pro-B cells of PIP4 transgenic mice which express 2.1-fold the pro-B level of IRF4, respectively, exhibit a minor upregulation of *Igλ* non-coding transcription and recombination (Bevington and Boyes, 2013). These studies together with the temporal analysis of IRF4 binding and IRF4 binding site mutation experiments strongly suggest that the level of IRF4 alone is the master regulator of *Igλ* locus non-coding transcription.

5.10 IRF4 is essential for *Igλ* locus contraction

Analysis of pro-B and pre-B cells by 3C identified five key interactions that are altered during locus activation. Remarkably, all five interactions were recapitulated in the 1D1-T215 cell line upon IRF4-ER_{T2} activation, implying that the binding of IRF4 to the enhancers drives the locus contraction observed. Interestingly, the *Igλ* locus contraction appears to be dramatically reduced when the PU.1/IRF4 binding site was removed from either HS6 or $Eλ_{3-1}$. The binding of other transcription factors, with the exception of MEF2C, to these enhancers is not altered by the removal of these binding sites. Furthermore, the timing of IRF4-ER_{T2} binding correlates well with the increased interaction frequency between the enhancers and vital regions such as HSCλ1 and *Jλ1*. These observations together suggest that IRF4 binding is crucial for *Igλ* locus contraction.

IRF4 has no known ability to directly regulate long range interactions therefore it is unclear how IRF4 binding results in the increased *Igλ* locus contraction observed. The binding of IRF4 could potentially mediate locus contraction by improving the binding of architectural proteins such as YY1. At HS6, YY1 binding was observed to increase in pre-B cells and this would correlate with increased IRF4 binding. Notably, YY1 can interact with enhancer RNAs, which aids the recruitment of YY1 to the enhancer (Sigova et al., 2015). Therefore, it is possible that IRF4 binding to the enhancer could regulate locus contraction by upregulating enhancer RNA production, resulting in increased YY1 recruitment to the enhancer and thereby aiding in the generation of the enhancer hub. Alternatively, IRF4 has been shown to interact with the MED23 subunit of the Mediator complex (Griffiths et al., 2013). As Mediator has been shown to regulate long-range interactions via cohesin (Kagey et al., 2010) it is therefore possible that IRF4 binding directs locus contraction via interaction with Mediator. Further study, however, is required to determine exactly how IRF4 regulates these long-range interactions.

Chapter 6 – Discussion

6.1 Generation of a system capable of inducing V(D)J recombination

Aberrant V(D)J recombination has been strongly linked to chromosome translocations that underpin the development of several types of leukaemia and lymphoma, including acute lymphoblastic leukaemia and follicular lymphoma (Lieber, 2016). In order to understand how errors in V(D)J recombination are able to occur, it is vital to explore the regulation of this process. Rearrangement events are subject to tight regulation to ensure that recombination only occurs at the correct locus at the correct developmental stage to prevent widespread genomic instability.

The tight regulation of V(D)J recombination has been a significant barrier in the study of this process due to recombination only occurring at transient stages during B lymphopoiesis. The mechanism of the recombination reaction has mainly been explored by *in vitro* studies, but to explore how V(D)J recombination is epigenetically regulated to ensure that the correct locus is rearranged at the correct developmental stage, *in vivo* model systems are required. The analysis of B cell progenitors from transgenic mice deficient in key transcription factors e.g. IRF4 has provided a wealth of information regarding the function of these transcription factors in B cell development. However, these studies are often confounded by the numerous roles of these key transcription factors. For example, the role of PAX5 in *Igκ* recombination is very difficult to examine as *Pax5* deficient mice also have defective pre-BCR signalling. This makes it very difficult to determine the exact function of a transcription factor at a stage of B cell development.

In order to examine the regulation of V(D)J recombination in further detail a system in which V(D)J recombination can be induced is required. In Chapter 3, I describe the generation of a PIP-ER pro-B cell-line, 1D1-T215, capable of activating the *Igλ* locus upon induction by 4-hydroxytamoxifen. Development of such a cell line is important since there are problems with the strategies previously used to explore this process. To prevent recombination from occurring prematurely, two main strategies have been utilised: mutation of critical factors required for V(D)J recombination e.g. *Rag1/Rag2* (Schwickert et al., 2014), *Irf4* (Johnson et al., 2008), E2A (Lazorchak et al., 2006) and/or infection with the Abelson Murine Leukaemia virus.

These strategies have provided many insights into signalling pathways regulating B cell development and V(D)J recombination, however, they are both not truly representative of the events that occur in wild-type cells. Without generating an immortalised cell line, the rescue of V(D)J recombination by the re-expression of the mutated factor is likely to result in a degree of heterogeneity in populations of primary cells, which could potentially confound results. Furthermore, in the case of the rescue of a *Rag2*$^{-/-}$ deficient cell line, 63-12 (Shinkai et al., 1992), the expression of RAG2 failed to recapitulate the frequency and timescale of recombination observed in primary pro-B cells (Lescale et al., 2016), likely due to the repression of immunoglobulin loci accessibility by v-abl signalling. Whilst the repressive effects of v-abl signalling can be ablated by the use of small molecular inhibitors or temperature shifting in the case of v-abl temperature sensitive mutants, inhibition of v-abl signalling results in dramatically increased apoptosis together with large-scale alterations to the transcriptome (Muljo and Schlissel, 2003). These effects are not truly representative of the events that occur in wild-type cells, and therefore the events preceding V(D)J recombination in wild-type pro-B cells cannot be definitely assessed in this manner.

The analysis of 1D1-T215 cell holds several advantages over the strategies discussed. One major advantage of this system is that it relies solely on the expression of a pre-B cell level of IRF4-ER_{T2} and its subsequent activation by 4-hydroxytamoxifen. Unlike the inactivation of v-abl which results in many non-physiological alterations in gene expression, in 1D1-T215 cells only IRF4 activity is altered by induction. Furthermore, as IRF4-ER_{T2} is expressed at near physiological levels, the transcriptome and genome topological changes observed should be highly representative of the events that occur in wildtype pre-B cells prior to V(D)J recombination. Moreover, as IRF4-ER_{T2} activity is tightly regulated (Figure 3.13B) the exact effects of IRF4 on *Igλ* recombination can be examined with a high degree of detail, aided by the rapid proliferation of the cell line and the low of variability between 1D1-T215 cells. Additionally, this system enables high resolution temporal analysis of the regulation of *Igλ* recombination to be performed for the first time.

Whilst there are multiple advantages to the use of the 1D1-T215 to explore the regulation of V(D)J recombination, there are several caveats with the current cell line. Firstly, in order to generate a cell line in which V(D)J recombination can be induced, I needed to both immortalise the primary pro-B cells and prevent recombination, I therefore resorted to using A-MuLV based immortalisation. The result of this immortalisation strategy, however, is the ~20-fold reduction of *Rag1* expression compared to wildtype pro-B cells. This is likely due to STAT5, which is constitutively activated by v-abl signalling, preventing FOXO1 binding to *Erag* (Amin and Schlissel, 2008). The expression of *Rag2* is also likely to be downregulated by the same mechanism due to the shared regulatory element.

The downregulation of *Rag* expression is both advantageous and disadvantageous to this system. The extensive proliferation required to generate the 1D1-T215 combined with a primary pro-B cell level of RAG1 and RAG2 would have likely resulted in the premature recombination of the light chain loci, due to the low but detectable level of non-coding transcription at these loci. The repression of *Rag* expression has also enabled the genome editing experiments, such as those performed in Chapter 5 to occur without the premature recombination of the locus. Conversely, the reduction in *Rag* expression also limits the utility of this cell line somewhat as it prevents the temporal analysis of *Igλ* recombination, due to the altered timescales of recombination in comparison to primary pro-B cells.

With modifications to the 1D1-T215 cell line, it is highly likely that inducible recombination can be achieved. *Rag1* and *Rag2* could be introduced under a promoter insensitive to STAT5 repression e.g. the EF1α promoter, using lentiviral vectors. As this could potentially induce premature *Igλ/Igκ* recombination, either *Rag* gene should be placed under the control of an inducible promoter e.g. a tetracycline responsive promoter or fused to a regulatory domain e.g. ecDHFR (Iwamoto et al., 2010) to enable the controlled expression or activation of the RAG proteins. Alternatively, dCas9-VP64 fusion proteins targeted to enhancers have been shown to upregulate gene expression (Simeonov et al., 2017) implying that targeting this transcriptional activator to *Erag* could mediate the upregulation of both *Rag1* and *Rag2*. As Cas9 proteins have also been fused to regulatory domains (Maji et al., 2017), upregulation of *Rag* expression could be controlled by the induction of a dCas9 fusion protein and this could potentially provide a more physiological method of inducing *Rag* expression.

In addition to the repression of *Rag* expression, v-abl signalling also appears to reduce non-coding transcription at the *Igλ* locus. The effect appears to be more pronounced on *Jλ1* non-coding transcription, which is reduced by approximately 1000-fold compared to primary pro-B cells, however, *Vλ1* non-coding transcription is also affected. The mechanism by which this repression is mediated is unknown but likely involves STAT5, and ChIP-seq analysis of STAT5 in a v-abl immortalised cell line could potentially determine if this factor is involved in the repression of *Igλ* non-coding transcription. Again, the repression of *Igλ* non-coding transcription has advantages and disadvantages to this system. Whilst the level of non-coding transcription is not truly representative of primary pro-B cells, the repression of *Igλ* non-coding transcription enables the effects of IRF4 activation on non-coding transcription and locus contraction to be readily observed.

The final caveat of the 1D1-T215 cell line is that the insertion site(s) of the IRF4-ER_{T2} transgene is unknown due to the use of random viral integration. Due to preliminary data from low transgene copy number cell lines, it was evident that a number of transgene copies were required to express IRF4-ER_{T2} at the same level as IRF4 in pre-B. I initially sought to integrate the IRF4-ER transgene into the ROSA26 locus using CRISPR/Cas9, however, due to the low infection percentage and time constraints it was not possible to perform a targeted insertion of the transgene. Whilst the exact location of the insertion site/sites are unknown, the IRF4-ER_{T2} transgene is stably expressed and is not prone to silencing. Thus, this strategy has generated a stable cell line that allows the initial steps in *Igλ* recombination to be examined and has the potential to be modified for the analysis of the later stages of recombination.

6.2 Coordinate activation of *Vλ1* and *Jλ1* gene segments by the parallel activation of sister enhancers

The first step in the activation of the *Igλ* locus for recombination is the upregulation of non-coding transcription, which is required to enable the recombinase access to the RSS. The analysis of PIP-ER pro-B cells and the generated cell line, 1D1-T215, has shown that non-coding transcription of *Vλ1* and *Jλ1* is coordinately upregulated and this coordinate activation is likely to play a role in promoting recombination between V and J gene segments. It would also help to prevent the recombination between V or J gene segments and cryptic RSSs elsewhere in the genome. Building on previous studies at the *Igλ* locus, I present a model by which the coordinate activation of *Vλ1* and *Jλ1* gene segments is regulated.

The data presented in Chapter 4 and Chapter 5 suggest the coordinate activation of *Vλ1* and *Jλ1* gene segments is regulated by the parallel activation of sister enhancers, mediated by the formation of long-range interactions between previously unidentified regulatory elements. I therefore propose a four-step model to explain how the coordinate regulation of non-coding transcription is regulated (Figure 6.1).

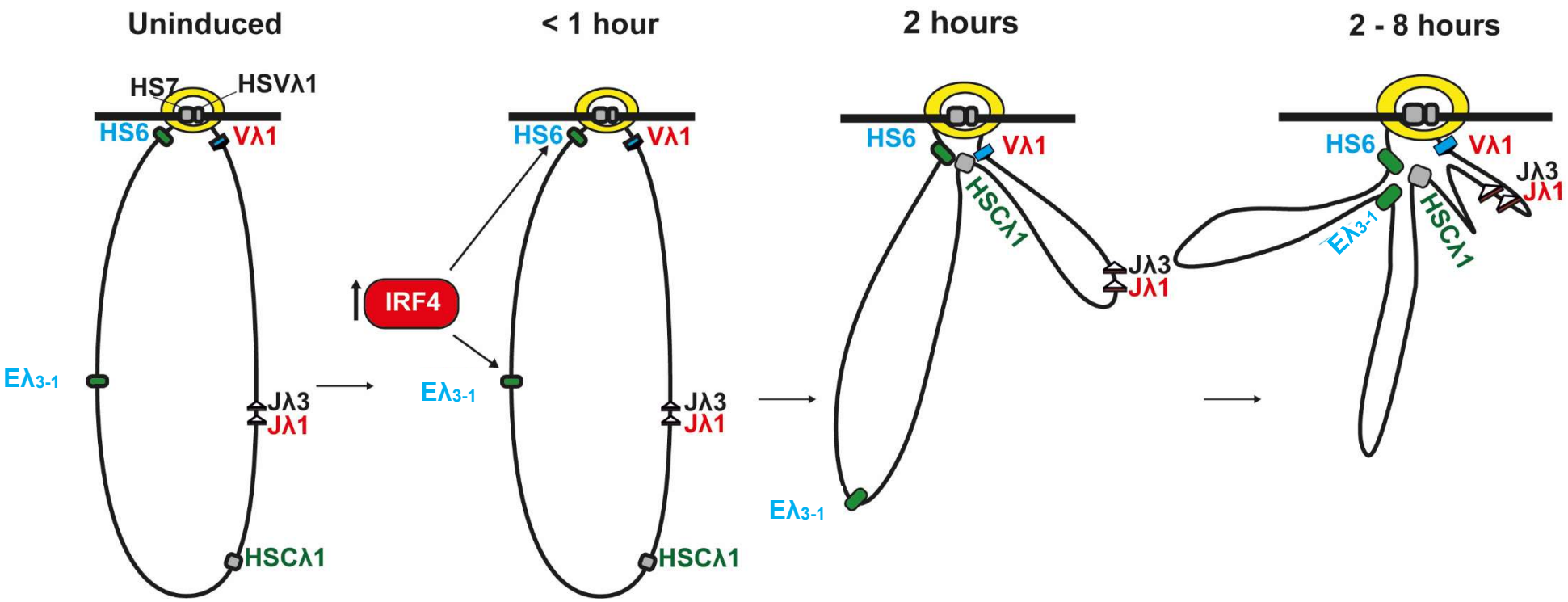


Figure 6.1: Proposed model for the coordinate regulation of *Vλ1* and *Jλ1* non-coding transcription. In untreated cells 1D1-T215 cell CTCF/cohesin (yellow) mediate the formation of an interaction between HS7 and HSVλ1 (grey). The activation of IRF4-ER_{T2} by 4-hydroxytamoxifen results in the simultaneous increase in IRF4 binding to HS6 and $Eλ_{3\text{-}1}$. This triggers the enhancers to interact with HSCλ1, forming a hub which facilitates the interaction between the enhancers and the *Vλ1* and *Jλ1* promoters, resulting in the upregulation of both Vλ1 and Jλ1 non-coding transcription.

Step 1 – Locus contraction mediated by CTCF

In Chapter 4, I identified multiple CTCF binding sites present in the *Igλ* locus by the analysis of published ChIP-seq data. In regard to the regulation of *Vλ1* and *Jλ1*, the CTCF binding sites HS*Vλ1* and HS7 appear to be the most important. HS*Vλ1* is located approximately 3 kb upstream of the *Vλ1* gene segment, whereas HS7 is located 24 kb downstream of the enhancer $Eλ_{3-1}$ and 2.2 kb downstream of HS6. The convergent orientation of the CTCF binding sites, together with examination of published pro-B Hi-C data and preliminary 3C analysis performed by James Scott, strongly suggests that these CTCF sites interact. This interaction is likely mediated by cohesin, as the presence of RAD21 can be observed in publicly available pro-B cell ChIP-seq data (Figure 4.12), and would result in the formation of an 85 kb domain encompassing the 3' half of the *Igλ* locus (Figure 4.18). Homologous CTCF binding sites can also be observed in the 5' half of the locus, upstream of Vλ2/VλX (HSVλ2 and HSVλX) and 34 kb downstream of the $Eλ_{2-4}$ enhancer (HS7-1). Furthermore, the orientation of these binding sites indicates that HSVλ2 or HSVλX can interact with HS7-1, via CTCF/cohesin, resulting in the formation of either a 117 kb or 96 kb domain.

The formation of separate 5' and 3' domains within the *Igλ* locus provides a potential explanation as to why recombination is rarely observed between gene segments present in the 5' and 3' halves of the *Igλ* locus (Sanchez et al., 1991). RAG recombinase activity has been shown to be tightly restricted within loop domains. DSBs generated within loop domains are re-joined with DBSs within topologically associated domains at a much higher frequency than elsewhere in the genome (Alt et al., 2013; Zhang et al., 2012), implying that the increased chromatin interaction frequency within loop domains results in increased likelihood of gene segment synapsis within the domain. Moreover, the synapsis of gene segments within separate domains appears to be inhibited (Hu et al., 2015). This is potentially due to the termination of RAG tracking along chromatin when the complex encounters CTCF/cohesin bound elements (Hu et al., 2015), which would prevent the capture of an RSS outside of a loop domain. The coordinate activation of V and J gene segments potentially has a role in increasing the efficiency of correct recombination events to prevent cryptic recombination within loop domains.

In addition to mediating the separation of the *Igλ* locus the CTCF/cohesin mediated interaction between HS*Vλ1* and HS7 results in the HS6 enhancer being relocated into close proximity to the promoter of *Vλ1*. Analysis of the interactions formed by HS6 by both 3C and Hi-C indicated an interaction between HS6 and *Vλ1*. Moreover, with the exception of the cell lines in which the PU.1/IRF4 binding sites are mutated (Figure 5.13), a substantial increase in interaction frequency between HS6 and *Vλ1*, in comparison to the control fragments, was observed in all cell types examined, implying that the interaction between HS6 and *Vλ1* occurs at the pro-B cell stage. In support of this hypothesis, CTCF and RAD21 binding can be detected at HS*Vλ1* and HS7 at the pro-B cell stage implying that the HS*Vλ1*-HS7 interaction is formed prior to locus activation. This is similar to studies at the *Igκ* locus which have indicated that the locus is contracted at the pro-B cell stage, but this is not sufficient for the initiation of V(D)J recombination. Instead the increased binding of IRF4 and E2A to the iEκ and 3'Eκ enhancers, mediated by pre-BCR signalling, is required to activate gene segments for recombination (Stadhouders et al., 2014).

Step 2 – The parallel binding of IRF4 to sister enhancers

Locus contraction mediated by the CTCF binding sites appears to be insufficient for *Igλ* activation. Analysis of the transcription factors binding to the *Igλ* enhancers at the pro-B and pre-B cell stages has indicated that IRF4 is the only factor to exhibit altered binding to both enhancers and furthermore, the level of IRF4 has been shown to be critical for *Igλ* recombination. In the pro-B cells of PIP3 transgenic mice, IRF4 expression is approximately three-fold that of non-transgenic mice, which is slightly above the pre-B cell level. Pro-B cells isolated from these mice exhibit a pre-B cell level of non-coding transcription and *Igλ* recombination (Bevington and Boyes, 2013). In contrast, the pro-B cells of PIP2 transgenic mice express 1.4-fold the level of IRF4 observed in non-transgenic pro-B cells and exhibit no increase in *Igλ* locus non-coding transcription or recombination compared to non-transgenic mice (Bevington and Boyes, 2013). Moreover, my data also confirm that the expression of a pre-B cell level of IRF4 is sufficient for the upregulation of *Igλ* locus non-coding transcription.

Whilst the level of IRF4 appears to vital for full activation of the Igλ locus, it is unclear how the overexpression of IRF4 results in increased *Igλ* locus non-coding transcription and recombination. One possibility is that a threshold level of IRF4 is required to achieve stable IRF4 binding to the *Igλ* enhancers and that only when this level is reached do the long-range interactions mediated by IRF4 occur, resulting in *Igλ* locus activation. The required level of IRF4 would not be present in the majority of pro-B cells, however, some pro-B cells do exhibit *Igλ* recombination. This could be explained by the heterogeneous expression of the transcription factor at the single cell level. This has been shown in pluripotent stem cells (Torres-Padilla and Chambers, 2014) and would potentially result in a small proportion of the pro-B cell population expressing sufficient IRF4 for *Igλ* recombination. Increasing the level of IRF4 could therefore increase the number of cells above the threshold required for activation, resulting in increased *Igλ* recombination occurring pre-B cells. Under this model, my data suggest that the threshold level of IRF4 to achieve stable binding to the enhancers is reached two-four hours post induction, in a proportion of 1D1-T215 cells, as interaction of both enhancers with HSCλ1 can be detected at this time. It is possible that this level

is reached earlier but 3C-qPCR is not sensitive enough to detect the alteration in locus conformation.

It is also possible that the large number of potential targets within the mouse genome and the instability of transcription factor-DNA interactions (Chen et al., 2014; McNally et al., 2000; Paakinaho et al., 2017; Voss et al., 2011) could potentially result in a low probability of *Igλ* enhancer occupation by IRF4 in pro-B cells. This is consistent with an investigation into the dynamics of the transcription factor SOX2 which indicated that the factor spent ~97% of the time in stochastic diffusion and only approximately 3% of SOX2 molecules are bound to recognition sites in embryonic stem cells with a mean residence time of ~12 seconds (Chen et al., 2014). Increasing the level of IRF4 would increase the probability of enhancer occupation and therefore *Igλ* locus activation. This model would suggest that the number of 1D1-T215 cells in which IRF4 is bound at the enhancers increases gradually after induction resulting in an increased number of cells displaying enhancer interactions and *Igλ* locus non-coding transcription.

Notably, the analysis of pro-B cell and pre-B cells suggested that IRF4 binding may increase simultaneously at both the $E\lambda_{3\text{-}1}$ and HS6 enhancers due to a similar enrichment in binding observed by ChIP-qPCR. This hypothesis was confirmed by temporal IRF4 ChIP-qPCR performed in the 1D1-T215 cell line (Figure 5.7) as a very similar pattern of IRF4 binding was observed at both enhancers following the induction of IRF4-$ER_{T2.}$ IRF4 occupation at both $E\lambda_{3\text{-}1}$ and HS6 appears to be required for the upregulation of *Igλ* locus non-coding transcription (Figure 5.6). It is therefore possible that the level of IRF4 expressed in pro-B cells is insufficient for the simultaneous activation of both enhancers or that the probability that both enhancers are occupied by IRF4 is very low. The enhancer-hub would therefore only be formed, and transcription upregulated in cells expressing the required level of IRF4.

It is difficult to identify which of the two models is the most plausible in the absence of data from single cells. Furthermore, determining transcription factor occupancy at the single cell level is technically challenging. While the analysis of histone modifications is possible by single cell ChIP (Clark et al., 2016) the frequency of transcription factor occupancy has not yet been assessed at the level of a single cell and current imaging techniques are not be appropriate for this analysis.

Step 3 – Enhancer recruitment to HSCλ1

The activation of both Eλ$_{3\text{-}1}$ and HS6 by IRF4 binding, which is likely to occur simultaneously, appears to result in the formation of a transcriptional hub centred on HSCλ1 before four hours post-induction. Notably, the removal of the PU.1/IRF4 binding site at either the HS6 or Eλ$_{3\text{-}1}$ enhancer dramatically reduces both *Vλ1* and *Jλ1* non-coding transcription, implying that both enhancers are required for gene segment activation. This implies that the formation of the enhancer hub may be essential for the regulation of *Vλ1* and *Jλ1* non-coding transcription. The first observable change in the three-dimensional structure of the *Igλ* locus following IRF4 activation is the increased interaction of Eλ$_{3\text{-}1}$ and HS6 with the hypersensitive site HSCλ1. This does not appear to be an artefact of the cell line as this interaction can also be observed in pre-B cells (Figure 5.9).

The function of HSCλ1 is poorly understood, this region appears to be highly accessible and be bound by E2A and the architectural protein YY1. As YY1 binding has been observed at HS6, it is likely that interactions mediated by YY1/cohesin result in the interaction of HSCλ1 with HS6. Interestingly, YY1 binding appears to increase slightly from the pro-B to pre-B cell stage, correlating with IRF4 binding. Whilst YY1 is likely involved in this interaction, the reason why the HS6-HSCλ1 interaction is not detectable prior to IRF4 binding is unclear. A possible explanation is that IRF4 mediates an increase in enhancer accessibility. $Eλ_{3\text{-}1}$ accessibility increases 1.47-fold from the pro-B to pre-B cell stage (Grange and Boyes, 2007). As IRF4 appears to be the only transcription factor to exhibit altered binding to the enhancer at the pre-B cell stage and the removal of the PU.1/IRF4 binding site at HS6 results in aberrant locus contraction, it is likely that IRF4 is responsible for this increased accessibility which could enhance YY1 binding. Alternatively, as discussed in Section 5.10, it is possible that IRF4 enhances the production of enhancer RNAs which increases the recruitment of YY1 to HS6. Additional studies are required, however, to determine if IRF4 binding increases YY1 recruitment and the mechanism by which this could occur.

Whilst 3C analysis strongly suggests the interaction of $Eλ_{3\text{-}1}$ with HSCλ1, it is unclear how this interaction is mediated. One possibility is that the Mediator complex or even H3K4me1 could facilitate this interaction via cohesin (Kagey et al., 2010; Local et al., 2018; Yan et al., 2018). IRF4 has been shown to interact with the Mediator subunit Med23 (Griffiths et al., 2013). This could potentially explain the link between IRF4 binding and the interaction between $Eλ_{3\text{-}1}$ and HSCλ1 as increased IRF4 binding would result in the increased recruitment of the Mediator complex, enabling a stronger interaction with HSCλ1. It is also possible that the HS6-HSCλ1 interaction is enhanced in this manner.

Interestingly, the temporal analysis of the interactions formed in the *Igλ* locus following IRF4 binding indicates that the interaction of HS6 with HSCλ1 occurs before two hours post-induction, prior to that of the interaction between $Eλ_{3\text{-}1}$ and HSCλ1 which occurs two to four hours post-induction. As these enhancers are both likely to be activated in parallel by IRF4 binding, it is unlikely that such a long delay exists. Instead it is more likely that this observation is due to the limited PCR amplification of the $Eλ_{3\text{-}1}$ containing fragment which limited the detection of interactions in comparison to the HS6 viewpoint.

Further work is required to determine why the generation of such a structure is necessary for gene segment activation. Enhancers bearing binding sites for different transcription factors have been shown to cluster, resulting in a transcription factor binding platform as large as 12.5 kb, termed stretch or super-enhancers (Hnisz et al., 2013; Whyte et al., 2012). The 3‘ *Igλ* enhancers, however, do not meet the criteria for a super-enhancer. A recent study by the de Laat group used multi-contact 4C sequencing to investigate the topology of the β-globin locus and observed that various elements of super-enhancers can aggregate to form tissue specific enhancer hubs (Allahyar et al., 2018), implying that hub formation is often involved in gene activation.

Step 4 – Coordinate activation of non-coding transcription by the parallel activation of sister enhancers.

The generation of the enhancer hub at HSCλ1 is highly correlated with the time at which the coordinate non-coding transcription of *Vλ1* and *Jλ1* is readily detectable. This suggests that the enhancer hub facilitates the formation of enhancer-promoter contacts between the two enhancers and the *Vλ1* and *Jλ1* promoters.

Analysis by 3C indicates that HS6 contacts *Vλ1* more frequently than *Jλ1*, which could be explained by the interaction of HS*Vλ1* with HS7 at the pro-B cell stage which would result in the close proximity of *Vλ1* with HS6 at the enhancer hub. Interestingly, this interaction appears to be insufficient for complete activation in the absence of the second enhancer, this is evidenced by both 3C analysis and PU.1/IRF4 binding site mutation at $Eλ_{3\text{-}1}$. Analysis of the interactions mediated by $Eλ_{3\text{-}1}$ indicated that this enhancer contacts *Jλ1* with increased frequency. The mechanism by which *Jλ1* is recruited to the enhancer hub is unclear as architectural protein binding cannot be detected. It is possible that the recruitment of *Jλ1* to the enhancer hub is mediated by the capture of the *Jλ1* promoter by $Eλ_{3\text{-}1,}$ which would explain the increased interaction frequency between these elements.

The coordinate activation of *Vλ1* and *Jλ1* gene segments could therefore be achieved by the parallel activation of HS6 and Eλ$_{3\text{-}1}$ by IRF4, resulting in the recruitment of the Mediator complex by the interaction of IRF4 with MED23. IRF4 binding results in the formation of a transcriptional hub at HSCλ1, potentially by the activities of YY1/cohesin and/or Mediator/cohesin resulting in a high degree of recruitment of the Mediator complex to HSCλ1. Interaction of HS6 with *Vλ1* and Eλ$_{3\text{-}1}$ with *Jλ1*, following their recruitment to the enhancer hub, would then result in a coordinate increase in non-coding transcription due to the activity of Mediator at both gene segments.

6.1 Implications of the coordinate activation model

In addition to enabling the coordinate upregulation of non-coding transcription, the recruitment of *Vλ1* and *Jλ1* to the enhancer hub could also play a role in facilitating recombination. The recruitment of the gene segments to the hub and their activation would aid the formation of a synaptic complex by RAGs whilst enhancing RSS cleavage due to the increased accessibility provided by non-coding transcription.

This model provides an explanation for the coordinate upregulation of *Vλ1* and *Jλ1*, but it is unclear how the bias in recombination between these two gene segments is achieved as *Jλ3* should also be present within the transcriptional hub and subject to the same upregulation of non-coding transcription. The temporal analysis of *Jλ3* non-coding transcription was not investigated and it is possible that the upregulation of *Jλ3* non-coding transcription occurs after *Jλ1* non-coding transcription or that differences in the *Jλ1* and *Jλ3* promoters results in a reduced level of *Jλ3* non-coding transcription compared to *Jλ1*. Alternatively, the three-dimensional architecture of the locus following contraction could potentially result in *Jλ1* being in closer proximity to *Vλ1*, which may explain the increased recombination frequency.

Whilst the model cannot currently explain the bias in recombination between *Vλ1* and *Jλ3*, it does provide an insight into why recombination is heavily biased towards the 3'*Igλ* domain. Recombination in the 5'*Igλ* domain is likely to be reduced due to the formation of two mutually exclusive chromatin domains and the absence of an HSCλ1 homologue. As CTCF mediated interactions can either be formed between HSVλ2 or HSVλX and HS7-1, this could prevent recombination of the V gene segment without an interacting upstream CTCF binding site. This would reduce the frequency of recombination of either V gene segment. In addition to the reduction of gene segment utilisation by the formation of alternative domains, no region bearing similarity to HSCλ1 can be observed in the 5' domain implying that the formation of the enhancer hub structure observed in the 3' domain is absent from the 5' domain. This is likely to dramatically reduce the frequency of recombination between V and J gene segments in the 5' domain.

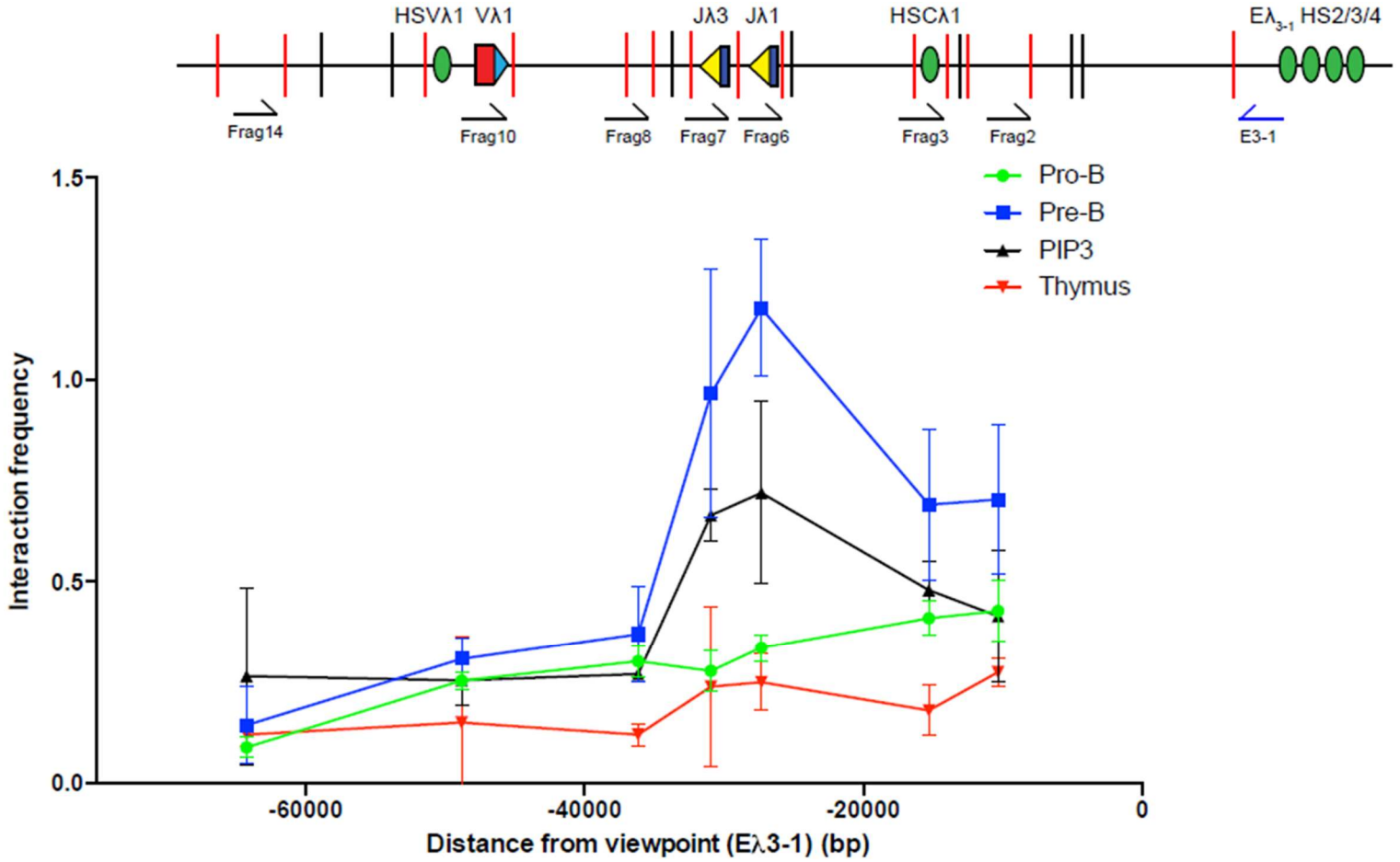
The proposed model bares similarities to that proposed for the regulation of the *Igκ* locus, where locus the locus is present in a contracted state at the pro-B cell stage (Lin et al., 2012; Ribeiro de Almeida et al., 2011; Stadhouders et al., 2014b) mediated by the actions of CTCF,YY1 and cohesin. These interactions are similar to the CTCF/cohesin mediated interactions formed in the *Igλ* locus. The iEκ and 3'Eκ enhancers are proposed to be active at this stage but act in an unfocused manner (de Almeida et al., 2015). Signalling via the pre-BCR results in the increased binding of IRF4 and E2A, in addition to the loss of STAT5 occupancy at the enhancers (Mandal et al., 2011). Further recruitment of E2A to both *Igκ* enhancers and Vκ genes (Lazorchak et al., 2006; Sakamoto et al., 2012) then allows the synergistic promotion of *Igκ* recombination by both enhancers, which now both operate in a highly focused and coordinated fashion to upregulate Vκ non-coding transcription (Stadhouders et al., 2014). This is similar to the proposed model as IRF4 binding appears to drive the formation of an enhancer hub which upregulates non-coding transcription. The major difference between these models is that at *Igκ* only V gene segments are proposed to be activated in this manner, whilst I propose that this results in the coordinated activation of two gene segments. However, at the *Igκ* locus there are over 100 V gene segments distributed over 3.2 Mb and therefore such a mechanism may not be feasible at this locus.

6.2 Conclusions and Future directions

In this thesis, I describe the generation and characterisation of a novel system to investigate the temporal regulation of V(D)J recombination and use this system, together with published ATAC-seq, ChIP-seq and Hi-C data to develop a model regarding how the coordinate activation of V and J gene segments is achieved. Whilst the generated cell line has the capacity to answer many questions regarding the regulation of V(D)J recombination, *Rag* expression must be upregulated in order to examine the later stages of V(D)J recombination and to explore the potentially carcinogenic effects of the recombination by-product, the ESC.

In the final results chapters, I describe the evidence for a model whereby the coordinate activation of *Vλ1* and *Jλ1* is regulated by the parallel activation of the sister enhancers HS6 and Eλ$_{3\text{-}1}$ via the formation of an enhancer hub focussed on HSCλ1. To provide additional support for this model, higher-resolution temporal analysis of long-range interactions should be performed. Ideally, these interactions should be assessed at an increased temporal resolution, for example 15- or 30-minute intervals, to better establish when the observed interactions occur. Furthermore, 3C is not an ideal technique to investigate these interactions as the *Igλ* locus is only recombined in approximately 5% of cells, this combined with the low amount of amplifiable material recovered by 3C and the analysis of this material by PCR, which can only detect one out of four possible ligation events, results in poor sensitivity. A method such as Capture-C (Davies et al., 2016), whereby a 3C sample is enriched for viewpoint containing fragments and sequenced to provide optimal resolution of a single locus, would provide increased sensitivity and increased interaction resolution to potentially identify rarer interactions.

Appendix



Appendix 1: Analysis of the interactions formed by Eλ3-1. Chromosome conformation capture analysis, using the restriction enzyme BamHI, of thymus cells (red), pro-B (green) and pre-B cells (blue) using Eλ3-1 as the viewpoint region. A substantial alteration in interaction frequency cannot be observed between Eλ3-1 and Vλ1 in any cell type examined. Error bars represent SEM. Data provided by Sarah Bevington.

Appendix 2: Interactions identified in the 3' half of the *Igλ* locus by Hi-C. Regions with a significant number of interaction reads identified by HOMER are displayed.

Interaction	Region(1)	Chr	Start	End	Region(2)	Start	End(2)	Interaction reads	Expected reads	Z-score	LogP	FDR
1	Jλ1	chr16	19049847	19069847	HSCλ1	19049198	19069198	340	0.005665	807.408446	-1722.84216	0
2	Cλ3	chr16	19049903	19069903	HSCλ1	19049225	19069225	337	0.005626	805.809807	-1704.510415	0
3	Jλ3	chr16	19050178	19070178	HSCλ1	19049343	19069343	315	0.00538	787.6476	-1570.35113	0
4	Eλ3-1	chr16	19012895	19032895	HS6	19002393	19022393	52	0.008732	80.120946	-158.497075	0
5	HSCλ1	chr16	19040482	19060482	Eλ3-1	19018272	19038272	26	0.007977	43.851562	-62.519797	0
6	Eλ3-1	chr16	19013875	19033875	HS7	18998362	19018362	25	0.007463	45.066827	-59.255407	0
7	HSCλ1	chr16	19044756	19064756	HS6	18997138	19017138	19	0.007265	35.187075	-40.439429	0
8	Jλ3	chr16	19054572	19074572	HS6	18998847	19018847	18	0.00589	41.117331	-37.507015	0
9	Cλ3	chr16	19054572	19074572	HS6	18998847	19018847	18	0.006159	39.319335	-37.499724	0
10	Jλ1	chr16	19053468	19073468	HS6	18998664	19018664	18	0.006201	39.049716	-37.498631	0
11	HSCλ1	chr16	19045477	19065477	HS7	18995652	19015652	18	0.006209	39.001052	-37.484776	0
12	Cλ1	chr16	19051199	19071199	HS6	18997789	19017789	18	0.006825	35.48131	-37.484177	0
13	Cλ1	chr16	19050728	19070728	HS7	18996954	19016954	16	0.005834	36.899576	-31.722063	0
14	Cλ1	chr16	19051956	19071956	Eλ3-1	19017868	19037868	16	0.007494	28.722966	-31.721601	0
15	HSVλ1	chr16	19078454	19098454	Cλ3	19055901	19075901	15	0.003651	55.275264	-29.006367	0
16	Jλ3	chr16	19054029	19074029	HS7	18997733	19017733	15	0.005034	40.088326	-28.945746	0
17	Jλ3	chr16	19055396	19075396	Eλ3-1	19016057	19036057	15	0.006467	31.205117	-28.945746	0
18	Cλ3	chr16	19054029	19074029	HS7	18997733	19017733	15	0.005264	38.335326	-28.940872	0
19	Cλ3	chr16	19055396	19075396	Eλ3-1	19016057	19036057	15	0.006763	29.840567	-28.940872	0
20	Jλ1	chr16	19052705	19072705	HS7	18997513	19017513	15	0.005301	38.072456	-28.940141	0
21	Vλ1	chr16	19076679	19096679	Cλ1	19051319	19071319	14	0.004046	46.554438	-26.29766	0
22	HSVλ1	chr16	19078122	19098122	Cλ1	19052383	19072383	14	0.004046	46.554438	-26.269448	0
23	HSVλ1	chr16	19078746	19098746	Jλ3	19056545	19076545	14	0.003491	53.949367	-26.269448	0
24	Jλ1	chr16	19054771	19074771	Eλ3-1	19016119	19036119	14	0.006809	27.660216	-26.212701	0
25	Vλ1	chr16	19077282	19097282	Cλ3	19055099	19075099	13	0.003651	47.905229	-23.62664	0

Interaction	Region(1)	Chr	Start	End	Region(2)	Start	End(2)	Interaction reads	Expected reads	Z-score	LogP	FDR
26	HSVλ1	chr16	19078174	19098174	Jλ1	19054591	19074591	13	0.003676	47.576736	-23.60278	0
27	Vλ1	chr16	19076753	19096753	HS6	18996279	19016279	12	0.004714	34.248987	-21.031201	0
28	Vλ1	chr16	19077532	19097532	Jλ1	19054299	19074299	12	0.003676	43.916987	-21.031201	0
29	Vλ1	chr16	19077525	19097525	Jλ3	19055783	19075783	12	0.003491	46.242315	-21.031201	0
30	HSVλ1	chr16	19076753	19096753	HS6	18996279	19016279	12	0.004714	34.248987	-21.011323	0
31	Vλ1	chr16	19077136	19097136	HS7	18993672	19013672	11	0.004029	36.731056	-18.517069	0
32	HSVλ1	chr16	19077136	19097136	HS7	18993672	19013672	11	0.004029	36.731056	-18.500804	0
33	Vλ1	chr16	19076150	19096150	HSCλ1	19044240	19064240	10	0.004306	31.241759	-16.090929	0
34	HSVλ1	chr16	19077314	19097314	HSCλ1	19044354	19064354	9	0.004306	28.117583	-13.750543	0.000002
35	Vλ1	chr16	19076049	19096049	Eλ3-1	19017203	19037203	8	0.005176	20.794028	-11.535803	0.000014
36	HSVλ1	chr16	19076049	19096049	Eλ3-1	19017203	19037203	8	0.005176	20.794028	-11.528181	0.000014

References

Abarrategui, I., and Krangel, M.S. (2006). Regulation of T cell receptor-alpha gene recombination by transcription. Nat. Immunol. *7*, 1109–1115.

Adams, B., Dörfler, P., Aguzzi, A., Kozmik, Z., Urbánek, P., Maurer-Fogy, I., and Busslinger, M. (1992). Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. Genes Dev. *6*, 1589–1607.

Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., and Thanos, D. (2000). Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter. Cell *103*, 667–678.

Agrawal, A., and Schatz, D.G. (1997). RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination. Cell *89*, 43–53.

Akashi, K., Kondo, M., Cheshier, S., Shizuru, J., Gandy, K., Domen, J., Mebius, R., Traver, D., And Weissman, I.L. (1999). Lymphoid Development from Stem Cells and the Common Lymphocyte Progenitors. Cold Spring Harb. Symp. Quant. Biol. *64*, 1–12.

Alarid, E.T., Bakopoulos, N., and Solodin, N. (1999). Proteasome-Mediated Proteolysis of Estrogen Receptor: A Novel Component in Autologous Down-Regulation. Mol. Endocrinol. *13*, 1522–1534.

Allahyar, A., Vermeulen, C., Bouwman, B.A.M., Krijger, P.H.L., Verstegen, M.J.A.M., Geeven, G., van Kranenburg, M., Pieterse, M., Straver, R., Haarhuis, J.H.I., et al. (2018). Enhancer hubs and loop collisions identified from single-allele topologies. Nat. Genet. *50*, 1151–1160.

de Almeida, C.R., Hendriks, R.W., and Stadhouders, R. (2015). Dynamic Control of Long-Range Genomic Interactions at the Immunoglobulin κ Light-Chain Locus. In Advances in Immunology, p.

Alt, F.W., Zhang, Y., Meng, F.-L., Guo, C., and Schwer, B. (2013). Mechanisms of programmed DNA lesions and genomic instability in the immune system. Cell *152*, 417–429.

Amin, R.H., and Schlissel, M.S. (2008). Foxo1 directly regulates the transcription of recombination-activating genes during B cell development. Nat. Immunol. *9*, 613–622.

Arnal, S.M., and Roth, D.B. (2007). Excised V(D)J recombination byproducts threaten genomic integrity. Trends Immunol. *28*, 289–292.

Atchison, M.L., and Perry, R.P. (1987). The role of the kappa enhancer and its binding factor NF-kappa B in the developmental regulation of kappa gene transcription. Cell *48*, 121–128.

Baarends, W.M., Hoogerbrugge, J.W., Roest, H.P., Ooms, M., Vreeburg, J., Hoeijmakers, J.H.., and Grootegoed, J.A. (1999). Histone Ubiquitination and Chromatin Remodeling in Mouse Spermatogenesis. Dev. Biol. *207*, 322–333.

Bailey, T.L., and Elkan, C. (1995). The value of prior knowledge in discovering motifs with MEME. Proceedings. Int. Conf. Intell. Syst. Mol. Biol. *3*, 21–29.

Banaszynski, L.A., Chen, L., Maynard-Smith, L.A., Ooi, A.G.L., and Wandless, T.J. (2006). A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules. Cell *126*, 995–1004.

Banerjee, A., and Rothman, P. (1998). IL-7 reconstitutes multiple aspects of v-Abl-mediated signaling. J. Immunol. *161*, 4611–4617.

Banerjee, J.K., and Schatz, D.G. (2014). Synapsis alters RAG-mediated nicking at Tcrb recombination signal sequences: implications for the “beyond 12/23” rule. Mol. Cell. Biol. *34*, 2566–2580.

Baniahmad, A., Steiner, C., Köhne, A.C., and Renkawitz, R. (1990). Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. Cell *61*, 505–514.

Barberis, A., Widenhorn, K., Vitelli, L., and Busslinger, M. (1990). A novel B-cell lineage-specific transcription factor present at early but not late stages of differentiation. Genes Dev. *4*, 849–859.

Barreto, V., and Cumano, A. (2000). Frequency and characterization of phenotypic Ig heavy chain allelically included IgM-expressing B cells in mice. J. Immunol. *164*, 893–899.

Baumann, B., Kistler, B., Kirillov, A., Bergman, Y., and Wirth, T. (1998). The mutant plasmacytoma cell line S107 allows the identification of distinct pathways leading to NF-kappaB activation. J. Biol. Chem. *273*, 11448–11455.

Beck, K., Peak, M.M., Ota, T., Nemazee, D., and Murre, C. (2009). Distinct roles for E12 and E47 in B cell specification and the sequential rearrangement of immunoglobulin light chain loci. J. Exp. Med. *206*, 2271–2284.

Bednar, J., Horowitz, R.A., Grigoryev, S.A., Carruthers, L.M., Hansen, J.C., Koster, A.J., and Woodcock, C.L. (1998). Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. Proc. Natl. Acad. Sci. *95*, 14173–14178.

Begum, N.A., Stanlie, A., Nakata, M., Akiyama, H., and Honjo, T. (2012). The histone chaperone Spt6 is required for activation-induced cytidine deaminase target determination through H3K4me3 regulation. J. Biol. Chem. *287*, 32415–32429.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482–485.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. Cell *98*, 387–396.

Belle, J.I., and Nijnik, A. (2014). H2A-DUBbing the mammalian epigenome: Expanding frontiers for histone H2A deubiquitinating enzymes in cell biology and physiology. Int. J. Biochem. Cell Biol. *50*, 161–174.

Benner, C., Isoda, T., and Murre, C. (2015). New roles for DNA cytosine modification, eRNA, anchors, and superanchors in developing B cell progenitors. Proc. Natl. Acad. Sci. *112*, 12776–12781.

Berek, C., and Ziegner, M. (1993). The maturation of the immune response. Immunol. Today *14*, 400–404.

Berek, C., Berger, A., and Apel, M. (1991). Maturation of the immune response in germinal centers. Cell *67*, 1121–1129.

Bertolino, E., Reddy, K., Medina, K.L., Parganas, E., Ihle, J., and Singh, H. (2005). Regulation of interleukin 7–dependent immunoglobulin heavy-chain variable gene rearrangements by transcription factor STAT5. Nat. Immunol. *6*, 836–843.

Bevington, S. (2009). The mechanism of enhancer mediated long range chromatin activation during V(D)J recombination.

Bevington, S., and Boyes, J. (2013). Transcription-coupled eviction of histones H2A/H2B governs V(D)J recombination. EMBO J. *32*, 1381–1392.

Bilokapic, S., Strauss, M., and Halic, M. (2018). Cryo-EM of nucleosome core particle interactions in trans. Sci. Rep. *8*, 7046.

Blackburn, E.H. (2005). Telomeres and telomerase: their mechanisms of action and the effects of altering their functions. FEBS Lett. *579*, 859–862.

Bolland, D.J., Wood, A.L., Johnston, C.M., Bunting, S.F., Morgan, G., Chakalova, L., Fraser, P.J., and Corcoran, A.E. (2004). Antisense intergenic transcription in V(D)J recombination. Nat. Immunol. *5*, 630–637.

Bolland, D.J., Wood, A.L., Afshar, R., Featherstone, K., Oltz, E.M., and Corcoran, A.E. (2007). Antisense Intergenic Transcription Precedes Igh D-to-J Recombination and Is Controlled by the Intronic Enhancer E. Mol. Cell. Biol. *27*, 5523–5533.

Boller, S., Li, R., and Grosschedl, R. (2018). Defining B Cell Chromatin: Lessons from EBF1. Trends Genet. *34*, 257–269.

Boudinot, P., Drapier, A.M., Cazenave, P.A., and Sanchez, P. (1994). Conserved distribution of lambda subtypes from rearranged gene segments to immunoglobulin synthesis in the mouse B cell repertoire. Eur. J. Immunol. *24*, 2013–2017.

Brandt, V.L., and Roth, D.B. (2009). Recent insights into the formation of RAG-induced chromosomal translocations. Adv. Exp. Med. Biol. *650*, 32–45.

Brass, A.L., Kehrli, E., Eisenbeis, C.F., Storb, U., and Singh, H. (1996). Pip, a lymphoid-restricted IRF, contains a regulatory domain that is important for autoinhibition and ternary complex formation with the Ets factor PU.1. Genes Dev. *10*, 2335–2347.

Brownlie, R.J., and Zamoyska, R. (2013). T cell receptor signalling networks: branched, diversified and bounded. Nat. Rev. Immunol. *13*, 257–269.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Calero-Nieto, F.J., Bert, A.G., and Cockerill, P.N. (2010). Transcription-dependent silencing of inducible convergent transgenes in transgenic mice. Epigenetics Chromatin *3*, 3.

Campisi, J. (1997). The biology of replicative senescence. Eur. J. Cancer *33*, 703–709.

Cao, J., and Yan, Q. (2012). Histone Ubiquitination and Deubiquitination in Transcription, DNA Damage Response, and Cancer. Front. Oncol. *2*, 26.

Chen, Y.Y., and Rosenberg, N. (1992). Lymphoid cells transformed by Abelson virus require the v-abl protein-tyrosine kinase only during early G1. Proc. Natl. Acad. Sci. U. S. A. *89*, 6683–6687.

Chen, J., Young, F., Bottaro, A., Stewart, V., Smith, R.K., and Alt, F.W. (1993). Mutations of the intronic IgH enhancer and its flanking sequences differentially affect accessibility of the JH locus. EMBO J. *12*, 4635–4645.

Chen, J., Zhang, Z., Li, L., Chen, B.-C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., et al. (2014). Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells. Cell *156*, 1274–1285.

Chen, L., Carico, Z., Shih, H.-Y., and Krangel, M.S. (2015). A discrete chromatin loop in the mouse Tcra-Tcrd locus shapes the TCRδ and TCRα repertoires. Nat. Immunol. *16*, 1085–1093.

Chen, L., Zhao, L., Alt, F.W., and Krangel, M.S. (2016). An Ectopic CTCF Binding Element Inhibits Tcrd Rearrangement by Limiting Contact between Vδ and Dδ Gene Segments. J. Immunol. *197*, 3188–3197.

Chen, Y.Y., Wang, L.C., Huang, M.S., and Rosenberg, N. (1994). An active v-abl protein tyrosine kinase blocks immunoglobulin light-chain gene rearrangement. Genes Dev. *8*, 688–697.

Chymkowitch, P., Le May, N., Charneau, P., Compe, E., and Egly, J.-M. (2011). The phosphorylation of the androgen receptor by TFIIH directs the ubiquitin/proteasome process. EMBO J. *30*, 468–479.

Ciubotaru, M., Trexler, A.J., Spiridon, L.N., Surleac, M.D., Rhoades, E., Petrescu, A.J., and Schatz, D.G. (2013). RAG and HMGB1 create a large bend in the 23RSS in the V(D)J recombination synaptic complexes. Nucleic Acids Res. *41*, 2437–2454.

Clark, M.R., Mandal, M., Ochiai, K., and Singh, H. (2014). Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling. Nat. Rev. Immunol. *14*, 69–80.

Clark, S.J., Lee, H.J., Smallwood, S.A., Kelsey, G., and Reik, W. (2016). Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. Genome Biol. *17*, 72.

Cleary, M.L., Smith, S.D., and Sklar, J. (1986). Cloning and structural analysis of cDNAs for bcl-2 and a hybrid bcl-2/immunoglobulin transcript resulting from the t(14;18) translocation. Cell *47*, 19–28.

Cogné, M., Lansford, R., Bottaro, A., Zhang, J., Gorman, J., Young, F., Cheng, H.-L., and Alt, F.W. (1994). A class switch control region at the 3′ end of the immunoglobulin heavy chain locus. Cell *77*, 737–747.

Coleclough, C., Perry, R.P., Karjalainen, K., and Weigert, M. (1981). Aberrant rearrangements contribute significantly to the allelic exclusion of immunoglobulin gene expression. Nature *290*, 372–378.

Corfe, S. a, Gray, A.P., and Paige, C.J. (2007). Generation and characterization of stromal cell independent IL-7 dependent B cell lines. J. Immunol. Methods *325*, 9–19.

Cote, J., Peterson, C.L., Workman, J.L., Chambon, P., Lopez, D., and Hagman, J. (1998). Perturbation of nucleosome core structure by the SWI/SNF complex persists after its detachment, enhancing subsequent transcription factor binding. Proc. Natl. Acad. Sci. *95*, 4947–4952.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Dallmeier, K., and Neyts, J. (2013). Simple and inexpensive three-step rapid amplification of cDNA 5' ends using 5' phosphorylated primers. Anal. Biochem. *434*, 1–3.

Danial, N.N., Pernis, A., and Rothman, P.B. (1995). Jak-STAT signaling induced by the v-abl oncogene. Science *269*, 1875–1877.

Danielian, P.S., White, R., Hoare, S.A., Fawell, S.E., and Parker, M.G. (1993). Identification of residues in the estrogen receptor that confer differential sensitivity to estrogen and hydroxytamoxifen. Mol. Endocrinol. *7*, 232–240.

Davies, J.O.J., Telenius, J.M., McGowan, S.J., Roberts, N.A., Taylor, S., Higgs, D.R., and Hughes, J.R. (2016). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nat. Methods *13*, 74–80.

de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. Mol. Cell *60*, 676–684.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing Chromosome Conformation. Science (80-.). *295*, 1306–1311.

Deng, Z., Liu, H., and Liu, X. (2015). RAG1-mediated ubiquitylation of histone H3 is required for chromosomal V(D)J recombination. Cell Res. *25*, 181–192.

Deniaud, E., and Bickmore, W.A. (2009). Transcription and the nuclear periphery: edge of darkness? Curr. Opin. Genet. Dev. *19*, 187–191.

Denker, A., and de Laat, W. (2016). The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev. *30*, 1357–1382.

Difilippantonio, M.J., McMahan, C.J., Eastman, Q.M., Spanopoulou, E., and Schatz, D.G. (1996). RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination. Cell *87*, 253–262.

Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T.J. (2003). Chromatin fiber folding: requirement for the histone H4 N-terminal tail. J. Mol. Biol. *327*, 85–96.

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., and Richmond, T.J. (2004). Nucleosome Arrays Reveal the Two-Start Organization of the Chromatin Fiber. Science (80-.). *306*, 1571–1573.

Drejer-Teel, A.H., Fugmann, S.D., and Schatz, D.G. (2007). The beyond 12/23 restriction is imposed at the nicking and pairing steps of DNA cleavage during V(D)J recombination. Mol. Cell. Biol. *27*, 6288–6299.

Duan, H., Xiang, H., Ma, L., and Boxer, L.M. (2008). Functional long-range interactions of the IgH 3' enhancers with the bcl-2 promoter region in t(14;18) lymphoma cells. Oncogene *27*, 6720–6728.

Eberharter, A., and Becker, P.B. (2002). Histone acetylation: a switch between repressive and permissive chromatin. EMBO Rep. *3*, 224–229.

Ebert, A., McManus, S., Tagoh, H., Medvedovic, J., Salvagiotto, G., Novatchkova, M., Tamir, I., Sommer, A., Jaritz, M., and Busslinger, M. (2011). The Distal VH Gene Cluster of the Igh Locus Contains Distinct Regulatory Elements with Pax5 Transcription Factor-Dependent Activity in Pro-B Cells. Immunity *34*, 175–187.

Eisen, H.N., and Reilly, E.B. (1985). Lambda chains and genes in inbred mice. Annu. Rev. Immunol. *3*, 337–365.

Eisenbeis, C.F., Singh, H., and Storb, U. (1993). PU.1 is a component of a multiprotein complex which binds an essential site in the murine immunoglobulin lambda 2-4 enhancer. Mol. Cell. Biol. *13*, 6452–6461.

Eisenbeis, C.F., Singh, H., and Storb, U. (1995). Pip, a novel IRF family member, is a lymphoid-specific, PU.1-dependent transcriptional activator. Genes Dev. *9*, 1377–1387.

Eng, F.C.S., Lee, H.S., Ferrara, J., Willson, T.M., and White, J.H. (1997). Probing the Structure and Function of the Estrogen Receptor Ligand Binding Domain by Analysis of Mutants with Altered Transactivation Characteristics. *17*, 4644–4653.

Engelke, U., Wang, D.-M., and Lipsick, J.S. (1997). Cells Transformed by a v-Myb-Estrogen Receptor Fusion Differentiate into Multinucleated Giant Cells. J. Virol. *71*, 3760–3766.

Farago, M., Rosenbluh, C., Tevlin, M., Fraenkel, S., Schlesinger, S., Masika, H., Gouzman, M., Teng, G., Schatz, D., Rais, Y., et al. (2012). Clonal allelic predetermination of immunoglobulin-κ rearrangement. Nature *490*, 561–565.

Feil, R., Wagner, J., Metzger, D., and Chambon, P. (1997). Regulation of Cre Recombinase Activity by Mutated Estrogen Receptor Ligand-Binding Domains. Biochem. Biophys. Res. Commun. *237*, 752–757.

Ferreirós-Vidal, I., Carroll, T., Taylor, B., Terry, A., Liang, Z., Bruno, L., Dharmalingam, G., Khadayate, S., Cobb, B.S., Smale, S.T., et al. (2013). Genome-wide identification of Ikaros targets elucidates its contribution to mouse B-cell lineage specification and pre-B-cell differentiation. Blood *121*, 1769–1782.

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc. Natl. Acad. Sci. U. S. A. *73*, 1897–1901.

Fisher, M.R., Rivera-Reyes, A., Bloch, N.B., Schatz, D.G., and Bassing, C.H. (2017). Immature Lymphocytes Inhibit Rag1 and Rag2 Transcription and V(D)J Recombination in Response to DNA Double-Strand Breaks. J. Immunol. *198*, 2943–2956.

Fitzsimmons, S.P., Bernstein, R.M., Max, E.E., Skok, J.A., and Shapiro, M.A. (2007). Dynamic changes in accessibility, nuclear positioning, recombination, and transcription at the Ig kappa locus. J. Immunol. *179*, 5264–5273.

Forcato, M., Nicoletti, C., Pal, K., Livi, C.M., Ferrari, F., and Bicciato, S. (2017). Comparison of computational methods for Hi-C data analysis. Nat. Methods *14*, 679–685.

Fugmann, S.D., Villey, I.J., Ptaszek, L.M., and Schatz, D.G. (2000). Identification of two catalytic residues in RAG1 that define a single active site within the RAG1/RAG2 protein complex. Mol. Cell *5*, 97–107.

Furlan-Magaril, M., Rebollar, E., Guerrero, G., Fernández, A., Moltó, E., González-Buendía, E., Cantero, M., Montoliu, L., and Recillas-Targa, F. (2011). An insulator embedded in the chicken α-globin locus regulates chromatin domain configuration and differential gene expression. Nucleic Acids Res. *39*, 89–103.

Fuxa, M., and Busslinger, M. (2007). Reporter gene insertions reveal a strictly B lymphoid-specific expression pattern of Pax5 in support of its B cell identity function. J. Immunol. *178*, 3031–3037.

Fuxa, M., Skok, J., Souabni, A., Salvagiotto, G., Roldan, E., and Busslinger, M. (2004). Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. Genes Dev. *18*, 411–422.

Garbe, J., Wong, M., Wigington, D., Yaswen, P., and Stampfer, M.R. (1999). Viral oncogenes accelerate conversion to immortality of cultured conditionally immortal human mammary epithelial cells. Oncogene *18*, 2169–2180.

Garcia, I.S., Kaneko, Y., Gonzalez-Sarmiento, R., Campbell, K., White, L., Boehm, T., and Rabbitts, T.H. (1991). A study of chromosome 11p13 translocations involving TCR beta and TCR delta in human T cell leukaemia. Oncogene *6*, 577–582.

Garrick, D., Sutherland, H., Robertson, G., and Whitelaw, E. (1996). Variegated expression of a globin transgene correlates with chromatin accessibility but not methylation status. Nucleic Acids Res. *24*, 4902–4909.

Garrick, D., Fiering, S., Martin, D.I.K., and Whitelaw, E. (1998). Repeat-induced gene silencing in mammals. Nat. Genet. *18*, 56–59.

Gebert, C., Correia, L., Li, Z., Petrie, H.T., Love, P.E., and Pfeifer, K. (2017). Chromosome choice for initiation of V-(D)-J recombination is not governed by genomic imprinting. Immunol. Cell Biol. *95*, 473–477.

Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. Annu. Rev. Biochem. *71*, 101–132.

van Gent, D.C., Hiom, K., Paull, T.T., and Gellert, M. (1997). Stimulation of V(D)J cleavage by high mobility group proteins. EMBO J. *16*, 2665–2670.

Georgopoulos, K. (2002). Haematopoietic cell-fate decisions, chromatin regulation and ikaros. Nat. Rev. Immunol. *2*, 162–174.

Gillies, S.D., Morrison, S.L., Oi, V.T., and Tonegawa, S. (1983). A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell *33*, 717–728.

Goebel, P., Janney, N., Valenzuela, J.R., Romanow, W.J., Murre, C., and Feeney, A.J. (2001). Localized gene-specific induction of accessibility to V(D)J recombination induced by E2A and early B cell factor in nonlymphoid cells. J. Exp. Med. *194*, 645–656.

Goff, S.P., Gilboa, E., Witte, O.N., and Baltimore, D. (1980). Structure of the Abelson murine leukemia virus genome and the homologous cellular gene: studies with cloned viral DNA. Cell *22*, 777–785.

Golding, A., Chandler, S., Ballestar, E., Wolffe, A.P., and Schlissel, M.S. (1999). Nucleosome structure completely inhibits in vitro cleavage by the V(D)J recombinase. EMBO J. *18*, 3712–3723.

Goldmit, M., Schlissel, M., Cedar, H., and Bergman, Y. (2002). Differential accessibility at the kappa chain locus plays a role in allelic exclusion. EMBO J. *21*, 5255–5261.

Gordon, S., Akopyan, G., Garban, H., and Bonavida, B. (2006). Transcription factor YY1: structure, function and therapeutic implications in cancer biology. Oncogene *25*, 1125–1142.

Gorman, J.R., van der Stoep, N., Monroe, R., Cogne, M., Davidson, L., and Alt, F.W. (1996). The Ig(kappa) enhancer influences the ratio of Ig(kappa) versus Ig(lambda) B lymphocytes. Immunity *5*, 241–252.

Grange, S., and Boyes, J. (2007). Chromatin opening is tightly linked to enhancer activation at the κ light chain locus. Biochem. Biophys. Res. Commun. *363*, 223–228.

Grawunder, U., Leu, T.M.J., Schatz, D.G., Werner, A., Rolink, A.G., Melchers, F., and Winkler, T.H. (1995). Down-regulation of RAG1 and RAG2 gene expression in PreB cells after functional immunoglobulin heavy chain rearrangement. Immunity *3*, 601–608.

Grazini, U., Zanardi, F., Citterio, E., Casola, S., Goding, C.R., and McBlane, F. (2010). The RING Domain of RAG1 Ubiquitylates Histone H3: A Novel Activity in Chromatin-Mediated Regulation of V(D)J Joining. Mol. Cell *37*, 282–293.

Greaves, M.F., and Wiemels, J. (2003). Origins of chromosome translocations in childhood leukaemia. Nat. Rev. Cancer *3*, 639–649.

Griffiths, S.J., Koegl, M., Boutell, C., Zenner, H.L., Crump, C.M., Pica, F., Gonzalez, O., Friedel, C.C., Barry, G., Martin, K., et al. (2013). A Systematic Analysis of Host Factors Reveals a Med23-Interferon-λ Regulatory Axis against Herpes Simplex Virus Type 1 Replication. PLoS Pathog. *9*, e1003514.

Grignani, F., Kinsella, T., Mencarelli, A., Valtieri, M., Riganelli, D., Grignani, F., Lanfrancone, L., Peschle, C., Nolan, G.P., and Pelicci, P.G. (1998). High-efficiency gene transfer and selection of human hematopoietic progenitor cells with a hybrid EBV/retroviral vector expressing the green fluorescence protein. Cancer Res. *58*, 14–19.

Grundy, G.J., Yang, W., and Gellert, M. (2010). Autoinhibition of DNA cleavage mediated by RAG1 and RAG2 is overcome by an epigenetic signal in V(D)J recombination. Proc. Natl. Acad. Sci. *107*, 22487–22492.

Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.-L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., et al. (2011a). CTCF-binding elements mediate control of V(D)J recombination. Nature *477*, 424–430.

Guo, C., Gerasimova, T., Hao, H., Ivanova, I., Chakraborty, T., Selimyan, R., Oltz, E.M., and Sen, R. (2011b). Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. Cell *147*, 332–343.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900–910.

Hagman, J., Rudin, C.M., Haasch, D., Chaplin, D., and Storb, U. (1990). A novel enhancer in the immunoglobulin lambda locus is duplicated and functionally independent of NF kappa B. Genes Dev. *4*, 978–992.

Hansen, A.S., Pustova, I., Cattoglio, C., Tjian, R., and Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. Elife *6*.

Haque, S.F.Y., Bevington, S.L., and Boyes, J. (2013). The Eλ(3-1) enhancer is essential for V(D)J recombination of the murine immunoglobulin lambda light chain locus. Biochem. Biophys. Res. Commun. *441*, 482–487.

Hardy, R.R., Carmack, C.E., Shinton, S.A., Kemp, J.D., and Hayakawa, K. (1991). Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. J. Exp. Med. *173*, 1213–1225.

Hawley, R.G. (1994). High-Titer Retroviral Vectors for Efficient Transduction of Functional Genes into Murine Hematopoietic Stem Cells. Ann. N. Y. Acad. Sci. *716*, 327–330.

Hayflick, L. (1965). The limited in vitro lifetime of human diploid cell strains. Exp. Cell Res. *37*, 614–636.

Hayflick, L., and Moorhead, P.S. (1961). The serial cultivation of human diploid cell strains. Exp. Cell Res. *25*, 585–621.

Hayzer, D.J. (1990). Immunoglobulin lambda light chain evolution: Igl and Igl-like sequences form three major groups. Immunogenetics *32*, 157–174.

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

Heizmann, B., Kastner, P., and Chan, S. (2013). Ikaros is absolutely required for pre-B cell differentiation by attenuating IL-7 signals. J. Exp. Med. *210*, 2823–2832.

Henikoff, S. (1998). Conspiracy of silence among repeated transgenes. BioEssays *20*, 532–535.

Herglotz, J., Unrau, L., Hauschildt, F., Fischer, M., Kriebitzsch, N., Alawi, M., Indenbirken, D., Spohn, M., Muller, U., Ziegler, M., et al. (2015). Essential control of early B-cell development by Mef2 transcription factors. Blood.

Hesslein, D.G.T., Pflugh, D.L., Chowdhury, D., Bothwell, A.L.M., Sen, R., and Schatz, D.G. (2003). Pax5 is required for recombination of transcribed, acetylated, 5' IgH V gene segments. Genes Dev. *17*, 37–42.

Hewitt, S.L., Wong, J.B., Lee, J.-H., Nishana, M., Chen, H., Coussens, M., Arnal, S.M., Blumenberg, L.M., Roth, D.B., Paull, T.T., et al. (2017). The Conserved ATM Kinase RAG2-S365 Phosphorylation Site Limits Cleavage Events in Individual Cells Independent of Any Repair Defect. Cell Rep. *21*, 979–993.

Hiom, K., and Gellert, M. (1998). Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. Mol. Cell *1*, 1011–1019.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hodawadekar, S., Park, K., Farrar, M.A., and Atchison, M.L. (2012). A Developmentally Controlled Competitive STAT5-PU.1 DNA Binding Mechanism Regulates Activity of the Ig E3' Enhancer. J. Immunol. *188*, 2276–2284.

Horowitz, R.A., Agard, D.A., Sedat, J.W., and Woodcock, C.L. (1994). The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. J. Cell Biol. *125*, 1–10.

Hsu, L.-Y., Lauring, J., Liang, H.-E., Greenbaum, S., Cado, D., Zhuang, Y., and Schlissel, M.S. (2003). A conserved transcriptional enhancer regulates RAG gene expression in developing B cells. Immunity *19*, 105–117.

Hu, J., Zhang, Y., Zhao, L., Frock, R.L., Du, Z., Meyers, R.M., Meng, F., Schatz, D.G., and Alt, F.W. (2015). Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. Cell *163*, 947–959.

Huang, S., Tao, X., Yuan, S., Zhang, Y., Li, P., Beilinson, H.A., Zhang, Y., Yu, W., Pontarotti, P., Escriva, H., et al. (2016). Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. Cell *166*, 102–114.

Hughes, M.M., Tillman, R.E., Wehrly, T.D., White, J.M., and Sleckman, B.P. (2003). The B12/23 restriction is critically dependent on recombination signal nonamer and spacer sequences. J. Immunol. *171*, 6604–6610.

Iacobucci, I., Storlazzi, C.T., Cilloni, D., Lonetti, A., Ottaviani, E., Soverini, S., Astolfi, A., Chiaretti, S., Vitale, A., Messa, F., et al. (2009). Identification and molecular characterization of recurrent genomic deletions on 7p12 in the IKZF1 gene in a large cohort of BCR-ABL1-positive acute lymphoblastic leukemia patients: on behalf of Gruppo Italiano Malattie Ematologiche dell'Adulto Acute Leuke. Blood *114*, 2159–2167.

Imamura, Y., Oda, A., Katahira, T., Bundo, K., Pike, K.A., Ratcliffe, M.J.H., and Kitamura, D. (2009). BLNK binds active H-Ras to promote B cell receptor-mediated capping and ERK activation. J. Biol. Chem. *284*, 9804–9813.

Inlay, M., Alt, F.W., Baltimore, D., and Xu, Y. (2002). Essential roles of the κ light chain intronic enhancer and 3′ enhancer in κ rearrangement and demethylation. Nat. Immunol. *3*, 463–468.

Inlay, M.A., Tian, H., Lin, T., and Xu, Y. (2004). Important roles for E protein binding sites within the immunoglobulin kappa chain intronic enhancer in activating Vkappa Jkappa rearrangement. J. Exp. Med. *200*, 1205–1211.

Inlay, M.A., Lin, T., Gao, H.H., and Xu, Y. (2006). Critical roles of the immunoglobulin intronic enhancers in maintaining the sequential rearrangement of IgH and Igk loci. J. Exp. Med. *203*, 1721–1732.

Ishihara, K., Medina, K., Hayashi, S., Pietrangeli, C., Namen, A.E., Miyake, K., and Kincade, P.W. (1991). Stromal-cell and cytokine-dependent lymphocyte clones which span the pre-B- to B-cell transition. Dev. Immunol. *1*, 149–161.

Iwamoto, M., Björklund, T., Lundberg, C., Kirik, D., and Wandless, T.J. (2010). A General Chemical Method to Regulate Protein Stability in the Mammalian Central Nervous System. Chem. Biol. *17*, 981–988.

Jäger, U., Böcskör, S., Le, T., Mitterbauer, G., Bolz, I., Chott, A., Kneba, M., Mannhalter, C., and Nadel, B. (2000). Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. Blood *95*, 3520–3529.

Jankovic, M., Casellas, R., Yannoutsos, N., Wardemann, H., and Nussenzweig, M.C. (2004). RAGs and regulation of autoantibodies. Annu. Rev. Immunol. *22*, 485–501.

Jankovic, M., Nussenzweig, A., and Nussenzweig, M.C. (2007). Antigen receptor diversification and chromosome translocations. Nat. Immunol. *8*, 801–808.

Johnson, K., Pflugh, D.L., Yu, D., Hesslein, D.G.T., Lin, K.-I., Bothwell, A.L.M., Thomas-Tikhonenko, A., Schatz, D.G., and Calame, K. (2004). B cell–specific loss of histone 3 lysine 9 methylation in the VH locus depends on Pax5. Nat. Immunol. *5*, 853–861.

Johnson, K., Hashimshony, T., Sawai, C.M., Pongubala, J.M.R., Skok, J.A., Aifantis, I., and Singh, H. (2008). Regulation of Immunoglobulin Light-Chain Recombination by the Transcription Factor IRF-4 and the Attenuation of Interleukin-7 Signaling. Immunity *28*, 335–345.

Jones, J.M., and Gellert, M. (2001). Intermediates in V(D)J recombination: a stable RAG1/2 complex sequesters cleaved RSS ends. Proc. Natl. Acad. Sci. U. S. A. *98*, 12926–12931.

Joshi, I., Yoshida, T., Jena, N., Qi, X., Zhang, J., Van Etten, R.A., and Georgopoulos, K. (2014). Loss of Ikaros DNA-binding function confers integrin-dependent survival on pre-B cells and progression to acute lymphoblastic leukemia. Nat. Immunol. *15*, 294–304.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kapitonov, V. V, and Jurka, J. (2005). RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. PLoS Biol. *3*, e181.

Kee, B.L., Rivera, R.R., and Murre, C. (2001). Id3 inhibits B lymphocyte progenitor growth and survival in response to TGF-β. Nat. Immunol. *2*, 242–247.

Khor, B., and Sleckman, B. (2005). Intra- and inter-allelic ordering of T cell receptor ? chain gene assembly. Eur. J. Immunol. *35*, 964–970.

Kiang, D.T., Kollander, R.E., Thomas, T., and Kennedy, B.J. (1989). Up-regulation of estrogen receptors by nonsteroidal antiestrogens in human breast cancer. Cancer Res. *49*, 5312–5316.

Kikuchi, K., Lai, A.Y., Hsu, C.-L., and Kondo, M. (2005). IL-7 receptor signaling is necessary for stage transition in adult B cell development through up-regulation of EBF. J. Exp. Med. *201*, 1197–1203.

Kim, Y.W., and Kim, A. (2017). Deletion of transcription factor binding motifs using the CRISPR/spCas9 system in the β-globin LCR. Biosci. Rep. *37*, BSR20170976.

Kim, D.R., Dai, Y., Mundy, C.L., Yang, W., and Oettinger, M.A. (1999). Mutations of acidic residues in RAG1 define the active site of the V(D)J recombinase. Genes Dev. *13*, 3070–3080.

Kim, M.-S., Lapkouski, M., Yang, W., and Gellert, M. (2015). Crystal structure of the V(D)J recombinase RAG1–RAG2. Nature *518*, 507–511.

Kim, M.-S., Chuenchor, W., Chen, X., Cui, Y., Zhang, X., Zhou, Z.H., Gellert, M., and Yang, W. (2018). Cracking the DNA Code for V(D)J Recombination. Mol. Cell *70*, 358–370.e4.

Kimura, E., Han, J.J., Li, S., Fall, B., Ra, J., Haraguchi, M., Tapscott, S.J., and Chamberlain, J.S. (2008). Cell-lineage regulated myogenesis for dystrophin replacement: a novel therapeutic approach for treatment of muscular dystrophy. Hum. Mol. Genet. *17*, 2507–2517.

Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., de Vries, S.S., Janssen, H., Amendola, M., Nolen, L.D., Bickmore, W.A., and van Steensel, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. Cell *153*, 178–192.

Kireeva, M.L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M., and Studitsky, V.M. (2002). Nucleosome Remodeling Induced by RNA Polymerase II. Mol. Cell *9*, 541–552.

Kirsch, I.R., Morton, C.C., Nakahara, K., and Leder, P. (1982). Human immunoglobulin heavy chain genes map to a region of translocations in malignant B lymphocytes. Science (80-.). *216*, 301–303.

Kishi, H., Wei, X.C., Jin, Z.X., Fujishiro, Y., Nagata, T., Matsuda, T., and Muraguchi, A. (2000). Lineage-specific regulation of the murine RAG-2 promoter: GATA-3 in T cells and Pax-5 in B cells. Blood *95*, 3845–3852.

Kitamura, D., Roes, J., Kühn, R., and Rajewsky, K. (1991). A B cell-deficient mouse by targeted disruption of the membrane exon of the immunoglobulin μ chain gene. Nature *350*, 423–426.

Knoblauch, R., and Garabedian, M.J. (1999). Role for Hsp90-associated cochaperone p23 in estrogen receptor signal transduction. Mol. Cell. Biol. *19*, 3748–3759.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Le Beau, M.M., Fisher, A.G., and Singh, H. (2002). Subnuclear Compartmentalization of Immunoglobulin Loci During Lymphocyte Development. Science (80-.). *296*, 158–162.

Kulaeva, O.I., Gaykalova, D.A., Pestov, N.A., Golovastov, V. V, Vassylyev, D.G., Artsimovitch, I., and Studitsky, V.M. (2009). Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. Nat. Struct. Mol. Biol. *16*, 1272–1278.

Kwon, J., Imbalzano, A.N., Matthews, A., and Oettinger, M.A. (1998). Accessibility of nucleosomal DNA to V(D)J cleavage is modulated by RSS positioning and HMG1. Mol. Cell *2*, 829–839.

Kwon, J., Morshead, K.B., Guyon, J.R., Kingston, R.E., and Oettinger, M.A. (2000). Histone acetylation and hSWI/SNF remodeling act in concert to stimulate V(D)J cleavage of nucleosomal DNA. Mol. Cell *6*, 1037–1048.

Landree, M.A., Wibbenmeyer, J.A., and Roth, D.B. (1999). Mutational analysis of RAG1 and RAG2 identifies three catalytic amino acids in RAG1 critical for both cleavage steps of V(D)J recombination. Genes Dev. *13*, 3059–3069.

Lauring, J., and Schlissel, M.S. (1999). Distinct factors regulate the murine RAG-2 promoter in B- and T-cell lines. Mol. Cell. Biol. *19*, 2601–2612.

Lazorchak, A.S., Schlissel, M.S., and Zhuang, Y. (2006). E2A and IRF-4/Pip promote chromatin modification and transcription of the immunoglobulin kappa locus in pre-B cells. Mol. Cell. Biol. *26*, 810–821.

Lee, J., and Desiderio, S. (1999). Cyclin A/CDK2 Regulates V(D)J Recombination by Coordinating RAG-2 Accumulation and DNA Repair. Immunity *11*, 771–781.

Lee, C.M., Cradick, T.J., and Bao, G. (2016). The Neisseria meningitidis CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells. Mol. Ther. *24*, 645–654.

Lee, J.W., Ryan, F., Swaffield, J.C., Johnston, S.A., and Moore, D.D. (1995). Interaction of thyroid-hormone receptor with a conserved transcriptional mediator. Nature *374*, 91–94.

Lescale, C., Abramowski, V., Bedora-Faure, M., Murigneux, V., Vera, G., Roth, D.B., Revy, P., de Villartay, J.-P., and Deriano, L. (2016). RAG2 and XLF/Cernunnos interplay reveals a novel role for the RAG complex in DNA repair. Nat. Commun. *7*, 10529.

Levin-Klein, R., Fraenkel, S., Lichtenstein, M., Matheson, L.S., Bartok, O., Nevo, Y., Kadener, S., Corcoran, A.E., Cedar, H., and Bergman, Y. (2017). Clonally stable Vκ allelic choice instructs Igκ repertoire. Nat. Commun. *8*, 15575.

Lewis, S.M. (1994). The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. Adv. Immunol. *56*, 27–150.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707–719.

Li, Z., Dordai, D.I., Lee, J., and Desiderio, S. (1996). A conserved degradation signal regulates RAG-2 accumulation during cell division and links V(D)J recombination to the cell cycle. Immunity *5*, 575–589.

Liang, H.-E., Hsu, L.-Y., Cado, D., and Schlissel, M.S. (2004). Variegated Transcriptional Activation of the Immunoglobulin κ Locus in Pre-B Cells Contributes to the Allelic Exclusion of Light-Chain Expression. Cell *118*, 19–29.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu. Rev. Biochem. *79*, 181–211.

Lieber, M.R. (2016). Mechanisms of human lymphoid chromosomal translocations. Nat. Rev. Cancer *16*, 387–398.

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

Lin, S.G., Guo, C., Su, A., Zhang, Y., and Alt, F.W. (2015). CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V(D)J recombination. Proc. Natl. Acad. Sci. U. S. A. *112*, 1815–1820.

Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat. Immunol. *13*, 1196–1204.

Liu, H., Schmidt-Supprian, M., Shi, Y.Y., Hobeika, E., Barteneva, N., Jumaa, H., Pelanda, R., Reth, M., Skok, J., Rajewsky, K., et al. (2007). Yin Yang 1 is a critical regulator of B-cell development. Genes Dev. *21*, 1179–1189.

Liu, Z.-M., George-Raizen, J.B., Li, S., Meyers, K.C., Chang, M.Y., and Garrard, W.T. (2002). Chromatin structural analyses of the mouse Igkappa gene locus reveal new hypersensitive sites specifying a transcriptional silencer and enhancer. J. Biol. Chem. *277*, 32640–32649.

Liu, Z., Widlak, P., Zou, Y., Xiao, F., Oh, M., Li, S., Chang, M.Y., Shay, J.W., and Garrard, W.T. (2006). A Recombination Silencer that Specifies Heterochromatin Positioning and Ikaros Association in the Immunoglobulin κ Locus. Immunity *24*, 405–415.

Lobanenkov, V. V, Nicolas, R.H., Adler, V. V, Paterson, H., Klenova, E.M., Polotskaja, A. V, and Goodwin, G.H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. Oncogene *5*, 1743–1753.

Local, A., Huang, H., Albuquerque, C.P., Singh, N., Lee, A.Y., Wang, W., Wang, C., Hsia, J.E., Shiau, A.K., Ge, K., et al. (2018). Identification of H3K4me1-associated proteins at mammalian enhancers. Nat. Genet. *50*, 73–82.

Loguercio, S., Barajas-Mora, E.M., Shih, H.-Y., Krangel, M.S., and Feeney, A.J. (2018). Variable Extent of Lineage-Specificity and Developmental Stage-Specificity of Cohesin and CCCTC-Binding Factor Binding Within the Immunoglobulin and T Cell Receptor Loci. Front. Immunol. *9*, 425.

Lonard, D.M., and Smith, C.L. (2002). Molecular perspectives on selective estrogen receptor modulators (SERMs): progress in understanding their tissue-specific agonist and antagonist actions. Steroids *67*, 15–24.

Lonard, D.M., Tsai, S.Y., and O'Malley, B.W. (2004). Selective estrogen receptor modulators 4-hydroxytamoxifen and raloxifene impact the stability and function of SRC-1 and SRC-3 coactivator proteins. Mol. Cell. Biol. *24*, 14–24.

Lu, R. (2008). Interferon regulatory factor 4 and 8 in B-cell development. Trends Immunol. *29*, 487–492.

Lu, R., Medina, K.L., Lancki, D.W., and Singh, H. (2003). IRF-4,8 orchestrate the pre-B-to-B transition in lymphocyte development. Genes Dev. *17*, 1703–1708.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251–260.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nat. Rev. Mol. Cell Biol. *13*, 436–447.

Ma, S., Turetsky, A., Trinh, L., and Lu, R. (2006). IFN regulatory factor 4 and 8 promote Ig light chain kappa locus activation in pre-B cell development. J. Immunol. *177*, 7898–7904.

Ma, S., Pathak, S., Trinh, L., and Lu, R. (2008). Interferon regulatory factors 4 and 8 induce the expression of Ikaros and Aiolos to down-regulate pre-B-cell receptor and promote cell-cycle withdrawal in pre-B-cell development. Blood *111*, 1396–1403.

Maji, B., Moore, C.L., Zetsche, B., Volz, S.E., Zhang, F., Shoulders, M.D., and Choudhary, A. (2017). Multidimensional chemical control of CRISPR-Cas9. Nat. Chem. Biol. *13*, 9–11.

Majumder, K., Koues, O.I., Chan, E.A.W., Kyle, K.E., Horowitz, J.E., Yang-Iott, K., Bassing, C.H., Taniuchi, I., Krangel, M.S., and Oltz, E.M. (2015). Lineage-specific compaction of Tcrb requires a chromatin barrier to protect the function of a long-range tethering element. J. Exp. Med. *212*, 107–120.

Malarkey, C.S., and Churchill, M.E.A. (2012). The high mobility group box: the ultimate utility player of a cell. Trends Biochem. Sci. *37*, 553–562.

Malin, S., McManus, S., Cobaleda, C., Novatchkova, M., Delogu, A., Bouillet, P., Strasser, A., and Busslinger, M. (2010). Role of STAT5 in controlling cell survival and immunoglobulin gene recombination during pro-B cell development. Nat. Immunol. *11*, 171–179.

Mandal, M., Powers, S.E., Maienschein-Cline, M., Bartom, E.T., Hamel, K.M., Kee, B.L., Dinner, A.R., and Clark, M.R. (2011). Epigenetic repression of the Igk locus by STAT5-mediated recruitment of the histone methyltransferase Ezh2. Nat. Immunol. *12*, 1212–1220.

Mandal, M., Hamel, K.M., Maienschein-Cline, M., Tanaka, A., Teng, G., Tuteja, J.H., Bunker, J.J., Bahroos, N., Eppig, J.J., Schatz, D.G., et al. (2015). Histone reader BRWD1 targets and restricts recombination to the Igk locus. Nat. Immunol. *16*, 1094–1103.

Marculescu, R., Le, T., Simon, P., Jaeger, U., and Nadel, B. (2002). V(D)J-mediated Translocations in Lymphoid Neoplasms: A Functional Assessment of Genomic Instability by Cryptic Sites. J. Exp. Med. *195*, 85–98.

Matheson, L.S., Bolland, D.J., Chovanec, P., Krueger, F., Andrews, S., Koohy, H., and Corcoran, A.E. (2017). Local Chromatin Features Including PU.1 and IKAROS Binding and H3K4 Methylation Shape the Repertoire of Immunoglobulin Kappa Genes Chosen for V(D)J Recombination. Front. Immunol. *8*, 1550.

Matthews, A.G.W., Kuo, A.J., Ramón-Maiques, S., Han, S., Champagne, K.S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D.N., et al. (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. Nature *450*, 1106–1110.

Matthias, P., and Baltimore, D. (1993). The immunoglobulin heavy chain locus contains another B-cell-specific 3' enhancer close to the alpha constant region. Mol. Cell. Biol. *13*, 1547–1553.

McBlane, F., and Boyes, J. (2000). Stimulation of V(D)J recombination by histone acetylation. Curr. Biol. *10*, 483–486.

McBlane, J.F., van Gent, D.C., Ramsden, D.A., Romeo, C., Cuomo, C.A., Gellert, M., and Oettinger, M.A. (1995). Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. Cell *83*, 387–395.

McKercher, S.R., Torbett, B.E., Anderson, K.L., Henkel, G.W., Vestal, D.J., Baribault, H., Klemsz, M., Feeney, A.J., Wu, G.E., Paige, C.J., et al. (1996). Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities. EMBO J. *15*, 5647–5658.

McMurry and Krangel, M.S. (2000). A Role for Histone Acetylation in the Developmental Regulation of V(D)J Recombination. Science (80-.). *287*, 495–498.

McNally, J.G., Müller, W.G., Walker, D., Wolford, R., and Hager, G.L. (2000). The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. Science *287*, 1262–1265.

Medvedovic, J., Ebert, A., Tagoh, H., Tamir, I.M., Schwickert, T.A., Novatchkova, M., Sun, Q., Huis In 't Veld, P.J., Guo, C., Yoon, H.S., et al. (2013). Flexible long-range loops in the VH gene region of the Igh locus facilitate the generation of a diverse antibody repertoire. Immunity *39*, 229–244.

Meyer, K.B., and Neuberger, M.S. (1989). The immunoglobulin kappa locus contains a second, stronger B-cell-specific enhancer which is located downstream of the constant region. EMBO J. *8*, 1959–1964.

Miyamoto, A., Kunisada, T., Yamazaki, H., Miyake, K., Nishikawa, S.I., Sudo, T., Shultz, L.D., and Hayashi, S.I. (1998). Establishment and characterization of pro-B cell lines from motheaten mutant mouse defective in SHP-1 protein tyrosine phosphatase. Immunol. Lett. *63*, 75–82.

Mombaerts, P., Iacomini, J., Johnson, R.S., Herrup, K., Tonegawa, S., and Papaioannou, V.E. (1992). RAG-1-deficient mice have no mature B and T lymphocytes. Cell *68*, 869–877.

Montalvo, E.A., Shi, Y., Shenk, T.E., and Levine, A.J. (1991). Negative regulation of the BZLF1 promoter of Epstein-Barr virus. J. Virol. *65*, 3647–3655.

Montefiori, L., Wuerffel, R., Roqueiro, D., Lajoie, B., Guo, C., Gerasimova, T., De, S., Wood, W., Becker, K.G., Dekker, J., et al. (2016). Extremely Long-Range Chromatin Loops Link Topological Domains to Facilitate a Diverse Antibody Repertoire. Cell Rep. *14*, 896–906.

Moreno-Cuevas, J.E., and Sirbasku, D.A. (2000). Estrogen Mitogenic Action. III. Is Phenol Red A “Red Herring”? Vitr. Cell. Dev. Biol. *36*, 447–464.

Morshead, K.B., Ciccone, D.N., Taverna, S.D., Allis, C.D., and Oettinger, M.A. (2003). Antigen receptor loci poised for V(D)J rearrangement are broadly associated with BRG1 and flanked by peaks of histone H3 dimethylated at lysine 4. Proc. Natl. Acad. Sci. U. S. A. *100*, 11577–11582.

Mostoslavsky, R., Kirillov, A., Ji, Y.H., Goldmit, M., Holzmann, M., Wirth, T., Cedar, H., and Bergman, Y. (1999). Demethylation and the establishment of kappa allelic exclusion. Cold Spring Harb. Symp. Quant. Biol. *64*, 197–206.

Mostoslavsky, R., Singh, N., Tenzen, T., Goldmit, M., Gabay, C., Elizur, S., Qi, P., Reubinoff, B.E., Chess, A., Cedar, H., et al. (2001). Asynchronous replication and allelic exclusion in the immune system. Nature *414*, 221–225.

Muljo, S. a, and Schlissel, M.S. (2003). A small molecule Abl kinase inhibitor induces differentiation of Abelson virus-transformed pre-B cell lines. Nat. Immunol. *4*, 31–37.

Mullighan, C.G., Miller, C.B., Radtke, I., Phillips, L. a, Dalton, J., Ma, J., White, D., Hughes, T.P., Le Beau, M.M., Pui, C.-H., et al. (2008). BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature *453*, 110–114.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553–563.

Murphy, K., and Weaver, C. (2017). Janeway’s Immunobiology. 9Th Edition. Garland Science.

Murre, C., McCaw, P.S., and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell *56*, 777–783.

Nagulapalli, S., and Atchison, M.L. (1998). Transcription factor Pip can enhance DNA binding by E47, leading to transcriptional synergy involving multiple protein domains. Mol. Cell. Biol. *18*, 4639–4650.

Nagulapalli, S., Goheer, A., Pitt, L., McIntosh, L.P., and Atchison, M.L. (2002). Mechanism of e47-Pip interaction on DNA resulting in transcriptional synergy and activation of immunoglobulin germ line sterile transcripts. Mol. Cell. Biol. *22*, 7337–7350.

Nawaz, Z., Lonard, D.M., Dennis, A.P., Smith, C.L., and O’malley, B.W. (1999). Proteasome-dependent degradation of the human estrogen receptor.

Nelsen, B., Tian, G., Erman, B., Gregoire, J., Maki, R., Graves, B., and Sen, R. (1993). Regulation of lymphoid-specific immunoglobulin μ heavy chain gene enhancer by ETS-domain proteins. Science (80-.).

Nichols, M., Rientjes, J.M., Logie, C., and Stewart, A.F. (1997). FLP recombinase/estrogen receptor fusion proteins require the receptor D domain for responsiveness to antagonists, but not agonists. Mol. Endocrinol. *11*, 950–961.

Nightingale, K.P., Baumann, M., Eberharter, A., Mamais, A., Becker, P.B., and Boyes, J. (2007). Acetylation increases access of remodelling complexes to their nucleosome targets to enhance initiation of V(D)J recombination. Nucleic Acids Res. *35*, 6311–6321.

Nikolajczyk, B.S., Sanchez, J.A., and Sen, R. (1999). ETS protein-dependent accessibility changes at the immunoglobulin mu heavy chain enhancer. Immunity *11*, 11–20.

Nirmala, P.B., and Thampan, R.V. (1995). Ubiquitination of the rat uterine estrogen receptor: Dependence on estradiol. Biochem. Biophys. Res. Commun. *213*, 24–31.

Notarangelo, L.D., Kim, M.-S., Walter, J.E., and Lee, Y.N. (2016). Human RAG mutations: biochemistry and clinical implications. Nat. Rev. Immunol. *16*, 234–246.

Nussenzweig, A., and Nussenzweig, M.C. (2010). Origin of chromosomal translocations in lymphoid cancer. Cell *141*, 27–38.

Nutt, S.L., Urbánek, P., Rolink, A., and Busslinger, M. (1997). Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus. Genes Dev. *11*, 476–491.

O'Shea, J.J., and Plenge, R. (2012). JAK and STAT Signaling Molecules in Immunoregulation and Immune-Mediated Disease. Immunity *36*, 542–550.

Ochodnicka-Mackovicova, K., Bahjat, M., Maas, C., van der Veen, A., Bloedjes, T.A., de Bruin, A.M., van Andel, H., Schrader, C.E., Hendriks, R.W., Verhoeyen, E., et al. (2016). The DNA Damage Response Regulates RAG1/2 Expression in Pre-B Cells through ATM-FOXO1 Signaling. J. Immunol. *197*, 2918–2929.

Oettinger, M.A., Schatz, D.G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science *248*, 1517–1523.

Ohnishi, K., and Melchers, F. (2003). The nonimmunoglobulin portion of λ5 mediates cell-autonomous pre-B cell receptor signaling. Nat. Immunol. *4*, 849–856.

Olaru, A., Patterson, D.N., Cai, H., and Livák, F. (2004). Recombination signal sequence variations and the mechanism of patterned T-cell receptor-β locus rearrangement. Mol. Immunol. *40*, 1189–1201.

Osipovich, O., Milley, R., Meade, A., Tachibana, M., Shinkai, Y., Krangel, M.S., and Oltz, E.M. (2004). Targeted inhibition of V(D)J recombination by a histone methyltransferase. Nat. Immunol. *5*, 309–316.

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., and O'Shea, C.C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science *357*, eaag0025.

Outters, P., Jaeger, S., Zaarour, N., and Ferrier, P. (2015). Long-Range Control of V(D)J Recombination & Allelic Exclusion: Modeling Views. In Advances in Immunology, p.

Paakinaho, V., Presman, D.M., Ball, D.A., Johnson, T.A., Schiltz, R.L., Levitt, P., Mazza, D., Morisaki, T., Karpova, T.S., and Hager, G.L. (2017). Single-molecule analysis of steroid receptor and cofactor action in living cells. Nat. Commun. *8*, 15896.

Palstra, R.-J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The β-globin nuclear compartment in development and erythroid differentiation. Nat. Genet. *35*, 190–194.

Pan, X., Papasani, M., Hao, Y., Calamito, M., Wei, F., Quinn, W.J., Basu, A., Wang, J., Hodawadekar, S., Zaprazna, K., et al. (2013). YY1 controls Igκ repertoire and B-cell development, and localizes with condensin on the Igκ locus. EMBO J. *32*, 1168–1182.

Papaemmanuil, E., Rapado, I., Li, Y., Potter, N.E., Wedge, D.C., Tubio, J., Alexandrov, L.B., Van Loo, P., Cooke, S.L., Marshall, J., et al. (2014). RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. Nat. Genet. *46*, 116–125.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. Cell *132*, 422–433.

Park, K., Atchison, M.L., Schaffer, L., Head, S.R., Schork, N.J., and Feeney, A.J. (1991). Isolation of a candidate repressor/activator, NF-E1 (YY-1, delta), that binds to the immunoglobulin kappa 3' enhancer and the immunoglobulin heavy-chain mu E1 site. Proc. Natl. Acad. Sci. U. S. A. *88*, 9804–9808.

Park, L.S., Friend, D.J., Schmierer, A.E., Dower, S.K., and Namen, A.E. (1990). Murine interleukin 7 (IL-7) receptor. Characterization on an IL-7-dependent cell line. J. Exp. Med. *171*, 1073–1089.

Parker, D.C. (1993). T Cell-Dependent B Cell Activation. Annu. Rev. Immunol. *11*, 331–360.

Parkinson, N.J., Roddis, M., Ferneyhough, B., Zhang, G., Marsden, A.J., Maslau, S., Sanchez-Pearson, Y., Barthlott, T., Humphreys, I.R., Ladell, K., et al. (2014). Violation of the 12/23 rule of genomic V(D)J recombination is common in lymphocytes. Genome Res.

Patenge, N., Elkin, S.K., and Oettinger, M.A. (2004). ATP-dependent remodeling by SWI/SNF and ISWI proteins stimulates V(D)J cleavage of 5 S arrays. J. Biol. Chem. *279*, 35360–35367.

Pelanda, R., Braun, U., Hobeika, E., Nussenzweig, M.C., and Reth, M. (2002). B cell progenitors are arrested in maturation but have intact VDJ recombination in the absence of Ig-alpha and Ig-beta. J. Immunol. *169*, 865–872.

Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F., and Scharff, M.D. (2008). The Biochemistry of Somatic Hypermutation. Annu. Rev. Immunol. *26*, 481–511.

Perkins, E.J., Kee, B.L., and Ramsden, D.A. (2004). Histone 3 lysine 4 methylation during the pre-B to immature B-cell transition. Nucleic Acids Res. *32*, 1942–1947.

Peschon, J.J., Morrissey, P.J., Grabstein, K.H., Ramsdell, F.J., Maraskovsky, E., Gliniak, B.C., Park, L.S., Ziegler, S.F., Williams, D.E., Ware, C.B., et al. (1994). Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice. J. Exp. Med. *180*, 1955–1960.

Pettersson, S., Cook, G.P., Brüggemann, M., Williams, G.T., and Neuberger, M.S. (1990). A second B cell-specific enhancer 3' of the immunoglobulin heavy-chain locus. Nature *344*, 165–168.

Pongubala, J.M., Nagulapalli, S., Klemsz, M.J., McKercher, S.R., Maki, R.A., and Atchison, M.L. (1992). PU.1 recruits a second nuclear factor to a site important for immunoglobulin kappa 3' enhancer activity. Mol. Cell. Biol. *12*, 368–378.

Pongubala, J.M., Van Beveren, C., Nagulapalli, S., Klemsz, M.J., McKercher, S.R., Maki, R.A., and Atchison, M.L. (1993). Effect of PU.1 phosphorylation on interaction with NF-EM5 and transcriptional activation. Science *259*, 1622–1625.

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

Predeus, A. V, Gopalakrishnan, S., Huang, Y., Tang, J., Feeney, A.J., Oltz, E.M., and Artyomov, M.N. (2014). Targeted chromatin profiling reveals novel enhancers in Ig H and Ig L chain Loci. J. Immunol. *192*, 1064–1070.

Puget, N., Leduc, C., Oruc, Z., Moutahir, M., Le Bert, M., and Khamlichi, A.A. (2015). Complete *cis* Exclusion upon Duplication of the Eμ Enhancer at the Immunoglobulin Heavy Chain Locus. Mol. Cell. Biol. *35*, 2231–2241.

Queen, C., and Baltimore, D. (1983). Immunoglobulin gene transcription is activated by downstream sequence elements. Cell *33*, 741–748.

Raghavan, S.C., Swanson, P.C., Wu, X., Hsieh, C.-L., and Lieber, M.R. (2004). A non-B-DNA structure at the Bcl-2 major breakpoint region is cleaved by the RAG complex. Nature *428*, 88–93.

Rawat, P., Jalan, M., Sadhu, A., Kanaujia, A., and Srivastava, M. (2017). Chromatin Domain Organization of the TCRb Locus and Its Perturbation by Ectopic CTCF Binding. Mol. Cell. Biol. *37*, MCB.00557-16.

Recillas-Targa, F., Pikaart, M.J., Burgess-Beusse, B., Bell, A.C., Litt, M.D., West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Position-effect protection and enhancer blocking by the chicken -globin insulator are separable activities. Proc. Natl. Acad. Sci. *99*, 6883–6888.

Reddy, E.P., Smith, M.J., and Srinivasan, A. (1983). Nucleotide sequence of Abelson murine leukemia virus genome: structural similarity of its transforming gene product to other onc gene products with tyrosine-specific kinase activity. Proc. Natl. Acad. Sci. U. S. A. *80*, 3623–3627.

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

Ribeiro de Almeida, C., Stadhouders, R., de Bruijn, M.J.W., Bergen, I.M., Thongjuea, S., Lenhard, B., van IJcken, W., Grosveld, F., Galjart, N., Soler, E., et al. (2011). The DNA-Binding Protein CTCF Limits Proximal Vκ Recombination and Restricts κ Enhancer Interactions to the Immunoglobulin κ Light Chain Locus. Immunity *35*, 501–513.

Ribeiro de Almeida, C., Stadhouders, R., de Bruijn, M.J.W., Bergen, I.M., Thongjuea, S., Lenhard, B., van Ijcken, W., Grosveld, F., Galjart, N., Soler, E., et al. (2011). The DNA-binding protein CTCF limits proximal Vκ recombination and restricts κ enhancer interactions to the immunoglobulin κ light chain locus. Immunity *35*, 501–513.

Ricci, M.A., Manzo, C., García-Parajo, M.F., Lakadamyali, M., and Cosma, M.P. (2015). Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. Cell *160*, 1145–1158.

Rickert, R.C. (2013). New insights into pre-BCR and BCR signalling with relevance to B cell malignancies. Nat. Rev. Immunol. *13*, 578–591.

Risca, V.I., Denny, S.K., Straight, A.F., and Greenleaf, W.J. (2017). Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping. Nature *541*, 237–241.

Rivera, R.R., Stuiver, M.H., Steenbergen, R., and Murre, C. (1993). Ets proteins: new factors that regulate immunoglobulin heavy-chain gene expression. Mol. Cell. Biol. *13*, 7163–7169.

Robbiani, D.F., and Nussenzweig, M.C. (2013). Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase. Annu. Rev. Pathol. *8*, 79–103.

Robinson, P.J.J., Fairall, L., Huynh, V.A.T., and Rhodes, D. (2006). EM measurements define the dimensions of the "30-nm" chromatin fiber: Evidence for a compact, interdigitated structure. Proc. Natl. Acad. Sci. *103*, 6506–6511.

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

Rolink, A.G., Winkler, T., Melchers, F., and Andersson, J. (2000). Precursor B cell receptor-dependent B cell proliferation and differentiation does not require the bone marrow or fetal liver environment. J.Exp.Med *191*, 23–32.

Romanow, W.J., Langerak, A.W., Goebel, P., Wolvers-Tettero, I.L., van Dongen, J.J., Feeney, A.J., and Murre, C. (2000). E2A and EBF act in synergy with the V(D)J recombinase to generate a diverse immunoglobulin repertoire in nonlymphoid cells. Mol. Cell *5*, 343–353.

Roque, M.C., Smith, P.A., and Blasquez, V.C. (1996). A developmentally modulated chromatin structure at the mouse immunoglobulin kappa 3' enhancer. Mol. Cell. Biol. *16*, 3138–3155.

Rosenberg, N. (1976). A quantitative assay for transformation of bone marrow cells by Abelson murine leukemia virus. J. Exp. Med. *143*, 1453–1463.

Rosenberg, N., Baltimore, D., and Schert, C.D. (1975). In Vitro Transformation of Lymphoid Cells by Abelson Murine Leukemia Virus (murine fetal liver/B-lynmphocytes/surface immunoglobulin/mercaptoethanol/ murine lymphosarcoma). *72*, 1932–1936.

Roth, D.B. (2003). Restraining the V(D)J recombinase. Nat. Rev. Immunol. *3*, 656–666.

Rouaud, P., Vincent-Fabert, C., Saintamand, A., Fiancette, R., Marquet, M., Robert, I., Reina-San-Martin, B., Pinaud, E., Cogné, M., and Denizot, Y. (2013). The IgH 3' regulatory region controls somatic hypermutation in germinal center B cells. J. Exp. Med. *210*, 1501–1507.

Ru, H., Chambers, M.G., Fu, T.-M., Tong, A.B., Liao, M., and Wu, H. (2015). Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. Cell *163*, 1138–1152.

Ru, H., Mi, W., Zhang, P., Alt, F.W., Schatz, D.G., Liao, M., and Wu, H. (2018). DNA melting initiates the RAG catalytic pathway. Nat. Struct. Mol. Biol.

Rudin, C.M., and Storb, U. (1992). Two conserved essential motifs of the murine immunoglobulin lambda enhancers bind B-cell-specific factors. Mol. Cell. Biol. *12*, 309–320.

Sabbattini, P., Georgiou, A., Sinclair, C., and Dillon, N. (1999). Analysis of mice with single and multiple copies of transgenes reveals a novel arrangement for the lambda5-VpreB1 locus control region. Mol. Cell. Biol. *19*, 671–679.

Sakamoto, S., Wakae, K., Anzai, Y., Murai, K., Tamaki, N., Miyazaki, M., Miyazaki, K., Romanow, W.J., Ikawa, T., Kitamura, D., et al. (2012). E2A and CBP/p300 act in synergy to promote chromatin accessibility of the immunoglobulin kappa locus. J Immunol *188*, 5547–5560.

Sanchez, P., Nadel, B., and Cazenave, P.-A. (1991). Vλ-Jλ rearrangements are restricted within a V-J-C recombination unit in the mouse. Eur. J. Immunol. *21*, 907–911.

Sato, H., Saito-Ohara, F., Inazawa, J., and Kudo, A. (2004). Pax-5 is essential for kappa sterile transcription during Ig kappa chain gene rearrangement. J. Immunol. *172*, 4858–4865.

Satyaraj, E., and Storb, U. (1998). Mef2 proteins, required for muscle differentiation, bind an essential site in the Ig lambda enhancer. J Immunol *161*, 4795–4802.

Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature *436*, 138–141.

Schatz, D.G., and Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. Nat. Rev. Immunol. *11*, 251–263.

Schatz, D.G., and Swanson, P.C. (2011). V(D)J recombination: mechanisms of initiation. Annu. Rev. Genet. *45*, 167–202.

Schatz, D.G., Oettinger, M.A., and Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. Cell *59*, 1035–1048.

Schatz, D.G., Oettinger, M.A., and Schlissel, M.S. (1992). V(D)J recombination: molecular biology and regulation. Annu. Rev. Immunol. *10*, 359–383.

Schlimgen, R.J., Reddy, K.L., Singh, H., and Krangel, M.S. (2008). Initiation of allelic exclusion by stochastic interaction of Tcrb alleles with repressive nuclear compartments. Nat. Immunol. *9*, 802–809.

Schlissel, M.S. (2003). Regulating antigen-receptor gene assembly. Nat. Rev. Immunol. *3*, 890–899.

Schlissel, M.S., and Baltimore, D. (1989). Activation of immunoglobulin kappa gene rearrangement correlates with induction of germline kappa gene transcription. Cell *58*, 1001–1007.

Schwickert, T.A., Tagoh, H., Gültekin, S., Dakic, A., Axelsson, E., Minnich, M., Ebert, A., Werner, B., Roth, M., Cimmino, L., et al. (2014). Stage-specific control of early B cell development by the transcription factor Ikaros. Nat. Immunol. *15*, 283–293.

Scott, E.W., Simon, M.C., Anastasi, J., and Singh, H. (1994). Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. Science *265*, 1573–1577.

Seitan, V.C., Hao, B., Tachibana-Konwalski, K., Lavagnolli, T., Mira-Bontenbal, H., Brown, K.E., Teng, G., Carroll, T., Terry, A., Horan, K., et al. (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. Nature *476*, 467–471.

Serwe, M., and Sablitzky, F. (1993). V(D)J recombination in B cells is impaired but not blocked by targeted deletion of the immunoglobulin heavy chain intron enhancer. EMBO J. *12*, 2321–2327.

Seto, E., Shi, Y., and Shenk, T. (1991). YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. Nature *354*, 241–245.

Shen, T., Horwitz, K.B., and Lange, C.A. (2001). Transcriptional hyperactivity of human progesterone receptors is coupled to their ligand-dependent down-regulation by mitogen-activated protein kinase-dependent phosphorylation of serine 294. Mol. Cell. Biol. *21*, 6122–6131.

Shimazaki, N., Tsai, A.G., and Lieber, M.R. (2009). H3K4me3 stimulates the V(D)J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: implications for translocations. Mol. Cell *34*, 535–544.

Shimizu, T., Mundt, C., Licence, S., Melchers, F., and Martensson, I.-L. (2002). VpreB1/VpreB2/ 5 Triple-Deficient Mice Show Impaired B Cell Development but Functional Allelic Exclusion of the IgH Locus. J. Immunol. *168*, 6286–6293.

Shin, H.Y., Wang, C., Lee, H.K., Yoo, K.H., Zeng, X., Kuhns, T., Yang, C.M., Mohr, T., Liu, C., and Hennighausen, L. (2017). CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. Nat. Commun. *8*, 15464.

Shinkai, Y., Rathbun, G., Lam, K.P., Oltz, E.M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, F., and Stall, A.M. (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. Cell *68*, 855–867.

Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science (80-.). *350*, 978–981.

Simeonov, D.R., Gowen, B.G., Boontanrart, M., Roth, T.L., Gagnon, J.D., Mumbach, M.R., Satpathy, A.T., Lee, Y., Bray, N.L., Chan, A.Y., et al. (2017). Discovery of stimulation-responsive immune enhancers with CRISPR activation. Nature *549*, 111–115.

Singh, H., Medina, K.L., and Pongubala, J.M.R. (2005). Contingent gene regulatory networks and B cell fate specification. Proc. Natl. Acad. Sci. *102*, 4949–4953.

Sleckman, B.P., Bassing, C.H., Alt, F.W., Hughes, M.M., D'Auteuil, M., Wehrly, T.D., Woodman, B.B., Gärtner, F., White, J.M., and Davidson, L. (2000). Recombination signal sequences restrict chromosomal V(D)J recombinationbeyond the 12/23 rule. Nature *405*, 583–586.

Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.-M., Zhu, P., and Li, G. (2014). Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. Science (80-.). *344*, 376–380.

Stadhouders, R., de Bruijn, M.J.W., Rother, M.B., Yuvaraj, S., Ribeiro de Almeida, C., Kolovos, P., Van Zelm, M.C., van Ijcken, W., Grosveld, F., Soler, E., et al. (2014a). Pre-B cell receptor signaling induces immunoglobulin κ locus accessibility by functional redistribution of enhancer-mediated chromatin interactions. PLoS Biol. *12*, e1001791.

Stadhouders, R., de Bruijn, M.J.W., Rother, M.B., Yuvaraj, S., de Almeida, C.R., Kolovos, P., Van Zelm, M.C., van Ijcken, W., Grosveld, F., Soler, E., et al. (2014b). Pre-B Cell Receptor Signaling Induces Immunoglobulin κ Locus Accessibility by Functional Redistribution of Enhancer-Mediated Chromatin Interactions. PLoS Biol. *12*, e1001791.

Stavnezer, J., Guikema, J.E.J., and Schrader, C.E. (2008). Mechanism and Regulation of Class Switch Recombination. Annu. Rev. Immunol. *26*, 261–292.

Stubbington, M.J.T., and Corcoran, A.E. (2013). Non-coding transcription and large-scale nuclear organisation of immunoglobulin recombination. Curr. Opin. Genet. Dev. *23*, 81–88.

Subrahmanyam, R., Du, H., Ivanova, I., Chakraborty, T., Ji, Y., Zhang, Y., Alt, F.W., Schatz, D.G., and Sen, R. (2012). Localized epigenetic changes induced by DH recombination restricts recombinase to DJH junctions. Nat. Immunol. *13*, 1205–1212.

Swanson, P.C. (2002). Fine structure and activity of discrete RAG-HMG complexes on V(D)J recombination signals. Mol. Cell. Biol. *22*, 1340–1351.

Taylor, B., Cobb, B.S., Bruno, L., Webster, Z., Fisher, A.G., and Merkenschlager, M. (2009). A reappraisal of evidence for probabilistic models of allelic exclusion. Proc. Natl. Acad. Sci. U. S. A. *106*, 516–521.

Teng, G., Maman, Y., Resch, W., Kim, M., Yamane, A., Qian, J., Kieffer-Kwon, K.-R., Mandal, M., Ji, Y., Meffre, E., et al. (2015). RAG Represents a Widespread Threat to the Lymphocyte Genome. Cell *162*, 751–765.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575–581.

Tora, L., Mullick, A., Metzger, D., Ponglikitmongkol, M., Park, I., and Chambon, P. (1989). The cloned human oestrogen receptor contains a mutation which alters its hormone binding properties. EMBO J. *8*, 1981–1986.

Torres-Padilla, M.-E., and Chambers, I. (2014). Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. Development *141*, 2173–2181.

Tsai, A.G., and Lieber, M.R. (2010). Mechanisms of chromosomal rearrangement in the human genome. BMC Genomics *11 Suppl 1*, S1.

Tsai, A.G., Lu, H., Raghavan, S.C., Muschen, M., Hsieh, C.-L., and Lieber, M.R. (2008). Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. Cell *135*, 1130–1142.

Vale, A.M., Kearney, J.F., Nobrega, A., and Schroeder, H.W. (2015). Molecular Biology of B Cells.

Verma-Gaur, J., Torkamani, A., Schaffer, L., Head, S.R., Schork, N.J., and Feeney, A.J. (2012). Noncoding transcription within the Igh distal V(H) region at PAIR elements affects the 3D structure of the Igh locus in pro-B cells. Proc. Natl. Acad. Sci. U. S. A. *109*, 17004–17009.

Volpi, S.A., Verma-Gaur, J., Hassan, R., Ju, Z., Roa, S., Chatterjee, S., Werling, U., Hou, H., Will, B., Steidl, U., et al. (2012). Germline deletion of Igh 3' regulatory region elements hs 5, 6, 7 (hs5-7) affects B cell-specific regulation, rearrangement, and insulation of the Igh locus. J. Immunol. *188*, 2556–2566.

Voss, T.C., Schiltz, R.L., Sung, M.-H., Yen, P.M., Stamatoyannopoulos, J.A., Biddie, S.C., Johnson, T.A., Miranda, T.B., John, S., and Hager, G.L. (2011). Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. Cell *146*, 544–554.

Wabl, M., and Steinberg, C. (1992). Allelic exclusion model questioned. Nature *359*, 370–371.

Weintraub, A.S., Li, C.H., Zamudio, A. V, Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573–1588.e28.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796–801.

Whitfield, J., Littlewood, T., Evan, G.I., and Soucek, L. (2015). The Estrogen Receptor Fusion System in Mouse Models: A Reversible Switch. Cold Spring Harb. Protoc. *2015*, pdb.top069815.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221–225.

Widom, J., and Klug, A. (1985). Structure of the 300A chromatin filament: X-ray diffraction from oriented samples. Cell *43*, 207–213.

Wijayaratne, A.L., Nagel, S.C., Paige, L.A., Christensen, D.J., Norris, J.D., Fowlkes, D.M., and McDonnell, D.P. (1999). Comparative Analyses of Mechanistic Differences Among Antiestrogens [1]. Endocrinology *140*, 5828–5840.

Wilson, T.M., Vaisman, A., Martomo, S.A., Sullivan, P., Lan, L., Hanaoka, F., Yasui, A., Woodgate, R., and Gearhart, P.J. (2005). MSH2–MSH6 stimulates DNA polymerase η, suggesting a role for A:T mutations in antibody genes. J. Exp. Med. *201*, 637–645.

Woloschak, G.E., and Krco, C.J. (1987). Regulation of kappa/lambda immunoglobulin light chain expression in normal murine lymphocytes. Mol. Immunol. *24*, 751–757.

Xiang, Y., and Garrard, W.T. (2008). The Downstream Transcriptional Enhancer, Ed, positively regulates mouse Ig kappa gene expression and somatic hypermutation. J. Immunol. *180*, 6725–6732.

Xiang, Y., Park, S.-K., and Garrard, W.T. (2013). Vκ gene repertoire and locus contraction are specified by critical DNase I hypersensitive sites within the Vκ-Jκ intervening region. J. Immunol. *190*, 1819–1826.

Xu, C.-R., and Feeney, A.J. (2009). The epigenetic profile of Ig genes is dynamically regulated during B cell differentiation and is modulated by pre-B cell receptor signaling. J. Immunol. *182*, 1362–1369.

Xu, Y., Davidson, L., Alt, F.W., and Baltimore, D. (1996). Deletion of the Igκ light chain intronic enhancer/matrix attachment region impairs but does not abolish VκJκ rearrangement. Immunity *4*, 377–385.

Yamane, H., and Paul, W.E. (2013). Early signaling events that underlie fate decisions of naive CD4 $^{+}$ T cells toward distinct T-helper cell subsets. Immunol. Rev. *252*, 12–23.

Yan, J., Chen, S.-A.A., Local, A., Liu, T., Qiu, Y., Dorighi, K.M., Preissl, S., Rivera, C.M., Wang, C., Ye, Z., et al. (2018). Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. Cell Res. *28*, 204–220.

Yancopoulos, G.D., and Alt, F.W. (1985). Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. Cell *40*, 271–281.

Ye, J. (2004). The immunoglobulin IGHD gene locus in C57BL/6 mice. Immunogenetics *56*, 399–404.

Yin, F.F., Bailey, S., Innis, C.A., Ciubotaru, M., Kamtekar, S., Steitz, T.A., and Schatz, D.G. (2009). Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. Nat. Struct. Mol. Biol. *16*, 499–508.

Yurchenko, V., Xue, Z., and Sadofsky, M. (2003). The RAG1 N-terminal domain is an E3 ubiquitin ligase.

Zachau, H.G. (1993). The immunoglobulin κ locus — or — what has been learned from looking closely at one-tenth of a percent of the human genome. Gene *135*, 167–173.

Zandi, S., Mansson, R., Tsapogas, P., Zetterblad, J., Bryder, D., and Sigvardsson, M. (2008). EBF1 Is Essential for B-Lineage Priming and Establishment of a Transcription Factor Network in Common Lymphoid Progenitors. J. Immunol.

Zhang, L., Reynolds, T.L., Shan, X., and Desiderio, S. (2011). Coupling of V(D)J recombination to the cell cycle suppresses genomic instability and lymphoid tumorigenesis. Immunity *34*, 163–174.

Zhang, Y., McCord, R.P., Ho, Y.-J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial Organization of the Mouse Genome and Its Role in Recurrent Chromosomal Translocations. Cell *148*, 908–921.

Zhang, Y., Yin, C., Zhang, T., Li, F., Yang, W., Kaminski, R., Fagan, P.R., Putatunda, R., Young, W.-B., Khalili, K., et al. (2015). CRISPR/gRNA-directed synergistic activation mediator (SAM) induces specific, persistent and robust reactivation of the HIV-1 latent reservoirs. Sci. Rep. *5*, 16277.

Zhang, Z., Espinoza, C.R., Yu, Z., Stephan, R., He, T., Williams, G.S., Burrows, P.D., Hagman, J., Feeney, A.J., and Cooper, M.D. (2006). Transcription factor Pax5 (BSAP) transactivates the RAG-mediated V(H)-to-DJ(H) rearrangement of immunoglobulin genes. Nat. Immunol. *7*, 616–624.

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

Zhou, X., Xiang, Y., and Garrard, W.T. (2010). The Ig Gene Enhancers, E3' and Ed, Are Essential for Triggering Transcription. J. Immunol. *185*, 7544–7552.

Zhou, X., Xiang, Y., Ding, X., and Garrard, W.T. (2012). A new hypersensitive site, HS10, and the enhancers, E3' and Ed, differentially regulate Igκ gene expression. J. Immunol. *188*, 2722–2732.