Dynamic Enhancer-Mediated Activation of the Murine Immunoglobulin Lambda Locus

Zeqian Gao

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 2020

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.

Acknowledgement

First and foremost, I would like to express my upmost respect and gratitude to my supervisor, Dr. Joan Boyes, for all of her training and guidance throughout my PhD. I would also like to thank Joan for highlighting the significance of asking relevant scientific questions and for the speed with which comments and suggestions on thesis drafts were given. I would also like to thank my co-supervisor, Dr. Ian Wood, for his guidance and support.

Many thanks to all members of the Boyes lab, past and current: In particular, Dr. James Scott for help in the lab during the early parts of this project and for kindly allowing me to use his ChIP samples and data for analysis in this project, Dr. Alastair Smith for help in characterization of the inducible system to investigate enhancer-promoter interactions, Dr. Xiaoling Wang for help with my PhD application and basic molecular assays, Dr. Daniel Thwaites for help with Taqman probe qPCR and western blotting assays, Matthew Burke and Laura Cook for telling me interesting stories about their PhD projects. These nice people make the lab a very pleasant place to work.

On a personal level, I also wish to thank my parents, my mother-in-law and my sister for their support. Finally, I would like to thank my wife, Miao Wang, and my son, Qisen Gao, for their unending love and support.

My study was financially supported by the University of Leeds – China Scholarship Council joint scholarship.

Abstract

Enhancers are the most dynamically utilized part of eukaryotic genomes, playing an essential role in the regulation of precise spatiotemporal patterns of gene expression during development. However, what enhancers deliver to promoters to activate transcription remains unclear. An ideal model to investigate enhancer-promoter contacts is the murine Igλ locus as over-expression of a single transcription factor, IRF4, which directly binds to the Eλ3-1 enhancer in pro-B cells, is sufficient to activate Igλ non-coding transcription. An inducible pro-B cell line that expresses IRF4-ER_{T2} was generated that allows the transcription activators that are delivered from Eλ3-1 to the Vλ1 and Jλ1 promoters to be deciphered. By using temporal ChIP and 3C technologies, here I present evidence that E2A, p300, Mediator and Integrator binding to Eλ3-1 are early events during Igλ activation, whereas YY1 binding to Eλ3-1 is a late event. Building on published ATAC-seq and ChIP-seq data, I found that an insulated neighbourhood domain (IND) is already present in pro-B cells, sealing the 3' half of the Igλ locus via binding of CTCF/cohesin at HS7 and HSVλ1. Increased binding of IRF4, E2A, p300, Mediator, Integrator and YY1 were also observed at two other enhancers, HS6 and HSCλ1, facilitating locus contraction. Furthermore, I present evidence that Eλ3-1 encodes bidirectional enhancer RNAs. Knock-down of the Eλ3-1 sense enhancer RNAs leads to disruption of Igλ activation. Intriguingly, knock-down of the Eλ3-1 anti-sense enhancer RNA results in a higher level of activation of the Igλ locus, suggesting an intrinsic repressive role of the anti-sense enhancer RNAs in the activation of gene transcription. This work provides the first evidence of the sequential order of recruitment of diverse transcription activators at enhancers and promoters of antigen receptor genes, and also identifies a unique role for the interplay between sense and anti-sense RNAs in the activation of gene transcription.

Contents

List of Figures

List of Tables

Abbreviations

3C	Chromosome conformation capture
A-MLV	Abelson murine leukaemia virus
ATAC-seq	Assay for transposase-accessible chromatin with high-throughput sequencing
BAC	Bacterial artificial chromosome
BCL-2	B cell lymphoma 2
BCR	B cell receptor
BLAST	Basic local alignment search tool
BLNK	B cell linker protein
bp	Base pair
BRE	TFIIB recognition element
BTK	Bruton tyrosine kinase
CBE	CTCF binding element
CDK	Cyclin dependent kinase
cDNA	Coding DNA
Cer	Contracting element for recombination
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation with high-throughput sequencing
CLP	Common lymphoid progenitor
COMPASS	Complex Proteins Associated with Set1
CPSF	Cleavage and polyadenylation specificity factor
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
CTD	Carboxy terminal domain

dH2O	Deionized water
DMEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribonucleic acid
DPE	Downstream promoter element
DSB	Double stranded break
DSIF	DRB sensitivity-inducing factor
EBF1	Early B cell factor 1
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced green fluorescent protein
ELISA	Enzyme linked immunosorbent assay
ER	Estrogen receptor
ERK	Extracellular signal-regulated kinase
EMCV	Encephalomyocarditis virus
EMSA	Electrophoretic mobility shift assay
ER	Estrogen receptor
ESC	Excised signal circle
EZH2	Enhancer of zeste homolog 2
FACS	Fluorescence activated cell sorting
FCS	Foetal calf serum
FITC	Fluorescein isothiocyanate
FISH	Fluorescence *in situ* hybridisation
FOXO1	Forkhead box O1
GFP	Green fluorescent protein
H3K4me3	Histone H3 lysine 4 trimethylation
H3K4me	Histone H3 lysine 4 methylation

H3K27me	Histone H3 lysine 27 methylation
H3K27ac	Histone H3 lysine 27 acetylation
HA	Human Influenza hemagglutinin
HBD	Hormone binding domain
HEPES	4-(2-hydroxyethyl) piperazine-1-ethanesulfonic acid
HD	Homeodomain
HOT	Highly occupied target region
HS	Hypersensitive site
HSC	Hematopoietic stem cell
IGCR1	Intergenic control region1
IgH	Immunoglobulin heavy chain
IgL	Immunoglobulin light chain
Igλ	Immunoglobulin lambda chain
Igκ	Immunoglobulin kappa chain
IL-7	Interleukin 7
IND	Insulated neighborhood domain
Inr	Initiator element
INT	Integrator
IRES	Internal ribosome entry site
IRF4	Interferon regulatory factor 4
ITAM	Immunoreceptor tyrosine-based activation motif
KD	Knock down
kDa	Kilodalton
KO	Knock out
KOAc	Potassium acetate
LCR	Locus control region

MACS	Model based analysis of ChIP-seq
Mb	Megabase
MED	Mediator
MCL-1	Myeloid cell leukaemia sequence 1
MLL	Mixed Lineage Leukemia
mRNA	Messenger ribonucleic acid
NaOAc	Sodium acetate
NELF	Negative elongation factor
NHEJ	Non-homologous end joining
NK	Natural killer
NTD	N-terminal domain
PAIR	Pax5 associated intergenic repeat
PAX5	Paired Box 5
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PE	Phycoerythrin
PEI	Polyethyleneimine
PHD	Plant homeodomain
PQ52	Promoter 5’ of DQ52
PRC2	Polycomb repressive complex 2
PI3K	Phosphoinositide 3- kinase
PIC	Pre-initiation complex
PLCγ	phospholipase Cγ
PMSF	Phenylmethylsulfonyl fluoride
qPCR	Quantitative PCR
RAG	Recombination activating gene

RNA	Ribonucleic acid
RNAP	RNA polymerase
RPMI	Roswell Park Memorial Institute
RSS	Recombination signal sequence
SDS	Sodium dodecyl sulphate
SEC	Super elongation complex
Sis	Silencer in the intervening sequence
SMC	Structural maintenance of chromosomes
STAT5	Signal transducer and activator of transcription 5
SV40	Simian virus 40
SYK	Spleen tyrosine kinase
TAE	Tris acetate EDTA
TAD	Topological associating domain
TBP	TATA box binding protein
TCR	T cell receptor
TE	Tris EDTA
TSS	Transcription start site
UP	upstream promoter element
YY1	Ying Yang 1
ZAP70	ζ chain associated protein kinase of 70 kDa

Chapter 1: Introduction

A) Gene transcription

1.1 Composition and assembly of DNA dependent RNA polymerase in prokaryotic organisms

Gene expression is the process by which the genetic information in DNA is copied, via transcription, into messenger RNA (mRNA), which is then translated to generate a new protein. The mRNA intermediate was first identified in *Escherichia coli* (*E.coli*). In prokaryotic cells, all genes are transcribed by a single RNA polymerase (RNAP), which is the principal enzyme that catalyses the polymerization of ribonucleotides by building the chain in the 5´ to 3´ direction. *E. coli* RNAP is a multiprotein, multifunctional complex that is comprised of five different types of subunits: α, β, β´, ω and δ. It is well established that the core RNAP consists of two α and one of each β, β' and ω subunits, whereas the δ subunit is relatively weakly bound to the core. The core RNAP is assembled sequentially, via the pathway: 2α→α2→α2β→α2ββ´ω (Ishihama, 1981; Ishihama et al., 1987). The α subunit plays an essential role in RNAP assembly, promoter DNA recognition and transcription regulation. The α subunit contains 329 amino acid (aa) residues, is well organized into two structural domains, the N-terminal domain (αNTD) and the C-terminal domain (αCTD), connected by a linker region. αNTD functions mainly in providing a docking platform for the other subunits and is necessary and sufficient for enzyme assembly (Igarashi et al., 1991; Kimura and Ishihama, 1995, 1996). In contrast, αCTD is not essential for the polymerase assembly and maintaining the basal level of transcription (Igarashi et al., 1991). The main function of αCTD is transcription regulation via directly and indirectly interacting with different transcription factors (Ebright and Busby, 1995; Ishihama, 1992). In addition, αCTD can recognize upstream promoter element (UP)-containing promoters via binding of the AT-rich DNA motifs within the UP (Ross et al., 1993). The β and β´ subunits are the two largest and most complicated subunits of *E. coli* RNAP. The active centre for RNA synthesis is built by these two subunits, which form the “catalytic unit” and three important channels for double-stranded DNA

(dsDNA) template entry and exit of the RNA product (Lee and Borukhov, 2016; Nudler, 2009). The ω subunit is the smallest subunit of *E. coli* RNAP and mainly binds to the double-psi-β-barrel (DPBB) domain within the β´ subunit which contains the RNA catalytic unit; the ω subunit is believed to play a role in maintaining the basic catalytic activity of RNAP and/or in preventing the DPBB domain from damage (Mathew and Chatterji, 2006; Sutherland and Murakami, 2018).

The core RNAP only exhibits low levels of specificity in recognition of promoter DNA sequences, leading to inefficient and non-specific transcription. It is the δ subunit that determines the specificity of engagement with different types of gene promoters. Multiple δ factors are present in *E. coli* cells and are distinguished by their characteristic molecular weights, including δ70, δ54, δ38, δ32, δ28, δ24 and δ19 (Gruber and Gross, 2003). Each δ factor is capable of directing the core RNAP to transcribe a specific set of genes. For instance, the first identified and largest δ factor, δ70, encoded by the *rpoD* gene, directs RNAP to transcribe housekeeping genes (Hawley and McClure, 1983). The second largest δ factor, δ54, directs RNAP to transcribe a set of genes involved in nitrogen metabolism (Hunt and Magasanik, 1985). Therefore, the *E. coli* cells can coordinately regulate transcription output by regulating the level of each δ factor.

1.2 Composition and assembly of RNAP in eukaryotic organisms

Compared to prokaryotic organisms that contain a single RNAP, eukaryotic organisms have a more complicated and efficient transcription machinery. Eukaryotic organisms contain at least three distinct RNAPs, RNAPI, RNAPII and RNAPIII, which share structural and mechanistic homology (Vannini and Cramer, 2012). Each of these RNAPs transcribes a specific set of genes. Type I RNAP (RNAPI) specifically transcribes the three largest species of ribosomal RNAs, namely 28S, 18S and 5.8S, which are the most abundant RNA species and act as the enzymatic scaffold for ribosome assembly (Moss and Stefanovsky, 2002). RNAPI accounts for 35-60% of all nuclear transcription in eukaryotic cells (Moss and Stefanovsky, 2002). Type III RNAP (RNAPIII) is

mainly responsible for the transcription of the smallest rRNA species (5S rRNA) and transfer RNA (tRNA) genes (Turowski and Tollervey, 2016). Type II RNAP (RNAPII) is specifically devoted to transcription of protein-coding genes, microRNAs and most small nuclear RNAs (snRNAs) (Yokoyama, 2018), and has been most rigorously studied.

Eukaryotic RNAPII is comprised of 12 subunits (Rpb1 through Rpb12) in yeast and humans (Myer and Young, 1998). Rpb1 is the largest and catalytic subunit of the RNAPII and its carboxyl terminal domain (CTD) contains multiple tandem conserved heptapeptide repeats Tyr_1-Ser_2-Pro_3-Thr_4-Ser_5-Pro_6-Ser_7 (Allison et al., 1985; Corden et al., 1985). Post-translational modifications of residues within the heptapeptide repeats are essential for the recruitment of different transcription and processing factors. For instance, serine 5 phosphorylated CTD recruits the 5' capping enzymes to the newly synthesized mRNAs via direct interactions (McCracken et al., 1997). Rpb2 is the second largest component of RNAPII and contains multiple domains for DNA and RNA binding. Eukaryotic Rpb1 and Rpb2 form the catalytic centre for RNA synthesis and show significant sequence and structural similarity to the β and β´ subunits, respectively, of prokaryotic RNAP (Woychik, 1998). Similarly, A190 and A135 subunits from RNAPI as well as C160 and C128 subunits from RNAPIII are homologues of the bacterial β and β´ subunits, respectively (Vannini and Cramer, 2012). RNAPII Rpb3 and Rpb11 subunits show sequence homology to prokaryotic RNAP α subunit and interact with Rpb1 and Rpb2 to form the structural and functional equivalent of prokaryotic core RNAP (Kimura et al., 1997; Zhang and Darst, 1998). Likewise, AC40 and AC19 subunits shared by RNAPI and RNAPIII are homologous to Rpb3 and Rpb11 (Vannini and Cramer, 2012). Notably, Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12 are shared among eukaryotic RNAPI, RNAPII and RNAPIII and thus exert their function in the transcription of all types of eukaryotic RNAs (Vannini and Cramer, 2012).

Compared to bacterial RNAP assembly, eukaryotic RNAP assembly is a more sophisticated process. Previous *in vitro* experiments demonstrated that independently expressed human RNAPII subunits do not assemble a complete

enzyme after mixing cell lysates, but instead, these studies revealed that Rpb3 and Rpb5 are major anchoring sites for other RNAPII subunits (Acker et al., 1997). Further *in vivo* studies showed that Rpb3 and Rpb2 form a sub-complex immediately after their synthesis (Kolodziej and Young, 1991). This Rpb2-Rpb3 subcomplex subsequently recruits Rpb1 (Kolodziej and Young, 1991); Rpb1 is the only subunit of RNAPII that can bind strongly to Rpb5, and Rpb5 enters the RNAPII complex in an Rpb1 dependent manner (Acker et al., 1997).

1.3 Mechanism of RNAPII transcription

Promoter recognition by RNAPII is the first step in gene transcription. Gene promoters are genomic DNA regions that lie upstream of the coding sequence of transcribed genes. The core promoter region resides most proximal to the start codon and contains transcription start site (TSS) and binding sites for RNAPII and general transcription factors, such as the TATA box, Initiator (Inr) element, TFIIB recognition element (BRE) and downstream promoter element (DPE). RNAPII is incapable of recognizing gene promoters to initiate transcription on its own. Instead, multiple general transcription factors are needed to form the pre-initiation complex (PIC) that recognizes gene promoters. In eukaryotic cells, the general transcription factors include transcription factor IIA (TFIIA), TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Woychik and Hampsey, 2002). TFIID and TFIIB are two general transcription factors that show sequence-specific DNA binding activity. TFIID exists as a multi-subunit protein complex, consisting of TATA box binding protein (TBP) and 14 TBP-associated factors (TAFs) (Dynlacht et al., 1991; Pugh and Tjian, 1991). TBP is a horseshoe-shaped protein complex that binds to the TATA box whereas other proteins in the TFIID complex bind a variety of DNA elements within core promoters, such as Inr and DPE, which are essential for promoter recognition (Peterson et al., 1990). TFIIB is a single polypeptide and its C-terminal domain shows sequence-specific DNA binding activity, with specificity for the BRE element, a DNA motif present in a subset of promoters immediately upstream of the TATA box (Lagrange et al., 1998). TFIIE and TFIIF are both hetero-tetramers that are comprised of two TFIIEα/β and TFIIF α/β subunits, respectively. TFIIE and TFIIF can also interact with template DNA but without sequence specificity

(Forget et al., 1997; Kim et al., 1997; Woychik and Hampsey, 2002). TFIIH is the largest and most sophisticated general transcription factor and has catalytic activities (Seroz et al., 1995). TFIIH contains the cyclin-dependent protein kinase 7 (CDK7) and two ATP-dependent DNA helicases XPB and XPD, which are responsible for phosphorylation of the CTD of the Rpb1 subunit of RNAPII and promoter DNA melting and clearance, respectively (Douziech et al., 2000; Feaver et al., 1994; Kim et al., 2000).

Assembly of a PIC at gene promoters is a rate-limiting step during the activation of gene transcription. Previous studies proposed a “step-wise” model for assembly of the RNAPII PIC (Buratowski et al., 1989). In this model, PIC assembly is nucleated by binding of the TBP subunit of TFIID to the TATA-box. TFIIB subsequently interacts with the TFIID-TATA complex and contributes to the transcription polarity by binding asymmetrically to the BRE motif (Lagrange et al., 1998). The unphosphorylated RNAPII, in association with TFIIF, then binds to the TFIIH-TFIID-TATA complex. TFIIE and TFIIH are the last two general transcription factors to be recruited to complete assembly of the PIC. Whilst TFIIA has been shown to stabilize TFIID-TATA interactions (Hampsey, 1998), the exact point during PIC formation at which TFIIA participates, remains elusive. Once PIC assembly is complete, TFIIH phosphorylates the Ser 5 residue of the CTD of RNAPII Rpb1 and unwinds the promoter DNA (Spangler et al., 2001); this allows activated RNAPII to start to synthesize nascent RNAs. However, Ser 5 phosphorylated RNAPII stalls after the synthesis of a nascent RNA of 20–60 nucleotides (Rasmussen and Lis, 1993; Saunders et al., 2006).

Promoter proximal pausing is another rate-limiting step during the activation of gene transcription. The pausing of initiated RNAPII 20-60 nucleotide (nt) downstream from the TSS is modulated by RNAPII physically interacting with the DRB sensitivity-inducing factor (DSIF) and the negative elongation factor (NELF) (Yamaguchi et al., 2013). While DSIF is a heterodimer comprised of SPT4 and SPT5 (Wada et al., 1998; Hartzog et al., 1998), NELF is a multi-subunit protein complex consisting of NELF-A, NELF-B, NELF-C/D and NELF-E (Yamaguchi et al., 2013). DSIF binds strongly to RNAPII and the interaction

interface spans from the DNA cleft to the RNA exit tunnel of RNAPII (Bernecky et al., 2017). NELF shows only weak affinity to DSIF and RNAPII alone but binds strongly to the DSIF-RNAPII complex (Yamaguchi et al., 2002). Paused RNAPII is activated by the positive elongation factor, p-TEFb, that causes its release from promoter proximal regions. p-TEFb is comprised of the cyclin-dependent kinase 9 (CDK9) and cyclin T (Grana et al., 1994; Peng et al., 1998). Recruitment of CDK9 to promoter proximal regions phosphorylates the Ser 2 residue of the CTD of Rpb1, the SPT5 subunit of DSIF and NELF-E, leading to the dissociation of NELF from promoters (Fujinaga et al., 2004; Isel and Karn, 1999; Ivanov et al., 2000). Phosphorylation of DSIF converts it to a positive elongation factor, which remains bound to the Ser 2 phosphorylated RNAPII (Fujinaga et al., 2004). This elongating from of RNAPII efficiently transcribes the entire gene body.

B) Different types of enhancer-like elements

The first enhancer, identified in 1981, was a 72 bp repeated sequence that resides upstream of the simian virus 40 (SV40) early region and can significantly increase ectopic expression of a reporter gene (Banerji et al., 1981; Moreau et al., 1981). Two years later, a non-viral enhancer was discovered within the mouse immunoglobulin heavy chain locus, followed by enhancers being documented in many diverse organisms (Banerji et al., 1983; Gillies et al., 1983). It is widely recognized that enhancers can recruit combinations of transcription factors that subsequently interact with subunits of TFIID, the Mediator complex or cohesin complex. Complexes anchored at the enhancer can facilitate the recruitment of RNAPII to target core promoters by looping out the intervening DNA sequences (Kagey et al., 2010; Malik and Roeder, 2010). In addition, transcription factors bound at enhancers can interact directly with ATP-dependent chromatin-remodelling complexes or enzymes that have histone modifying activities, altering three-dimensional chromatin structures and increasing the accessibility of enhancer sequences to other transcription cofactors at the cognate promoters (Bajpai et al., 2010; Zippo et al., 2009).

Recent genome-wide studies have indicated that there are a great number of enhancers located in metazoan genomes (for instance, more than one million enhancers in the human genome) and these enhancers are distributed throughout the genome, including intergenic regions, introns and exons of protein-coding genes. Enhancers are characterized by increased chromatin accessibility, deposition of specific histone modifications, such as H3K27ac and H3K4me1, binding of lineage-specific transcription factors and enrichment of the RNAPII machinery (Shlyueva et al., 2014). These enhancers can be classified into different groups, including the locus control region (LCR), super-enhancers/stretch enhancers, shadow enhancers and highly occupied target (HOT) region, according to their mechanism of action.

1.4 LCR

LCRs refer to genomic regions that are sufficient to fully activate a linked gene in a tissue-specific, copy number dependent manner, independent of its position of integration (Li et al., 2002). The most widely studied example of LCR-regulated gene expression is the *β-globin* locus. The human β-globin LCR resides 6-22 kb upstream to the first globin gene in the locus. It is comprised of five DNase I hypersensitive sites (HS), which are a typical feature of enhancers (Li et al., 2002). HS1 to HS4 are formed only in erythroid cells, while HS5 is present in different lineages (Li et al., 1999). LCR enhancer activity resides in HS2, HS3 and HS4, but not in HS1 or HS5 (Fraser and Grosveld, 1998). HS2 acts as a traditional enhancer, which means its activity can be detected in transient reporter assays. HS3 and HS4 only display enhancer activity when these two regulatory elements are integrated into chromatin (Hardison et al., 1997), indicating that alteration of chromatin structure may be involved in mediating the activities of these enhancers. Subsequent studies showed that the mouse β-globin LCR behaves in a very similar way to its human counterpart, except the mouse β-globin LCR has six HS, rather than five (Bulger et al., 1999; Kim and Dean, 2012).

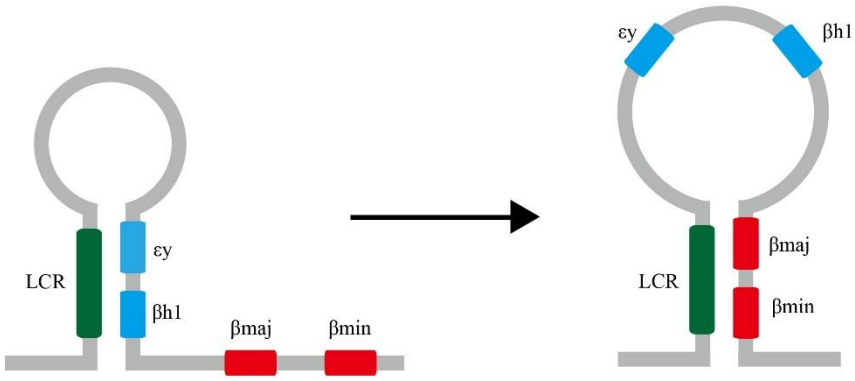

Figure 1.1 – Chromatin loop formation at mouse β-globin locus during different development stages

The mouse β-globin locus consists of four functional genes, *εy*, *βh1*, *βmaj* and *βmin*. In the early embryo, the first two genes are arranged into a transcription hub with LCR, whilst in the adult, the latter two genes are arranged into the transcription hub. Adapted from de Laat et al., 2003 and Kim et al., 2012.

Transcription of mouse *β-globin* genes is orchestrated by the LCR in a tissue and developmental stage-specific manner. In the early embryo, the mouse *εy* and *βh1* genes are actively transcribed whilst in the adult, first the mouse *βmaj* and *βmin* are efficiently expressed (Kim and Dean, 2012). Specifically, in early embryonic cells, the transcriptionally active *εy* and *βh1* genes are positioned close to the active transcription hub which is formed by LCR HSs (Figure 1.1), whilst in adult cells, *βmaj* and *βmin* genes are brought into close proximity of the LCR chromatin hub (Figure 1.1, Carter et al., 2002). In the latter case, the inactive globin genes and the intervening chromatin are looped out to enable LCR/gene interactions (Palstra et al., 2003). Subsequent studies showed that LCRs not only control the multiple-gene loci, but also the single genes, such as the lysozyme gene (Bonifer et al., 1990).

1.5 Super enhancers

Super-enhancers represent a class of regulatory genomic regions that are unusually strongly enriched for the binding of Mediator, an RNAPII coactivator that, together with cohesin, is involved in enhancer-promoter communication

(Witte et al., 2015). Bioinformatic analyses, based largely on enhancer bound transcription factors, lead to the identification of super-enhancers. For instance, super-enhancers in mouse embryonic stem cells were defined as follows: (Whyte et al., 2013). Firstly, the chromatin regions that bind all three master transcription regulators of ES cells, Oct4, Sox2 and Nanog, on basis of ChIP-seq data, were considered as enhancers. Secondly, enhancer-like elements within ~20 kb of each other were concatenated to define a single entity spanning a chromatin fragment. Finally, such “stitched” enhancer entities with high levels of enrichment of Med1 (screened by ChIP-seq) were considered to be super-enhancers.

Similarly, stretch enhancers were identified by bioinformatic analysis on the basis of the presence of specific epigenetic modifications and transcription factor binding, such as H3K4me1 and H3K27ac, that spanned unexpectedly long (>3 kb) chromatin fragments (Parker et al., 2013). According to the comparative analysis of super-enhancers and stretch enhancers, super-enhancers were believed as a subset of stretch enhancers. For instance, 6,426 stretch enhancers and 683 super-enhancers were identified in the H1 human embryonic stem cell line, respectively (Hnisz et al., 2013; Parker et al., 2013), and greater than 70% of super-enhancers overlapped with stretch enhancers. In addition, some super-enhancers and stretch enhancers also correspond to characterized LCRs, the original super-enhancers as defined by functional studies. Owing to the absence of functional studies relating to the majority of super-enhancers and stretch enhancers, it remains unknown whether they constitute novel LCRs.

1.6 Shadow enhancers

It is curious that extended LCRs, super-enhancers and stretch enhancers have been identified in mammalian genomes, but not in Drosophila. To some extent, this could reflect a difference in genome organization. It might be more efficient to distribute enhancers around any available regions in the relatively compact Drosophila genome, whereas in mammalian genomes, the clustering of regulatory elements within an extended enhancer could occur in gene deserts

located large distances from the target genes. This clustering could lead to enhanced chromatin accessibility and behave cooperatively to contact with cognate promoters, giving a more robust transcription pattern. In Drosophila, this robust transcription pattern may be similarly achieved by shadow enhancers (Barolo, 2012; Perry et al., 2010).

The term “shadow enhancers” was coined by Mike Levine and colleagues in 2008, to describe the discovery of remote regulatory elements of Drosophila genes, *brinker* and *sog* (Hong et al., 2008). Before the development of genome-wide ChIP, enhancers were generally identified by trial and error via cloning genomic segments (usually HS) from within or around genes into a reporter construct. The expression of the reporter was subsequently compared with the gene’s endogenous expression as determined by *in situ* hybridization or other reporter assays. By using genome-wide profiling of transcription factors, redundant regulatory elements were identified for genes for which known enhancers existed. The transcription factor binding pattern at these redundant enhancers showed a remarkably similar pattern (like a shadow) to the previously characterized enhancer (Hong et al., 2008). When used in reporter assays, shadow enhancers caused the same increased transcription pattern as the previously known enhancer (Hong et al., 2008). Shadow enhancers usually reside either within an intron of, or on the far side of, a neighbouring gene (Zeitlinger et al., 2007).

1.7 HOT region

A HOT region is a novel class of regulatory genomic regions that has been recently defined in several model organisms, including *Caenorhabditis elegans,* Drosophila and humans, via genome-wide ChIP analyses of the binding of a variety of transcription factors (Farley and Levine, 2012). Chromatin regions of traditional enhancers are characterized by low occupancy and are generally bound by one or several different transcription factors. Consistent with this, traditional enhancers are enriched for sequence motifs that are recognized by specific transcription factors, suggesting direct DNA-protein interactions. In contrast, HOT regions are characterized by indirect loading of a number of

different transcription factors. The recruitment of transcription factors to HOT regions is independent of sequence motifs because of a dearth of such specific transcription factor motifs (Farley and Levine, 2012). HOT regions contain several sequence features, including the Zelda binding motif, GAGA binding elements and TAGteam sequence motif (Liang et al., 2008; Satija and Bradley, 2012). These sequence motifs are recognized by transcription factors Zelda and GAGA which act to potentiate transcription factor binding by catalysing development of regions of open chromatin (Harrison et al., 2011; Satija and Bradley, 2012). Similar to other sequences encompassing these elements, HOT regions display increased nucleosome turnover and are enriched in the histone variant, histone H3.3, indicative of “accessible chromatin” (Jin et al., 2005). It is likely that both Zelda anchoring and interactions between transcription factors play important roles in the formation of numerous, different transcription factor complexes at HOT regions.

C) Models of enhancer-promoter interactions

In mammalian genomes, genes can span hundreds of kilobases and can be controlled by distant enhancers. It is generally agreed that enhancers can increase the transcription of target genes by delivering coactivators to their cognate promoters. However, the mechanism by which enhancers specifically communicate with their correct promoters is not entirely understood. To date, four hypotheses have been proposed to describe this communication: chromatin looping, linking, tracking and facilitated tracking.

1.8 Chromatin looping

The currently favoured model for enhancer-promoter communications involves homotypic or heterotypic interactions between enhancer-bound transcription factors and promoter-bound transcription factors to form a chromatin loop that juxtaposes enhancer and promoter regions at the base of the loop and that loops out the long intervening genomic sequences (Figure 1. 2) (Su et al., 1991). Chromosome engineering studies have revealed that forcing an enhancer-promoter loop is sufficient to activate gene transcription at the β-globin locus in erythroid cells, in which other essential transcription factors required for β-

globin expression are already bound (Deng et al., 2014). The technique of chromosome conformation capture (3C), and its variants further provided strong evidence for the physical interactions between enhancers and promoters (Jin et al., 2013). In addition, fluorescence *in situ* hybridization (FISH) has demonstrated the spatial juxtaposition of distant enhancers and promoters (Lettice et al., 2014).

1.9 Linking

The linking model proposes that the binding of facilitators between enhancers and their cognate promoter mediates enhancer activity (Figure 1.2). Homeodomain (HD) containing proteins are known to bind *in vitro* to a broad range of distinct genomic DNA sequences with a similar preference (Walter and Biggin, 1996). Consistent with this, *in vivo* experiments have shown that HD proteins bind uniformly throughout their target gene loci, and at lower levels to enormous active genes in the Drosophila embryo (Walter et al., 1994). In contrast, transcription factors generally bind only to short genomic regulatory regions, such as enhancers and promoters. Further studies revealed that HD transcription factors are directly involved in the control of nearly all active genes in the Drosophila embryo (Liang and Biggin, 1998). Therefore, numerous HD proteins binding throughout a chromatin region might be important to communication between enhancers and promoters. The Chip protein, originally isolated in Drosophila, plays a vital role in the control of HD transcription factor activities (Morcillo et al., 1997). Chip family proteins can physically interact with HD transcription factors and in Xenopus, the Chip homolog Xhbd1, was found to regulate the DNA binding activity of HD transcription factors (Breen et al., 1998). From these observations, it was suggested that communication between an enhancer and its cognate promoter is mediated by a chain of complexes containing Chip related proteins that are anchored to the intervening chromatin regions by physically interacting with HD containing factors.

However, assembly of the facilitator complex on the intervening sequence between the enhancer and the promoter could involve at least as great an energy expenditure as would the formation of direct enhancer-promoter

contacts. Furthermore, studies showed that the mouse chip protein, LDB1, can form a complex with GATA-1, Tal-1 and LMO2. The complex occupies both the LCR and the promoters of the mouse *β-globin* gene when the gene is actively transcribed (Song et al., 2007; Wadman et al., 1997), indicating that Chip family proteins can facilitate the formation of enhancer-promoter loops through interactions with other transcription factors.

1.10 Tracking

The original view of enhancer function is that it provides specific binding sites for RNAPII and other components of the transcription machinery, followed by tracking of these factors on the chromatin fibre until they encounter their correct core promoter (Figure 1.2). Tracking-like mechanisms are supported by studies of several loci. For instance, some studies have revealed unidirectional spreading of H3K4ac, CBP/p300 acetyltransferase, or RNAPII and TBP with accompanying synthesis of intergenic non-coding RNAs at the β-globin locus (Kim and Dean, 2004; Wang et al., 2005; Zhao and Dean, 2004; Zhu et al., 2007). Furthermore, insertion of a terminator or insulator between the enhancer and the promoter traps RNAPII and blocks the corresponding long-range enhancer-promoter interactions. These studies implicate tracking as the primary step for enhancer-promoter interactions and raise the possibility that a stable enhancer-promoter loop is only formed when the tracking step is completed. Notably, the tracking process will not alter the proximity between the enhancer and promoter.

1.11 Facilitated tracking

The facilitated tracking model incorporates aspects of both the looping and tracking models and suggests that the juxtaposition of enhancers and promoters represents only the final stage in enhancer-promoter interactions (Figure 1.2). This model is well documented for the human *ε-globin* gene. Transient transfection studies showed that HS2 enhancer complex, containing the enhancer DNA region, RNAPII and TBP tracks along the intervening chromatin fibre, synthesizing short, polyadenylated, intergenic RNAs before eventually looping to the *ε-globin* promoter (Zhu et al., 2007). An insulator

inserted in the intervening chromatin fibre between enhancers and their cognate promoter traps the enhancer DNA and associated RNAPII and TBP at the interfering site, impeding facilitated tracking mid-stream and finally blocking long-range enhancer function (Zhu et al., 2007).

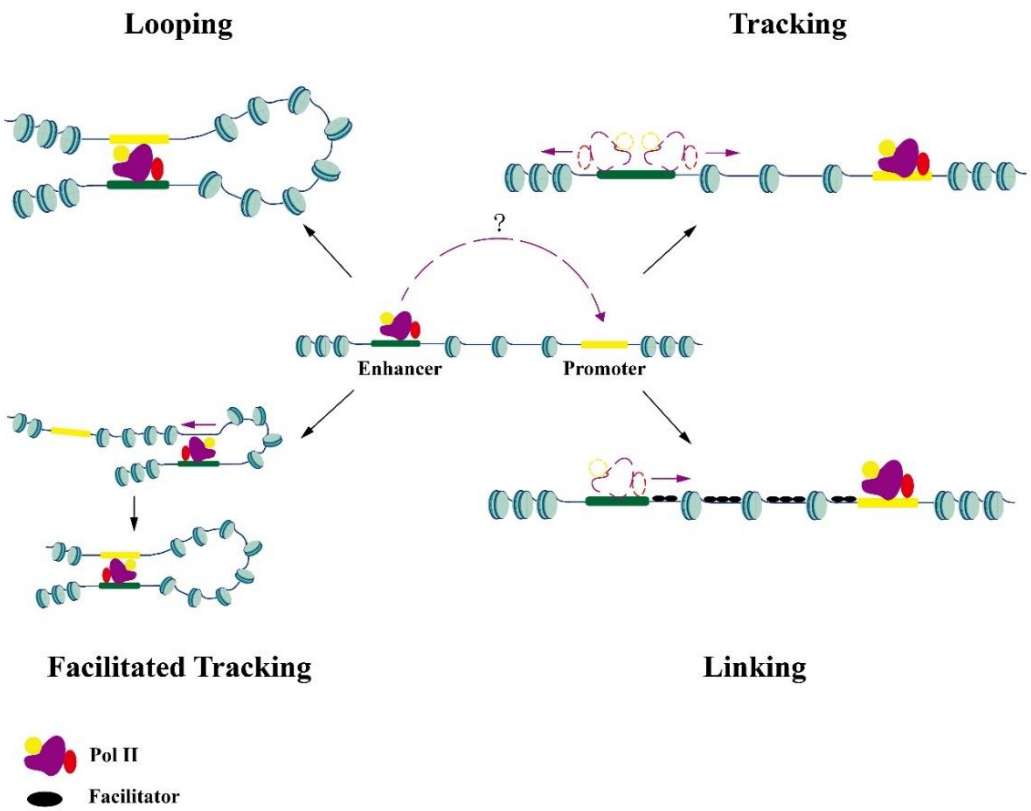

Figure 1.2 – Existing models for enhancer-promoter interactions

In the linking model, “facilitator” proteins are initially recruited at enhancers and then spreads out towards cognate promoters. In the tracking model, RNAPII recruited by transcription activators at enhancers may track along the genome DNA towards target promoters. In the looping model, enhancers may be brought into close proximity with cognate promoters due to interactions between enhancer and promoter bound proteins, and this loops out the intervening chromatin region. In the facilitated tracking model, RNAPII tracks along the chromatin DNA towards cognate promoters and the intervening chromatin region between the enhancer and tracking RNAPII is looped out. Figure adapted from Li et al., 2006 and Vernimmen et al., 2015.

D) The physical basis of enhancer-promoter interactions

1.12 Formation of topologically associating domains

Enhancer-promoter interactions have been proven to increase transcription efficiency at different gene loci. Chromatin organization is essential for the communication between cis-acting sequence elements and many enhancers and their corresponding promoter in mammalian genomes are arranged into a topological associating domain (TAD). It has demonstrated that physical contacts between genomic segments within the same TAD are relatively frequent, whereas chromatin interactions across TAD boundaries occur relatively infrequently (Dixon et al., 2012). TADs can be further partitioned into smaller units of insulated neighbourhoods (Figure 1.3) (Hnisz et al., 2016a). There are approximately 13,000 insulated neighbourhoods in human embryonic stem cells, ranging from 25 kb to 940 kb in length with each containing at least one gene (Dowen et al., 2014). The majority of enhancer-promoter loops (approximate 90%) are fully constrained within the insulated neighbourhood boundaries in human ESCs (Hnisz et al., 2016b). Further studies have shown that the insulating property of these boundaries is achieved by the CCCTC binding factor (CTCF) homodimer and its associated cohesin complex (Giles et al., 2010).

CTCF is highly conserved and ubiquitously expressed zinc finger protein in mammals and was initially identified as a transcriptional regulator of the *c-myc* oncogene (Klenova et al., 1993, Filippova et al., 1996). There are approximately 55,000 binding sites for CTCF in mammalian genomes, which are commonly present in nucleosome linker regions, surrounded by well-positioned nucleosomes (Wang et al., 2012). Approximately only 10% of CTCF binding sites are conserved between mammalian species and tissues, whereas around half of these binding sites show a cell-specific distribution (Wang et al., 2012). The majority of CTCF sites have been demonstrated to have insulating properties that facilitate enhancer-promoter specificity within insulated neighbourhoods (Hnisz et al., 2016a). In addition, a minor fraction of CTCF binding (approximate 19% in human ESCs) is enriched at enhancer and

promoter regions themselves, indicating their potential role in the establishment of enhancer-promoter loops (Ong and Corces, 2014b). Notably, CTCF binding polarity is essential for establishing CTCF mediated chromatin structures and the majority of CTCF binding sites that mediate long range interactions, are in a convergent orientation (de Wit et al., 2015). Both deletion or inversion of such CTCF binding sites disrupt chromatin interactions with the mutated CTCF binding site (de Wit et al., 2015).

The cohesin loop complex is an essential constituent of interphase and mitotic chromosomes where one of its roles is to hold sister chromatids together following DNA replication. This multiprotein complex is comprised of four main subunits: Smc1, Smc3, Rad21/Scc1 and SA/Scc3 (Losada et al., 1998). Two of these subunits, Smc1 and Smc3, are members of the Structural Maintenance of Chromosomes (SMC) family which is a class of chromosomal ATPases that regulate different aspects of the three-dimensional chromosomal structure. Each of these two SMC proteins contains one 50 nm-long intramolecular antiparallel coiled-coil, which forms a rod-shaped protein with a globular hinge domain at one terminus and an ATP nucleotide binding domain (NBD) at the other. Heterotypic interactions between the hinge domains of Smc1 and Smc3 result in the formation of V–shaped heterodimer with a Smc1 NBD at the end of one coiled-coil arm and Smc3 NBD at the other end. This V-shaped heterodimer is further complemented by Rad21 and SA to form a ring-shaped structure, which can topologically embrace two chromatin fibres (Nasmyth and Haering, 2009).

Recent evidence has shown that strong cohesin binding sites commonly coincide with the binding of CTCF, indicating that cohesin could be involved in the establishment of boundaries of chromatin domains, such as the insulated neighbourhoods (Hnisz et al., 2016a; Wendt et al., 2008). Moreover, numerous and often weaker cohesin binding sites are present at active enhancers and promoters (Merkenschlager and Odom, 2013). This therefore suggests that the cohesin loop complex could embrace two regions of chromatin fibres that

contain an enhancer and its cognate promoter respectively, thereby strengthening the interactions between enhancers and promoters.

1.13 Characteristic chromatin modifications at enhancers and promoters

Chromatin mediates extensive packaging of genomic DNA via the interaction with basic proteins in the nucleus of eukaryotic cells. The fundamental subunit of chromatin is the nucleosome, which is comprised of 146 bp of genomic DNA wrapped around a core histone octamer (Finch et al., 1977; Luger et al., 1997). The nucleosome octamer is a globular protein complex that made up of two of each core histone, H2A, H2B, H3 and H4, which are arranged into a central H3/H4 tetramer and two more peripheral histone H2A/H2B dimers (Kornberg, 1974). Histone H1 binds to DNA at the entry/exit points of the nucleosome and is essential for the formation of higher order chromatin structures. In addition to the core histones, there are many histone variants, such as H3.3 and H2A.Z, and replacement of core histone proteins with histone variants can impact the chromatin structure and function (Jin et al., 2005).

The “Histone code” hypothesis was initially put forward by Allis in 2000 (Strahl and Allis, 2000) and has been demonstrated to play an essential role in the regulation of genome structure and gene transcription (Bannister and Kouzarides, 2011; Jenuwein and Allis, 2001). The flexible N-terminal histone tails that protrude from the globular nucleosome core are subject to a number of post-translational modifications. These modifications constitute a “histone code” which is written by multiple histone modifying enzymes (for example, histone acetyltransferases) that catalyze the addition of specific chemical modifications, including methyl, acetyl and phosphoryl groups, in a residue-specific manner on histone tails. Conversely, histone modifications can be erased by specific enzymes (for example, histone deacetylases) that remove chemical groups from histone tails. “Reader” proteins translate the histone code into biological outcomes, such as transcription activation or repression, by directly interacting with individual, or combinations of histone modifications.

Histone acetylation and methylation are two of the most well studies post-translational modifications and these can occur at multiple lysine and arginine residues on histone tails. Histones interact with negatively charged DNA through electrostatic attractions to establish a highly condensed chromatin architecture. Acetylation of lysine residues results in charge neutralisation and induces chromatin decompaction (Hizume et al., 2010). The steady state level of histone acetylation is maintained by the combined action of histone acetyltransferases (HATs) and histone deacetylases (HDACs). HATs can be divided into three main families: GCN5-related N-acetyltransferases family (GNAT), MOZ, YBF2/SAS3, SAS2 and TIP60 protein family (MYST) and p300/CBP (Sterner and Berger, 2000). In general, HATs, such as Gcn5, p300 and Tip60, modify lysine residues residing in the N-terminal tails. However, lysine 56, which is located in the core region of histone H3, can also be acetylated (Xu et al., 2005). As lysine 56 resides facing towards the major groove of DNA, acetylation of this residue can affect core histone-DNA interactions. Removal of acetyl groups from histones is catalyzed by HDACs. HDACs can be grouped into three distinct families: class I and class II histone deacetylases, and the third class, the NAD-dependent enzymes of the Sir family. In S. cerevisiae, the class I histone deacetylase complex, Rpd3, containing HDAC1 and HDAC2, is recruited to promoters and leads to localized deacetylation (Wu and Grunstein, 2000). Although class I HDACs are ubiquitously expressed, class II HDACs display tissue-specific expression. For instance, certain class II HDACs are involved in the regulation of muscle differentiation by interacting with members of the myocyte enhancer factor II (MEF2) (Black and Olson, 1998). This interaction is mediated by domains located in the N-terminal region of muscle-specific class II HDACs which are absent in other class II HDACs. The class III Sir proteins have been shown to generate a hypoacetylated state on histone H3 and H4 tails, leading to the spreading of heterochromatin in yeast (Gottschling, 2000).

Lysine residues within histone tails can also be modified by different methyltransferases, leading to the generation of three methylation states,

namely mono-, di-, and tri-methylation (Hyun et al., 2017). Each state may be involved in either transcription activation or repression, depending on which lysine is methylated. As methylation does not neutralize the charge of the lysine residue, its primary function is to recruit other non-histone proteins. Similar to histone acetylation, the state of histone lysine methylation is maintained by the combined action of histone methyltransferases (HMTs) and histone demethylases (HDMs). HMTs are mainly catalyzed by two families of enzymes: the SET domain containing proteins and the Dot1-like proteins. The SET domain contains a catalytic center that is responsible for methyltransferase activity (Xiao et al., 2003). Dot1-like proteins do not contain SET domain but their N-terminal region contains a catalytic core consisting of a S-adenosyl-L-methionine (SAM) binding pocket and a lysine binding channel (Min et al., 2003). Notably, Dot1-like proteins only methylate lysine residues located in H3 due to structural constraints (Min et al., 2003). Removal of methyl groups from histone lysine residues is catalyzed by histone lysine demethylases (KDMs). KDMs are a large protein family that can be structurally divided into two sub-classes. One subclass includes the lysine specific demethylase (LSD)1 and 2. LSD1/2 contains amine oxidase domains that can remove mono- and di-methyl groups from histone lysine residues (Lee et al., 2005; Shi et al., 2004). The other subclass of lysine demethylases are the JmjC domain containing proteins, which remove mono-, di- and tri-methyl groups from specific histone lysine residues via an oxygenase mechanism (Klose et al., 2006).

Reader proteins generally contain evolutionary conserved domains, such as the bromodomain, chromodomain and plant homeodomain (PHD) finger, that specifically recognize certain histone modifications. There are two different reading models for the recognition of histone modifications (Yun et al., 2011). The monovalent recognition model suggests that a single reader domain-histone modification interaction can orchestrate the recruitment of reader protein complexes to their target regions. This is evidenced by the fact that mutations of the reader domains or histone modification sites abolish the proper recruitment of the reader protein complexes. However, a single histone modification can be recognized by different reader complexes and a single

reader protein complex generally contains multiple different recognition modules. This therefore suggests that the multivalent recognition model could predominate in the recognition of histone modifications. Increasing evidence has demonstrated that multivalent recognition is essential for the regulation of the transcription machinery. For instance, the largest subunit of TFIID, TAF1, contains two tandem bromodomains that recognize the dual-acetylated histone peptide, H4K5acK12ac (Jacobson et al., 2000). Another subunit of TFIID, TAF3, contains a PHD finger domain that recognizes H3K4me3. Interestingly, TFIID interacts with H3K4me3 more strongly when it is flanked by a dual acetylation mark, H3K9acK14ac (Vermeulen et al., 2010), indicating the synergistic effect between bromodomains and PHD finger domains during the recognition of histone modifications.

Extensive studies have examined the role of the histone code in the regulation of gene transcription. Knowledge of the specific role of individual histone modifications has enabled their genome distribution to facilitate the identification of cis-acting elements. For instance, H3K4me3 is only present at gene promoters, whereas H3K27ac is specifically enriched at transcriptional enhancers. In addition, generation of histone modifications correlates with PIC assembly. For example, TFIIH is the last GTR to be recruited during PIC assembly (He et al., 2013). Once TFIIH is loaded, the Ser 5 residue of the CTD of RNAPII Rpb1 is phosphorylated by the kinase subunit of TFIIH, CDK7 (Spangler et al., 2001). Ser5 phosphorylated RNAPII subsequently recruits the Set1 COMPASS complex to facilitate the generation of H3K4me3 at gene promoters (Ng et al., 2003). I therefore capitalized on knowledge of these modifications to identify cis-acting elements within the Igλ locus as described in Chapter 4, as well as to elucidate the temporal order of events during promoter activation.

H3K4me3 at promoters

H3K4me3 is an activating chromatin modification that is highly enriched at the core promoters of RNAPII transcribed genes (Bernstein et al., 2005). H3K4me3 is tightly associated with transcription initiation and H3K4me3 can directly

interact with a subunit of TFIID, TAF3, to facilitate the recruitment of TFIID to active promoters (Lauberth et al., 2013). Tri-methylation of H3K4 at core promoters is catalysed by the complex of proteins associated with Set1, the COMPASS complex, which is highly conserved from yeast to human (Herz et al., 2012). In yeast, the COMPASS complex is comprised of eight subunits which include the Set1 methyltransferase and core structural components Swd1, Swd3 and Bre2 (Shilatifard, 2012). Set1 alone is inactive but within the COMPASS complex it can catalyse mono-, di-, and tri-methylation of H3K4 in yeast (Shilatifard, 2012). By contrast, there are six Set1 homologues (Set1A, Set1B, MLL1, MLL2, MLL3 and MLL4) that have been identified in mammals (Shilatifard, 2012). Each of these Set1 homologues can form a methyltransferase complex with similar core components to catalyse methylation at H3K4 (Dou et al., 2006; Steward et al., 2006). Like the yeast Set1 complex, the Set1A and Set1B complexes catalyse addition of the bulk of trimethylation at H3K4 at mammalian promoters (Ardehali et al., 2011; Mohan et al., 2011; Wu et al., 2008). The COMPASS complex is recruited to promoters by the RNAPII elongation machinery due to direct interactions between Set1 and Ser 5 phosphorylated RNAPII as well as the transcription elongation factors, Paf1 and Rtf1 (Ng et al., 2003).

H3K4me1 at enhancers

Analysis of genome-wide histone modifications by ChIP has identified common chromatin marks at enhancers. The chromatin signature that most reliably predicts enhancers is the relatively high level of mono-methylated H3 lysine 4 (Heintzman et al., 2007). In contrast to the relatively sharp peak of H3K4me3 at core promoter regions, H3K4me1 occupancy at enhancers can be very broad, extending a kilobase or more either side of the transcription factor binding region (Heintzman et al., 2007). Mono-methylation at H3K4 is implemented by the MLL3 and MLL4 COMPASS complexes (Shilatifard, 2012). High levels of enrichment of MLL3 and MLL4 at enhancers is attributed to the physical interactions between MLL3/4 and lineage-specific transcription factors bound at enhancers (Lee et al., 2013). The presence of MLL3/4 at enhancers can

further regulate enhancer activities through the recruitment of the acetyltransferase, p300 (Wang et al., 2016).

H3K27ac at enhancers

H3K27 is another important histone signature at enhancers. The presence H3K27ac together with H3K4me1 is associated with a higher level of expression of nearby genes (Creyghton et al., 2010). H3K27 acetylation in mammalian genomes is carried out by two highly similar proteins, CBP and p300, which have both co-activator function as well as histone acetyltransferase activity (Bannister and Kouzarides, 1996; Ogryzko et al., 1996). These two acetyltransferases contain many functional domains that are involved in interplay with several transcription factors, including cyclic AMP response element-binding protein (CREB) and E1A, which are influenced by several cellular signalling pathways, leading to the activation of a variety of genes (Roth et al., 2003). Although CBP and p300 are found at promoter regions, their enrichment at intergenic or intragenic regions can be a useful marker to identify mammalian enhancers.

H3K27me3 at enhancers

H3K27me3 is a repressive histone mark but can be found at a fraction of lineage-specific enhancers in stem cells. This histone modification is implemented by a multiprotein complex, polycomb repressive complex 2 (PRC2) (Di Croce and Helin, 2013). Transcriptional enhancers can be classified as three groups, namely primed, poised and active enhancers, according to their epigenetic states (Calo and Wysocka, 2013; Creyghton et al., 2010; Rada-Iglesias et al., 2011). Primed enhancers only drive basal levels of gene transcription and are characterized by H3K4me1 and p300 binding. In contrast, poised enhancers are not only marked by H3K4me1 and p300 binding but also by H3K27me3 and PRC2 binding. Lastly, active enhancers substantially promote gene expression and are marked by H3K4me1, H3K27ac and p300 binding. Transition from “poised” to “active” requires disassociation of PRC2 and H3K27me3, which may be aided by the H3K27 demethylase UTX (Herz et

al., 2012). K3K27ac is subsequently deposited by the acetyltransferase p300 at enhancers, resulting in the production of “active” enhancers.

1.14 Transcription factors involved in formation enhancer-promoter loops

Pioneer transcription factors

Pioneer transcription factors are capable of disrupting chromatin organization and can bind to their specific binding sites irrespective of nucleosomes, although this does depend on the context of other transcription factors. For instance, the DNA binding domain of the transcription factor, FoxA, resembles that of linker histones H1 and H5, which could thus alter nucleosome structure (Cirillo et al., 2002). Furthermore, the purine-rich transcription factor, PU.1, is involved in the incorporation of the histone variant H3.3 into nucleosomes, leading to an altered nucleosome structure (Wang et al., 2014, Stopka et al., 2005). Likewise, NF-Y, a CCAAT box binding factor, has a core histone-like structure and has been proposed to be involved in the increased accessibility of chromatin fibres by nucleosome replacement and facilitating the binding of numerous master regulators, such as Oct4 and Sox2, to enhancers in ES cells (Romier et al., 2003, Oldfield et al., 2014). Enhancers that have been primed by pioneer transcription factors in specific cell lineages can provide a chromatin landscape that can subsequently control cell-specific responses to other transcription factors that act downstream of generic signalling pathways.

Lineage-specific transcription activators

In contrast to general transcription factors, lineage-specific transcription factors are only expressed in certain cell types and are essential for cell development. Lineage-specific transcription factors can be recruited by pioneer factors to tissue-specific enhancers, leading to enhancer activation and formation of enhancer-promoter interactions (Schoenfelder and Fraser, 2019). For instance, the lymphocyte-specific transcription factor IRF4, and/or the highly-related factor IRF8, are essential for the differentiation of pro-B to pre-B cells (Lu et al., 2003; Muljo and Schlissel, 2003). IRF4 is robustly recruited to B cell-specific

enhancers, such as the immunoglobulin light chain enhancers, via direct interactions with the pioneer transcription factor PU.1 (Bevington and Boyes, 2013).

YY1

YY1 is a ubiquitously expressed transcription factor that belongs to the zinc finger family of DNA binding proteins. Four zinc finger domains that are responsible for genomic DNA binding, are located at its carboxyl terminus. Further domains and motifs, including the REPO domain, the glycine-rich region, the proline-rich region, the glutamine-rich region and the histidine stretch, are situated in the central and N-terminal part of the YY1 (Atchison, 2014) and were shown to interact with diverse transcription-related factors and complexes, such as cohesin and TBP (Pan et al., 2013, Riquet et al., 2001), indicating its potential role in the regulation of gene transcription. Indeed, ChIP analysis showed that numerous YY1 proteins are recruited to enhancer and promoter regions (Sigova et al., 2015). Further studies revealed that YY1 is tightly associated with the formation of chromatin loops in immunoglobulin gene loci as evidenced by a YY1 conditional knock-out which led to a decrease in chromatin looping events (Atchison, 2014). In addition, a more recent study demonstrated that YY1 is capable of binding single stranded RNAs (Wai et al., 2016). Interactions between RNAs encoded by transcriptional enhancers and YY1 is essential for YY1 recruitment to the corresponding regulatory elements and also facilitate enhancer-promoter interactions (Sigova et al., 2015). Together, these data indicate that YY1 could facilitate the long-range communication between enhancers and promoters to affect the expression of target genes.

Mediator

The Mediator complex is a large multiprotein complex that was initially identified in *Saccharomyces cerevisiae* as a vital regulator of gene transcription (Nonet and Young, 1989). Comparative genomics demonstrated that the Mediator complex contains approximately 30 distinct subunits in mammals and that the majority of Mediator subunits are conserved from yeast to humans (Nagulapalli

et al., 2016). Structural studies reveal that the approximately 30 subunits within the Mediator complex can be divided into four distinct modules, termed the head, middle, and tail modules, that form a relatively stable core structure, and the cyclin-dependent kinase (CDK8) module, that is comprised of CDK8 (or its paralog CDK19), cyclin C, MED12 (or MED12-like) and MED13 (or MED13-like) (Plaschka et al., 2016). Notably, the CDK8 module associates reversibly with the core structure of Mediator, leading to the production of two main isoforms, the larger Mediator and smaller Mediator complex, distinguished by the presence or absence of the CDK8 module (Plaschka et al., 2016). Mediator is able to contact RNAPII and general transcription factors via its head and middle modules. These interactions between Mediator and RNAPII are essential for its function in PIC assembly and stimulating RNAPII CTD phosphorylation at gene promoters (Plaschka et al., 2015, Robinson et al., 2016, Esnault et al., 2008, Eychenne et al., 2016). Mediator is also capable of interacting with a number of tissue-specific transcription factors through its different subunits and participates in transmitting regulatory signals from tissue-specific transcription factors to the basal RNAPII machinery (Poss et al., 2013). These physical interactions are mainly established between tissue-specific transcription factors and the Mediator tail module and also explain the recruitment of Mediator to transcription factor bound enhancers (Malik and Roeder, 2010, Allen and Taatjes, 2015). These data together establish a novel model of enhancer-promoter communications: Mediator provides a physical bridge between transcription factors bound at enhancers and components of the PIC bound at promoters. The latest research in yeast further reveals that a single Mediator complex associates with the enhancer and core promoter *in vivo* (Petrenko et al., 2016), indicating that it indeed physically bridges these *cis*-acting elements.

Enhancer RNAs

Transcriptional enhancers are short regulatory genomic regions that were first demonstrated to be transcribed by the RNAPII machinery in 2010 via genome-wide transcriptome analysis (De Santa et al., 2010, Kim et al., 2010). The products of enhancer transcription - enhancer RNAs - are a subclass of non-coding RNAs. Enhancer RNAs are synthesized by active enhancers which are

characterized by enrichment of specific histone modifications, binding of lineage-specific transcription factors and enrichment of the RNAPII machinery (Shlyueva et al., 2014, Kaikkonen et al., 2013). Enhancer RNAs were originally demonstrated to be bidirectionally transcribed and non-polyadenylated RNAs (Kim et al., 2010). Subsequent studies identified a few enhancer RNAs that are unidirectionally transcribed and polyadenylated (Koch et al., 2011). Whilst unidirectional and bidirectional enhancer RNAs are both transcribed by RNAPII machinery, the 3' ends of enhancer RNAs are processed by different protein complexes. Similar to mRNAs, the 3' ends of unidirectional enhancer RNAs are processed by the cleavage and polyadenylation specificity factor (Mandel et al., 2006, Murthy and Manley, 1995). By contrast, the bidirectional enhancer RNAs contain 3' end processing signals (3' box) that are recognized and processed by the Integrator complex (Lai et al., 2015). Enhancer RNAs do not work in isolation and they exert functions via interacting with different RNA binding proteins. For example, enhancer RNAs have been demonstrated to be involved in the regulation of chromatin accessibility through interacting with the acetyltransferase, p300 (Bose et al., 2017). By contrast, enhancer RNAs can display inhibitory effects on the establishment of open chromatin via interacting with the polycomb repressive complex (Rinn et al., 2007, Yuan et al., 2012). Enhancer RNA binding partners also include architecture factors, such as Mediator (Lai et al., 2013), cohesin (Tsai et al., 2018), and YY1 (Sigova et al., 2015). Depletion of enhancer RNAs leads to the decrease in target gene transcription, which is accompanied by a reduced enrichment of architecture factors at enhancers and promoters and disruption of enhancer-promoter interactions (Tsai et al., 2018, Lai et al., 2013, Sigova et al., 2015). In addition, enhancer RNAs can directly interact with RNAPII machinery to activate gene transcription. For instance, enhancer RNAs interact with p-TEFb and NELF to activate pause-release of RNAPII and facilitate gene transcription (Shii et al., 2017, Schaukowitch et al., 2014).

Integrator

Integrator complex is a large multi-subunit protein complex that possesses catalytic RNA endonuclease activity, which is required for 3' end processing of

non-polyadenylated, RNAPII dependent, uridylate-rich and small nuclear RNA transcripts, including enhancer RNAs (Rienzo and Casamassimi, 2016). Proteomic analysis demonstrated that the Integrator complex consists of at least 14 subunits in humans (INTS1 through INTS14) (Baillat and Wagner, 2015, Chen et al., 2012). The most common predicted motifs within the Integrator complex are alpha-helical repeats, such as HEAT, ARM and VWA domains, indicative of protein-protein interaction surfaces (Rienzo and Casamassimi, 2016). Evolutionary analysis showed that INTS11 shares substantial sequence homology with CPSF-73 (cleavage and polyadenylation specificity factor, 73 kDa) which is the endonuclease subunit for CPSF and which is responsible for 3’ end processing of pre-mRNAs (Millevoi and Vagner, 2010, Romeo and Schumperli, 2016, Wu et al., 2017). Likewise, INTS11 is the endonuclease of the Integrator complex and contains the catalytic activity for the cleavage reaction at the 3’ ends of enhancer RNAs (Lai et al., 2015). An increasing number of publications demonstrated that apart from its role in 3’ end processing of non-coding RNAs, Integrator is tightly associated with the RNAPII machinery at promoter proximal regions as well as in the establishment of chromatin contacts. For example, Ser 7 phosphorylation of RNAPII by TFIIH has been shown to be essential for interactions with Integrator at promoter proximal regions of snRNA genes (Akhtar et al., 2009, Baillat and Wagner, 2015). Integrator can be also recruited to promoters of protein-coding genes by the negative elongation factors, NELF and DSIF, via direct interactions (Stadelmayer et al., 2014, Skaar et al., 2015). Depletion of Integrator does not change the level of binding of NELF and DSIF at gene promoters but instead leads to the disruption of RNAPII pause release (Skaar et al., 2015, Stadelmayer et al., 2014). Moreover, Integrator facilitates transcription of immediate early genes via recruiting the super elongation complex, which is a large multi-subunit protein complex comprising of p-TEFb and other elongation factors (Jonkers and Lis, 2015), to promoters of the corresponding genes in HeLa cells following activation by epidermal growth factor (EGF) (Gardini et al., 2014). Furthermore, depletion of Integrator disrupts the recruitment of components of the super elongation complex, such as ELL2 and AFF4, to EGF responsive genes, leading to decreased transcription (Gardini et al., 2014).

Integrator has also been shown to be involved in the establishment of correct enhancer-promoter contacts during cell development. For example, knock-down of subunits of Integrator complex abolishes chromatin contacts between promoters of immediate early genes and their corresponding enhancers in HeLa cells, following activation with EGF (Lai et al., 2015).

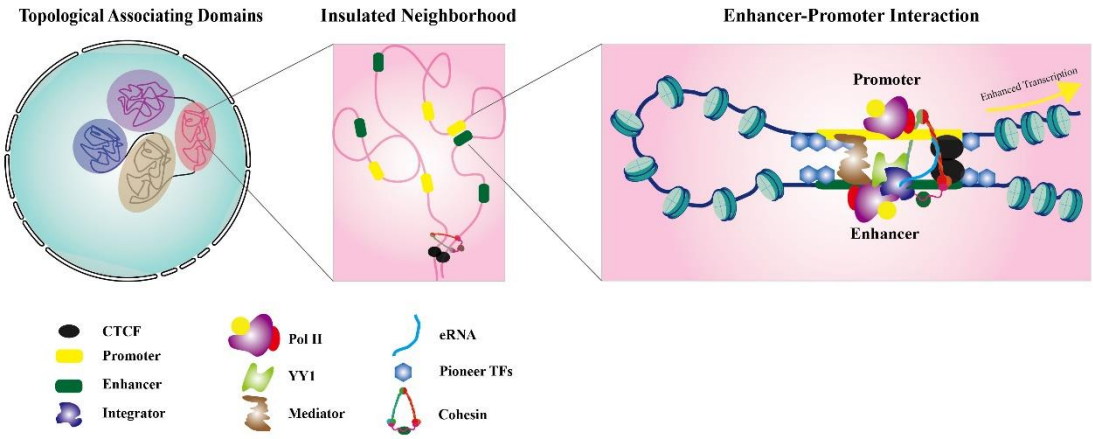

Figure 1.3 – Physical contacts between an enhancer and its cognate promoter occur when the gene is transcribed

Pioneer transcription factors bound at cis-acting sequences lead to increased chromatin accessibility. The Mediator and cohesin complexes are involved in the stability of enhancer-promoter loops. Some enhancer RNAs facilitate the looping through interactions with subunits of the cohesin complex. YY1 is tightly associated with the loop formation, probably through interactions with subunits of the cohesin complex and the RNAPII complex. Adapted from Kim et al., 2015 and Hnisz et al., 2016a.

E) Activation of antigen receptor loci is tightly coupled with early B cell development

1.15 Overview of B and T cell development

Pluripotent hematopoietic stem cells (HSCs), present in the bone marrow are capable of giving rise to all blood cell lineages, including common lymphoid progenitors (CLPs). CLPs then differentiate to the progenitors of B, T and natural killer (NK) cells (Kondo et al., 1997). The generation of B cells in the bone marrow critically relies on the expression of the transcription factor Pax5

which is termed as the guardian of B cell identity and cements commitment to the B cell lineage (Nutt et al., 1999). Activation of antigen receptor loci is intrinsically associated with B cell development as V(D)J recombination and expression of a pre-BCR are essential for the proliferation and survival of B cell progenitors (Rolink et al., 2000). V(D)J recombination first occurs at the IgH locus with D_H to J_H recombination in intermediate pro-B cells, followed by V_H to D_H joining at the late pro-B cell stage (Figure 1.4). Once the heavy chain V-D-J rearrangement has occurred, the μ chain product is expressed and forms the pre-BCR complex with surrogate light chains, VpreB and λ5 (Martensson and Ceredig, 2000). Expression of the pre-BCR complex on the cell surface leads to a cascade of signalling events, stimulating the expansion of large pre-B cells (Geier and Schlissel, 2006). Following differentiating into small pre-B cells, recombination is initiated at one of the two light chain loci, Igκ and Igλ (Gorman and Alt, 1998). Once either Igκ or Igλ has successfully recombined, the rearranged light chain product pairs with the μ chain to form the B cell receptor, IgM, which is subsequently displayed on the cell surface of immature B cells (Vale et al., 2015).

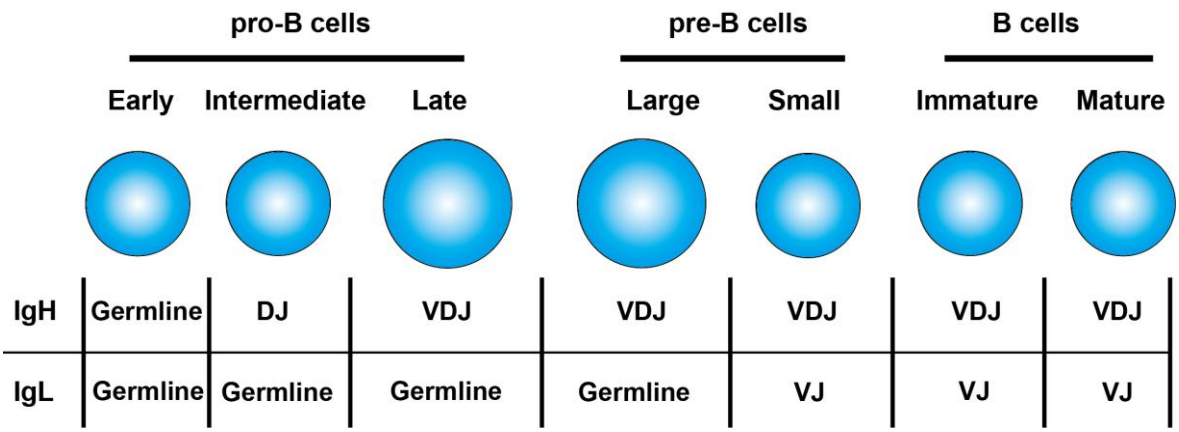

Figure 1.4 – Overview of B cell development

The rearrangement status of V, D and J gene segments at the IgH locus and IgL loci at different stages of B cell development are shown (Hardy et al., 1991).

Commitment to T cell lineage occurs in the thymus and is dependent on the expression of the transcription factor Notch 1 (Pui et al., 1999). Most T cells

express the TCRαβ receptor while the remaining ~5% of T cells express the TCRγδ receptor. Similar to rearrangement events that occur in developing B cells, recombination of TCR loci is coupled with T cell differentiation (Krangel, 2009). During the development of TCRαβ cells, V(D)J recombination initially takes place at the TCRβ locus in pro-T cells at the $CD4^-/CD8^-$ double negative stage, with D_β to J_β joining preceding V_β to D_β-J_β joining. Productive VDJ rearrangements at the TCRβ locus leads to the expression of the β chain. The β chain interacts with CD3 as well as the surrogate α chain to form the pre-T cell receptor (pre-TCR). Pre-TCR expression on T cell surface results in cell expansion and differentiation to $CD4^+/CD8^+$ double positive T cells (pre-T cells). Recombination then occurs at the TCRα locus in pre-T cells and leads to the generation of an α chain that pairs with the β chain to form the TCRαβ cell receptor.

1.16 Mechanism of activation of antigen receptor loci

Activation of V(D)J recombination at antigen receptor loci is the first step in the generation of a highly diverse set of antigen receptor genes, namely the immunoglobulin (Ig) and T cell receptor (TCR) genes, that enable vertebrates to combat a vast range of potential pathogens. There are seven antigen receptor loci in mammalian genomes, the IgH, Igκ and Igλ light chain loci, as well as the TCRα, β, γ, and δ loci. Ig and TCR loci contain many copies of the variable (V), diversity (D) and joining (J) gene segments. Each locus has multiple discontinuous V and J gene segments, and the IgH, TCRβ and δ loci additionally contain D gene segments located between the V and J segments. During V(D)J recombination, one each of the V, D, and J gene segments are joined to create a variable region exon, which encodes the antigen binding portion of the receptor. All V, D and J gene segments are flanked by conserved recombination signal sequences (RSSs), which consist of conserved heptamer and nonamer sequences, separated by a 12 or 23 bp non-conserved spacer. Efficient recombination only occurs between gene segments that are flanked by RSSs with different spacer lengths (Tonegawa, 1983). The V(D)J recombination reaction is initiated by two lymphocyte-specific proteins, RAG1

and RAG2, that directly bind to the RSSs to form a synaptic complex, thereby bringing two recombining gene segments into close proximity (Curry et al., 2005). A double-stranded DNA break is introduced precisely at the heptamer/RSS junction (Curry et al., 2005) and results in covalently sealed hairpin structures at the coding ends, whilst the signal ends are blunt and 5' phosphorylated. These ends are then ligated by the classical non-homologous end joining pathway (cNHEJ) (Malu et al., 2012) to form the excised signal circle (ESC). The coding ends are extensively processed to add and delete nucleotides before they too are ligated by the cNHEJ pathway.

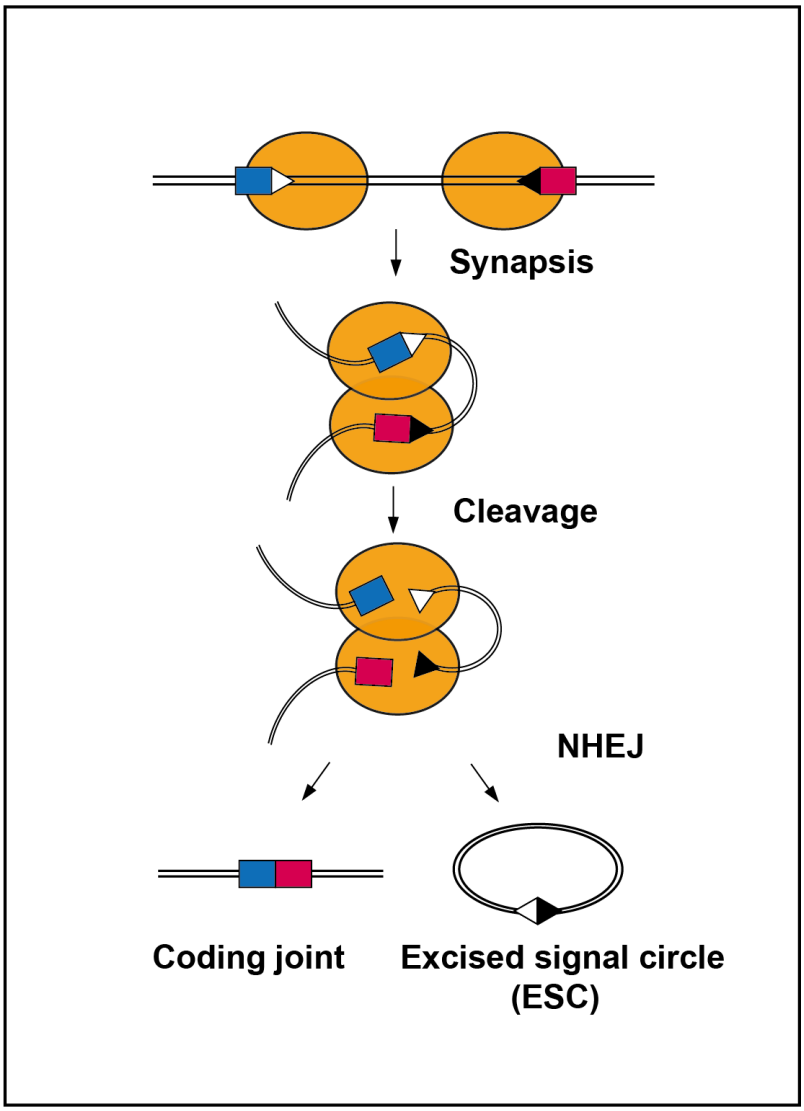

Figure 1.5 – The V(D)J recombination reaction

The blue and red rectangles represent the recombining gene segments. RSSs, depicted by white and black triangles, are bound by the RAG1 and RAG2 complex (orange circles) and then brought together to form a synaptic complex. A double-stranded DNA break is introduced at the heptamer/coding region junction by the recombinase. The coding ends are sealed to generate a hairpin structure, whilst the signal ends are blunt. These ends are then ligated by the non-homologous end joining pathway (NHEJ). Adapted from Arnal and Roth (2007).

Regulation of RAG expression

V(D)J recombination is regulated in a strict lymphocyte lineage-specific manner as RAG1 and RAG2 are only expressed in developing B and T cells. Moreover, rearrangement events occur at different stages of lymphocyte development and

this can be partially explained by the modulation of RAG expression levels during cell differentiation (Grawunder et al., 1995). RAG1 and RAG2 are intronless genes that lie next to each other on mouse chromosome 2 with the Erag enhancer regulating both. Regulation of expression is achieved by communication between the *RAG* gene promoters and the Erag enhancer via a number of lymphoid specific transcription factors such as FOXO1, E2A, NF-Y, LEF-1, Ikaros, PAX5, and GATA (Hsu et al., 2003; Kuo and Schlissel, 2009). During early B and T cell development, there are two separate waves of RAG expression, that correspond to recombination (Grawunder et al., 1995; Wilson et al., 1994). The first wave of RAG expression occurs at the pro-B stage and CD4-/CD8- double negative pro-T stage to enable recombination of the IgH and TCRβ loci, respectively. Following the formation of the pre-BCR or pre-TCR complex and subsequent signalling from the cell surface, RAG expression decreases substantially prior to a period of rapid cell expansion of large pre-B or $CD4^+/CD8^+$ double positive pre-T cells (Grawunder et al., 1995). Upon exit from the cell cycle, the second wave of RAG expression occurs to enable rearrangement of the Ig light chain and TCRα loci. Following this, the expression of IgM reduces the level of RAG expression in immature B cells and positive selection results in decreased RAG expression in T cells (Brandle et al., 1992). RAG expression can also be upregulated in B cells at later stages of development if the BCR recognizes a self-antigen, to enable continued rearrangement of the Ig light chain loci until a functional BCR without auto-reactivity generated (Nemazee, 2006).

RAG activity is also regulated by post-translational modification. The cyclin dependent kinase 2 (CDK2)/CyclinE complex has been demonstrated to phosphorylate RAG2 at threonine-490, leading to the degradation of RAG2 via the ubiquitin-proteasome pathway (Lee and Desiderio, 1999; Li et al., 1996). The CDK2/Cyclin E complex mainly exerts its function at the G1 to S phase transition and remains active during S phase (Geng et al., 2003). This suggests that RAG2 activity is restricted to the G0 and G1 phases of the cell cycle (Lee and Desiderio, 1999). Because both RAG1 and RAG2 are required to catalyse the V(D)J recombination reaction, this restricts recombination events to the G0

and G1 phases of cell cycle. This cell cycle regulation of RAG2 ensures that RAG-mediated DNA double-stranded breaks (DSBs) are not generated during the DNA replication and cell division which may result in abnormally repaired DNA breaks or segregation of broken chromosomes.

Regulation of RSS accessibility

Lymphoid-specific RAG expression explains how V(D)J recombination is restricted to different stages of B and T cell development. However, it cannot explain how specific antigen receptor loci only undergo rearrangement at a certain stage of cell development. The ordered regulation of recombination can be explained by “accessibility hypothesis”; this suggests that RSSs only become accessible for recombination at the correct developmental stages and in the correct lymphocyte types (Yancopoulos and Alt, 1985). Evidence to support this hypothesis came from studies where the nuclei from B and T cells at different developmental stages were isolated and subjected to *in vitro* RAG cutting: RAG-mediated DNA cleavage only occurred at RSSs of antigen receptor loci that were rearranged at the respective stage of B and T cell development (Stanhope-Baker et al., 1996). In addition, nucleosomes have been shown to repress V(D)J recombination by occluding RSSs via association with the histone octamer; this leads to the RSS being inaccessible to RAG binding (Golding et al., 1999; Kwon et al., 1998; McBlane and Boyes, 2000). Therefore, nucleosome remodelling is required to cause RSSs to become accessible to the RAG machinery prior to recombination (Bevington and Boyes, 2013).

Histone modifications

As described previously, histone modifications such as acetylation and methylation are essential for the regulation of chromatin architecture and gene transcription. Histone acetylation is involved in the regulation of V(D)J recombination as loci undergoing rearrangement show an enrichment for histone H3 and H4 acetylation. For example, during recombination of the TCRα locus, the nucleosomes associated with the gene segments that are undergoing rearrangement are marked by increased levels of histone 3 acetylation

(McMurry and Krangel, 2000). Moreover, histone acetylation has been shown to be associated with stage-specific regulation of V(D)J recombination. In early pro-B cells, rearrangement occurs firstly between D and J gene segments that are hyperacetylated. By contrast, V gene segments only become acetylated once D to J rearrangements is complete (Chowdhury and Sen, 2001). Mechanistically, histone acetylation has been shown to facilitate the activation of antigen receptor loci via increasing the chromatin accessibility to the recombinase (Nightingale et al., 2007).

Association between histone methylation and the activation of antigen receptor loci was first confirmed by the enrichment of H3K4me3 at gene segments that are actively involved in rearrangement at the IgH and TCRβ loci (Morshead et al., 2003). Subsequent, and more detailed studies, revealed that deposition of H3K4me3 is strongly associated with stage-specific initiation of V(D)J recombination (Fitzsimmons et al., 2007; Goldmit et al., 2005; Perkins et al., 2004). The links between this histone modification and the RAG machinery was confirmed by the discovery of physical interactions between H3K4me3 and the RAG2 PHD finger domain (Elkin et al., 2005; Matthews et al., 2007; Ramon-Maiques et al., 2007). Binding of RAG2 to this histone mark facilitates the recruitment of RAG2 to the RSSs of gene segments that are to be rearranged (Ji et al., 2010; Schatz and Ji, 2011) and also induces structural changes of the RAG complex, resulting in increased RAG binding and cleavage activities (Bettridge et al., 2017; Lu et al., 2015).

Non-coding transcription

The accessibility hypothesis mentioned above suggests the activation of antigen receptor loci coincides with non-coding transcription through the unrearranged gene segments (Yancopoulos and Alt, 1985). This was supported by the finding that initiation of V(D)J recombination at the Igκ locus correlates with the activation of transcription through unrearranged gene segments (Schlissel and Baltimore, 1989). Furthermore, blocking non-coding transcription of the Jα61-50 gene segments in the TCRα locus via targeted insertion of a transcription terminator led to the complete elimination of

recombination of these specific gene segments (Abarrategui and Krangel, 2006). It is thought that RNAPII mediated non-coding transcription within antigen receptor loci facilitates V(D)J recombination through recruiting RAG proteins. For instance, the histone methyltransferase Set1 is recruited by RNAPII machinery and this can deposit H3K4me3 at these actively transcribed regions (Ng et al., 2003), to facilitate recruitment of the RAG machinery via physical interactions between H3K4me3 and the RAG2 PHD finger domain (Abarrategui and Krangel, 2009; Bettridge et al., 2017; Matthews et al., 2007). In addition, the RNAPII machinery also interacts with histone acetyltransferases to facilitate the deposition of acetylation on histone tails in actively transcribed regions (Wittschieben et al., 1999), which will lead to the generation of a more open chromatin architecture. This, however, cannot explain how nucleosomes are remodelled to enable RAG access to the target RSSs. Further studies proposed a model of how the RSSs that are occluded by nucleosomes are transiently released for cutting by the RAG machinery. Specifically, non-coding transcription could lead to the transient eviction of a H2A/H2B dimer that in turn releases 35-40 bp of nucleosomal DNA. If the RSS lies within this released region, then this will enable RAG binding and cleavage of the RSS; ChIP and *in vivo* accessibility studies provided evidence for this model (Bevington and Boyes, 2013). Notably, transcriptional enhancers have been shown to be essential for both non-coding transcription and antigen receptor locus recombination (Krangel, 2003). Consequently, locus activation critically relies on activation of these enhancers that is mediated via lineage-specific transcription factors.

1.17 Transcription factors involved in the activation of antigen receptor loci

PU.1

PU.1 belongs to the Ets family of transcription factors and is required for the development of cells of haematopoietic lineage (Scott et al., 1994; McKercher et al., 1996). Transgenic mice containing a homozygous mutation in the DNA binding domain of PU.1 are embryonic lethal, lacking mature macrophage,

neutrophil, B and T cells (McKercher et al., 1996). PU.1 has been shown to be important for the activation of both heavy and light chain immunoglobulin loci as PU.1 binds to the IgH intronic enhancer (Nelsen et al., 1993; Rivera et al., 1993), the Igκ and Igλ light chain enhancers (Eisenbeis et al., 1993; Pongubala et al., 1992) and at specific promoters of unrearranged gene segments (Shin and Koshland, 1993). Specifically, PU.1 binds the Eμ enhancer and activates transcription and chromatin accessibility of the IgH locus in cooperation with other transcription factors (Nelsen et al., 1993; Rivera et al., 1993). At the light chain loci, PU.1 binds to the Igκ enhancer κE3' and the two Igλ enhancers Eλ2-4 and Eλ3-1 enhancers in conjunction with IRF4 to stimulate activation of light chain recombination at the pre-B cell stage (Eisenbeis et al., 1995; Pongubala et al., 1992).

IRF4

IRF4 is a lymphoid-specific transcription factor that belongs to the interferon regulatory factor family. IRF4 was initially demonstrated to be critical at the late stages of lymphocyte development as knock-out of the *IRF4* gene in mice by homologous recombination leads to a defect in late B and T cell function (Mittrucker et al., 1997). Whilst these *IRF4* knock-out animals can still produce surface immunoglobulins and T cell receptors, these mice cannot generate antibodies in response to pathogens (Mittrucker et al., 1997). Later studies revealed that IRF4 is also essential for control of pre-B cell development. For example, B cell development is blocked at the large pre-B cell stage in IRF4$^{-/-}$/IRF8$^{-/-}$ mice, and this is accompanied by a disruption of sterile transcription and recombination at the light chain loci (Lu et al., 2003). As IRF4 expression increases at the pro-B to pre-B transition and IRF4 binding sites are located in the κE3', Eλ2-4 and Eλ3-1 enhancers within Ig light chain loci (Eisenbeis et al., 1995; Pongubala et al., 1992), it is highly likely that IRF4 plays an essential role in the activation of Ig light chain loci. In support of this hypothesis, IRF4 has been demonstrated to induce chromatin modifications and activation of Igκ germline transcription (Johnson et al., 2008; Lazorchak et al., 2006). Also consistent with these findings, a more recent study from Boyes lab showed that equipping the pro-B cells with a pre-B level of IRF4 is sufficient to activate the

non-coding transcription and V(D)J recombination at the Igλ locus (Bevington & Boyes, 2013).

Notably, IRF4 alone has only a minimal affinity to its binding motifs because of the presence of an autoinhibitory domain (Eisenbeis et al., 1995) and requires physical interactions with PU.1 to bind strongly to its recognition sites within both Igκ and Igλ enhancers (Eisenbeis et al., 1995; Pongubala et al., 1993). Interactions have also been observed between IRF4 and E2A and binding of the IRF4-E2A complex to the κE3' enhancer was shown to activate the Igκ locus (Nagulapalli and Atchison, 1998; Nagulapalli et al., 2002). Furthermore, depletion of IRF4 leads to a reduced level of recruitment of E2A to enhancers (Lazorchak et al., 2006). Likewise, binding motifs for IRF4 and E2A are also present at the IgH intronic enhancer and binding of the complex here can induce IgH sterile transcription (Nagulapalli and Atchison, 1998).

E2A

E2A belongs to the basic helix-loop-helix (bHLH) class of transcription factors and plays a critical role in the regulation of early B cell development. Knockout of E2A results in a complete block of progression of B cell development beyond the pro-B cells stage in mice, which fail to initiate D_H to J_H recombination (Bain et al., 1994; Zhuang et al., 1994). The *E2A* gene encodes two protein products, E12 and E47 through alternative RNA splicing. These two proteins bind to E-box motifs, which are present in various *cis*-acting elements within antigen receptor loci (Bain et al., 1994; Zhuang et al., 1994). Binding of E2A to these regulatory elements appears to be essential for the regulation of non-coding transcription and V(D)J recombination. This was verified by over-expression of E2A in a murine pre-T cell line which induced activation of non-coding transcription and D to J recombination at the IgH locus (Schlissel et al., 1991). Similarly, E2A is critical for the activation of light chain loci. For instance, over-expression of E2A with RAG1 and RAG2 in a human kidney cell line, BOSC23, triggered recombination not only between D_H to J_H but also between V_κ and J_κ (Romanow et al., 2000). Consistent with this, mutations introduced into E-boxes at IgH and Igκ loci led to decreased levels of V(D)J recombination (Fernex et

al., 1995, Inlay et al., 2004). The functionality of E2A may be attributed to its ability to recruit histone modifiers since E2A can interact with the histone acetyltransferases, p300 and SAGA (Eckner et al., 1996; Ogryzko et al., 1996). Therefore, the binding of E2A at enhancers within antigen receptor loci could facilitate the deposition of active chromatin modifications. In support of this, knockout of E2A in pre-B cells impairs the level of histone acetylation at the κE3' enhancer (Lazorchak et al., 2006).

STAT5

STAT5 represents two highly related transcription factors, STAT5A and STAT5B. Although STAT5A and STAT5B are encoded by two different genes, the proteins share substantial sequence homology at the amino acid level (Grimley et al., 1999). STAT5 activation, which is orchestrated by IL-7 signalling, has been demonstrated to be essential for the control of early B cell survival and for ordered antigen receptor gene rearrangement (Malin et al., 2010). Further studies showed that STAT5 can facilitate the deposition of the repressive histone mark, H3K27me3, at iEκ via recruitment of the polycomb protein enhancer of zeste homolog2 (EZH2), leading to the inhibition of E2A binding at iEκ and reduced Igκ non-coding transcription (Mandal et al., 2011). In addition, STAT5 is capable of decreasing IRF4 binding at Igκ enhancers via displacing PU.1 (Hodawadekar et al., 2012).

1.18 Signalling pathways orchestrating early B cell development

IL-7 signalling

IL-7 is an essential cytokine that has been shown to be involved in the regulation of B cell proliferation, survival and differentiation (Milne and Paige, 2006). The different roles of IL-7 in B cells are achieved through its interaction with the IL-7 receptor which is present on the surface of early B cells including pre-pro-B, pro-B and pre-B cells (Milne and Paige, 2006). STAT5 is phosphorylated and activated in pro-B cells through the Janus kinase which in turn is regulated by IL-7 stimulation (O'Shea and Plenge, 2012). Activated STAT5 facilitates the continued expansion and survival of B cell progenitors via

activating expression of multiple factors involved in cell proliferation and survival, such as cyclin D3, myeloid cell leukaemia sequence 1 (MCL-1) and B cell lymphoma 2 (BCL-2) (Clark et al., 2014). Moreover, IL-7 signalling is important for B cell commitment. IL-7 signalling has been shown to promote the expression of the early B cell factor 1 (EBF1) which is an essential transcription factor for determining B cell lineage fate together with E2A and PAX5. In support of this, overexpression of EBF1 is capable of restoring normal B cell differentiation in IL-7$^{-/-}$ mice (Kikuchi et al., 2005).

IL-7 signalling is also involved in the developmental stage-specific regulation of V(D)J recombination. In pro-B cells, STAT5, activated by IL-7, binds to the IgH locus and facilitates V_H rearrangement, whereas STAT5 binding to the Igκ locus represses its recombination (Bertolino et al., 2005; Johnson et al., 2008). IL-7 signalling is also able to activate phosphoinositide-3-kinase (PI3K) signalling to repress RAG expression via promoting degradation of a transcription activator of RAG expression, FOXO1 (Amin and Schlissel, 2008).

Pre-BCR signalling

Productive rearrangement of the IgH locus leads to the generation of a pre-BCR which is comprised of Igμ, Igα, Igβ and the surrogate light chain proteins, λ5 and VpreB (Bankovich et al., 2007). Pre-BCR signalling is essential for the proliferation, survival and maturation of pre-B cells. Interactions between charged and glycosylated residues within pre-BCR molecules facilitate pre-BCR auto-crosslinking on the cell surface of pre-B cells (Ohnishi and Melchers, 2003). Pre-BCR crosslinking promotes SCR family kinase mediated phosphorylation of tyrosine residues within the immunoreceptor tyrosine-based activation motif (ITAM) of CD79A and CD79B (also known as Igα and Igβ), leading to signal amplification and recruitment of spleen tyrosine kinase (SYK) and ζ chain associated protein kinase of 70 kDa (ZAP70) (Rickert, 2013). These tyrosine kinases, in turn, phosphorylate the B cell linker protein (BLNK) in pre-B cells, which recruits Bruton tyrosine kinase (BTK) and phospholipase Cγ (PLCγ) to facilitate the activation of the extracellular signal-regulated kinase (ERK). This leads to the continued proliferation of pre-B cells (Imamura et al.,

2009). After this proliferative burst, large pre-B cells exit the cell cycle and differentiate into small pre-B cells where BLNK inhibits the activation of AKT and thus promotes the nuclear translocation of the activator of *RAG* genes, FOXO1 (Rickert, 2013). BLNK also represses JAK3-mediated activation of STAT5 (Nakayama et al., 2009) and promotes the expression of transcription factors that are involved in activation of light chain V(D)J recombination such as E2A, IRF4 and Ikaros (Heizmann et al., 2013; Johnson et al., 2008; Lazorchak et al., 2006), allowing light chain recombination to occur.

F) Regulation of long-range chromatin interactions of antigen receptor loci

Tissue-specific long-range chromatin contacts are essential for the regulation of expression of developmentally regulated genes. These developmentally regulated genes are expressed only in the appropriate lineage, whilst they remain silent in other lineages. Haematopoiesis is an excellent system to investigate the regulation of gene transcription as it is very accessible and generates multiple lineages in bone marrow. Also, differentiated lineages and their progenitor cells can be easily separated by using different cell surface markers. B and T cell lineages are of particular interest due to V(D)J recombination of antigen receptor loci which generates a vast antigen receptor repertoire to combat a range of pathogens. Therefore, substantial effort has been expended on understanding the mechanism by which long-range chromatin interactions of antigen receptor loci are regulated to facilitate V(D)J recombination.

1.19 Long-range chromatin organization of the IgH locus

Structure of the murine IgH locus

The murine IgH locus spans approximately 2.8 megabases (Mb) on chromosome 12 and is comprised of 100 functional V_H gene segments, 8-12 D_H gene segments (depending on the mouse strain) and four J_H gene segments (Figure 1.6) (Ye, 2004).

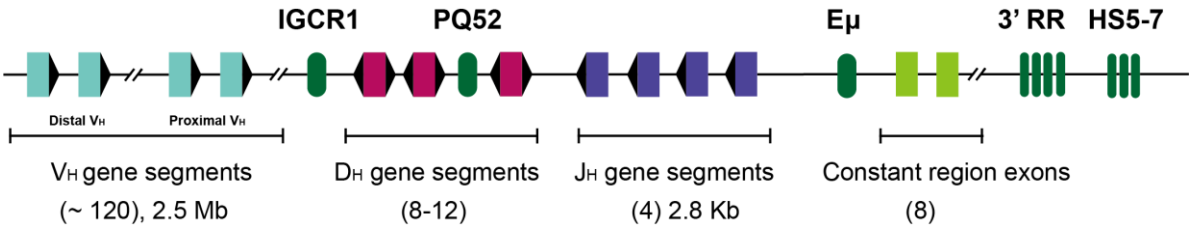

Figure 1.6 – Structure of the murine IgH locus

A schematic diagram of the murine IgH locus. The IgH locus includes approximately 110 functional V_H gene segments (cyan), eight to twelve D_H gene segments (purple) and four J_H gene segments (blue) and eight constant region exons (light green). RSSs are shown as black triangles. Regulatory elements are shown in green. IGCR1 – intergenic control region 1. PQ52 – promoter of 5' DQ52. Eμ – heavy chain enhancer. 3' RR – 3' regulatory region. HS5-7 – hypersensitive sites 5 to 7.

The first regulatory element identified in the IgH locus was the enhancer, Eμ, which resides in the intron between the J_H gene segments and constant gene segments (Figure 1.6) (Gillies et al., 1983). This enhancer is essential for IgH activation as replacement of Eμ with a neomycin resistance gene or a short nucleotide resulted in a decrease in non-coding transcription and rearrangement (Chen et al., 1993; Serwe and Sablitzky, 1993). As a low level of IgH recombination was still observed in Eμ deleted cells, this suggested additional regulatory elements could compensate for the absence of Eμ. Indeed, two additional regulatory elements, Eα and 3'E_H, were found to reside at the 3' end of the IgH locus (Matthias and Baltimore, 1993; Pettersson et al., 1990). However, more recent studies have demonstrated that these 3' enhancers are not required for V(D)J recombination but instead, appear to be essential for class switch recombination and somatic mutation that occur at a later stage of B cell development (Cogne et al., 1994; Rouaud et al., 2013). In addition to enhancers, several genomic regions that contain multiple CTCF binding sites, such as IGCR1 and HS5-7, were identified and appear to be involved in the regulation of long-range chromatin organization of the IgH locus (Figure 1.6) (Guo et al., 2011b).

Regulation of D_H to J_H joining

Rearrangement of IgH initiates with D_H to J_H joining in early pro-B cells, followed by V_H to D_H-J_H joining in late pro-B cells (Figure 1.4) (Kumari and Sen, 2015). Recombination among these gene segments involves ordered chromatin organization that brings distant V, D and J gene segments into close proximity. The rearrangement of D_H to J_H gene segments is mediated by long-range DNA interactions among regulatory elements that lead to the formation of separate chromatin domains (Figure 1.7). Three regulatory elements including IGCR, PQ52 (a promoter 5' of DQ52) and HS5-7 are associated with chromatin organization of the IgH locus. 3C data from pro-B cells initially revealed chromatin interactions of the Eμ enhancer with PQ52 and HS5-7 (Guo et al., 2011b). The Eμ-PQ52 interaction results in the generation of a ~5 kb chromatin loop structure that contains all four J_H gene segments, whereas the Eμ-HS5-7 interaction forms a ~200 kb chromatin loop structure that constrains the constant gene segments (Guo et al., 2011b). Further studies identified another chromatin loop domain formed between IGCR1 and Eμ, which is approximately 70 kb in length and contains the D_H gene segments (Verma-Gaur et al., 2012). Formation of these chromatin loops is a prerequisite for the ordered D_H to J_H joining. The Eμ-PQ52 chromatin domain is believed to be the "recombination centre" and displays the highest level of H3K4me3 and of RAG1/RAG2 binding (Schatz and Ji, 2011; Teng et al., 2015). The Eμ-PQ52 chromatin domain is positioned closest to the PQ52-IGCR1 chromatin domain which contains the D_H gene segments, promoting RSS capture to enable D-J recombination to occur (Figure 1.7).

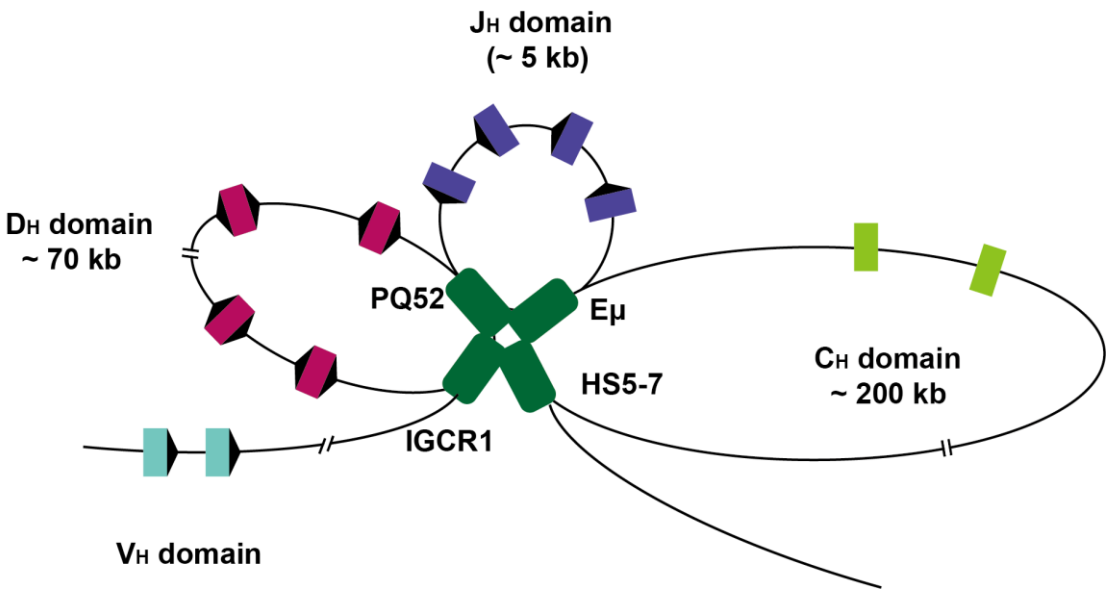

Figure 1.7 – Chromatin loops formed by the Eμ enhancer facilitates DH to JH joining

Chromatin interactions between Eμ and PQ52, IGCR1 and HS5-7 lead to the generation of three chromatin loops in the 3' IgH domain. The smallest J_H domain formed between Eμ and PQ52 is proposed to be the recombination centre for D_H to J_H joining. The D_H domain formed between PQ52 and IGCR1 is positioned close to the J_H recombination centre, facilitating D_H to J_H rearrangement via bringing D_H and J_H gene segments into close proximity. The largest V_H chromatin domain formed between HS5-7 and Eμ includes all constant region exons, but it is still unclear if this domain has a function in D_H to J_H joining. Adapted from Kumari et al., 2015.

Formation of the IGCR1-PQ52 domain is also thought to separate the V_H gene segments from the recombination centre to ensure correct D_H to J_H recombination. In support of this, knock out of IGCR1 facilitates unordered rearrangement of proximal V_H gene segments to unrearranged D_H gene segments (Featherstone et al., 2010). The role of the Eμ-HS5-7 chromatin domain in the regulation of D_H-J_H recombination is less clear as knock out of HS5-7 does not affect the IgH repertoire substantially (Volpi et al., 2012). However, it is possible that other regulatory elements located at 3' end of IgH could potentially compensate for the deletion of HS5-7 (Volpi et al., 2012). In addition, a recent study proposed a model of how D_H to J_H joining is mediated by the cohesion loop complex (Zhang et al., 2019). In this model, the J_H

recombination centre is formed via RAG binding to a J_H 23 RSS. The J_H recombination centre serves as a dynamic sub-loop anchor and scans convergent CTCF binding element (CBE) anchored chromatin loops, potentially formed by cohesion-mediated loop extrusion. This facilitates shortening of the distance between D_H 12-RSSs and the J_H recombination centre, thereby promoting D_H to J_H recombination (Zhang et al., 2019).

Regulation of V_H to D_H-J_H joining

Once D_H-J_H rearrangement is complete, a number of epigenetic alterations occur, specifically at the rearranged D_H-J_H segment, to increase chromatin accessibility and facilitate V_H to D_H-J_H joining (Subrahmanyam et al., 2012). There are more than 100 V_H gene segments spanning 2.5 Mb in the IgH locus. To achieve V_H to D_H-J_H joining, V_H gene segments and rearranged D_H-J_H gene segments must be brought into close proximity and to ensure the production of a diverse antigen receptor repertoire, all V_H gene segments should have an equal opportunity to recombine with the rearranged D_H-J_H gene segment.

Early DNA fluorescence *in situ* hybridisation (DNA-FISH) experiments demonstrated that the IgH locus undergoes large-scale chromatin contraction during V(D)J recombination, which was thought to promote rearrangement between distant gene segments (Kosak et al., 2002). Further studies revealed that the transcription factor PAX5 is essential for IgH locus folding. Depletion of PAX5 results in an extended chromatin configuration of IgH, accompanied by inhibition of rearrangement of distant V_H gene segments to the rearranged D_H-J_H gene segment (Fuxa et al., 2004). Consistent with this, a recent study revealed that cohesin-mediated loop extrusion within CBE anchored chromatin domain is essential for V_H to D_H-J_H rearrangement and PAX5 is capable of regulating cohesin binding via inhibiting expression of the cohesin-release factor, WAPL (Hill et al., 2020).

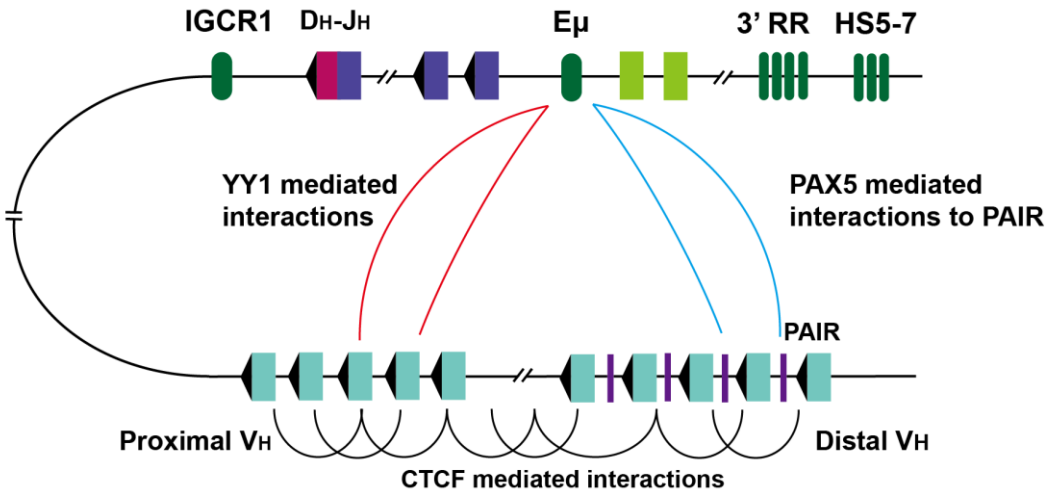

Figure 1.8 – Chromatin interactions mediate V_H to D_H-J_H joining

A schematic diagram of chromatin loops formed at the IgH locus after D_H to J_H joining. Firstly, multiple chromatin loops mediated by CTCF are established within the V domain, which are shown as black lines. Chromatin folding is then mediated by interactions established between Eμ and V_H5'7183 as well as V_H3'558 in the proximal V_H domain, which are dependent on YY1 and shown as red lines. Moreover, PAX5 dependent interactions, shown in light blue, established between Eμ and PAIR elements (purple) function to recruit distal V_H gene segments.

Moreover, previous publications showed that CTCF mediates the formation of a number of chromatin loops between sites within the V_H region (Guo et al., 2011b), whereas YY1 facilitates the establishment of chromatin interactions between the Eμ enhancer and two sites within the V_H domain, including V_H5'7183 and V_H3'558 (Figure 1.8) (Liu et al., 2007). Therefore, it is possible that YY1 facilitates bringing CTCF-mediated chromatin loops formed within the V_H domain to within a close distance of D_H gene segments. In addition, chromatin interactions established within the V_H region are tightly associated with the PAX5-associated intergenic repeat (PAIR) elements. There are ten PAIR elements identified in the distal V_H region, which are highly occupied by multiple transcription factors including PAX5, E2A, CTCF and cohesin (Ebert et al., 2011). Further 3C experiments revealed that the Eμ enhancer forms chromatin loops via interacting PAIR4 and PAIR6 in pro-B cells (Verma-Gaur et al., 2012).

The above data have resulted in the proposal of a model of how V_H to D_H-J_H joining is achieved. Initially, the multiple CTCF mediated chromatin loops are formed within the V_H domain (Guo et al., 2011a; Jhunjhunwala et al., 2008). Next, large scale chromatin loops are established between the 5' end of the V_H chromatin domain to the Eμ enhancer (Guo et al., 2011b; Verma-Gaur et al., 2012). Specifically, these chromatin interactions include YY1 mediated Eμ-V_H5'7183 and Eμ-V_H3'558 (Guo et al., 2011b) as well as PAX5 mediated Eμ-PAX4 and Eμ-PAX6 interactions (Verma-Gaur et al., 2012). Ordered formation of these chromatin loops brings all V_H gene segments to the rearranged D_H-J_H gene segment with equal opportunity to recombine (Kumari and Sen, 2015), thus ensuring the generation of a diverse antigen receptor repertoire.

1.20 Long range chromatin organization of the Igκ locus

Structure of the murine Igκ locus

The murine Igκ locus spans approximately 3.2 Mb on chromosome 6, and is comprised of 160 Vκ gene segments, approximately 100 of which are functional, five Jκ gene segments, four of which are functional, and a single constant gene segment (Figure 1.9). Notably, Vκ gene segments are present in both forward and reverse orientations, thus enabling deletional and inversional rearrangement to occur.

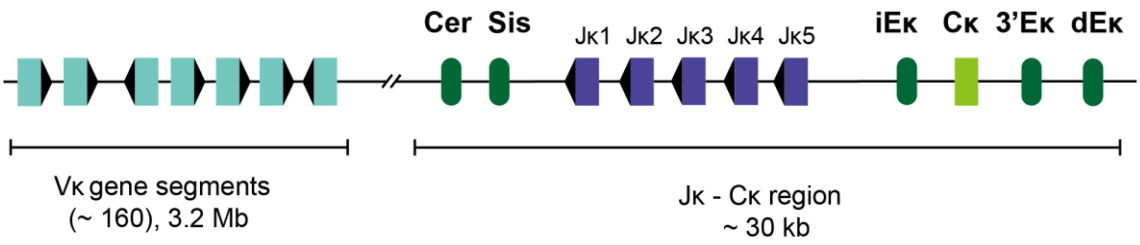

Figure 1.9 – Structure of the murine Igκ locus

A schematic diagram of the murine Igκ locus. The Igκ locus includes approximately 160 functional Vκ gene segments (cyan), five Jκ gene segments (blue) and a single constant exon Cκ (light green). RSSs are shown as black triangles. Regulatory elements are shown in green. Cer – contacting element for recombination. Sis – silencer in the intervening sequence. IEκ – intronic κ enhancer. 3'Eκ – 3' κ enhancer. dEκ – distal κ enhancer.

Regulation of Igκ recombination

V(D)J recombination at the Igκ is orchestrated by two enhancers: the intronic κ enhancer, iEκ, and the 3'κ enhancer, 3'Eκ (Inlay et al., 2002). These two enhancers are activated by diverse transcription factors. Specifically, iEκ contains binding motifs for multiple transcription factors including E2A and NF-κB and is activated by binding of these transcription factors in pro-B cells (Johnson et al., 2008), whereas the 3'Eκ enhancer contains a binding motif for IRF4 which is dramatically increased at the pre-B cell stage (Stadhouders et al., 2014). These studies indicate that the stage-specific activation of rearrangement of Igκ is mainly orchestrated by the 3'Eκ enhancer as large scale chromatin interactions to iEκ are already established in pro-B cells (Stadhouders et al., 2014). Furthermore, a distal κ enhancer (dEκ) was identified which is located downstream of 3'Eκ (Liu et al., 2002). The main function of dEκ appears to be the regulation of mature Igκ transcription and somatic hypermutation at later stages of B cell development (Xiang and Garrard, 2008).

In addition to the three enhancer elements mentioned above, there are two further regulatory elements, named silencer in the intervening sequence (Sis) and contracting element for recombination (Cer), located between Vκ gene

segments and Jκ gene segments, that are occupied by high levels of CTCF (Degner et al., 2009; Xiang et al., 2013). Knock out of Sis leads to increased proximal Vκ rearrangement (Xiang et al., 2011), implying that, similar to the role of IGCR1 in IgH recombination, Sis ensures a diverse Igκ repertoire by inhibiting the Igκ enhancers from contacting the proximal Vκ gene segments. Cer has been demonstrated to be the only regulatory element involved in Igκ locus contraction (Xiang et al., 2013). Deletion of Cer inactivates Igκ locus organization, resulting in a disrupted recombination repertoire where both Jκ and Vκ choice is altered (Xiang et al., 2013).

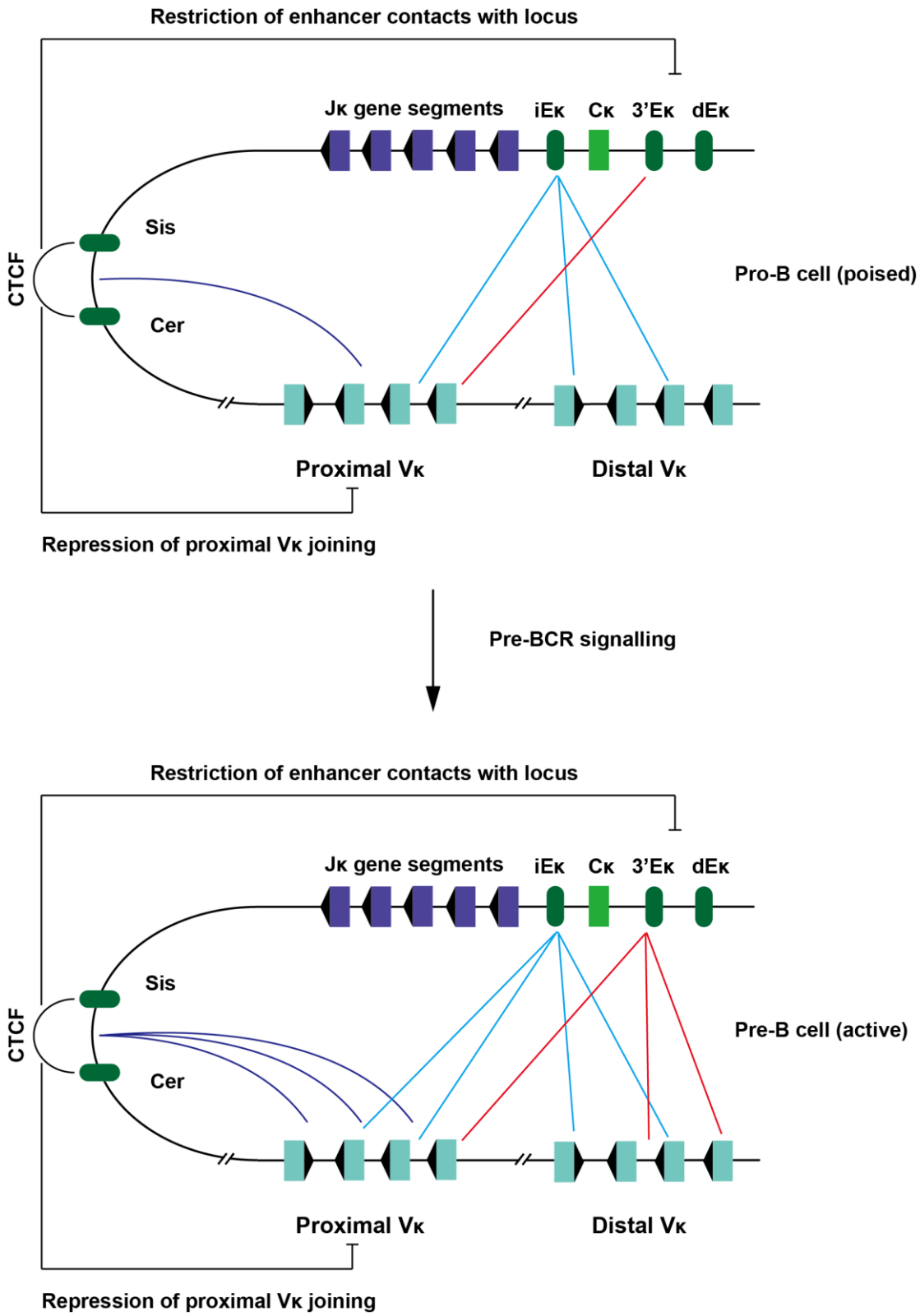

Figure 1.10 – Chromatin interactions within the Igκ locus

A schematic diagram of chromatin interactions established within the murine Igκ locus. Firstly, chromatin loops are formed at the pro-B cell stage, where chromatin

interactions between Vκ domain and iEκ are already established. Furthermore, interactions between the Vκ domain to Sis/Cer as well as 3'Eκ are at a lower frequency in pro-B cells. Upon activation of pre-BCR signalling, chromatin interactions between the Vκ domain and regulatory elements, including Cer, Sis, iEκ and 3'Eκ, increase. In addition, CTCF mediated Cer-Sis interactions repress the joining of proximal Vκ gene segments to Jκ gene segments and restrict enhancer contacts within the locus. Adapted from Ribeiro de Almmeida, 2015a.

More recent studies together have resulted in a proposed model of how the Igκ locus is activated. The Igκ locus is already in a compressed state in pro-B cells and this appears to mediated by chromatin interactions formed between sites bound by E2A and CTCF within the Vκ domain to iEκ, Sis and Cer, leading to a basal level of Igκ rearrangement (Stadhouders et al., 2014; Xiang et al., 2014). Once differentiation occurs from pro-B to pre-B cells, IRF4 is activated by pre-BCR signalling, which then binds to the 3'Eκ enhancer, resulting in the formation of increased chromatin contacts with the Vκ domain and a dramatic increase in Igκ recombination (Stadhouders et al., 2014).

1.21 Long range chromatin organization of the Igλ locus

Structure of the murine Igλ locus

The murine immunoglobulin λ locus is ~230 kb in length; it is the smallest antigen receptor locus and is located on chromosome 16. It consists of three variable (V) and four joining (J) gene segments, where each J gene segment precedes a constant (C) region (Gerdes and Wabl, 2002). The order of these gene segments is shown in Figure 1.11. Expression of the Igλ light chain requires the recombinational joining of V and J gene segments. Approximately 70% of recombination at the Igλ locus occurs between the Vλ1 and Jλ1 gene segments (Boudinot et al., 1994).

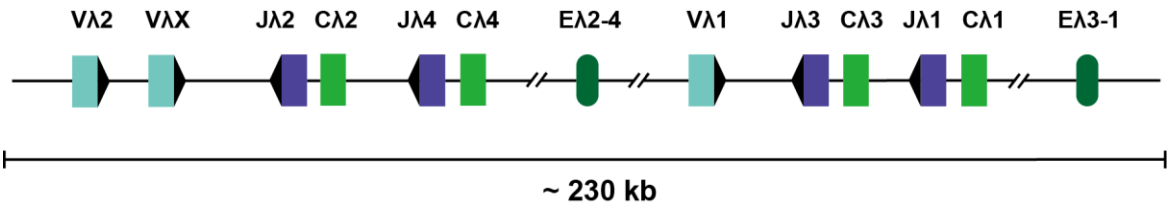

Figure 1.11 – Structure of the murine Igλ locus

A simplified schematic of the murine Igλ locus. The murine Igλ locus includes three functional Vλ gene segments (cyan), four Jλ gene segments (blue) and four constant exons (light green). RSSs are shown as black triangles. Enhancer elements are shown in green.

Regulation of Igλ recombination

Igλ recombination is orchestrated by two enhancers Eλ3-1 and Eλ2-4. These enhancers share approximately 90% sequence homology and contain binding sites for multiple transcription factors, such as PU.1, IRF4 and E2A. A recent study demonstrated that Eλ3-1 is essential for Vλ1-Jλ1 rearrangement (Haque et al., 2013). Although previous studies identified binding sites for structural transcription factors, such as CTCF within the Igλ locus, the mechanism by which Igλ locus folding is regulated, and how this leads to activation of recombination remains unknown.

G) Aims

Enhancers are the most dynamically utilised part of the eukaryotic genome, playing a critical role in the regulation of the majority of developmentally expressed genes. Despite their discovery nearly 30 years ago, and their essential role in regulating huge numbers of genes, exactly how these elements function remains poorly understood. Fundamental questions remain at a number of levels such as how the enhancer finds and commits to the correct promoter and what the enhancer delivers to the promoter to trigger increased transcription output.

An ideal gene model to investigate enhancer-mediated transcription activation is the murine Igλ locus. Approximately 70% of recombination occurs between

Vλ1 and Jλ1 gene segments and activation of non-coding transcription of unrearranged Vλ1 and Jλ1 gene segments is a prerequisite for recombination. Previous data from Boyes lab showed that activation of Vλ1 and Jλ1 transcription is tightly regulated by the Eλ3-1 enhancer. This enhancer contains binding sites for IRF4 and equipping pro-B cells with a pre-B level of IRF4 is sufficient to activate the Vλ1 and Jλ1 gene segments. Therefore, Eλ3-1 mediated activation of Vλ1 or Jλ1 is an excellent system to investigate enhancer-promoter interactions. To temporally dissect the mechanism by which Eλ3-1 mediated transcription activation is triggered by IRF4, I sought to develop and characterise a pro-B cell line that expresses an inducible IRF4, namely IRF4-ER. Using this inducible IRF4 pro-B cell line, I aimed to address the following questions:

1) What do enhancers deliver to promoters to increase transcriptional output?
The traditional view is that enhancers deliver important accessory factors to potentiate either formation of the pre-initiation complex or the transition to elongation. However, the precise details of what an enhancer delivers, in what temporal order, and how this results in increased firing of the promoter, remains incompletely understood. I aim to test the role of IRF4 in mediating the specific Eλ3-1/Vλ1 interaction and to investigate what additional factors are needed to establish Eλ3-1-Vλ1 interactions.

2) How does the enhancer specifically commit to the correct promoter?
Whole-genome analyses have shown that enhancers don't necessarily interact with the closest promoter and, whilst transcription factors are important in mediating the specificity of enhancer-promoter interactions, a number of transcription factors are bound throughout the genome. These need to be avoided to prevent the costly mistake of transcribing the wrong DNA. I aim to investigate how chromatin folding of Igλ is achieved to enable correct enhancer-promoter interactions.

These aims are addressed in the following results chapters. In the first results chapter, an inducible IRF4 pro-B cell line, 1D1-T215, was developed using a

retroviral system. Using this inducible system, I describe key transcription factors delivered to the target gene promoter to activate its transcription. In the next results chapter, I investigate how the correct chromatin environment is established to facilitate correct enhancer-promoter interactions. Finally, in the third results chapter, I investigate how enhancer RNAs encoded by Eλ3-1 are involved in the activation of gene transcription and chromatin organization.

Chapter 2: Materials and Methods

A) Common Buffers

Alkaline lysis buffer I

50 mM	Glucose
25 mM	Tris-HCl (pH 8.0)
10 mM	EDTA (pH 8.0)

Alkaline lysis buffer II

0.2 M	NaOH
1 % (w/v)	SDS

Alkaline lysis buffer III

3 M	KOAc
5 M	Acetic acid

DNA sample loading buffer (6x)

15% (w/v)	Ficoll®-400
20 mM	Tris-HCl (pH 8.0)
60 mM	EDTA (pH 8.0)
0.48% (w/v)	SDS
0.03% (w/v)	Xylene Cyanol
0.03% (w/v)	Bromophenol Blue

Phosphate buffered saline (PBS)

17 mM	NaCl
0.33 mM	KCl
1 mM	Na_2HPO_4
0.18 mM	KH_2PO_4

Adjusted to pH 7.4

Protein sample loading buffer (6x)

6% (w/v)	SDS
50% (v/v)	Glycerol
60 mM	Tris-HCl (pH 6.8)
640 mM	β-Mercaptoethanol
0.06% (w/v)	Bromophenol Blue

Tris-acetate-EDTA buffer (TAE)

40 mM	Tris-acetate
1 mM	EDTA

TE (Tris-EDTA) buffer

10 mM	Tris-HCl (pH 8.0)
1 mM	EDTA (pH 8.0)

TBE (Tris-borate-EDTA) buffer

90 mM	Tris base
90 mM	Boric acid
2 mM	EDTA (pH 8.0)

TBS (Tris-buffered saline) buffer

50 mM	Tris-HCl (pH 7.4)
150 mM	NaCl
2 mM	KCl

TGS (Tris-glycine-SDS) buffer

25 mM	Tris base
192 mM	Glycine
0.1% (w/v)	SDS

Western blot semi-dry transfer buffer

48 mM	Tris base
39 mM	Glycine
20% (v/v)	Methanol
0.04% (w/v)	SDS

B) Media

Dulbecco's Modified Eagle Medium (DMEM) medium

DMEM medium (Sigma D5671) supplemented with:

10 %	Foetal calf serum (PAA Laboratories)
2 mM	L-Glutamine
50 U/ml	Penicillin
50 µg/ml	Streptomycin

Lysogeny Broth (LB)

1 % (w/v)	Bacto-tryptone
0.5 % (w/v)	Yeast extract
0.5 % (w/v)	NaCl

LB Agar

1.5 % (w/v)	Agar
1 % (w/v)	Bacto-tryptone
0.5 % (w/v)	Yeast extract
0.5 % (w/v)	NaCl

Ampicillin or Kanamycin was added to a final concentration of 50 µg/ml to make this media selective.

Pro-B cell medium

McCoy's 5A medium (Gibco 26600-023) supplemented with:

15 %	Foetal calf serum gold (PAA Laboratories)
100 U/ml	Penicillin
100 µg/ml	Streptomycin
50 nM	β-Mercaptoethanol (Added fresh before use)
0.1 % (w/v)	$NaHCO_3$
1 mM	Sodium Pyruvate
1.6 mM	L-Glutamine
0.16 mg/ml	L-Asparagine
0.16 mg/ml	L-Serine
0.4 ×	Essential amino acids (50x stock, Life Technologies, 11130)
0.4 ×	Non-essential amino acids (100x stock, Life Technologies, 11140)
1 ×	Vitamin mix (Life Techology, 11120)
50 µM	β-Mercaptoethanol (Added fresh before use)

RPMI-1640 medium

RPMI-1640 medium (Sigma R8758) supplemented with:

10 %	Foetal calf serum (PAA Laboratories)
2 mM	L-Glutamine
50 U/ml	Penicillin
50 µg/ml	Streptomycin
50 nM	β-Mercaptoethanol (Added fresh before use)

Super Optimal Broth (SOB)

2 % (w/v)	Bacto-tryptone
0.5 % (w/v)	Yeast extract
0.5 % (w/v)	NaCl
0.02 % (w/v)	KCl

Virus Production Medium

DMEM medium (Sigma D5671) supplemented with:

5 %	Foetal bovine serum (Capricorn FBS-12A)
2 mM	L-Glutamine
50 U/ml	Penicillin
50 µg/ml	Streptomycin

C) Manipulation of DNA and RNA

2.1 Conventional polymerase chain reaction (PCR)

A PCR reaction typically contained 1× ThermolPol Reaction Buffer (20 mM Tris pH 8.8, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1 % Triton X-100), 0.2 mM of each dNTP, 1~10 ng of DNA template, 0.2 µM of each primer, and 2 units of Taq DNA polymerase (NEB #M0267S) in a 25 µl reaction volume. Thermocycling conditions were as follows:

95 °C	2 min	
95 °C	30s	30 ~ 35 cycles
Tm	30s	30 ~ 35 cycles
72 °C	1 kb/min	30 ~ 35 cycles
72 °C	10 min	
4 °C	Infinite	

2.2 Real-time PCR using SYBR Green

The PCR reactions were carried out in a Corbett RotorGene 6000 qPCR machine and analysed using the corresponding software. A typical quantitative PCR reaction contained 5 µl 2×SensiFAST SYBR No-Rox mix (Bioline #BIO-98080), 2~10 ng DNA template, 400 nM of each primer in a total volume of 10 µl. All PCR reactions were conducted in duplicate. In each case, a standard curve of the amplicon was analysed concurrently to evaluate the amplification efficiency and to calculate the relative amount of amplicon in unknown samples. Finally, a melt curve was generated to evaluate the specificity of the PCR reaction.

2.3 Restriction enzyme digestion

Restriction enzymes were purchased from New England Biolabs (NEB). Digestion reactions were carried out in the recommended buffers and at the recommended temperatures. For diagnostic digestions, 0.5 µg DNA was incubated with 4 units of enzyme in a 10 µl reaction volume at 37 °C for 1 hour. To prepare DNA fragments for cloning, 2 µg DNA was incubated with 10 units of enzyme in a 100 µl reaction volume at 37 °C overnight. In addition, 1 unit of the calf intestinal phosphatase (NEB #M0290S) was added to digestions of vector DNA to remove the phosphate group at the 5' end of the linear vector to minimize recircularization. Following digestion, contaminants were removed from the DNA by phenol/chloroform extraction and ethanol precipitation.

2.4 Separation of DNA fragments on agarose gels

PCR products or digested plasmid DNAs were separated by agarose gel electrophoresis. The percentage of the agarose gel varied between 0.5~2% according to the size of the DNA fragments to be separated. The gels were electrophoresed in a BioRAD sub-cell GT tank, submerged in 1 × TAE buffer. 6× loading buffer (20 mM Tris pH 8.0, 15 % Ficoll®-400, 66 mM EDTA pH 8.0, 0.1 % SDS, 0.09 % bromophenol blue) was added to DNA samples before electrophoresis. Ethidium bromide (2 ng/ml) was added to the gel to visualize the DNA fragments using a BioRAD Gel DocTM XR+ system.

2.5 Phenol-chloroform extraction and ethanol precipitation of DNA

An equal volume of phenol-chloroform solution was added to the DNA solution with vigorous vortexing for 45 seconds, followed by centrifugation at 16,000 g at room temperature for 2 minutes. The upper phase was transferred to a fresh tube and DNA was then precipitated by addition of a one-tenth volume of 3 M sodium acetate pH 5.2 and two volumes of ethanol with gentle inversion, followed by incubation on dry ice for 5 minutes. After centrifugation at 20,000 g for 10 minutes at 4 °C, the DNA pellet was washed twice with 1 ml of 70 % ethanol and centrifuged at 20,000 g for 5 minutes. The pellet was then air-dried for 5 minutes and resuspended in a suitable volume of TE or ddH_2O.

2.6 Ligation of DNA fragments

Ligation reactions were carried out using T4 DNA ligase (NEB #M0202S). Typically, the ligation reaction contained 1× T4 ligase buffer (50 mM Tris pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT), 1 µl of T4 DNA ligase, ~50 ng linear dephosphorylated vector DNA, and a 3-fold molar excess of digested insert DNA in a 10 µl reaction volume, followed by incubation at room temperature for 2 hours.

2.7 Transformation of plasmid DNA into DH5α competent *E.coli* cells

Transformation was carried out using DH5α competent cells (Sambrook et al., 1989). 5 µl of the ligation reaction or 2 ng of plasmid DNA was added to a 50 µl aliquot of chemically competent cells, followed by incubation for 30 minutes on ice. The cell-DNA mixture was heat-shocked at 42 °C for 90 seconds and then quenched on ice for 1 minute. Following addition of 300 µl of SOB medium, cells were incubated at 37 °C for 1 hour. 150 µl of the bacterial culture was plated on a LB-agar plate supplemented with Ampicillin (50 µg/ml).

2.8 Site-directed mutagenesis

The mutagenesis was carried out using the Q5 high fidelity polymerase (NEB #E0554). The exponential amplification reaction of mutagenesis typically contained 1× Q5 Reaction Buffer (25 mM Tris pH 9.3, 50 mM KCl, 2 mM $MgCl_2$, 1 mM β-mercaptoethanol), 200 µM dNTPs, 0.5 µM of forward and reverse primer, 5 ng of plasmid DNA and 0.5 units of Q5 polymerase in a 25 µl reaction volume. Thermocycling conditions were as follows:

98 °C	30 s	
98 °C	5 s	25 cycles
57 °C	20 s	
72 °C	30 s/kb	
72 °C	2 min	
10 °C	Infinite	

The PCR product was checked by agarose gel electrophoresis. 1 µl of DpnI (NEB #R0176S) was added directly to the remaining product, followed by incubation at 37 °C for 2 hours. The ligation reaction typically included 1× T4 ligation buffer (50 mM Tris pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT), 0.5 µl T4 PNK (NEB #M0201S), 0.5 µl T4 ligase and 2 µl DpnI-treated PCR product in a 10 µl reaction volume and was incubated at room temperature for 2 hours to achieve phosphorylation of the PCR product and its ligation in the same reaction tube. 5µl of ligated product was transformed into DH5α competent *E.coli* cells.

2.9 Small-scale preparation of plasmid DNA from *E.coli* cells

Small-scale preparation of plasmid DNA was conducted according to Sambrook et al. A single bacterial colony was inoculated into 2 ml of LB medium, supplemented with Ampicillin (50 µg/ml). The culture was then incubated at 37 °C overnight with shaking at 200 rpm. 1.5 ml of bacterial culture was transferred into a 1.8 ml Eppendorf tube and centrifuged at 16,000 g for 30 seconds at room temperature. The supernatant was removed carefully to leave the bacterial pellet as dry as possible, followed by resuspension in 100 µl of ice cold Alkaline Lysis Solution I (25 mM Tris pH 8.0, 50 mM Glucose, 10 mM EDTA pH 8.0) by vigorous vortexing. 200 µl of Lysis Solution II (0.2 M NaOH, 1% SDS) was then added and mixed by inversion, followed by incubation on ice for 5 minutes. This was followed by addition of 150 µl of ice cold Alkaline Lysis Solution III (3 M potassium acetate, 11.5 % glacial acetic acid) with gentle vortexing and incubation on ice for 5 minutes. Following centrifugation at 16,000 g for 5 minutes at room temperature, the supernatant was decanted into a fresh tube. An equal volume of phenol-chloroform solution was added, followed by vigorous vortexing and centrifugation at 16,000 g for 5 minutes at room temperature. The supernatant was transferred into a fresh tube and the plasmid DNA was precipitated with two volumes of ethanol at room temperature, followed by gentle vortexing and centrifugation at 16,000 for 5 minutes at room temperature. The pellet was washed twice with 1 ml of 70 % ethanol and then resuspended in 50 µl TE containing RNase A (20 µg/ml).

2.10 Large-scale preparation of plasmid DNA from *E.coli* cells

100 ml of LB medium containing Ampicillin (50 µg/ml) was inoculated with 0.5 ml of bacterial culture from a small-scale plasmid preparation, followed by incubation at 37 °C overnight with shaking at 200 rpm. Bacterial cells were harvested by centrifugation at 6,000 g for 15 minutes. Plasmid DNA was extracted using the Qiagen Plasmid Maxi Kit (QIAGEN #12162) according to manufacturer's instructions. The DNA pellet was resuspended in 500 µl of TE.

2.11 Extraction of genomic DNA from mammalian cells

1×10^6 mammalian cells were resuspended in 500 µl of Lysis Buffer (200 mM NaCl, 10 mM Tris pH 8.0, 0.2 % SDS, 200 ng/ml Proteinase K) in a 1.8 ml Eppendorf tube and incubated at 56 °C overnight with rotation. Genomic DNA was phenol-chloroform extracted and ethanol precipitated, followed by resuspension in 500 µl of TE. The sample was then incubated at 37 °C for 10 minutes. After addition of 3 µl of RNase A (10 mg/ml) and 16 µl NaOAc (3 M pH 5.2), the sample was incubated at 37 °C for a further 30 minutes. Following addition of 10 µl of Proteinase K (10 mg/ml) and 10 µl of SDS (20%), the sample was incubated at 37 °C for 30 minutes. Finally, genomic DNA was purified by phenol-chloroform extraction and ethanol precipitation, and was resuspended in a suitable volume of TE. The concentration was determined using a NanoDrop spectrophotometer.

2.12 Total RNA extraction from mammalian cells

Total RNA was extracted using TRIzol (Invitrogen #3289) according to the manufacturer's instructions. Briefly, 5×10^6 cells were washed twice with ice cold PBS and then resuspended in 1 ml of TRIzol reagent with vigorous vortexing for 45 seconds, followed by incubation at room temperature for 5 minutes. 200 µl of chloroform was then added to the cell lysate with vigorous shaking for 15 seconds. Following incubation at room temperature for 2 minutes, the lysate was centrifuged at 20,000 g for 15 minutes at 4 °C. After transfer of the aqueous phase to a fresh tube, RNA was precipitated by addition of 0.5 ml of isopropanol with gentle inversion. Following incubation at room temperature for 10 minutes, RNA was aggregated to form a visible pellet by centrifugation at 20,000 g for

10 minutes at 4 °C. The RNA pellet was then washed twice with 1 ml of 80 % EtOH for 5 minutes at 20,000 g. After removal of the supernatant, the RNA pellet was air-dried for 3 minutes before resuspension in 30 µl of RNAase-free water. Contaminating genomic DNA was removed by addition of 2 units of DNase I (NEB #M0303S) in digestion buffer (10 mM Tris pH 7.5, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$) at 37 °C for 1.5 hours. Total RNA was phenol/chloroform extracted and ethanol precipitated, before being resuspended in 30 µl of RNase-free water. The concentration of total RNA was determined using a NanoDrop spectrophotometer (DeNovix Incorporated).

2.13 Synthesis of Complementary DNA (cDNA)

Typically, reactions contained 1 µg of DNase I-treated total RNA, 1 µl of 50 µM random hexamer or oligo dT primer, 1 µl 10 mM dNTP mix and ddH_2O to 12 µl, which were heated at 65 °C for 5 minutes. Reactions were then quenched on ice before addition of 4 µl of first strand buffer (250 mM Tris pH 8.3, 375 mM KCl, 15 mM $MgCl_2$), 2 µl 0.1 M DTT and 1 µl RNasin (Promega), followed by incubation at 37 °C for 2 minutes. 1 µl of M-MLV reverse transcriptase was then added to reactions before incubation at 37 °C for 50 minutes. After inactivation of the enzyme through incubation at 70 °C for 15 minutes, cDNA was phenol/chloroform extracted and ethanol precipitated, and finally resuspended in 30 µl of ddH_2O. The concentration was determined using a NanoDrop spectrophotometer.

D) Common protein-based methods

2.14 Preparation of whole cell protein extracts

Pro-B cells were harvested by centrifugation at 600 g for 3 minutes at 4 °C. Cell pellets were washed twice with ice cold 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 % NaDOC, 0.1% SDS) and lysis buffer (5% SDS, 150 mM Tris pH 6.7, 30% glycerol) supplemented with protease inhibitors (Complete™, Mini Inhibitor Cocktail Tablets with EDTA, Roche). For blotting phospho-proteins, NaF was added at a final concentration of 5mM in lysis buffer.

Lysates were boiled immediately for 5 minutes and then cleared by centrifugation at the top speed for 10 minutes at 4 °C.

2.15 SDS-PAGE

SDS-PAGE gels were prepared using a Mini-Protein casting system (BioRAD) using gel solutions made according to Laemmli (1970). Separating gels were generally 10 % (37.5:1 acrylamide:bis-acrylamide), with a stacking gel of 4% (37.5:1 acrylamide:bisacrylamide). Before loading on the gel, extracted proteins were mixed with a one-fifth volume of 6 x protein loading buffer and boiled for 2 minutes. Gels were submerged in 1 x TGS running buffer and electrophoresed for 1 hour at 170 V in a Mini-Protein Tetra Cell gel tank (BioRAD) until the dye front was at the bottom of the gel or to a point appropriate for the molecular weight of the protein being detected.

2.16 Western blotting

Following electrophoresis, proteins were transferred to PVDF membrane (Immobilon-P, IPVH00010, Millipore) for Western blotting analysis. PVDF membrane was washed in methanol and rinsed with dH_2O before being soaked with SDS-PAGE gels and blotting papers (Whatman 3MM) in semi-dry transfer buffer. Three pieces of pre-soaked blotting papers were firstly placed on the centre of the cassette base of Trans-Blot Turbo transfer system (BioRAD). The PVDF membrane was then placed on the blotting paper. The SDS-PAGE gel was placed on top of the membrane, followed by three more pieces of pre-soaked blotting paper. The assembled transfer pack then placed into an electroblotter and blotted for 30 minutes at 25 V. Following blotting, the PVDF membrane was blocked with 10 ml of a solution of 5 % non-fat milk powder in TBS-T blocking buffer (50 mM Tris pH 7.6, 150 mM NaCl, 5% milk, 0.05% Tween-20) for 1 hour at room temperature. All primary antibody hybridisations were conducted overnight at 4 °C (Table 2.1), whereas secondary or tertiary antibody hybridisations were performed at room temperature for an hour. At each hybridisation, membranes were washed with changes every five minutes with TBS-T for an hour. Following this, membranes were developed by incubation with enhanced chemiluminescence substrate (Thermo Scientific) for

2 minutes at room temperature and imaged using a G:BOX ChemiXT4 system (Syngene).

2.17 Co-immunoprecipitation

3×10^6 293T cells were plated in a 10 cm dish of 10 ml complete DMEM medium 24 hours before transfection. Three hours prior to transfection, the medium was changed to fresh serum-free DMEM medium. Plasmids (10 ug) were mixed well with 500 µl of OptiMEM medium by gentle vortexing. Concomitantly, 30 µl of PEI stock solution (1 mg/ml) was diluted in 500 µl of OptiMEM medium. These two solutions were then mixed well with gentle vortexing for 15 seconds, followed by incubation at room temperature for 15 minutes. The mixture was then added to cells dropwise, followed by gentle swirling to mix. Cells were incubated at 37 °C for 36 hours prior to harvest. The media was aspirated and the cells were washed twice with ice cold PBS. Cells were then scraped off and transferred into a 1.5 ml Eppendorf tube in 1 ml ice cold PBS. Cells were harvested by centrifugation at 500 g for 2 minutes. The supernatant was aspirated and the cells were resuspended in 257 µl of IP Buffer A (10 mM Hepes-KOH pH 7.9, 10 mM HCl, 1.5 mM $MgCl_2$, 0.1 % NP-40) supplemented with protease inhibitors (CompleteTM, Mini Inhibitor Cocktail Tablets with EDTA, Roche). Following this, 18 µl of 5M NaCl was added into the cell lysate and mixed well by vortexing; subsequently, 25 µl of 40 % glycerol was added into cell lysate, followed by rotating for 20 minutes at 4 °C. The supernatant was collected by centrifugation at 20,000 g for 15 minutes and transferred into a 1.5 ml Eppendorf tube containing 300 µl of IP Buffer B (10 mM Hepes pH 7.9) supplemented with protease inhibitors. To pre-clear the cell lysate, 15 µl of protein-G beads (Sigma Sepharose® Fast Flow) was added, followed by rotating at 4 °C for 20 minutes. The pre-cleared cell lysate was then transferred into a 2.0 ml siliconized tube, followed by addition of 1 µg antibody. The hybridisation was performed at 4 °C overnight with rotation. Following this, 20 µl of protein-G beads was added and mixed by rotating at 4 °C for an hour. Target proteins were then precipitated with beads by centrifugation at 500 g at 4 °C for 2 minutes and then washed three times with IP Buffer C (10 mM Hepes pH 7.9, 5 mM KCl, 0.75 mM $MgCl_2$, 0.05 % NP-40, 150 mM NaCl)

supplemented with protease inhibitors. After the final wash, the pellet was resuspended into 40 µl of IP Buffer C, followed by mixing with one fifth 6 x protein loading buffer. Immunoprecipitated samples were boiled for 5 minutes and then resolved by 10 % SDS-PAGE and western blotting analysis was conducted with the indicated antibodies.

2.18 Luciferase assay

The luciferase assay was carried out using the Dual-Luciferase Kit (Promega) according to manufacturer's instructions. Cells were washed twice with ice cold PBS and then resuspended in 1 ml Passive Lysis Buffer, followed by gentle shaking at room temperature for 15 minutes. After transfer to a fresh Eppendorf tube, the lysate was subject to a vigorous vortexing for 15 seconds and then centrifuged at 16,000 g for 10 minutes at 4°C. 100 µl of the Luciferase Assay substrate was predispensed into a luminometer tube. 20 µl of the lysate was then added, followed by determination of firefly luciferase activity using the SIRIUS luminometer V3.0. Determination of the Renilla luciferase activity was achieved by addition of 100 µl of Stop & GlO reagent.

E) Cell Culture

2.19 Culture of adherent cells

293T and COS-7 cells were maintained in complete DMEM media. Generally, adherent cells were cultured in 15 ml of media in a T75 flask. Cells were passaged 1:5 when they were 80-90% confluent. To split cells, the media was aspirated and cells were washed twice with 10 ml PBS. The cells were then detached by adding 1 ml trypsin/EDTA (Sigma Aldrich T3924). Detached cells were resuspended in 10 ml complete DMEM media and 3 ml of cells were added to 15 ml of fresh media in a new T75 flask.

2.20 103/BCL-2 and 1D1-T215 cell culture

103/BCL-2 and 1D1-T215 cells were maintained at a density of 0.5 – 2 x 10^6 cells /ml, in complete RPMI medium supplemented with β-mercaptoethanol to a final concentration of 50 nM and incubated at 33 °C with 5 % CO_2.

2.21 Preparation of interleukin-7 (IL-7)

IL-7 secreting cells, Mo-IL-7, were a kind gift from Prof. A. Rolink. Mo-IL-7 cells were cultured in complete DMEM medium supplemented with 12.5 µg/ml xanthine and 4 µg/ml mycophenolic acid. Cells were incubated for 2 days after they reach confluency. The supernatant containing IL-7 was then harvested by centrifugation at 600 g for 3 minutes, followed by filtration through a 0.22 µM filter to remove any cell debris.

2.22 Determination of the concentration of IL-7

IL-7 concentration was determined using the Mouse IL-7 Quantikine ELISA® Kit (R&D Systems) according to manufacturer's instructions. In brief, 50 µl of assay diluent RD1-21 was added to each well in a 96-well plate, followed by addition of 50 µl of standard and IL-7 sample. After incubation at room temperature for 2 hours with shaking at 500 rpm on a horizontal shaker, each well was washed 5 times with the Wash Buffer. Following addition of 100 µl of Mouse IL-7 Conjugate, samples were incubated on the shaker for another 2 hours. After five washes with Wash Buffer, 100 µl of Substrate Solution was added, followed by incubation in the dark for 30 minutes at room temperature. Finally, 100 µl of Stop Solution was added, followed by determination of the concentration of IL-7 using a microplate reader set at 450 nm.

2.23 Primary pro-B cell culture

Bone marrow was flushed from the femurs of 5-week old mice using a 1 ml syringe into 10 ml of pro-B cell medium. Primary cells were cultured at 33 °C, 5 % CO_2 for 7 days with addition of an additional of 5 ml fresh medium on the 4th day.

2.24 Transfection of COS-7 cell using polyethyleneimine (PEI)

Transfection of COS-7 cells was carried out using PEI (Alfa Aesar #043896.01). 1.2×10^5 cells were plated per well of a 6-well plate in 1 ml of complete DMEM medium 24 hours before transfection. Three hours prior to transfection, the medium was changed to fresh serum-free DMEM medium. Plasmid DNA (2 ng)

was mixed well with 100 µl of OptiMEM medium by gentle vortexing. Concomitantly, 6 µl of PEI stock solution (1 mg/ml) was diluted with 100 µl of OptiMEM medium. These two solutions were then mixed well with gentle vortexing for 15 seconds, followed by incubation at room temperature for 15 minutes. The mixture was then added to cells dropwise, followed by gentle swirling to mix. Cells were incubated at 37 °C for 48 hours prior to harvest.

2.25 Transfection of 103/BCL-2 cells by electroporation

Electroporation was carried out using the NucleofectorTM Kit (LONZA Catalog #VPA1010) according to manufacturer's instructions. Briefly, 4×10^6 cells were washed twice with ice cold PBS and then resuspended in 100 µl of transfection reagent (82 µl nucleofector plus 12 µl supplement 2), followed by addition of the plasmid DNA to be transfected. Cells were then transferred to a cuvette and electroporated using the setting Z01 of the AMAXA electroporator. Following addition of 500 µl complete RPMI medium, cells were decanted to a well of a 6-well plate using a sterile pastette; an additional 1400 µl of RPMI medium was added to cells, followed by incubation at 33 °C under 5 % CO_2.

2.26 4-hydroxytamoxifen treatment of cell lines

4-hydroxytamoxifen was used to activate the IRF4-ER_{T2} protein in 1D1-T215 cell lines. Cells at a density of 1 x 10^6 cell/ml were induced by 4-hydroxytamoxifen (Insight Biotechnology; Cat HY-16950-2mg) at a concentration of 2 µM. Treated cells were incubated at 37 °C in 5% CO_2 for the number of hours indicated.

2.27 Semi-solid agar assay

The semi-solid agar media used in this study is the complete RPMI medium containing 0.3 % agar for growing single cell clones. 2.5 ml of heated 1.2 % agar solution was well mixed with 7.5 ml of 1.33 x complete RPMI media. 500 cells were added to the semi-solid agar media when the media temperature had cooled to ~37 °C. Cells were well mixed with the media and transferred into a 10 cm dish, followed by incubation at 37 °C for 10-12 days in 5 % CO_2. Cells were fed with 5 ml of fresh complete RPMI-agar medium every five days. Single

cell clones were picked using 200 µl tips and expanded in complete RPMI media in 12-well plates.

F) Virus based methods

2.28 Production of retroviral particles

Retroviral particles were generated using the Phoenix system. The Phoenix cell line was created by stably transfecting 293T cells with constructs capable of producing gag-pol and envelope proteins for retrovirus packaging (Grignani et al., 1998). 3 x 10^6 Phoenix cells were plated in a 10 cm dish with 10 ml complete DMEM media 24 hours before transfection. Three hours prior to transfection, the medium was changed to fresh virus production medium. 4 µg of MSCV-IRF4-ER_{T2}-GFP construct was mixed with 500 µl of OptiMEM medium by gentle vortexing. Concomitantly, 12 µl of PEI stock solution (1 mg/ml) was diluted with 500 µl of OptiMEM medium. These two solutions were then mixed well with gentle vortexing for 15 seconds, followed by incubation at room temperature for 15 minutes. The mixture was then added to cells dropwise, followed by gentle swirling to mix. Cells were incubated at 37 °C for 48 and 72 hours prior to harvest. The retrovirus containing supernatant was filtered through a 0.45 µm syringe filter and flash frozen on dry ice and stored at -80 °C until use.

2.29 Production of lentiviral particles

Lentiviral particles were produced in 293T cells which were transfected with the lentiviral backbone constructs, the packaging construct (pCMVR8.74, Addgene #22036) and the envelope construct (pMD2.G, Addgene #12259). For lentiviral backbone constructs, pLKO.1-puro (Addgene #10878) was used to produce shRNA-mediated knock-down viral particles, whereas lentiCRISPRv2 (Addgene #98290) was used to produce Cas9-mediated knock-out viral particles. 3 x 10^6 293T cells were plated in a 10 cm dish with 10 ml complete DMEM media 24 hours before transfection. Three hours prior to transfection, the medium was changed to fresh virus production medium. 2 µg of pLKO.1 shRNA plasmid or lentiCRISPRv2 gRNA plasmid, 1.5 µg of pCMVR8.74 packaging plasmid and 0.5 µg of pMD2.G envelope plasmid were mixed with

500 µl of OptiMEM medium by gentle vortexing. Concomitantly, 12 µl of PEI stock solution (1 mg/ml) was diluted with 500 µl of OptiMEM medium. These two solutions were then mixed well by gentle vortexing for 15 seconds, followed by incubation at room temperature for 15 minutes. The mixture was then added to cells dropwise, followed by gentle swirling to mix. Cells were incubated at 37 °C for 48 and 72 hours prior to harvest. The lentivirus containing supernatant was filtered through a 0.45 µm syringe filter and flash frozen on dry ice and stored at -80 °C until use.

2.30 Determination of the optimal puromycin concentration

1D1-T215 cells were plated at a density of 5 x 10^5 cell/ml in each well of a 12-well plate. Puromycin was diluted and added to the cells at a final concentration from 0 to 5 µg/ml, in 0.5 µg/ml increments. Cells were examined each day and fresh puromycin-containing media was added every other day. After 5 days, complete cell death was observed in wells containing puromycin at a concentration of 2 µg/ml and above. Thus, 2 µg/ml was used for cell selection for further experiments.

2.31 Spinfection

5 x 10^5 cells were centrifuged at 315 x g for 3 minutes and resuspended in 1 ml of filtered retrovirus/lentivirus supplemented with polybrene (Sigma-Aldrich, TR-1003-G) at a final concentration of 4 µg/ml. Spinfection of cells was conducted by centrifugation at 800 g for 30 minutes at 32 °C in 12 well plates. Following spinfection, 1 ml of fresh complete RPMI media was added to each well and cells were maintained at 37 °C for 24 hours before analysis or puromycin selection.

2.32 Knock-down of Med23, Med1, YY1 and eRNAs using shRNA

shRNA targeting the Med23, Med1, YY1, sense Eλ3-1 and antisense Eλ3-1 eRNAs were designed using the The RNAi consortium database (TRC, BROAD Institute) and then cloned into the pLKO.1-puro vector. This vector was co-transfected into 293T cells together with the packaging plasmids, pCMVR8.74 and pMD2.G, to produce lentiviral particles. Spinfection of 1D1-T215 cells was

carried out by addition of lentivirus to the cells, followed by centrifugation at 800 g for 30 minutes at 32 °C. Puromycin was added immediately to the culture medium at a final concentration of 2 µg/ml, followed by incubation at 33 °C for 7 days. Cells were then plated in 10 ml of semi-solid agar medium in a 10 cm dish. Cells were fed with 5 ml of fresh complete RPMI-agar medium every five days. After 25 days, single visible colonies were selected and dispersed in 0.5 ml of complete RPMI medium. After growth for two weeks, genomic DNA was isolated and analysed by sequencing to check if the knockout had been generated.

2.33 Knock-out of the binding site of YY1 within HSCλ1 using CRISPR-Cas9

sgRNA targeting the YY1 binding site within HSCλ1 was designed using the online design software (http://crispr.mit.edu) and then cloned into the LentiCRISPRv2 vector. This vector was co-transfected into 293T cells together with the packaging plasmids, pCMVR8.74 and pMD2.G, to produce lentiviral particles. The lentiviral infection, puromycin selection and single cell isolation were performed as described above (Section 2.33).

G) Flow Cytometry

2.34 Cell staining for fluorescence-activated cell sorting (FACS)

Pro-B and pre-B cells were harvested by centrifugation at 312 g for 3 minutes. The cell pellet was resuspended in 10 ml of 0.168 M NH_4Cl, followed by a 10 minutes' incubation at room temperature to lyse the erythrocytes. Following addition of 30 ml ice cold PBS, cells were centrifuged at 312 g for 3 minutes. The cell pellet was then resuspended in 1 ml of ice cold FACS staining buffer (2% FCS, 25 mM pH 8.0 Hepes, 1 mM EDTA), followed by transfer to a FACS tube. Antibodies were then added to the cells with mixing, followed by incubation at room temperature in dark for 10 minutes. Pro-B cells were typically stained with phycoerythrin (PE) labelled anti-CD43 (BD Pharmingen cat no. 553271) and fluorescein isothiocyanate (FITC) labelled anti-CD19 (BD Pharmingen cat no. 553785). Cells were then washed by addition of 2 ml of

FACS staining buffer, followed by centrifugation at 312 g for 3 minutes. Prior to sorting, the cell pellet was resuspended in 0.6 ml ice cold staining solution and filtered through a mesh filter to generate a single cell suspension. Notably, CD43 is cell marker that is expressed by pro-B cells and which is lost upon differentiation to pre-B cells (Hardy et al., 1991b; Löffert et al., 1994). Surface expression of CD19 is observed on pro-B, pre-B cells, immature B, plasmablasts, as well as short- and long-lived plasma cells (Tedder, 2009). Therefore, B cells purified using $CD43^{-}$ /$CD19^{+}$ strategy are not pure pre-B cells

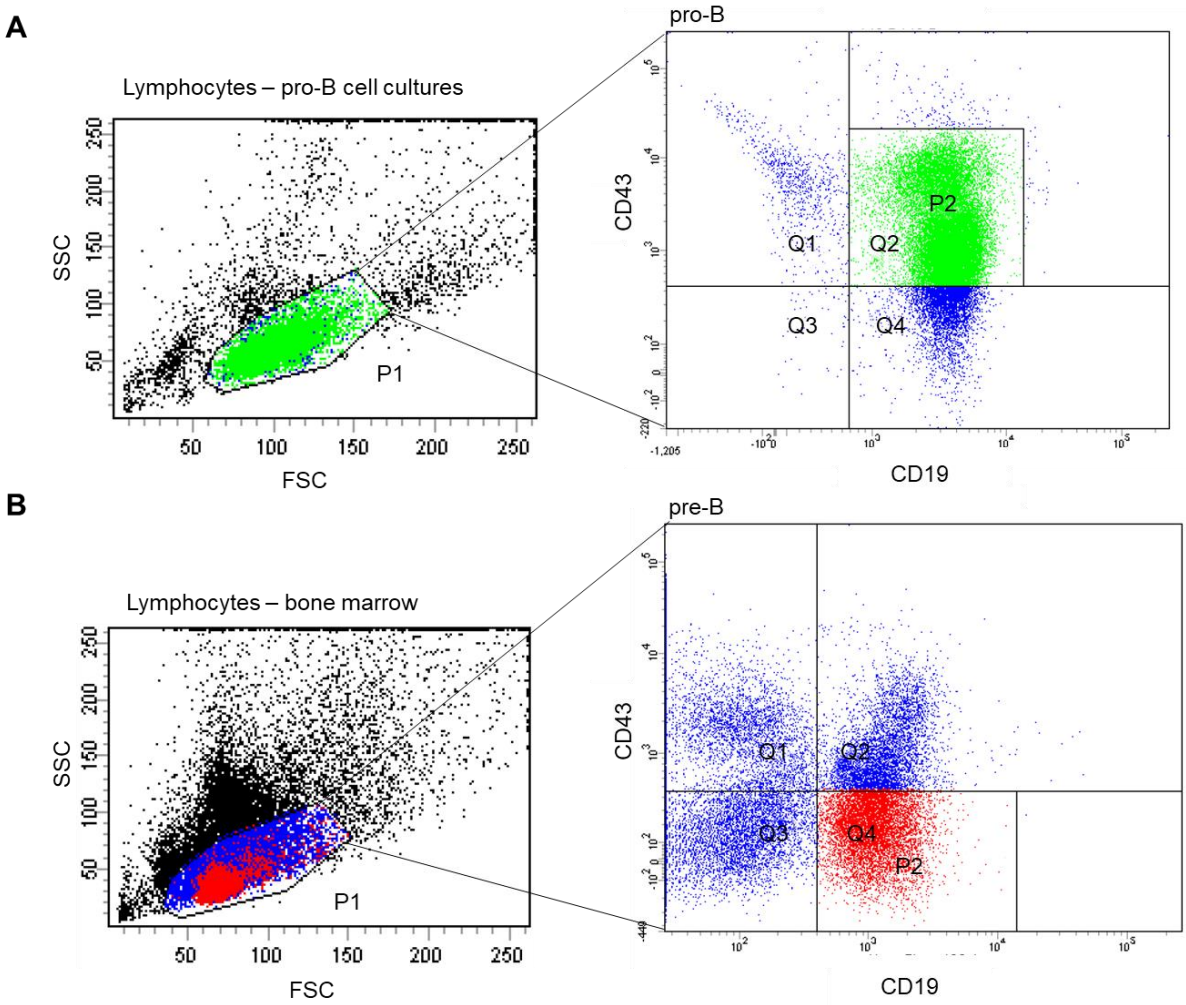

Figure 2.1 – Sort template for isolating primary pro-B/pre-B cells

A) Primary pro-B cells were purified from a bone marrow culture grown for 7 days. Lymphocytes were isolated based on the forward scatter (FSC) and side scatter (SSC), P1 (left). From these lymphocytes, pro-B cells were gated based on their staining with anti-CD43-PE and anti-CD19-FITC conjugated antibodies, P2 (right).

B) Primary pre-B cells were purified directly from bone marrow. Lymphocytes were isolated based on FSC and SSC, P1 (left). From these lymphocytes, pre-B cells were gated based on their staining with anti-CD43-PE and anti-CD19-FITC conjugated antibodies, P2 (right).

2.35 Flow cytometry to isolate cell populations

EGFP expressing 1D1-T215 cells were separated from untransduced cells by flow cytometry using a FACSMelody™ cell sorter (Becton Dickinson, New Jersey, USA). Primary Pro-B and pre-B cells were stained with FITC and PE conjugated antibodies before sorting, as described above. Sorted cells were

collected to 1 ml of sort buffer, followed by centrifugation at 285 g for 10 minutes at 4 °C. Cells were then recovered to 2 ml of complete RPMI medium supplemented with β-mercaptoethanol to a final concentration of 50 nM and incubated at 33 °C for 1 hour.

2.36 Flow cytometry for analysis

EGFP expressing 1D1-T215 cells were analysed by flow cytometry using a CytoFLEX flow cytometer (Beckman Counlter, USA) to determine the percentage of cells that had successfully been transduced. Cells were prepared for flow cytometry by washing with, and resuspension in, ice cold PBS.

H) ChIP and 3C

2.37 Chromatin Immunoprecipitation (ChIP)

ChIP was carried out according to Boyd and Farnham (Boyd and Farnham, 1999) with modifications. 2×10^7 cells were washed with ice cold PBS and then resuspended in 25 ml PBS in a 50 ml falcon tube. Crosslinking was achieved by addition of formaldehyde to a final concentration of 0.8 % followed by gentle agitation at room temperature for 10 minutes. Crosslinking was quenched by addition of glycine to a final concentration of 0.125 M and agitation at room temperature for 5 minutes. The cells were centrifuged at 700 g for 4 minutes at 4°C and then washed three times with ice cold PBS. The cell pellet was resuspended in 2 ml lysis buffer 1 (10 mM Tris pH 8.0, 10 mM NaCl, 0.2 % NP-40, 10 mM sodium butyrate supplemented with PMSF and Roche complete protease inhibitor cocktail), followed by incubation on ice for 10 minutes. The cell lysate was centrifuged at 600 g for 5 minutes and then resuspended in 1 ml of lysis buffer 2 (50 mM Tris pH 8.1, 10 mM EDTA, 1% SDS, 10 mM sodium butyrate, supplemented with PMSF and Roche complete protease inhibitor cocktail). The chromatin was sheared into ~500 bp fragments using a sonicator at an amplitude of 10 µm for six times 15 seconds with 1-minute incubation on ice between each burst. Cells debris were then removed by centrifugation at 1,200 g for 10 minutes at 4°C. The supernatant was diluted 5-fold in dilution buffer (20 mM Tris pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.01 % SDS, 1 % Triton

X-100, 10 mM sodium butyrate, supplemented with PMSF and Roche complete protease inhibitor cocktail). Chromatin was precleared by addition of 30 µl of protein G-beads and agitation at 4°C for 30 minutes, followed by centrifugation at 300 g for 2 minutes. The supernatant was transferred to a fresh 15 ml falcon tube, and a 150 µl aliquot was taken as the input control. Antibody was added to bind target fragmented chromatin at a concentration according to manufacturer's recommendations, followed by rotation overnight at 4 °C. The antibodies used are shown in Table 2.1. The chromatin sample was then incubated with 70 µl of protein G-beads at 4 °C with rotation for 2 hours and then centrifuged at 300 g for 2 minutes at 4 °C. Bound chromatin fragments were washed twice with wash buffer 1 (20 mM Tris pH 8.0, 50 mM NaCl, 2 mM EDTA, 0.1 % SDS and 1 % Triton X-100), twice with high salt buffer (20 mM Tris pH 8.0, 2 mM EDTA, 500 mM NaCl, 0.01 % SDS and 1 % Triton X-100), once with wash buffer 2 (10 mM Tris pH 8.0, 0.25 M LiCl, 1 mM EDTA, 1 % NP40 and 1 % sodium deoxycholate) and twice with TE, followed by elution in 200 µl of elution buffer (0.1 M NaHCO3 and 1 % SDS) with rotation for 15 minutes. The volume of input and bound sample was both increased to 400 µl with TE. RNase A and NaCl were added to a final concentration of 30 µg/ml and 0.375 M respectively, followed by incubation at 65 °C for 5 hours to reverse the crosslinks. Proteinase K (100 µg/ml) was then added and the reaction was incubated at 45 °C overnight. Chromatin samples were then subject to thrice phenol/chloroform extraction, once EtOH precipitation and thrice 70% EtOH wash, followed by resuspension in 30 µl ddH_2O.

2.38 Preparation of BAC template for 3C analysis

Bacterial artificial chromosome (BAC) Rp23-24i11 was obtained from Children's Hospital Oakland Research Institute and contains the 3' half of the murine Igλ locus. Because DpnII (NEB, R0543M) is blocked by Dam methylation, the BAC DNA was digested by its isoschizomer Sau3AI (NEB, R0169S) and ligated at a high concentration to generate all possible ligation products to give a normalisation control for 3C. 20 µg of BAC DNA was treated with 25 U of Sau1AI in a total volume of 500 µl at 37 °C overnight. The digested BAC DNA was cleaned by phenol-chloroform extraction and recovered by

ethanol precipitation and finally resuspended in 40 µl of TE. Following this, the BAC DNA was ligated with 2000 cohesive end units/ml of T4 DNA ligase in a total volume of 60 µl at 16 °C overnight. The ligated products were then purified by phenol/chloroform extraction and ethanol precipitation and resuspended in 100 µl.

2.39 Chromatin Conformation Capture (3C)

3C was carried out according to Dekker et al (Dekker et al., 2002) with modifications. 1×10^7 1D1-T215 cells were washed twice with ice cold PBS and resuspended in 10 ml PBS supplemented with 10 % FCS, followed by filtering through a 40 µM cell strainer to generate a single-cell suspension. Crosslinking was achieved by addition of formaldehyde to a final concentration of 2% and further agitation at room temperature for 10 minutes. The crosslinker was then quenched by addition of 1M glycine (ice cold) to a final concentration of 0.125 M. After incubation on ice for 5 minutes, cells were centrifuged at 500 g for 5 minutes at 4 °C and the supernatant was carefully removed. Following lysis in 5 ml lysis buffer (10 mM Tris pH 8.0, 10 mM NaCl, 0.2 % NP-40, 50 µg/ml PMSF, 1x Roche complete protease inhibitor cocktail) on ice for 45 minutes, the nuclei were centrifuged at 750 g for 5 minutes at 4 °C and then resuspended in 1 ml of ice cold PBS before centrifugation at 1000 g for 1 minutes. Nuclei were flash frozen in liquid nitrogen and stored in – 80 °C. Stored nuclei were resuspended in 500 µl 1.2 × NEB DpnII buffer (50 mM Bis-Tris-HCl pH 6.0, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT) in a screw capped Eppendorf tube. SDS was added to a final concentration of 0.3 % followed by vigorous pipetting. The nuclei were shaken at 200 rpm for 60 minutes at 37 °C with pipetting every 15 minutes, to avoid aggregation. Following addition of Triton X-100 to a final concentration of 3 %, the nuclei were incubated at 37°C for 60 minutes with shaking at 200 rpm. The nuclei were digested by addition of 100 units of DpnII (NEB, R0543M)) and then incubated at 37 °C for 4 hours with shaking at 200 rpm. An additional 100 units of DpnII was added, followed by overnight digestion. Finally, 100 units of DpnII were added to the sample, followed by a 4-hour incubation at 37°C with shaking, to achieve sufficient digestion. After inactivation of the restriction enzyme by incubation at 65 °C for 20 minutes, the digested nuclei were

transferred into a fresh 50 ml falcon tube. Ligation was performed in 7 ml of 1 x ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, 5 mM DTT) with 25 U T4 DNA ligase (Roche) at 16 °C overnight. RNase A was then added to a final concentration of 10 µg/ml to degrade RNA at 37 °C for 30 minutes. Crosslinks were reversed by addition of proteinase K to a final concentration of 100 µg/ml and incubation at 65 °C for at least 4 hours. The ligated DNA sample was phenol/chloroform extracted and precipitated with EtOH, and finally resuspended in 100 µl TE.

2.40 Nested PCR assay to detect 3C interactions

The Eλ3-1 was used as a viewpoint to determine interactions within the Igλ locus. Nested PCR assay was used to detect 3C interactions between Eλ3-1 and other *cis*-acting elements. Nested PCR reactions were also performed on the BAC control template to allow for normalisation in differences in primer efficiency. The first round of PCR was performed using Taq DNA Polymerase using primers and conditions listed in Table 2.6. For the second round, TaqMan qPCR was conducted in duplicate in 10 µl volumes with 5 µl of 1:10 diluted first round PCR product, 400 pM each primer, 100 pM 5' nuclease probe and 5 µl qPCRBIO probe mix (PCRBIO PB20.21-05). 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).

2.41 Analysis of pro-B and pre-B ChIP-seq data

Publicly available ChIP-seq datasets from pro-B and pre-B cells were downloaded from the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/), NCBI and listed in Table 2.2. Raw reads were processed to remove adapter sequences and low-quality reads using Trim_Galore. Clean reads were mapped to the *Mus musculus* (mm10) genome using Bowtie2 with default parameters. Peaks of mapped reads were called using the Model-based Analysis of ChIP-seq (MACS), generating output files in wig format. The MACS output files were finally uploaded to the Integrated Genomics Viewer (IGV) for visualising ChIP-seq traces on the reference

genome. Read processing before visualization was carried out on the Galaxy project webserver (https://usegalaxy.org/).

I) Antibodies

Table 2.1 – Antibodies used in the different applications and amounts per experiment

Antibody	Application	Amount	Supplier	Catalogue No.
CTCF	ChIP	5 µg	Millipore	07-729
SMC1A	ChIP	5 µg	Bethyl Laboratories, Inc.	A300-055A
P300	ChIP	4 µg	Santa Cruz Biotechnology	sc-484
E2A	ChIP	3 µg	Santa Cruz Biotechnology	sc-763
YY1	ChIP/WB	4.8 µg	Proteintech	22156-1-AP
Med1	ChIP/WB	4 µg	Bethyl Laboratories, Inc.	A300-793A
IntS11	ChIP	4 µg	Bethyl Laboratories, Inc.	A301-274A
Ser5p RNAPII	ChIP	2.7 µg	Abcam	ab5131
Ser2p RNAPII	ChIP	4 µg	Abcam	ab5095
Rpb1	ChIP	5 µg	Santa Cruz Biotechnology	sc-900
IRF4	ChIP	4 µg	Proteintech	11247-2-AP
PU.1	ChIP	16 µg	Santa Cruz Biotechnology	sc-3525
Ser5p RNAPII	WB	1 : 1,000	Santa Cruz Biotechnology	sc-47701
HA	WB	1 : 2,500	Abcam	ab9110
Myc	WB	1 : 2,500	Abcam	ab9132
β-tubulin	WB	1 : 5,000	OriGene	TA503129
CD19	FACS	8 µl/ml	BD Bioscience	553785
CD43	FACS	8 µl/ml	BD Bioscience	553271

J) Next generation sequencing datasets

Table 2.2 – Published next generation sequencing datasets analysed in this study

Factor	Cell type	Accession number	Additional notes
ATAC-seq	Pro-B	GSM1635407	N/A
H3K27ac	Pro-B	GSM1463433; GSM1463439	N/A
H3K4me1	Pro-B	GSM1463434; GSM1463439	N/A
H3K4me3	Pro-B	GSM1463434; GSM1463439	N/A
IRF4	Pro-B	GSM1296534; GSM1296537	Rag2$^{-/-}$
PU.1	Pro-B	GSM1290093	Haftl derived line c10
Med1	Pro-B	GSM1038263; GSM1038264	v-Abl immortalized 38B9
P300	Pro-B	GSM1290115	Haftl derived line c10
E2A	Pro-B	GSM546523; GSM546540	Rag1$^{-/-}$
YY1	Pre-B	GSM1897389; GSM1897390	Express Igμ
CTCF	Pro-B	GSM672401	Rag2$^{-/-}$
Rad21	Pro-B	GSM672403	Rag2$^{-/-}$
CTCF	Liver	GSM722759	N/A
CTCF	Heart	GSM722692	N/A
CTCF	Lung	GSM722859	N/A
CTCF	Kidney	GSM722698	N/A
CTCF	Spleen	GSM722990	N/A

K) Oligonucleotides

All primers used in this project were synthesized by Integrated DNA Technologies (IDT) and were diluted in ddH_2O to a concentration of 100 µM before use.

Table 2.3 - Oligonucleotides used for cloning

All oligonucleotides used for cloning are shown with their respective Tm

Oligonucleotide	**Sequence 5'-3'**	**T_m**
IRES_F	ATACCATGGAACTACGGGCTGCAGGAATTC	62 °C
IRES_R	ATGTTGCCGTCCTCCTTGAAGTCGATGC	62 °C
Jλ1p_F	AGCGCTAGCGTTCCCATATCTATGCAACACC	63 °C
Jλ1p_R	AGCACTCGAGGCACTGTGATATAGACTCATGC	63 °C
Eλ3-1_F	GCAGTCGACTTCCACTATCATCTCCTGAGATG	63 °C
Eλ3-1_R	GATGGATCCGGAAGGGTGTTTACAACCTTC	63 °C
Eλ3-1_IRF4mut_F	AATAGGAACTTCAACCAAGTCC	55 °C
Eλ3-1_IRF4mut_R	ATTTCTCTTTTTCTGTGACC	56 °C
Eλ3-1_PU.1mut_F	AGAAATAATACCAACTGAAACCAAGTCCATTAG	56 °C
Eλ3-1_PU.1mut_R	CTTTTTCTGTGACCATGAG	57 °C
Med23_F	CACCGGGCATAGTGATCGCTAAATC	60 °C
Med23_R	AAACGATTTAGCGATCACTATGCCC	60 °C
YY1_F	ATCAGAATTCATGGCCTCGGGCGAC	56 °C
YY1_R	AGTCCTCGAGTCGAGAAGGTCTTCTCTCTTC	56 °C
IRF4_F	CCGCTCGAGATGAACTTGGAGACGGGCAGC	58 °C
IRF4_R	GCTCTAGACTCTTGGATGGAAGAATGACGG	58 °C
YY1_shRNA_F	CCGGCGACGGTTGTAATAAGAAGTTCTCGAGAACTTCTTATTACAACCGTCGTTTTTG	N/A
YY1_shRNA_R	AATTCAAAAACTGTCCGGTGGAGGCATTAAACTCGAGTTTAATGCCTCCACCGGACAG	N/A

Med23_shRNA_F	CCGGTCTAAGAGATAAGTAAGTTACCTCGAGGTAACTTACTTATCTCTTAGATTTTTG	N/A
Med23_shRNA_R	AATTCAAAAATCTAAGAGATAAGTAAGTTACCTCGAGGTAACTTACTTATCTCTTAGA	N/A
Med1_shRNA_F	CCGGCCAGAAAGCAATGAATAAATTCTCGAGAATTTATTCATTGCTTTCTGGTTTTTG	N/A
Med1_shRNA_R	AATTCAAAAACCAGAAAGCAATGAATAAATTCTCGAGAATTTATTCATTGCTTTCTGG	N/A
CDK7_shRNA_F	CCGGCTGTCCGGTGGAGGCATTAAACTCGAGTTTAATGCCTCCACCGGACAGTTTTTG	N/A
CDK7_shRNA_R	AATTCAAAAACTGTCCGGTGGAGGCATTAAACTCGAGTTTAATGCCTCCACCGGACAG	N/A
YY1HSCλ1_sgRNA_F	CACCGATTCTTGCTCACAAGGGATA	N/A
YY1HSCλ1_sgRNA_R	AAACTATCCCTTGTGAGCAAGAATC	N/A
SenseEλ3-1_shRNA_F	CCGGCTCCTCCACAGAGCTTGTAATCTCGAGATTACAAGCTCTGTGGAGGAGTTTTTG	N/A
SenseEλ3-1_shRNA_R	AATTCAAAAACTCCTCCACAGAGCTTGTAATCTCGAGATTACAAGCTCTGTGGAGGAG	N/A
AntisenseEλ3-1_shRNA_F	CCGGTCTGTACTTTCATTCACATTCCTCGAGGAATGTGAATGAAAGTACAGATTTTTG	N/A
AntisenseEλ3-1_shRNA_R	AATTCAAAAATCTGTACTTTCATTCACATTCCTCGAGGAATGTGAATGAAAGTACAGA	N/A

Table 2.4 - Oligonucleotides used for qPCR analysis of cDNA and genomic DNA
All oligonucleotides used for PCR and qPCR assays of cDNA and genomic DNA are shown with their respective T_m.

Oligonucleotide	**Sequence 5'-3'**	**T_m**
HRPT_F	GGGGGCTATAAGTTCTTTGC	57 °C
HRPT_R	TCCAACACTTCGAGAGGTCC	57 °C
U6_F	CGCTTCGGCAGCACATATAC	59 °C
U6_R	TTCACGAATTTGCGTGTCAT	59 °C
GAPDH_F	ACTTTCTTGTGCAGTGCCAGC	56 °C
GAPDH_R	GCACACTTCGCACCAGCATC	56 °C
Vλ1GT_F	GTGAATTATGGCCTGGATTTCACT	58 °C
Vλ1GT_R	GAGCGACAAGTGAGTGTGAC	58 °C
Jλ1GT_F	ACTTGAGAATAAAATGCATGCAAGG	58 °C
Jλ1GT_R	TGTGGCCTTGTTAGTCTCGA	58 °C
HSCλ1_YY1BS_F	TGGGTCGACGATAGGCATGGAGATAGGGAGTG	58 °C
HSCλ1_YY1BS_R	TGGGTCGACGATAGGCATGGAGATAGGGAGTG	58 °C

Table 2.5 - Oligonucleotides used for qPCR analysis of chromatin immunoprecipitation samples

All oligonucleotides used for qPCR analysis of ChIP samples are shown with their respective T_m.

Oligonucleotide	**Sequence 5'-3'**	**T_m**
IntgeneIII_F	CAAGGAAAGGCCAACCAATA	53 °C
IntgeneIII_R	TAACCCTTTCCCCAGCTCTT	53 °C
Vλ1promoter_ChIP_F	GAGTTATATTATGTCTGTCTCACAGC	55 °C
Vλ1promoter_ChIP_R	GCATTGTTGCATACCCACTGC	55 °C
Jλ1promoter_ChIP_F	GGCAATGATTCTACCTTGTGTAGG	57 °C
Jλ1promoter_ChIP_R	CCACCAGCTGTGTAAAGTCTATGC	57 °C
HSCλ1_ChIP_F	GCCAGGTGTTCAGGAAGTC	58 °C
HSCλ1_ChIP_R	GCTGCCATATCCCTTGTGAG	58 °C
Eλ3-1_ChIP_F	GACATTACAAGCTCTGTGGAG	56 °C
Eλ3-1_ChIP_R	GCTAATGGACTTGGTTTCAGTTCC	56 °C
HS6_ChIP_F	AGGCAGCATCAGGCCTTAGGACTA	60 °C
HS6_ChIP_R	AGCATGACAAACAGAACCAGGTGT	60 °C
HS7_ChIP_F	ACCTTCTCTTTGCTCTGCAGGCA	58 °C
HS7_ChIP_R	ACCCAGAGGCTTTCCTGCAATGT	58 °C
HSVλ1_ChIP_F	ACACTGTAAGGGGCCAATGA	58 °C
HSVλ1_ChIP_R	GCAGCTTGGCAAATAAATGTAGG	58 °C

Table 2.6 - Oligonucleotides used for qPCR analysis of chromosome conformation capture samples

The primers for the nested PCR and qPCR assay are shown, with their respective T_m. In addition, the number of cycles used in the first round of PCR for the outer primers are shown.

Oligonucleotide	**Sequence 5'-3'**	**T_m**
ERRC3_out_F	CCAGAACTTCAAGCACAACCC	60 °C
ERRC3_out_R	GGAAAATGTATCTCAACAGTGGCTG	60 °C
ERRC3_in_F	GGGACTGTTTGTTTGGAAAACC	60 °C
ERRC3_in_R	AGGTGGAGTGACTCATTAGAAGG	60 °C
Eλ3-1_out_F	GACATTACAAGCTCTGTGGAG	60 °C
Eλ3-1_in_R	GAGGGTCAGGGGCCAGTTTT	60 °C
HS6_out_F	AGGCAGCATCAGGCCTTAGGACTA	60 °C
HS6_in_F	CCAAAGTGGCCAACAGAAATCTTG	60 °C
HSCλ1_out_F	CCAGGACTTAGCCAGTTCAG	60 °C
HSCλ1_in_R	ATCTTCAGTCCAGAGACAACCATCC	60 °C
Vλ1_out_F	ACCCTTTTCAGACCATTTCCC	60 °C
Vλ1_in_R	AACAGTCACACTCACTTGTCGC	60 °C
Jλ1_out_F	TGAATTGCTATCTCATGGAGAAGG	60 °C
Jλ1_in_R	AGAGCACAGAACATTCAGCACAG	60 °C
Control_out_F	AAGGAGGTAACTGCGTTGGAG	60 °C
Control_in_F	AAGGTGGAGGAATGGAGAGCATC	60 °C
ERCC3_probe	AGCCCTTTACTCTGAGGTAGTGTCTG	60 °C
Eλ3-1_probe	TGAATCCTGGAAGGTCATGTCCCA	60 °C

L) Plasmid maps

Maps of plasmids used during this the work presented in this thesis are shown below. Plasmid maps were generated using SnapGene Viewer (Version 4.3.10).

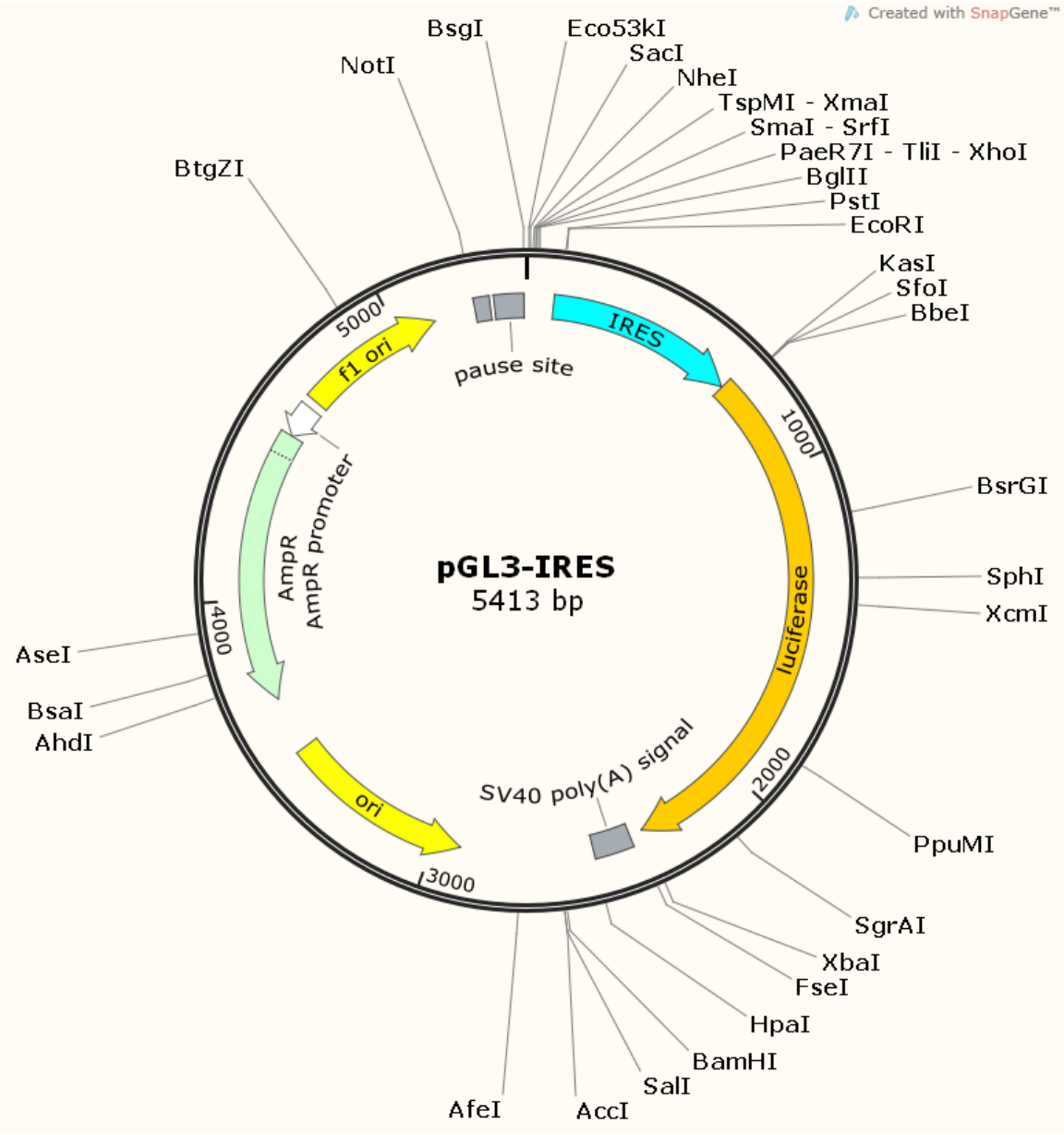

Figure 2.2 – Map of pGL3-IRES

The sequence of IRES from EMCV was cloned in front of luciferase reporter gene in pGL3-Basic (Promega). All unique restriction sites with a recognition sequence of 6 base pairs or more are shown.

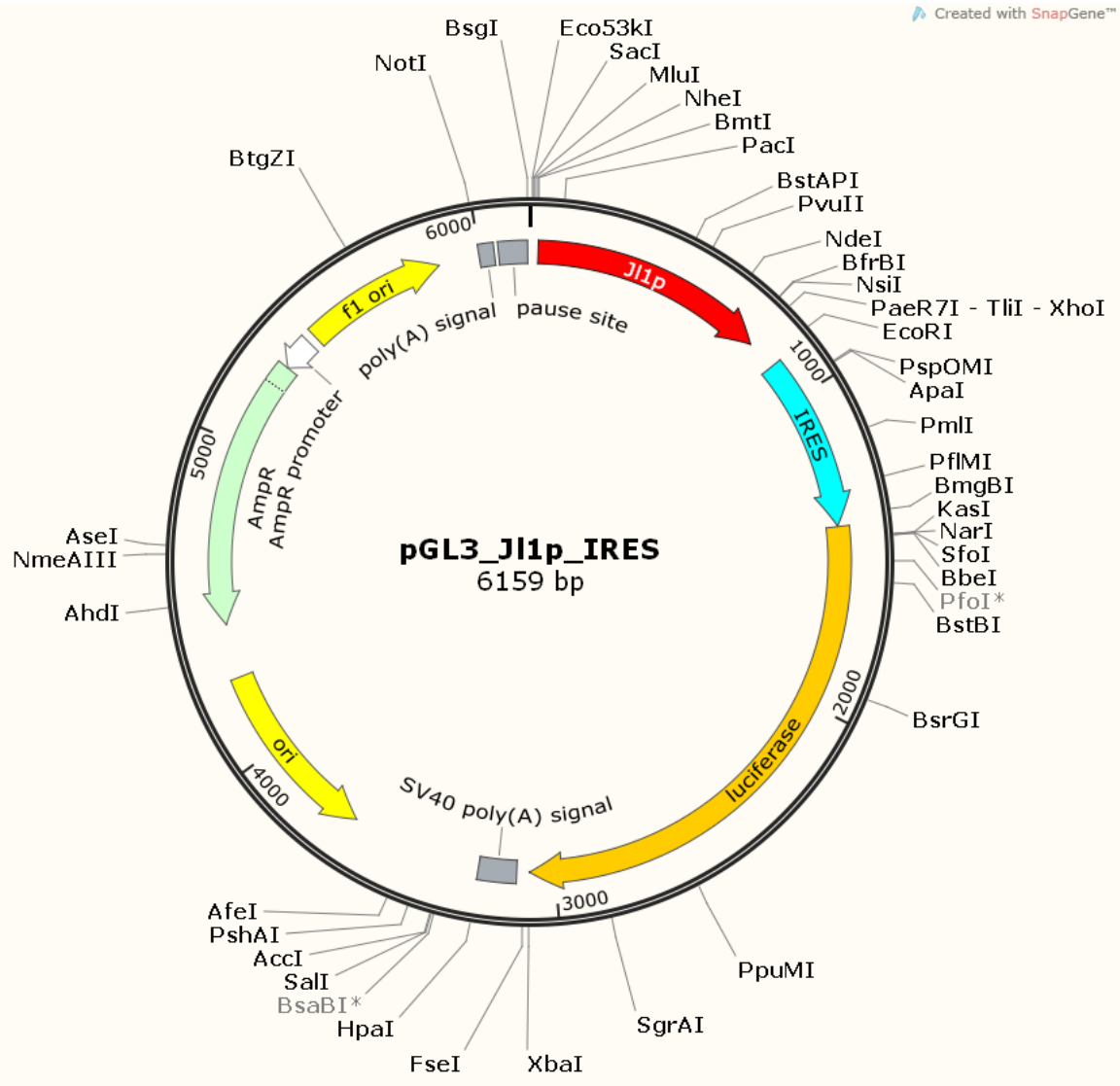

Figure 2.3 – Map of pGL3-IRES-Jλ1p

The sequences of Jλ1 promoter and IRES were cloned in front of luciferase reporter gene in pGL3-Basic (Promega). All unique restriction sites with a recognition sequence of 6 base pairs or more are shown.

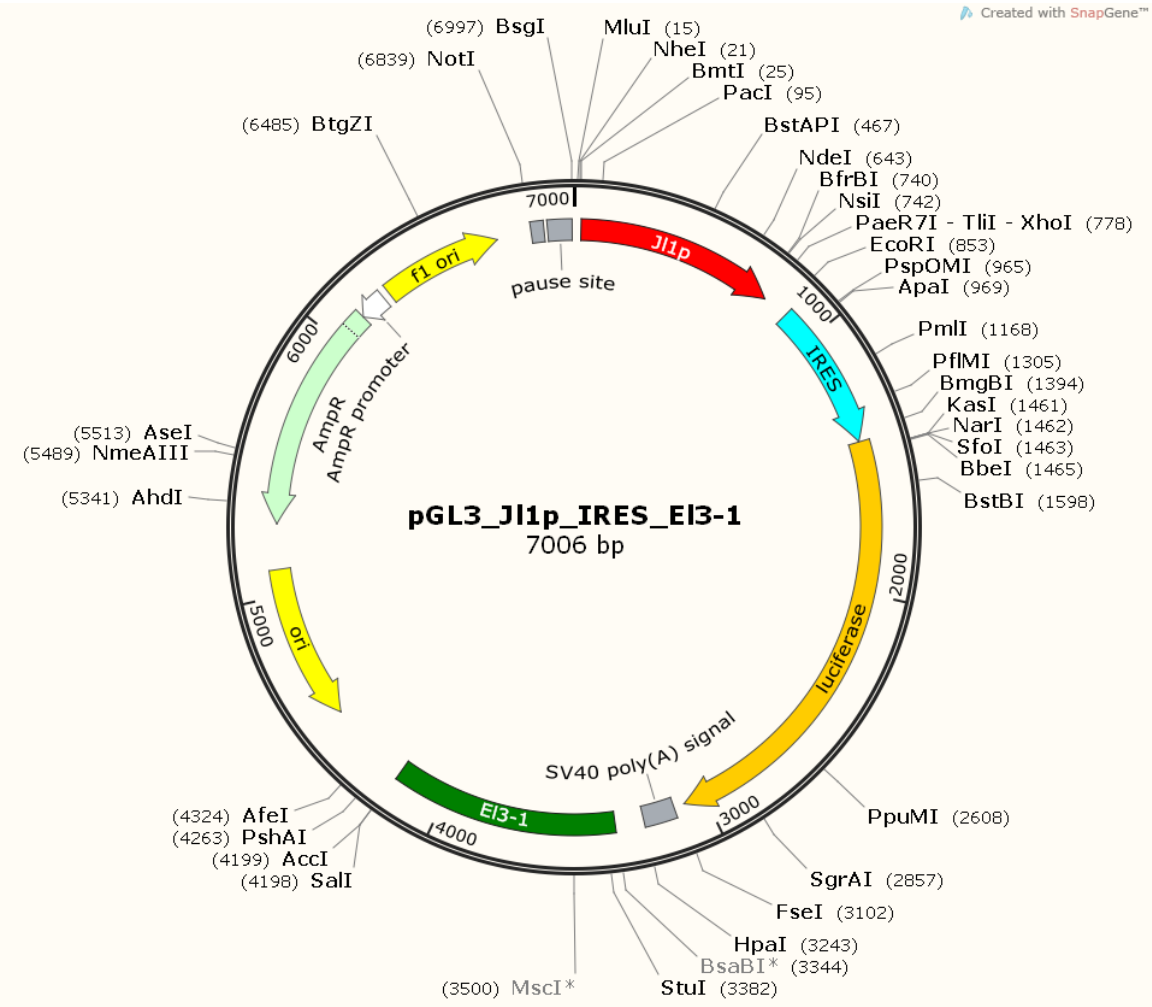

Figure 2.4 – Map of pGL3-IRES-Jλ1p-Eλ3-1

The sequences of Jλ1 promoter, IRES and Eλ3-1 were cloned in pGL3-Basic (Promega). All unique restriction sites with a recognition sequence of 6 base pairs or more are shown.

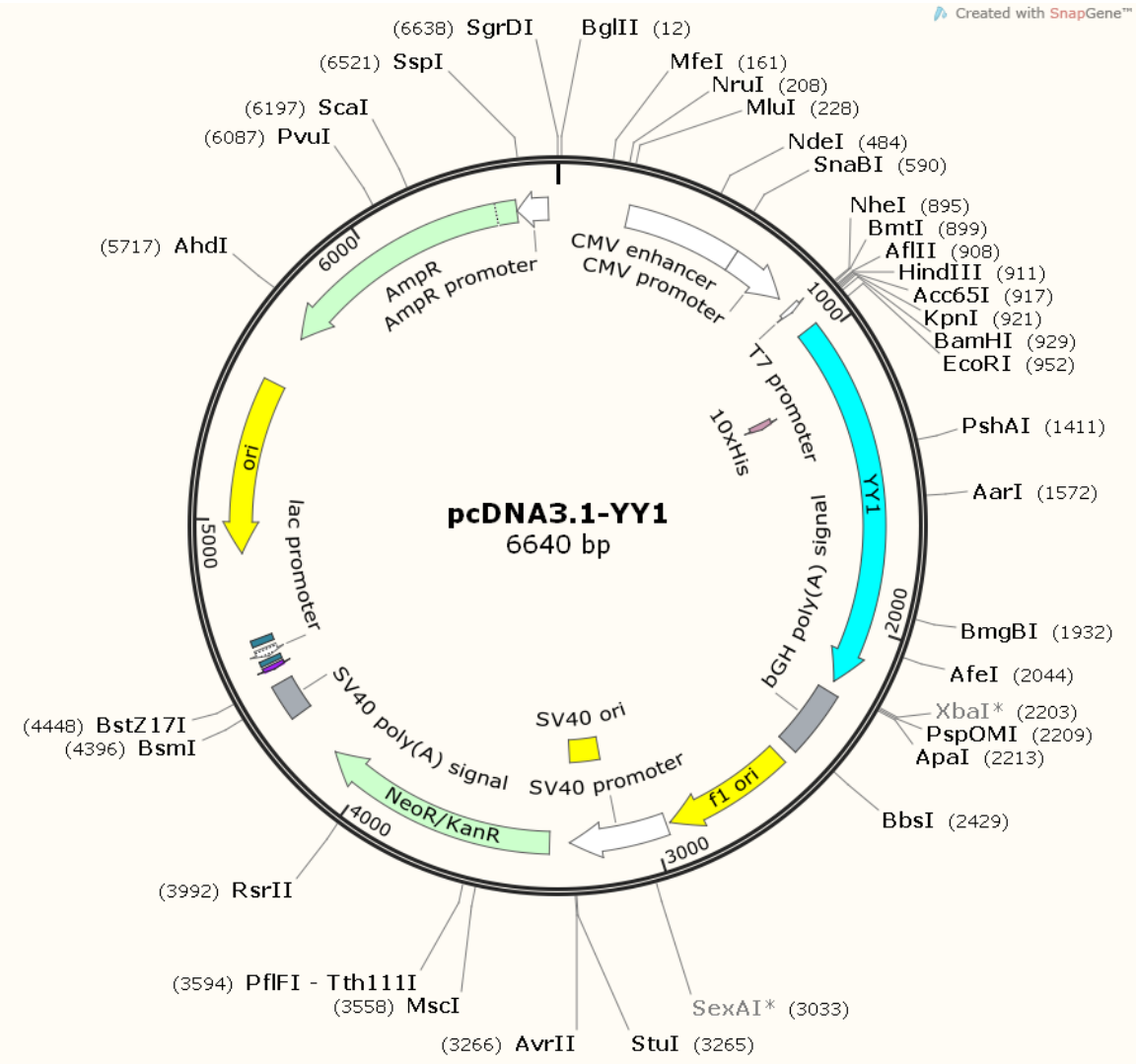

Figure 2.5 – Map of pDNA3.1-YY1

Yy1 cDNA is expressed under the control of the cytomegalovirus (CMV) promoter, present in the pcDNA3.1 vector (Invitrogen). All unique restriction sites with a recognition sequence of 6 base pairs or more are shown.

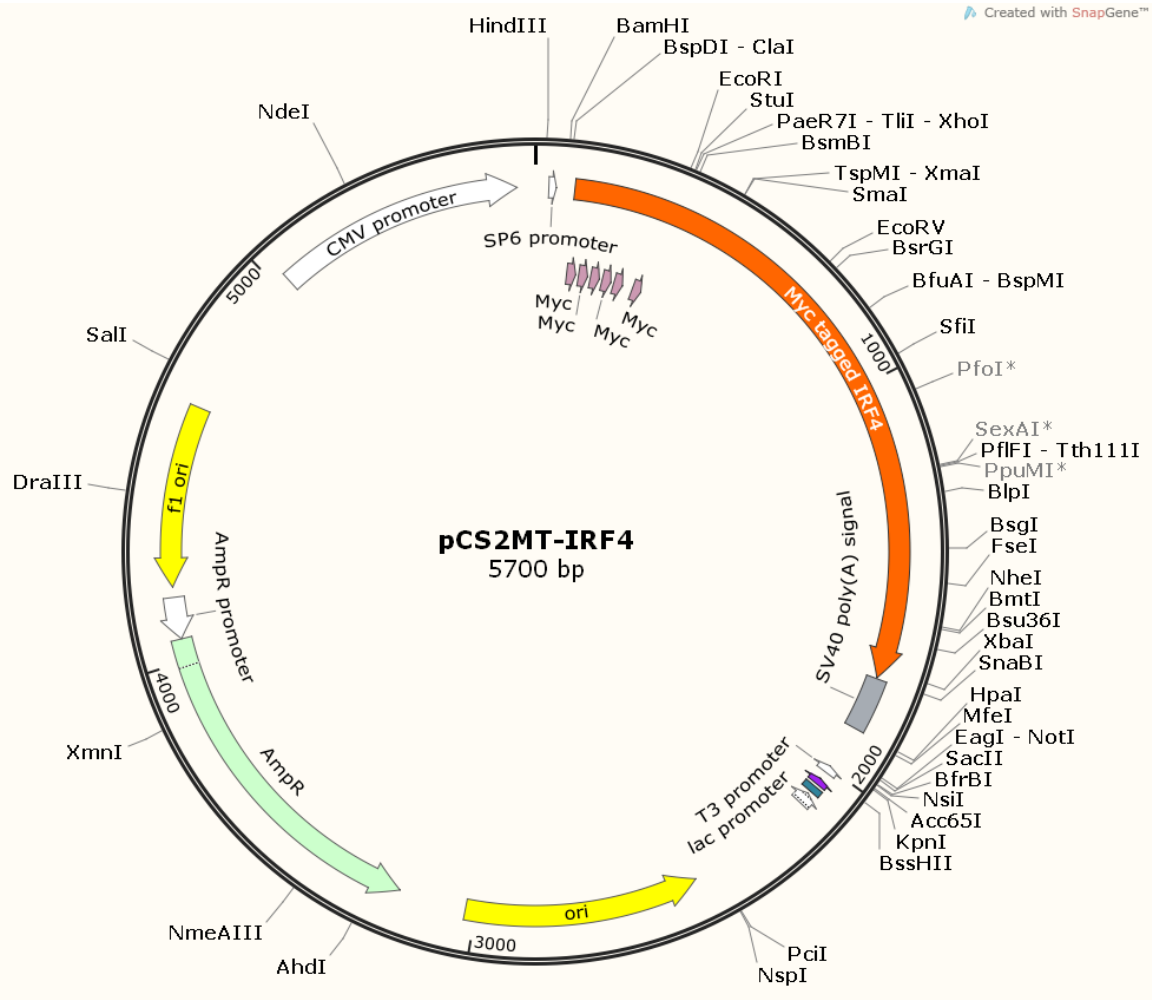

Figure 2.6 – Map of pCS2MT-IRF4

Irf4 cDNA was cloned into the pCS2MT vector (Rupp et al., 1994)., in frame with 6 x Myc tag, under the control of the cytomegalovirus (CMV) promoter. All unique restriction sites with a recognition sequence of 6 base pairs or more are shown.

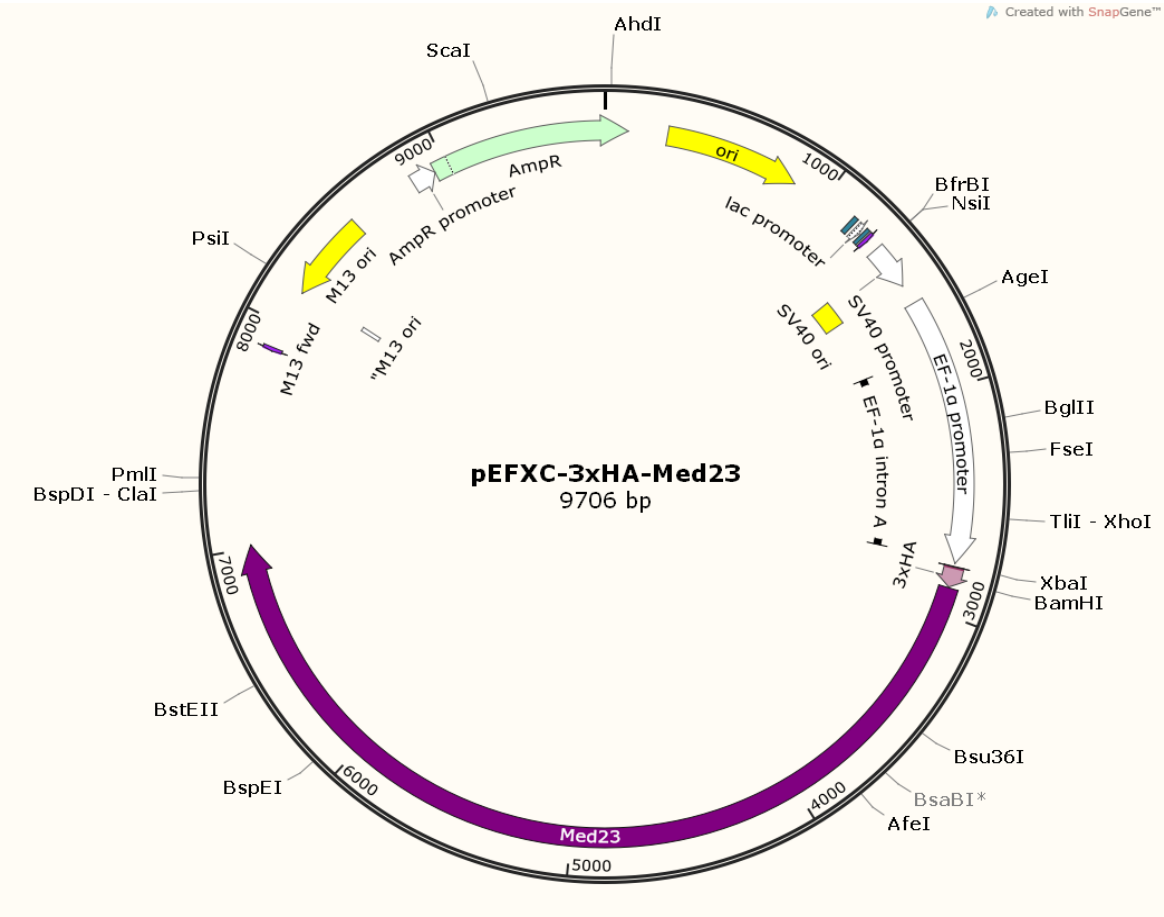

Figure 2.7 – Map of pEFXC-Med23

Med23 cDNA was cloned into the pEFXC vector (Mizushima et al., 1994)., in frame with 3 x HA tag, under the control of the elongation factor 1 alpha (EF-1α) promoter. All unique restriction sites with a recognition sequence of 6 base pairs or more are shown.

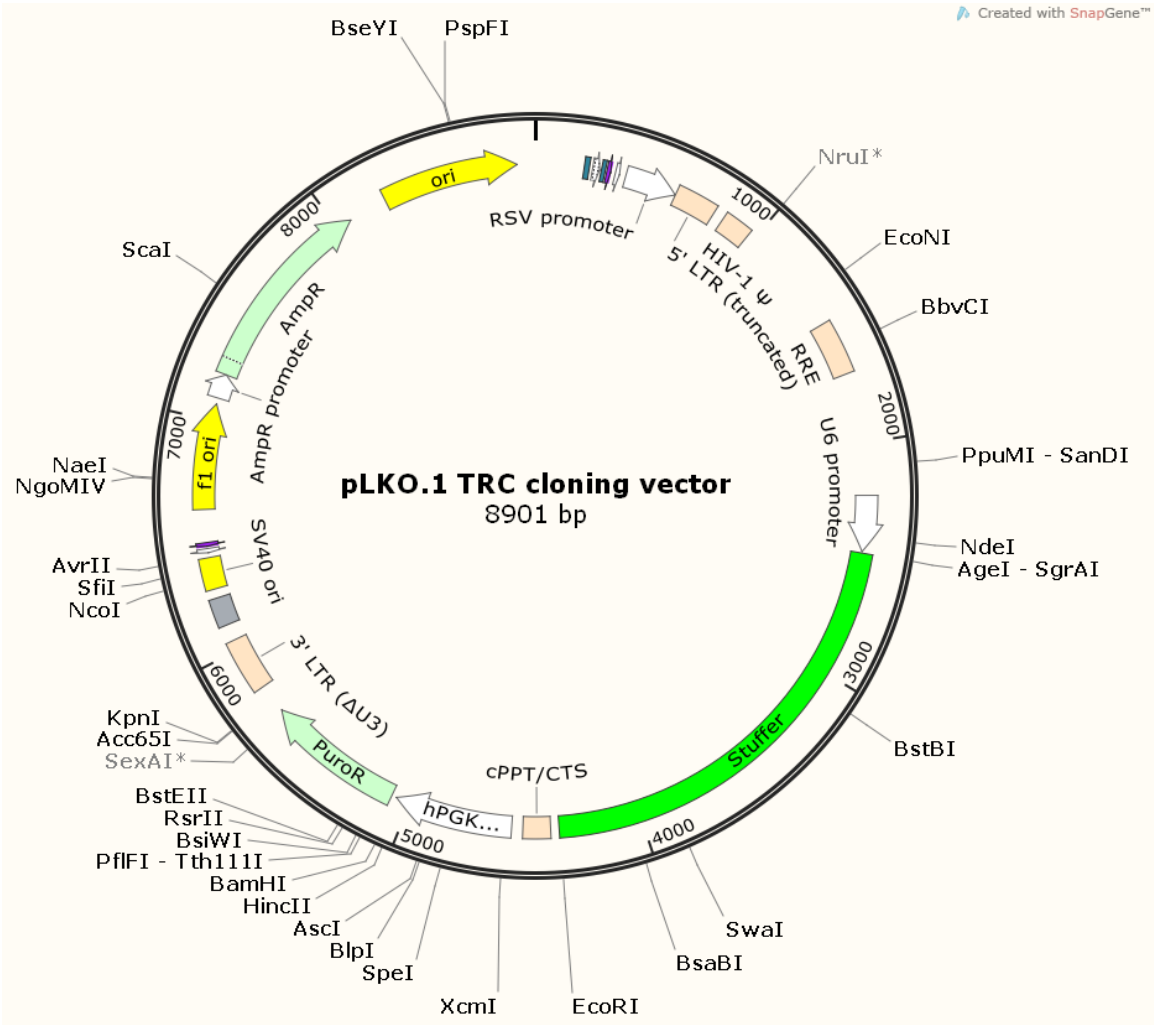

Figure 2.8 – Map of pLKO.1-puro

pLKO.1 puro was a gift from Bob Weinberg (Addgene plasmid # 8453; http://n2t.net/addgene:8453; RRID:Addgene_8453). The pLKO.1-TRC cloning vector contains a 1.9 kb stuffer that is released upon digestion with EcoRI and AgeI. shRNA oligonucleotides are designed so that they are flanked with sequences that are compatible with the EcoRI and AgeI sticky ends. Forward and reverse oligos are annealed and ligated into the pLKO.1 vector, producing a final plasmid that expresses the shRNA of interest. All unique restriction sites with a recognition sequence of 6 base pairs or more are shown.

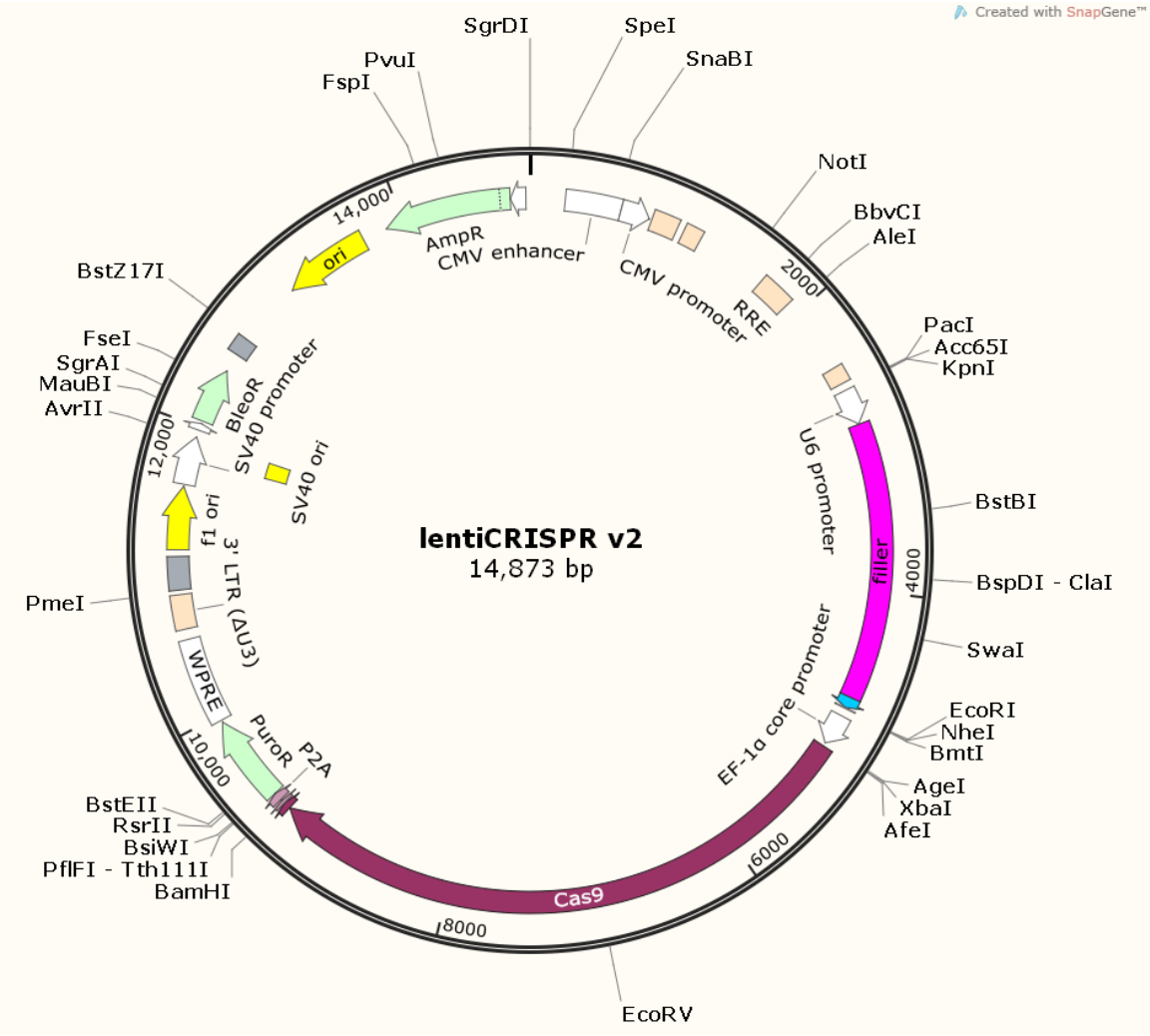

Figure 2.9 – Map of lentiCRISPR V2

lenti-CRISPR V2 was a gift from Feng Zhang (Addgene plasmid # 52961; http://n2t.net/addgene:52961; RRID: Addgene_52961).). The lentiCRISPR v2 vector contains a 1.9 kb filler that is released upon digestion with BsmBI. sgRNA oligonucleotides are designed so that they are flanked by sequences that are compatible with the sticky ends of BsmBI. Forward and reverse oligos are annealed and ligated into the lentiCRISPR v2 vector, producing a final plasmid that expresses the Cas9 and a sgRNA of interest. All unique restriction sites with a recognition sequence of 6 base pairs or more are shown.

Chapter 3: Temporal analysis of Eλ3-1 mediated activation of unrearranged gene segments in the murine Igλ locus

A) Introduction

The spatiotemporal control of gene transcription is a highly intricate and tightly regulated process that is crucial for cell development in all eukaryotic organisms. Dysregulation of this process is associated with many diseases, such as MonoMAC syndrome which is characterized by monocytopenia as well as B, natural killer, and dendritic cell lymphopenia (Hsu et al., 2013; Mathelier et al., 2015). Gene transcription starts with regulatory events at promoters, where transcription factors bind to specific motifs at core promoters that lie immediately upstream of TSSs and activate the assembly of the RNA polymerase II transcription pre-initiation complex. Whilst promoters play a role in controlling the basal transcription activity (Zabidi and Stark, 2016), much greater regulation relies on a second class of regulatory element, known as transcriptional enhancers. Transcriptional enhancers can reside many thousands of bases from the cognate gene promoters, either upstream or downstream, and are composed of concentrated clusters of recognition motifs for diverse transcription factors, often including pioneer factors, architecture factors and transcription activators (Long et al., 2016). Transcriptional enhancers physically interact with gene promoters and vastly increase the level at which the gene is transcribed (Vernimmen and Bickmore, 2015). Even though enhancers were discovered almost 30 years ago, and even though they outnumber promoters in the genome by approximately 50:1 and play such a crucial role in gene regulation, some very basic questions remain concerning how transcriptional enhancers work. For instance, how the enhancer specifically finds its correct promoter over huge distances within the densely packaged cell nucleus, how it then commits to activating transcription from a specific promoter and how it then vastly increases its transcription, are poorly understood. Given the huge energetic cost of transcribing a gene, to say nothing of the potentially life-threatening consequences of transcribing the wrong gene, at the wrong level and at the wrong time, it is vital that enhancers carry out their functions in an error-free way.

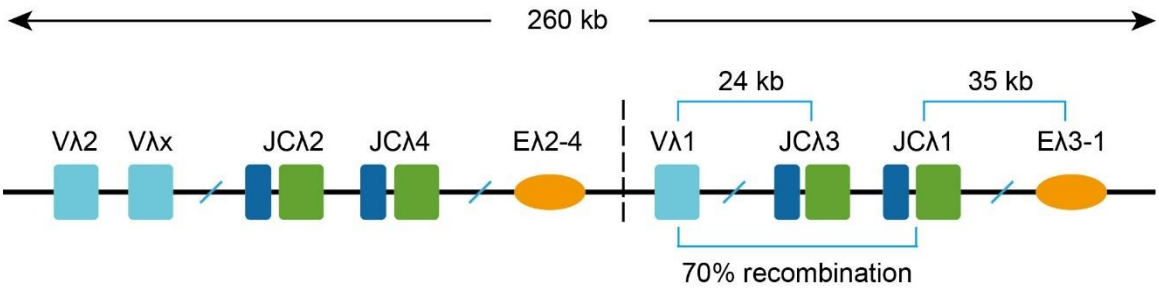

Figure 3.1 – The murine immunoglobulin lambda locus

A simplified schematic of the murine Igλ locus; this appears to have arisen from gene duplication indicated by the dashed line. The green rounded rectangles depict constant (C) exons, cyan rounded rectangles depict V gene segments and blue rounded triangles depict J gene segments. Orange ovals depict enhancers. 70% of recombination occurs between the Vλ1 and Jλ1 gene segments.

To address how the enhancer finds and commits to its cognate promoter, we used the murine Igλ locus as a model. The murine Igλ locus is ~ 230 kb in length; it is the smallest antigen receptor locus and is located on chromosome 16. It consists of three V and four J gene segments, where each J gene segment precedes a constant (C) region (Gerdes and Wabl, 2002). Expression of the Igλ light chain requires the recombinational joining of the V and J gene segments. Approximately 70% of recombination at the Igλ locus occurs between the Vλ1 and Jλ1 gene segments (Boudinot et al., 1994). Previous studies showed that IRF4 is a transcription factor that plays essential roles at different stages of B lymphocyte development (Acquaviva et al., 2008). IRF4 has been shown to be enriched at Eλ3-1 in pro-B cells and its binding increases by ~3-fold as cells progress to the pre-B stage. Interestingly, overexpression of IRF4 alone in murine pro-B cells can activate the Igλ locus completely (Bevington and Boyes, 2013). The mechanism by which IRF4 binding to the Eλ3-1 enhancer leads to the activation of non-coding transcription of Igλ is unknown.

In this chapter, I describe the development of a system where I can induce the activity of an enhancer, which allows me to turn on the enhancer and follow its

activation of transcription temporally. I further describe the temporal analysis of activator binding to build a detailed picture of the stages of Igλ locus activation.

B) Results

3.1 The distal enhancer, Eλ3-1, activates Jλ1 non-coding transcription in the presence of PU.1 and IRF4

Activation of non-coding transcription of unrearranged gene segments is crucial for V(D)J recombination as this is thought to be central to increasing chromatin accessibility and to add crucial epigenetic marks (Jhunjhunwala et al., 2009). Jλ1 is involved in 70% of recombination in the murine Igλ locus and consequently is a good model to follow promoter activation. Eλ3-1 is a B cell-specific transcriptional enhancer that is located within the 3' half of the Igλ locus and has been demonstrated to be essential for Igλ activation (Hagman et al., 1990; Haque et al., 2013). RT-qPCR was used to analyse the change of Jλ1 transcription in murine pro-B and pre-B cells and, as can be seen in Figure 3.2A, Jλ1 transcription increases by ~12-fold from pro-B to pre-B cells. Recent published Hi-C data demonstrated that Eλ3-1 physically interacts with the Jλ1 promoter through looping out of the ~35 kb intervening sequence between these two *cis*-acting elements (Krijger et al., 2016). This is also confirmed by 3C-qPCR which showed that the interaction frequency between Eλ3-1 and Jλ1 increased by ~4-fold from pro-B to pre-B cells (Figure 3.2B).

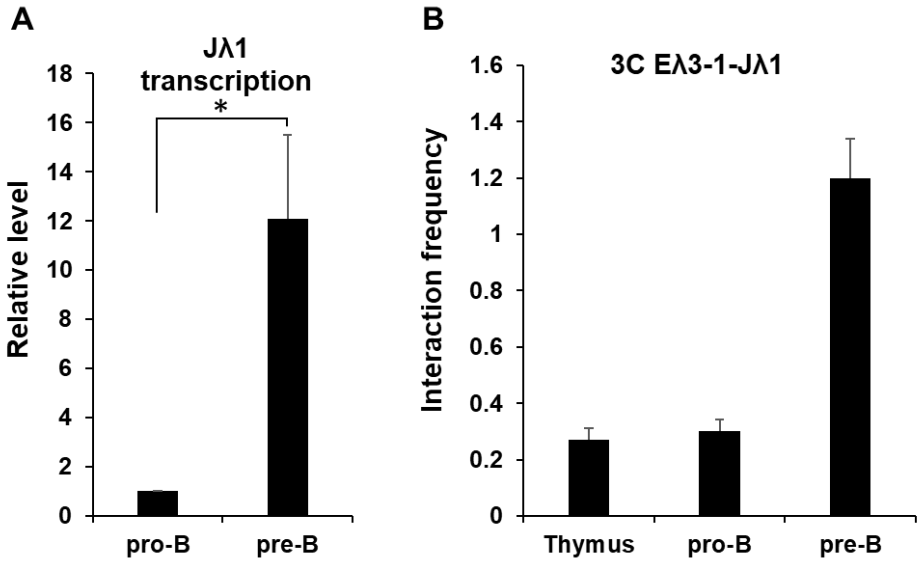

Figure 3.2 – The increase of Jλ1 transcription from pro-B to pre-B cells correlates with an elevated interaction frequency between Eλ3-1 and Jλ1

A) The transcription level of Jλ1 in non-transgenic pro-B and pre-B cells was analysed by quantitative PCR. Data were normalized to the expression level of the housekeeping gene, *Hprt*. Error bars show standard error of the mean (SEM) from three biological replicates. * represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

B) The interaction frequency between Eλ3-1 and Jλ1 was determined by the 3C assay. Error bars show standard error of the mean (SEM) from two replicates for each cell type. The 3C experiments shown in this figure were performed by Sarah Bevington.

Previous electrophoretic mobility shift assay (EMSA) demonstrated that PU.1 (Eisenbeis et al., 1993), IRF4 (Eisenbeis et al., 1995) and E-box proteins (Rudin and Storb, 1992) directly bind to Eλ3-1. This was confirmed by ChIP data from the Boyes lab, who also showed that increased levels of IRF4 in pro-B cells lead to activation of the murine Igλ locus (Bevington and Boyes, 2013). It was previously demonstrated that PU.1 physically interacts with IRF4 to form a complex (Eisenbeis et al., 1995) and these proteins were further shown to play regulatory roles in the of transcription of IgL chain loci and B cell development (Batista et al., 2017). Furthermore, knockout of E-box proteins in mice reduces Jλ1 transcription (Beck et al., 2009). Therefore, from these data, it appears that Jλ1 transcription is activated by Eλ3-1 using these enhancer-bound

transcription factors. To test this idea, the dual luciferase assay was applied, which is a widely used method for studying regulated gene expression. To this end, sequences from the Jλ1 promoter and Eλ3-1 enhancer were cloned into the luciferase construct, pGL3-basic, which was then transfected into COS-7 cells. To augment the expression of Firefly luciferase, the internal ribosome entry site (IRES) sequence from encephalomyocarditis virus (EMCV) (Martinez-Salas, 1999) was cloned in front of the Firefly luciferase gene in all luciferase reporter constructs. Viral IRESs are unique RNA sequences that enable ribosome recruitment and mRNA translation (Balvay et al., 2009). EMCV IRES is widely used in expression vectors due to its ability to increase the expression of target genes (Balvay et al., 2009). Because the Jλ1 promoter sequence has multiple transcription start sites (Engel et al., 2001) followed by translational start and stop codons, it seemed likely that these latter elements could prevent translation of the Firefly luciferase gene (data not shown). Therefore, to enable firefly luciferase expression levels to better reflect the true level of Jλ1 promoter activity, EMCV IRES was cloned into the all luciferase reporter plasmids just upstream of the Firefly luciferase cDNA. Expression constructs for IRF4, PU.1 and E47 were co-transfected into COS-7 cells together with the Firefly and Renilla luciferase reporter constructs. The results showed that the increase of luciferase activity driven by Eλ3-1 is very modest, indicating that other B cell-specific transcription factors could be involved (Figure 3.3A).

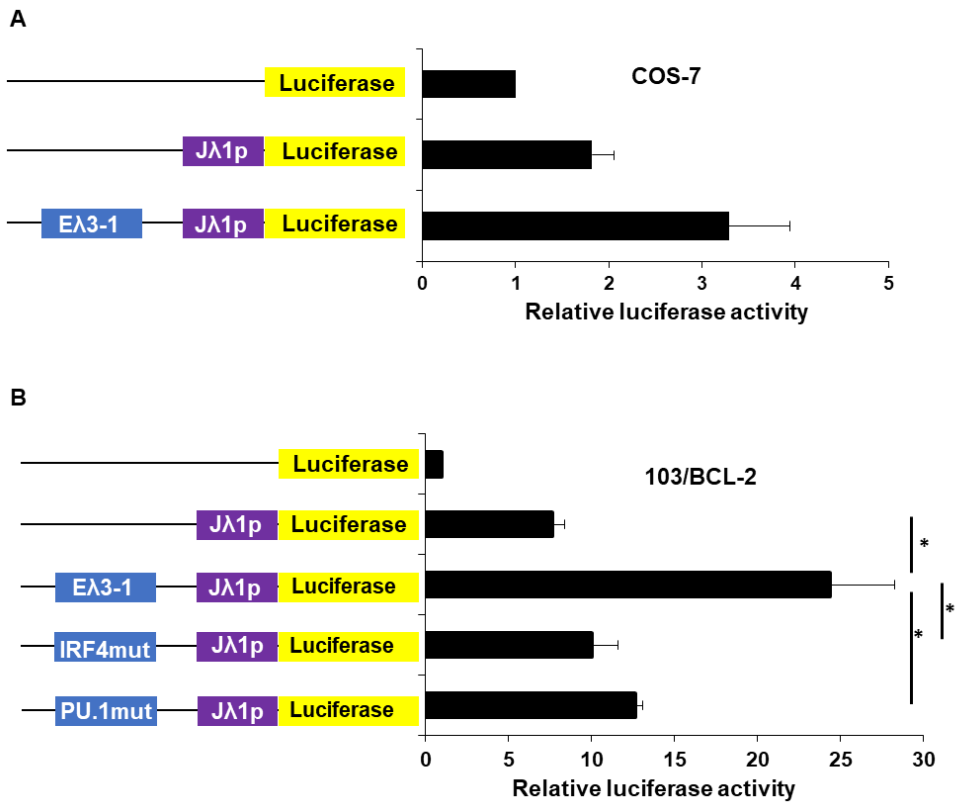

Figure 3.3 – EΛ3-1 is an enhancer of JΛ1 transcription

A) Luciferase activity driven by EΛ3-1 in COS-7 cells in the presence of PU.1, IRF4 and E47 transcription factors. This shows only a limited increase of ~ 1.8-fold. The results of dual luciferase assay are expressed as the Firefly luciferase activity normalized to the reference, Renilla luciferase activity (FL/RL).

B) Luciferase activity driven by wild type EΛ3-1 and mutant EΛ3-1 in temperature shifted 103/BCL-2 cells. The JΛ1 promoter increases luciferase activities by ~ 7-fold compared with the empty vector; EΛ3-1 gives a further 3-fold increase over the JΛ1 promoter. To generate the PU.1mut and IRF4mut constructs, the core consensus of the PU.1 binding site “GGAA” was mutated to “TCAA” and the core consensus of the IRF4 binding site “GAAA” was mutated into “CCAA” within EΛ3-1. Both mutations lead to a significant decrease or even loss of luciferase activity compared to the wild type EΛ3-1.

Error bars show standard error of the mean (SEM) from three experimental repeats.

* represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

103/BCL-2 is a B cell line that was transformed with a temperature-sensitive Abelson murine leukaemia virus (A-MuLV). This cell line undergoes a transition from pro-B cells to pre-B cells when shifted from 33 °C to the non-permissive

temperature of 39 °C as the temperature sensitive Abl kinase is inactivated (Chen et al., 1994; Cocea et al., 1999; Shah et al., 2008). IRF4 levels increase upon temperature shift, leading to the activation of transcription and recombination of Ig light chain loci (Xu and Feeney, 2009). To test if natural levels of B cell transcription factors, found in 103/BCL-2 cells enable Eλ3-1 to activate Jλ1 transcription, 103/BCL-2 cells were electroporated with the luciferase reporter constructs and then subjected to temperature shift for 20 hours at which time the Igλ locus is fully activated (Xu and Feeney, 2009). The results showed that the luciferase activity driven by the Jλ1 promoter increases by ~ 7-fold compared with the empty vector; addition of the enhancer, Eλ3-1, gives a further ~ 3-fold increase over the Jλ1 promoter alone (Figure 3.3B).

Previous data from our lab indicated that Eλ3-1 exerts its function through the recruitment of PU.1 and IRF4. Therefore, the binding sites for PU.1 and IRF4 within Eλ3-1 were mutated to examine the effects on Eλ3-1 activity. According to motif analysis using the Pfam database (El-Gebali et al., 2019), the core consensus motif sequences of PU.1 and IRF4 are “GGAA” and “GAAA” respectively. Previous studies showed that even single point mutations of these two core consensus motif sequences can lead to a significant decrease in the binding enrichment of the corresponding factors (Foxler et al., 2011; Li et al., 2016a). In this study, the core consensus of the PU.1 binding site “GGAA” was mutated to “TCAA” and the core consensus of the IRF4 binding site “GAAA” was mutated to “CCAA” within the luciferase reporter construct. These mutant constructs were then electroporated into 103/BCL-2 cells followed by temperature shift for 20 hours. The results suggest that both mutations within the core consensus of these two TF binding sites lead to a significant decrease or even loss of luciferase activity driven by Eλ3-1 compared to the wild type enhancer (Figure 3.3B).

3.2 IRF4 bound to Eλ3-1 facilitates Ser5 phosphorylation of the C-terminal domain (CTD) of RNAPII recruited to the Jλ1 promoter

The luciferase assay above shows that binding of IRF4 and PU.1 to Eλ3-1 is important for Jλ1 transcription. Previous publications showed that equipping

pro-B cells with the elevated pre-B levels of IRF4 activates the Igλ locus completely in transgenic mice (Bevington and Boyes, 2013). To test if PU.1 also plays a role in the activation of the Igλ locus, chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR) was conducted. As can be seen in Figure 3.4A, the level of PU.1 binding to Eλ3-1 does not change significantly from pro-B cells to pre-B cells. PU.1 shows a high level of occupancy at its genomic binding motif sequences, whereas IRF4 only shows low levels of occupancy at its cognate genomic DNA binding site in absence of PU.1 (Escalante et al., 2002). These data therefore suggest that the main role of PU.1 in the activation of Jλ1 transcription is to facilitate IRF4 binding at Eλ3-1.

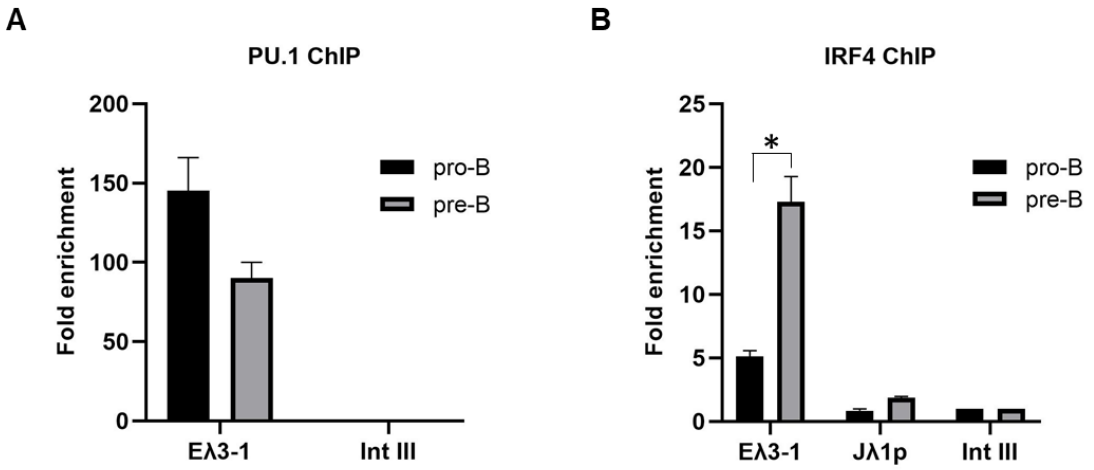

Figure 3.4 – Binding of PU.1 and IRF4 in Eλ3-1 in pro-B and pre-B cells

A) PU.1 binding was analysed by ChIP-qPCR in non-transgenic mouse pro-B and pre-B cells. The fold enrichment over input DNA at Eλ3-1 and Intgene III (negative control region) is shown. Intgene III is an intergenic region located approximately 2 kb downstream of the Vλ1 gene segment. No transcription and transcription factor binding was observed within the Intgene III region, and therefore this region was used as a negative control to normalize the binding level of the transcription factors of interest. All values are normalized to binding at Intgene III as a negative control. The ChIP experiments shown in this Figure were performed by James Scott.

B) IRF4 binding was analysed by ChIP-qPCR as above. The ChIP experiments shown in this Figure were performed by James Scott.

Error bars show standard error of the mean (SEM) from three experimental repeats.

* represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

Previous studies showed that enhancers can deliver increased concentrations of transcription activators and components of basal transcription machinery to promoters (Calo and Wysocka, 2013; Deng et al., 2012; Pennacchio et al., 2013). ChIP-qPCR analysis showed that IRF4 binding to Eλ3-1 increases ~3-fold from pro-B to pre-B cells (Figure 3.4B). By contrast, only a limited level of IRF4 is enriched at the Jλ1 promoter in pro-B cells that does not increase dramatically in pre-B cells (Figure 3.4B). Therefore, to determine how the enhancer influences protein binding to the promoter, ChIP-qPCR analysis of RNAPII binding was initially carried out. The results show that RNAPII is already present at the Jλ1 promoter in pro-B cells where Jλ1 has only low levels of activity and its level also do not increase in pre-B cells, where Jλ1 is active (Figure 3.5A).

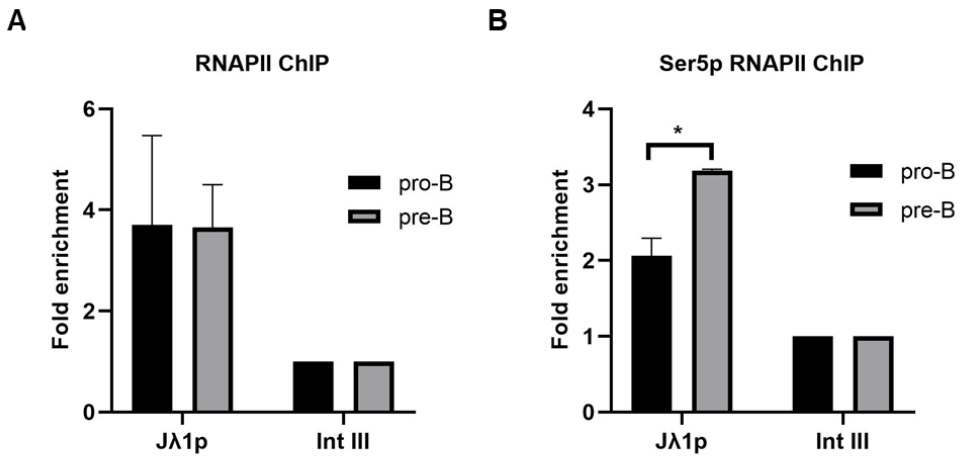

Figure 3.5 – Binding of basal transcription machinery to the Jλ1 promoter in pro-B and pre-B cells

A) Total RNAPII binding was analysed by ChIP-qPCR in non-transgenic mouse pro-B and pre-B cells. The fold enrichment at the Jλ1 promoter and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. The ChIP experiments shown in this figure were performed by Sarah Bevington.

B) Ser5 phosphorylated RNAPII binding to the Jλ1 promoter was analysed by ChIP-qPCR in pro-B and pre-B cells as above.

Error bars show standard error of the mean (SEM) from three experimental repeats.

* represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

Jλ1 transcription increases significantly from pro-B to pre-B cells despite the fact that there is no dramatic change in the enrichment of total RNAPII at the Jλ1 promoter from pro-B to pre-B cells (Figures 3.2A and 3.5A). Previous studies showed that activation of the murine Igλ locus leads to changes of epigenetic modifications at the unrearranged gene segments. Specifically, the level of H3K4me3 at the Jλ1 promoter increases significantly from pro-B to pre-B cells (Bevington and Boyes, 2013). It is notable that TFIIH mediates serine 5 (Ser 5) phosphorylation of RNAPII, that in turn is responsible for activating SETD1A/B methyltransferase activity, which methylates H3K4 at promoters (Ebmeier et al., 2017). Furthermore, phosphorylation of the Ser 5 residue of CTD of RNAPII is a pre-requisite for the release of the transcription initiation complex from TSSs (Phatnani and Greenleaf, 2006). Taken together, these data suggest that Ser 5 phosphorylated Pol II may be low in pro-B cells and increase in pre-B cells, activating SETD1A/B methyltransferase that in turn augments the level of H3K4me3 at promoters. To test this idea, ChIP-qPCR analysis of Ser 5 phosphorylated RNAPII was conducted and as can be seen in Figure 3.5B, the binding level of Ser 5 phosphorylated RNAPII indeed increases significantly from pro-B to pre-B cells, suggesting that Ser 5 phosphorylation of RNAPII at the Jλ1 promoter constitutes part of the activation of Jλ1 transcription that is induced by increased levels of IRF4.

3.3 Development of an inducible pro-B cell line to investigate the activation of the Igλ locus

To test the hypothesis by which Jλ1 transcription is activated by IRF4, I generated an inducible pro-B cell line, in collaboration with my colleagues, which allows the temporal investigation of Jλ1 transcription. To establish a pro-B cell line, bone marrow was firstly extracted from six-week-old mice and then immediately infected with the Abelson murine leukaemia virus (A-MuLV) for immortalization. Individual cell colonies were isolated using the semi-solid agar assay and viable colonies were then transferred from agar plates to RPMI media for expansion. To ensure the generated cell lines were at the pro-B stage of development, pro-B specific cell surface markers, CD19 and CD43, were used to screen target cells by flow cytometry (Figure 3.6A). Subsequently, the

expression levels of IRF4 and PU.1 were analysed in CD19^{+}/CD43^{+} double-positive pro-B cell lines. This is important as the correct levels of IRF4 and PU.1 expression are essential for the activation of the Igλ locus. A cell line, 1D1, that exhibits similar levels of IRF4 and PU.1 expression to primary pro-B cells was selected for further experiments.

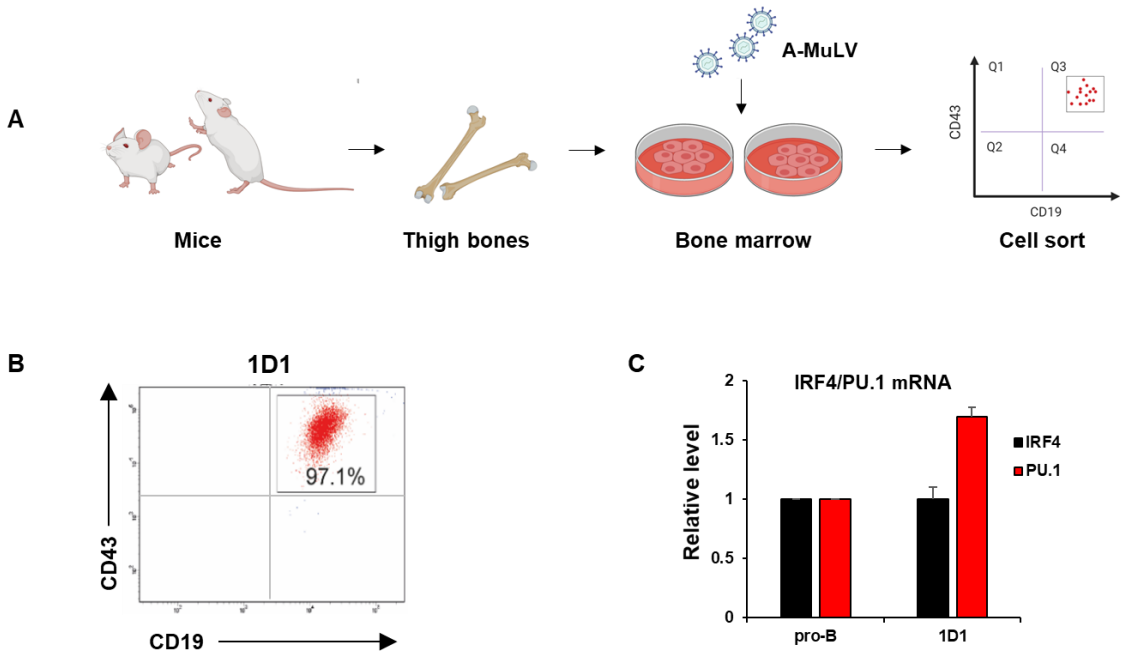

Figure 3.6 – Generation and analysis of a pro-B cell line, 1D1

A) Schematic diagram of generation of A-MuLV immortalized pro-B cell lines

B) Cell surface marker analysis of 1D1 by flow cytometry. 1D1 cells express CD19 and CD43, which are characteristic of pro-B cells. The data shown in this figure was generated by Alastair Smith.

C) Analysis of the levels of IRF4 and PU.1 expression using RT-qPCR. 1D1 cells displayed similar levels of expression of IRF4 and PU.1 to primary pro-B cells. The data shown in this figure was generated by Alastair Smith.

To enable inducible activation of the Igλ locus in the 1D1 cell line, the estrogen receptor ligand binding domain (ER_{T2}) was fused to the IRF4 cDNA and then cloned into the retroviral vector, MSCV-IRES-GFP. The corresponding retrovirus was produced by transfection of the construct, MSCV-IRF4-ER_{T2}-IRES-GFP, into the Phoenix packaging cell line (Grignani et al., 1998). The

resulting retroviruses were introduced into 1D1 cells by spinfection, to generate pro-B cell lines in which IRF4-ER_{T2} is stably expressed. The transduced 1D1 cells with the highest expression of GFP were purified by flow cytometry and monoclonal clones were isolated by seeding cells into semi-solid agar; viable cell colonies were then transferred to RPMI media for further analysis. Twenty monoclonal cell lines were obtained and analysed by flow cytometry. The cell clones with the highest level of GFP expression were selected to analyse the activation of Jλ1 non-coding transcription following induction with the oestrogen mimic, 4-hydroxytamoxifen, for 8 hours. All cell lines displayed an increased level of Jλ1 non-coding transcription post induction (Figure 3.7). The 1D1-IRF4-ER_{T2} pro-B cell clone 15 (referred to as 1D1-T215) shows the most substantial increase of Jλ1 non-coding transcription and was therefore chosen for further analysis.

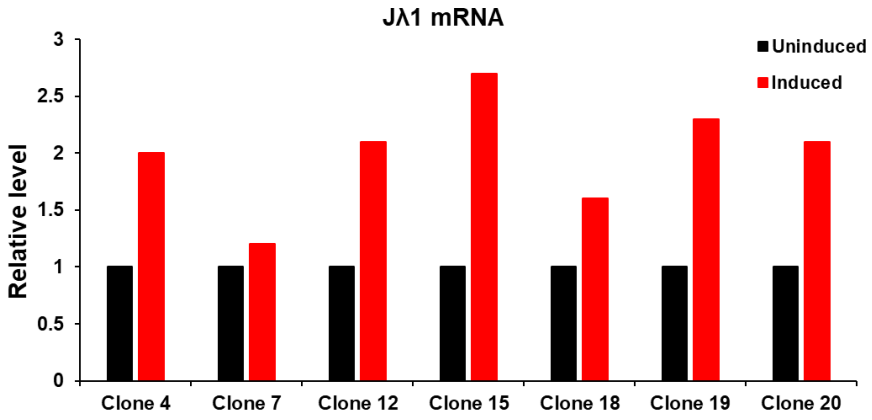

Figure 3.7 – Alteration in Jλ1 transcription in the 1D1 derivatives expressing inducible IRF4-ER_{T2} after induction

Jλ1 non-coding transcription was analysed in 1D1 pro-B cell lines with high GFP expression using RT-qPCR. All cell lines exhibited an increased level of Jλ1 transcription at 8 hours post induction (hpi) with 2 μM 4-hydroxytamoxifen. Clone 15 displayed the highest increase.

The level of IRF4 expression was also examined using RT-qPCR to test if the activation of Igλ in 1D1-T215 cells might be caused by the inducible IRF4. Remarkably, the results show that the 1D1-T215 cell line displayed similar

levels of IRF4 expression to primary pre-B cells (A. Smith, PhD thesis, 2018). Next, I sought to confirm that the 1D1-T215 cell line was a suitable model to investigate the enhancer-mediated activation of non-coding transcription of unarranged gene segments. Initially, I analysed the level of Jλ1 non-coding transcription in 1D1-T215 cells harvested at the time points post induction shown in Figure 3.8A using RT-qPCR. The results show that there is a modest increase of the level of Jλ1 transcription from 0 to 8 hours after IRF4 induction and a relatively sharp increase from 8 to 12 hours post induction (Figure 3.8A). Notably, the IRF4-ER_{T2} transgene in the 1D1-T215 cell line translocates to the nucleus following induction with 4-OH tamoxifen and reaches its highest level in nucleus at just 2 hours post induction (A. Smith, PhD thesis, 2018). I next performed the temporal chromatin immunoprecipitation (ChIP) analysis of IRF4 binding to the Eλ3-1 enhancer to test the correlation between enhancer binding of the transcription activator and target gene activation. Remarkably, I found that the IRF4 binding to Eλ3-1 increases dramatically from 0 to 4 hpi, followed by only a slight increase from 4 to 12 hpi (Figure 3.8B). This suggests that IRF4 binding in the Eλ3-1 enhancer is an early event during the activation of Jλ1 non-coding transcription. Because the chromatin contraction occurs between the enhancer and its cognate promoter during the activation of gene transcription, the interaction frequency between the Eλ3-1 enhancer and Jλ1 promoter should be increased in 1D1-T215 cells post induction. Temporal chromatin conformation capture (3C) analysis was therefore conducted and the results demonstrate that the physical interaction between Eλ3-1 and Jλ1 displayed a substantial increase at 8 hpi just before enhanced Jλ1 transcription is observed (Figure 3.8C). This is consistent with the idea that bringing the enhancers and promoters into close proximity is a prerequisite for efficient transcription of target genes. The close relationship between increased IRF4 binding and enhanced Jλ1 transcription suggest that this is a suitable system to perform temporal ChIP analysis of candidate transcription activators to determine which is important for the activation of the Igλ locus.

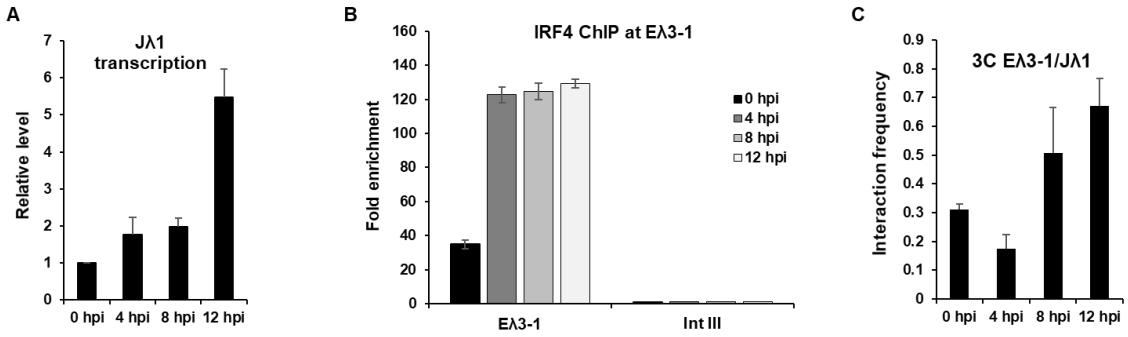

Figure 3.8 – The increase of Jλ1 non-coding transcription correlates with IRF4-mediated Eλ3-1 - Jλ1 interactions

A) The level of Jλ1 non-coding transcription was analysed by RT-qPCR in 1D1-T215 cells following induction. This shows a sharp increase from 8 to 12 hours post-induction, indicating that transcription activation is a relatively late event. Data were normalized to the expression level of the housekeeping gene, *Hprt.* Error bars show standard error of the mean (SEM) from three experimental repeats.

B) IRF4 binding to the Eλ3-1 enhancer was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at Eλ3-1 and Intgene III (negative control region) is shown. IRF4 binding at Eλ3-1 exhibited a sharp increase from 0 to 4 hours following induction and is relatively an early event in the activation of Igλ. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

C) The interaction between Eλ3-1 and Jλ1 was analysed by 3C-qPCR in 1D1-T215 cells following induction. The interactions between Eλ3-1 and Jλ1 show a clear increase from 4 to 8 hpi just before the large increase in Jλ1 transcription. Data were normalized using an interaction within the *ERCC3* locus. Error bars show standard error of the mean (SEM) from three replicates.

In light of the temporal changes in 3C interactions and Jλ1 transcription, it appears that an increased level of IRF4 binding at the Eλ3-1 enhancer triggers an interaction of the enhancer with the Jλ1 promoter, to cause efficient Jλ1 transcription. If this is the case, we might expect to observe IRF4 binding at the Jλ1 promoter where a Eλ3-1-Jλ1 loop has been demonstrated to form in pre-B cells (Figure 3.4B). To verify the IRF4 binding in the Jλ1 promoter in 1D1-T215 cells, temporal ChIP-qPCR analysis was conducted following induction. Unfortunately, however, a significant enrichment of IRF4 at the Jλ1 promoter

was not detected in induced 1D1-T215 cells (Figure 3.9A). To investigate why this might be the case, comparative analysis of the Jλ1 mRNA level in 1D1-T215 cells and primary pre-B cells was carried out using RT-qPCR. The results show that in 1D1-T215 cells, Jλ1 transcription is substantially repressed by approximately 10,000-fold compared with pre-B cells (Figure 3.9B). Subsequent sequence analysis showed that there is a binding site for the transcription repressor, STAT5, located in the Jλ1 promoter. STAT5 has been shown to repress chromatin accessibility of the Igκ locus by binding as a tetramer and recruiting polycomb repression complex 2 (Mandal et al., 2011). STAT5 is activated by v-Abl kinase through the JAK1/3 signalling (Danial and Rothman, 2000) and it seems likely that this mechanism leads to repression of Jλ1 non-coding transcription in 1D1-T215 cells, which are an Abl-kinase-derived cell line.

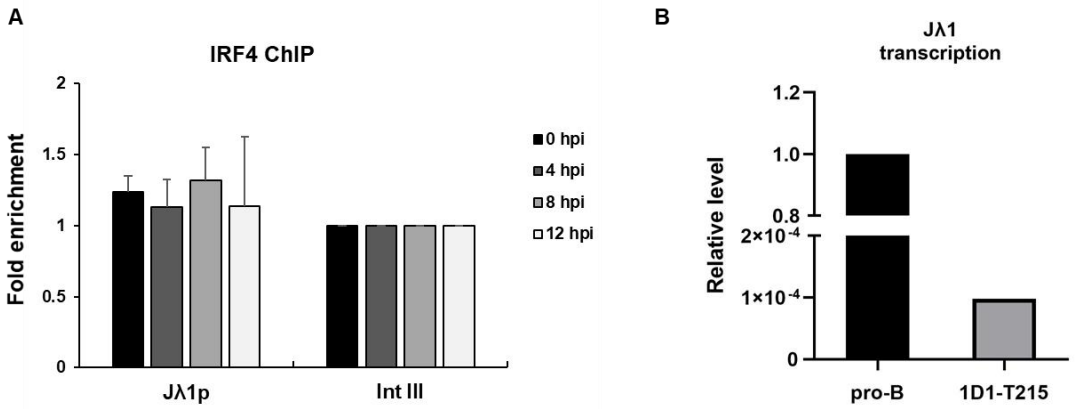

Figure 3.9 – Jλ1 non-coding transcription is strongly repressed in A-MuLV immortalized pro-B cell lines

A) IRF4 binding to the Jλ1 promoter was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at Jλ1p and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

B) The level of Jλ1 non-coding transcription was analysed by RT-qPCR in uninduced 1D1-T215 cells and pro-B cells. Data were normalized to the expression level of the housekeeping gene, *Hprt.*

Vλ1-Jλ1 recombination accounts for approximately 70% of the total recombination events of the Igλ locus (Boudinot et al., 1994) and consequently, Vλ1 non-coding transcription is as important as Jλ1 transcription for Igλ recombination. Therefore, it seemed possible that Vλ1 might be used instead as a target gene for investigating Eλ3-1-mediated gene activation. The Vλ1 promoter therefore was subjected to sequence analysis and conserved binding motifs for STAT5 were not found. The level of Vλ1 transcription in 1D1-T215 cells and pre-B cells was examined and the results showed that the 1D1 cell line displays a similar level of Vλ1 transcription to pre-B cells (A. Smith, PhD thesis, 2018). To test if Vλ1 non-coding transcription is controlled by Eλ3-1, temporal 3C analysis was carried out in 1D1-T215 cells which suggested that the physical distance between Eλ3-1 and Vλ1 is reduced following induction (Figure 3.10A). This implies that Eλ3-1 is also an enhancer of Vλ1 transcription. Furthermore, temporal analysis of the level of Vλ1 transcription was conducted by RT-qPCR and the data demonstrate that Vλ1 transcription is increased similarly to Jλ1 following induction (Figure 3.10B) and correlates with the changes in 3C interaction between Eλ3-1 and Vλ1. Combined with the evidence that IRF4 binding in the Vλ1 promoter displays a gradual increase following induction (Figure 3.10C), it appears that Vλ1 is a more robust system than Jλ1 to dissect the mechanism by which enhancer-mediated gene activation occurs.

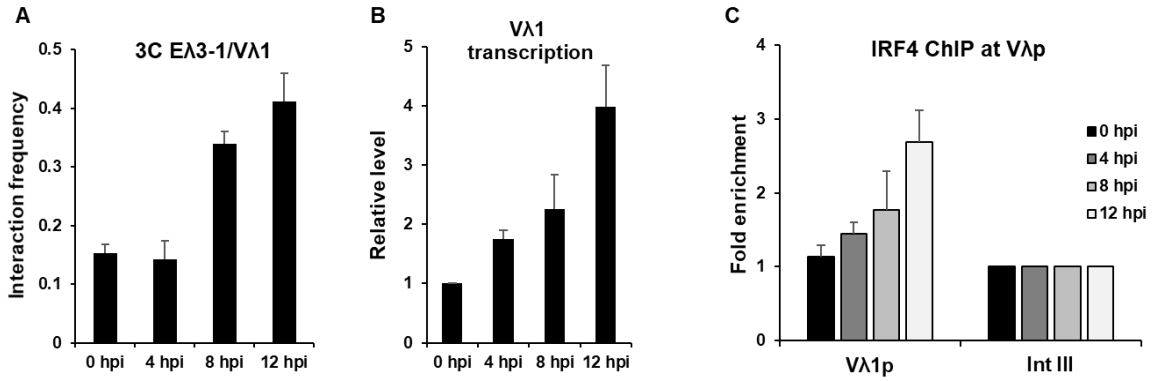

Figure 3.10 – Vλ1 is an ideal model to investigate Eλ3-1-mediated promoter activation in 1D1-T215 cells

A) The interaction between Eλ3-1 and Vλ1 was analysed by 3C-qPCR in 1D1-T215 cells at the time indicated following induction. A clear increase is observed from 4 to 8 hpi, which is just before Vλ1 becomes efficiently transcribed. Data were normalized by detecting an interaction within the *ERCC3* locus. Error bars show standard error of the mean (SEM) from three replicates.

B) The level of Vλ1 non-coding transcription was analysed by RT-qPCR in 1D1-T215 cells following induction. Jλ1 transcription shows a relatively sharp increase from 8 to 12 hours following induction. Data were normalized to the expression level of the housekeeping gene, *Hprt*. Error bars show standard error of the mean (SEM) from three experimental repeats.

C) IRF4 binding at the Vλ1 promoter was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at Vλ1p and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

3.4 IRF4 increases the chromatin accessibility of the enhancer and promoter through recruiting E2A and p300

The basic helix-loop-helix (bHLH) transcription factor E2A is known to interact with IRF4 (Lazorchak et al., 2006) and play a crucial role in promoting non-coding transcription of unarranged Igλ gene segments in pre-B cells as demonstrated by knock-out studies (Beck et al., 2009) which also results in a significant decrease in the number of surface Igλ$^+$ cells in bone marrow (Beck et al., 2009). To determine if E2A is involved in the regulation of Vλ1 non-coding

transcription, temporal ChIP-qPCR analysis of E2A binding to the Eλ3-1 enhancer and Vλ1 promoter was performed in 1D1-T215 cells. As can be seen in Figure 3.11, E2A is clearly enriched at the Eλ3-1 enhancer and its binding increases gradually following induction. Although E2A binding is at low levels in the Vλ1 promoter, the enrichment in the promoter is reproducible and correlates with E2A binding to the enhancer (Figure 3.11).

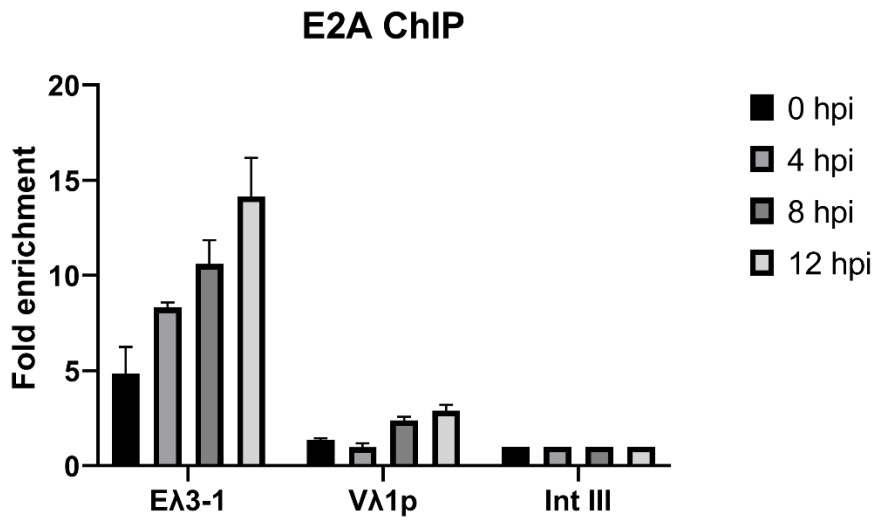

Figure 3.11 – E2A is recruited in both Eλ3-1 and Vλ1p in 1D1-T215 cells
E2A binding to Eλ3-1 and the Vλ1 promoter was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at Eλ3-1 and Vλ1p and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

E2A proteins play an essential role in recruiting histone-modifying activities (Sakamoto et al., 2012). p300 is a histone acetyltransferase that has been shown to exert its function in concert with numerous transcription factors and can acetylate histones close to transcriptional enhancers and promoters, facilitating the generation of more flexible and accessible chromatin (Vo and Goodman, 2001). Co-immunoprecipitation experiments demonstrated that E2A directly interacts with several histone acetyltransferases, including p300, that were also shown to act in synergy with p300 to activate the Igκ locus (Bradney

et al., 2003; Qiu et al., 1998). Consistent with this, E2A depletion in pre-B cells reduced the level of histone acetylation at enhancers within the Igκ locus (Lazorchak et al., 2006). Therefore, temporal ChIP analysis was performed to determine p300 binding at the Eλ3-1 enhancer and Vλ1 promoter to investigate if p300 is involved in the activation of Igλ. This showed that p300 is greatly enriched at the Eλ3-1 enhancer and displayed the largest increase from 0 to 4 hpi following induction (Figure 3.12). A moderate but reproducible increase of binding was also observed at the Vλ1 promoter (Figure 3.12), indicating that chromatin accessibility at the enhancer and promoter is likely increased. In turn, this is expected to facilitate the recruitment of more diverse transcription factors.

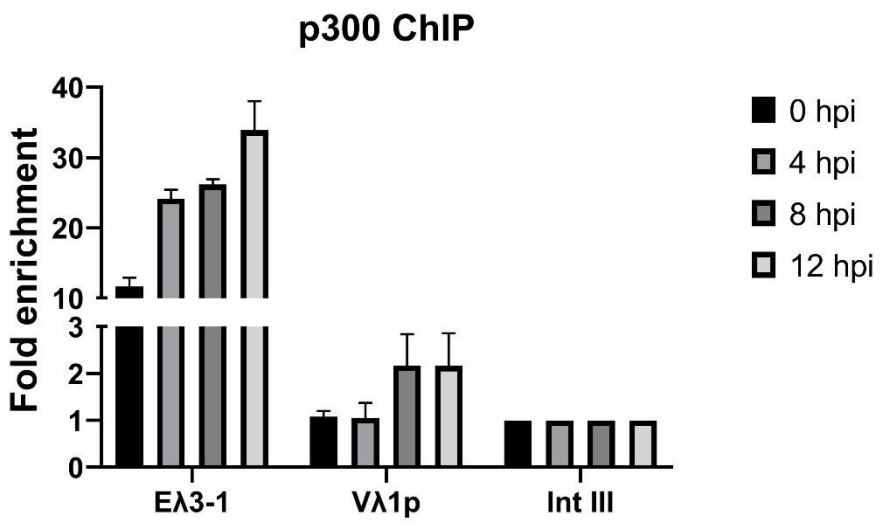

Figure 3.12 – p300 is recruited to both Eλ3-1 and Vλ1p in 1D1-T215 cells
p300 binding to the Eλ3-1 and Vλ1 promoter was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at Eλ3-1 and Vλ1p and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

3.5 IRF4 directly interacts with the Mediator complex to activate the Vλ1 non-coding transcription

The Mediator complex is an evolutionarily conserved, multi-subunit protein complex that plays an essential role in the regulation of enhancer-promoter

communications and PIC assembly (Lin et al., 2011; Malik and Roeder, 2016). The Mediator complex consists of more than 30 subunits which are organized into four distinct modules, termed the head, middle, tail and kinase modules (Allen and Taatjes, 2015). The head and middle modules carry out the most basic functions via interplay with RNAPII and other components of the preinitiation complex (Esnault et al., 2008; Robinson et al., 2012). Subunits of tail module physically interact with enhancer bound transcription activators (Ansari and Morse, 2012). Thus, it is thought that Mediator provides a physical bridge between transcription activators bound at enhancers and components of the preinitiation complex bound at promoters (Malik and Roeder, 2016). Med23 is the largest subunit in the tail module and has been shown to be essential for early B cell development (Chen et al., 2018). To investigate if Med23 is involved in the activation of Igλ non-coding transcription, co-immunoprecipitation experiments were conducted, which revealed a physical interaction between IRF4 and Med23 (Figure 3.13A). To determine if Med23 is required for Igλ activation, knock-down of Med23 expression was performed by using a shRNA lentiviral system, pLKO.1 (Moffat et al., 2006). Western blotting analysis demonstrated that compared to 1D1-T215 cells expressing a scrambled shRNA (shSCR), Med23 protein levels are diminished dramatically in 1D1-T215 cells expressing a shRNA targeting Med23 (shMed23, Figure 3.13B). Consistent with a role for Mediator in the regulation of transcription, I find that compared with shSCR 1D1-T215 cells, Vλ1 non-coding transcription is decreased significantly in shMed23 1D1-T215 cells (Figure 3.13C). Likewise, 3C analysis revealed that the interaction between Eλ3-1 and Vλ1 is disrupted in shMed23 1D1-T215 cells (Figure 3.13D). These data indicate that the Mediator complex is indispensable for the activation of Vλ1 non-coding transcription.

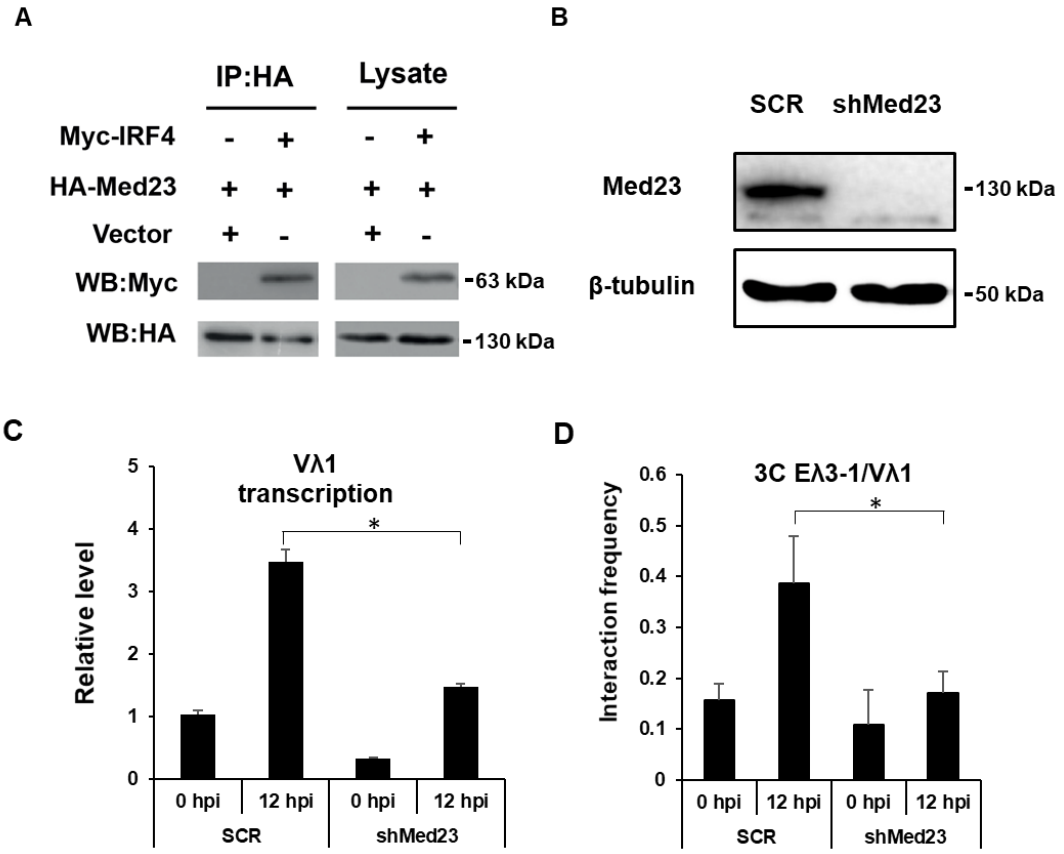

Figure 3.13 – Med23 is essential for activation of Vλ1 non-coding transcription

A) IRF4 physically interacts with Med23. IRF4 was overexpressed in 293T cells as a Myc-tagged fusion protein with HA-tagged Med23. Cells were harvested at 36 hours post transfection and approximately 5% of the cell lysate was used as input whilst the remaining lysate was immunoprecipitated with 1 µg of anti-HA antibody. Cell lysates and immunoprecipitated samples were resolved by 10% SDS-PAGE and western blotting analysis was conducted with the indicated antibodies.

B) Western blot analysis of the level of Med23 expression in shSCR and shMed23 1D1-T215 cells. The protein level of Med23 is diminished dramatically in shMed23 cells. β-tubulin was used as a loading control.

C) RT-qPCR analysis of Vλ1 non-coding transcription in shSCR and shMed23 1D1-T215 cells. Vλ1 non-coding transcription is reduced in Med23 knock-down cells. Data were normalized to the expression level of the housekeeping gene, *Hprt*.

D) 3C analysis of interactions between Eλ3-1 and Vλ1 in shSCR and shMed23 1D1-T215 cells. The interaction frequency between Eλ3-1 and Vλ1 is decreased in Med23 knock-down cells. Data were normalized by detecting an interaction with the *ERCC3* locus.

Error bars show standard error of the mean (SEM) from three replicates. * represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

The data above demonstrate that Med23 is essential for the activation of Vλ1 non-coding transcription, and to further investigate the role of Med23, temporal ChIP is required to test if Med23 binding to the enhancer and promoter correlates with the activation of transcription. Unfortunately, a commercial ChIP-grade antibody is not available that can be used for this analysis. Med1 is the largest subunit of the Mediator complex and belongs to the middle module (Tsai et al., 2014). Given that a ChIP-grade antibody against Med1 is available and the middle module, to which Med1 belongs connects to both the head and tail module, I examined Med1 binding by ChIP analysis. To verify that Med1 is required for the activation of Vλ1 non-coding transcription and has similar effects to Med23, knock-down of Med1 expression was carried out using the shRNA lentiviral system outlined above. Western blotting analysis demonstrated that Med1 protein levels are diminished dramatically in shMed1 1D1-T215 cells (Figure 3.14A). Combined with the reduced level of Vλ1 non-coding transcription in Med1 knock-down cells (Figure 3.14B), Med1 appears to have a similar role in the activation of Vλ1 non-coding transcription to Med23. Therefore, temporal ChIP analysis of Med1 binding was conducted in 1D1-T215 cells. The results show that Med1 occupancy at Eλ3-1 is significant and displays the biggest relative increase from 0 to 4 hpi following induction (Figure 3.14C). Compared to the enhancer, Med1 binding to the Vλ1 promoter is low but it is reproducible and correlates with Vλ1 non-coding transcription (Figure 3.14C). The high level of Med1 binding to the Eλ3-1 enhancer is possibly due to the strong binding of IRF4 to the enhancer, which leads to its recruitment. Together, these data suggest that Mediator complex recruitment by IRF4 to enhancers and promoters is indispensable for Igλ activation.

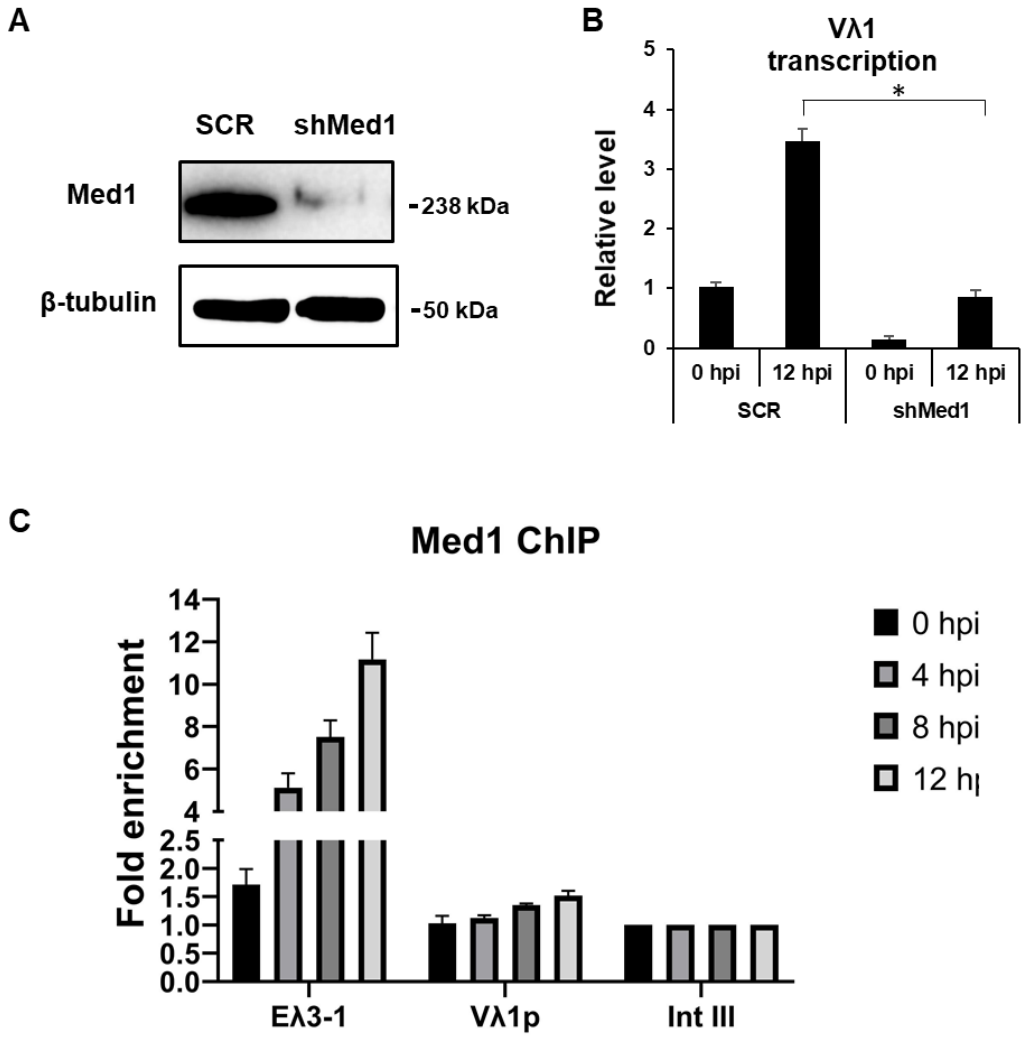

Figure 3.14 – Med1 is essential for the activation of Vλ1 non-coding transcription

A) Western blot analysis of Med1 expression levels in sh-SCR and shMed1 1D1-T215 cells. The protein level of Med23 is diminished dramatically in Med23 knock-down cells. β-tubulin was used as a loading control.

B) RT-qPCR analysis of Vλ1 non-coding transcription in shSCR and shMed1 1D1-T215 cells. Vλ1 non-coding transcription is significantly reduced in Med1 knock-down cells. Data were normalized to the expression level of the housekeeping gene, *Hprt*.

C) Med1 binding in the Eλ3-1 and Vλ1 promoter were analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at Eλ3-1 and Vλ1p and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control.

Error bars show standard error of the mean (SEM) from three experimental repeats.

* represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

3.6 CDK7 directs Vλ1 non-coding transcription via phosphorylating the Ser 5 residue of C-terminal domain (CTD) of RNAPII

The data presented above indicate that the Mediator complex could be delivered from Eλ3-1 to the Vλ1 promoter. This then leads to the question of how increased levels of Mediator at the promoter facilitate transcription. Before addressing this, it is important to analyse the level of Ser 5 phosphorylated RNAPII at target gene promoters. Previously, I showed data using primary pre-B cells that activation of non-coding transcription of unarranged gene segments of Igλ is tightly associated with the increased level of Ser 5 phosphorylated RNAPII binding at promoters (Figure 3.5B). It is well documented that CDK7 is responsible for phosphorylating the Ser 5 residue of CTD of RNAPII at the promoters of protein-coding genes (Valay et al., 1995). CDK7 is a subunit of the TFIIH complex which can be recruited by the Mediator complex to the promoters of target genes (Esnault et al., 2008). However, the Mediator complex itself contains a kinase subunit, CDK8, which forms part of the kinase domain. Previous studies have demonstrated that CDK8 is capable of catalysing the phosphorylation at the Ser 5 residue of CTD of RNAPII in *vitro* (Liao et al., 1995; Sun et al., 1998). To determine which kinase contributes to the activation of RNAPII in the Vλ1 promoter, knock-down of CDK7 expression was initially performed using the shRNA lentiviral system in 1D1-T215 cells. Western blotting analysis demonstrated that compared with the 1D1-T215 cells expressing a scrambled shRNA, CDK7 protein levels are diminished in 1D1-T215 cells that express a shRNA targeting CDK7 (Figure 3.15A). However, the global level of Ser 5 phosphorylated RNAPII does not change significantly (Figure 3.15A). Consistent with this, analysis of the level of Vλ1 non-coding transcription showed that only a limited decrease is observed in the CDK7 knock-down cells (Figure 3.15B). This is possibly due to the fact that the residual low level of CDK7 expression is sufficient to phosphorylate Ser 5 of RNAPII CTD.

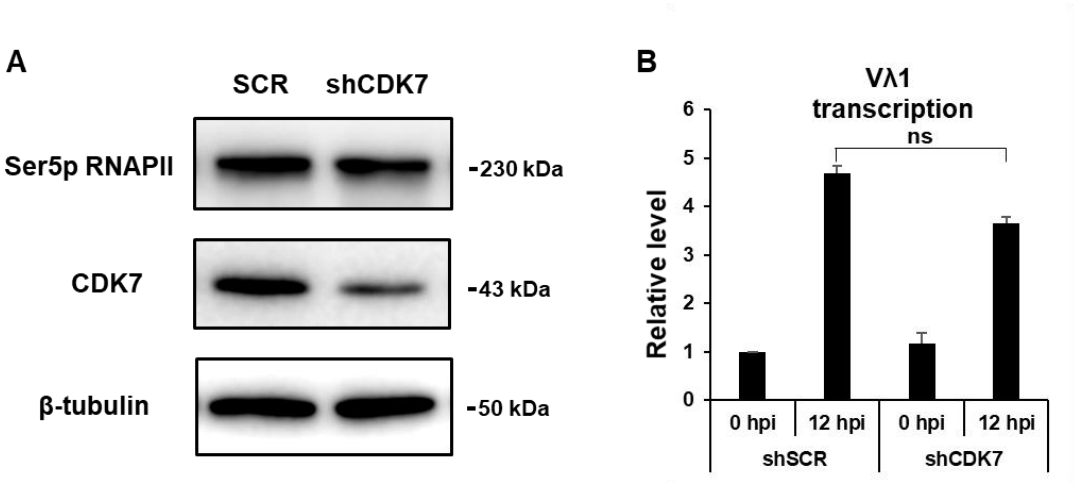

Figure 3.15 – Knock-down of CDK7 leads to a limited decrease of Vλ1 non-coding transcription after induction

A) Western blot analysis of the levels of CDK7 and Ser 5 phosphorylated RNAPII in sh-SCR and shCDK7 1D1-T215 cells. The protein level of CDK7 is clearly reduced in shCDK7 cells but no significant change of the level of Ser 5 phosphorylated RNAPII is observed in CDK7 knock-down cells. β-tubulin was used as a loading control.

B) RT-qPCR analysis of Vλ1 non-coding transcription in shSCR and shCDK7 1D1-T215 cells. Data were normalized to the expression level of the housekeeping gene, *Hprt*. Error bars show standard error of the mean (SEM) from three experimental repeats.

To inhibit the catalytic activity of CDK7 to a much greater extent, a small molecule inhibitor, THZ1, was used to treat 1D1-T215 cells. THZ1 is a covalent CDK7 inhibitor which has the unprecedented ability to target a cysteine residue residing outside of the canonical kinase domain (Kwiatkowski et al., 2014). To optimise the amount used, 1D1-T215 cells were initially treated with different concentrations of the inhibitor for 8 hours. Western blotting analysis revealed that treating the cells with 125 nM of THZ1 does not change the protein level of CDK7 but is sufficient to diminish nearly all of the Ser 5 phosphorylated RNAPII (Figure 3.16A). Therefore, 1D1-T215 cells treated with 125 nM of THZ1, were harvested at different time points and subject to Western blotting analysis. The results showed that treating cells with the inhibitor for 2 hours dramatically reduces Ser 5 phosphorylated RNAPII and almost all of the Ser 5 phosphorylated RNAPII is depleted after 8 hours treatment (Figure 3.16B). To

analyse the influence of CDK7 inhibitor on Vλ1 non-coding transcription, 1D1-T215 cells were firstly treated with 125 nM THZ1 for 2 hours. After removal of the CDK7 inhibitor, cells were treated with 2 μM 4-hydroxytamoxifen for 12 hours to activate IRF4-ER_{T2}. RT-qPCR results demonstrate that Vλ1 non-coding transcription is substantially repressed in inhibitor-treated cells (Figure 3.16C), indicating that CDK7 is essential for RNAPII Ser 5 phosphorylation during the activation of Igλ locus transcription.

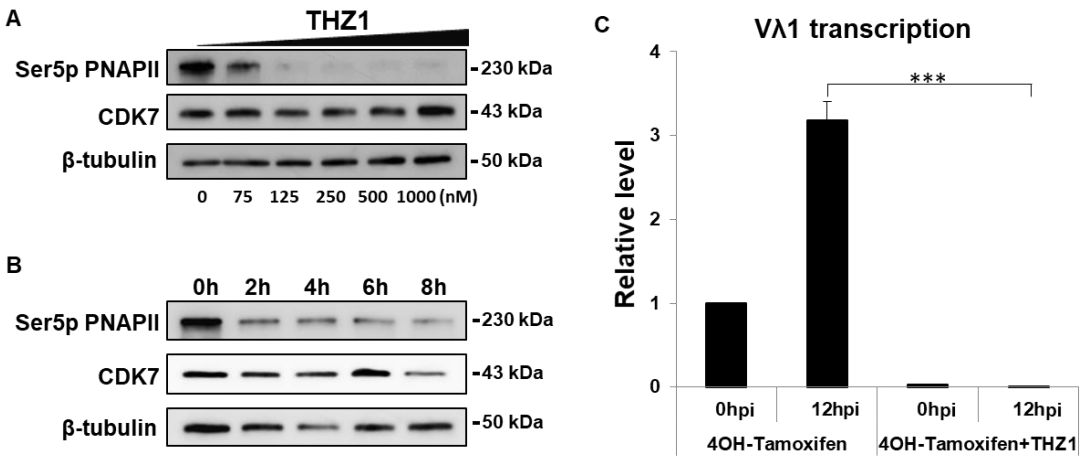

Figure 3.16 – Inhibition of CDK7 severely impairs the Vλ1 non-coding transcription

A) Western blot analysis of the level of CDK7 and Ser 5 phosphorylated RNAPII in cells treated with different concentrations of the CDK7 inhibitor THZ1 for 8 hours. THZ1 does not change the protein level of CDK7 but is enough to diminish nearly all of the Ser 5 phosphorylated RNAPII at a concentration of 125 nM. β-tubulin was used as a loading control.

B) Western blot analysis of the level of CDK7 and Ser 5 phosphorylated RNAPII in cells treated with the same amount of THZ1 inhibitor (125 nM) but harvested at different time points. THZ1 significantly reduces the level of Ser 5 phosphorylated RNAPII in cells treated with 125 nM THZ1 for 2 hours. β-tubulin was used as a loading control.

C) RT-qPCR analysis of Vλ1 non-coding transcription in 1D1-T215 cells treated with CDK7 inhibitor, THZ1. Before induction with 4-hydroxytamoxifen, 1D1-T215 cells were treated with 125 nM THZ1 to repress the level of Ser 5 phosphorylated RNAPII. Vλ1 non-coding transcription is substantially reduced in CDK7 inhibitor-treated cells both

before and after induction. Data were normalized to the expression level of the housekeeping gene, *Hprt*.
Error bars show standard error of the mean (SEM) from three experimental repeats.
* represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

To investigate if CDK8 catalyses the phosphorylation of the Ser 5 residue of CTD of RNAPII in 1D1-T215 cells, 1D1-T215 cells were treated with a small molecule CDK8 inhibitor, SEL120-34A, followed by induction with tamoxifen and analysis of the level of Vλ1 non-coding transcription. SEL120-34A is a potent, selective and competitive inhibitor that interferes with ATP binding to CDK8 (Rzymski et al., 2017). Low concentrations of the inhibitor (1-100 nM) are sufficient to substantially inhibit the catalytic activity of CDK8 in different cancer cells (Rzymski et al., 2017). To test if this CDK8 inhibitor inhibits the activation of Igλ in 1D1-T215 cells, 100 nM SEL120-34A was used to treat cells for 12 hours followed by the analysis of the level of Ser 5 phosphorylated RNAPII. Western blotting showed that there is no change of the global level of Ser 5 phosphorylated RNAPII in 1D1-T215 cells treated with SEL120-34A (data not shown). Analysis of the level of Vλ1 non-coding transcription in SEL120-34A treated cells demonstrated that SEL120-34A treatment does not change the level of Vλ1 non-coding transcription (Figure 3.17A). By contrast, the mRNA level of CDK8 target genes, including STAT1 and IRF9 (Rzymski et al., 2017), decreased significantly in 1D1-T215 cells treated with SEL120 (Figure 3.17B). These data suggest that CDK7 mediates phosphorylation of the Ser 5 residue of CTD of RNAPII during Igλ locus activation.

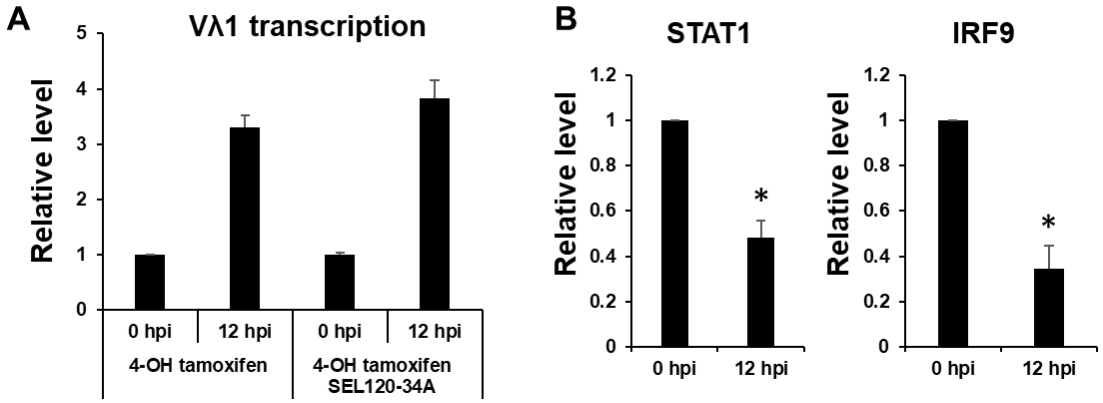

Figure 3.17 – Inhibition of CDK8 does not change Vλ1 non-coding transcription

A) RT-qPCR analysis of Vλ1 non-coding transcription in 1D1-T215 cells treated with CDK8 inhibitor, SEL120-34A. Before induction with 4-hydroxytamoxifen, cells were treated with a high level of SEL120-34A (100 nM) for 12 hours.

B) RT-qPCR analysis of the mRNA level of CDK8 target genes STAT1 and IRF9 in 1D1-T215 cells induced by the CDK8 inhibitor, SEL120-34A (100 nM). Data were normalized to the expression level of the housekeeping gene, *Hprt*.

Error bars show standard error of the mean (SEM) from three experimental repeats.

* represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

3.7 The change between the activating and elongating form of RNAPII in the enhancer-promoter loop during the activation of transcription

Ser5 phosphorylated RNAPII represents the activated form of RNAPII which my data suggest is catalysed by CDK7 during the activation of Vλ1 non-coding transcription (Figure 3.16). To determine the point at which the activated form of RNAPII changes to the elongating form in the enhancer-promoter loop, temporal ChIP analysis of Ser 5 phosphorylated RNAPII binding was performed in 1D1-T215 cells. Remarkably, the level of Ser 5 phosphorylated RNAPII binding is enriched both at Eλ3-1 and Vλ1p. The binding of Ser 5 phosphorylated RNAPII at Eλ3-1 displayed a gradual increase following induction (Figure 3.18), correlating with Mediator binding to the Eλ3-1 enhancer (Figure 3.14C). This suggests that the high level of binding of Ser 5 phosphorylated RNAPII at Eλ3-1 may be due to the strong binding of Mediator to the enhancer. Compared to the gradual increase of binding at Eλ3-1, Ser 5

phosphorylated RNAPII bound at the Vλ1 promoter is highest at 8 hours post induction (Figure 3.18).

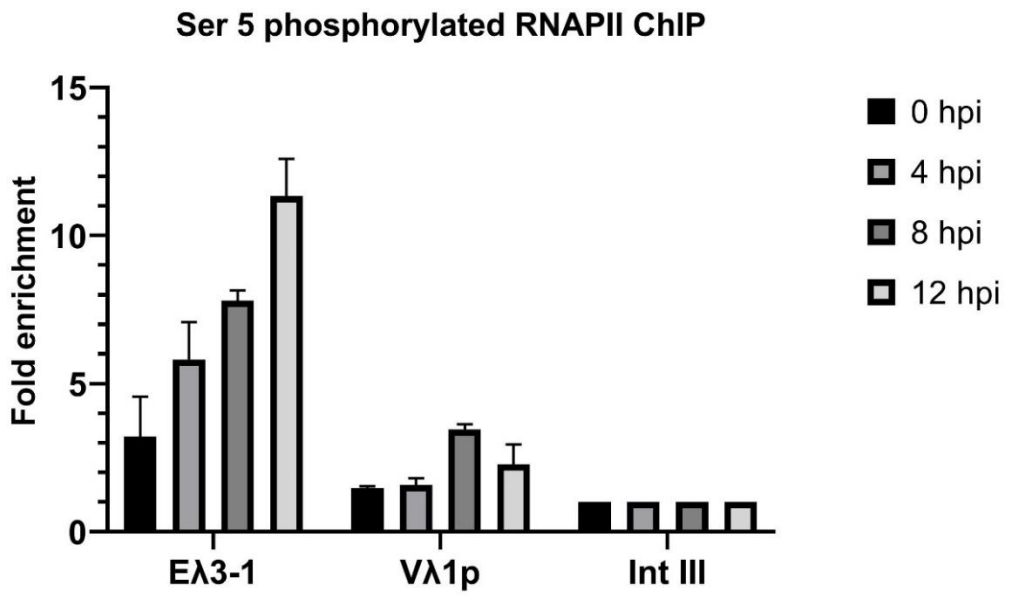

Figure 3.18 – Phosphorylation of the Ser 5 residue of RNAPII CTD is activated at the Eλ3-1 enhancer and Vλ1 promoter in 1D1-T215 cells following induction
Ser 5 phosphorylated RNAPII binding to the Eλ3-1 and Vλ1 promoter was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at Eλ3-1 and Vλ1p and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

Temporal analysis of Vλ1 non-coding transcription showed that the efficient transcription of Vλ1 begins at around 8 hours post induction and increases further by 12 hours (Figure 3.10B). These data suggest that the activating form of RNAPII at the Vλ1 promoter could be converted to the elongating form, namely Ser 2 phosphorylated RNAPII, from around 8 hours post induction. To test if this is the case, temporal ChIP analysis of Ser 2 phosphorylated RNAPII binding to the Vλ1 promoter was performed. Consistent with the observed changes in transcription, Ser 2 phosphorylated RNAPII bound to the Vλ1 promoter undergoes the greatest change from 8 to 12 hpi (Figure 3.19). As the Ser 2 phosphorylated RNAPII is catalysed by the positive elongation factor, p-

TEFb (Price, 2000) it is highly likely that a transcription activator functions in the late stages of the activation of Vλ1 non-coding transcription via activating p-TEFb.

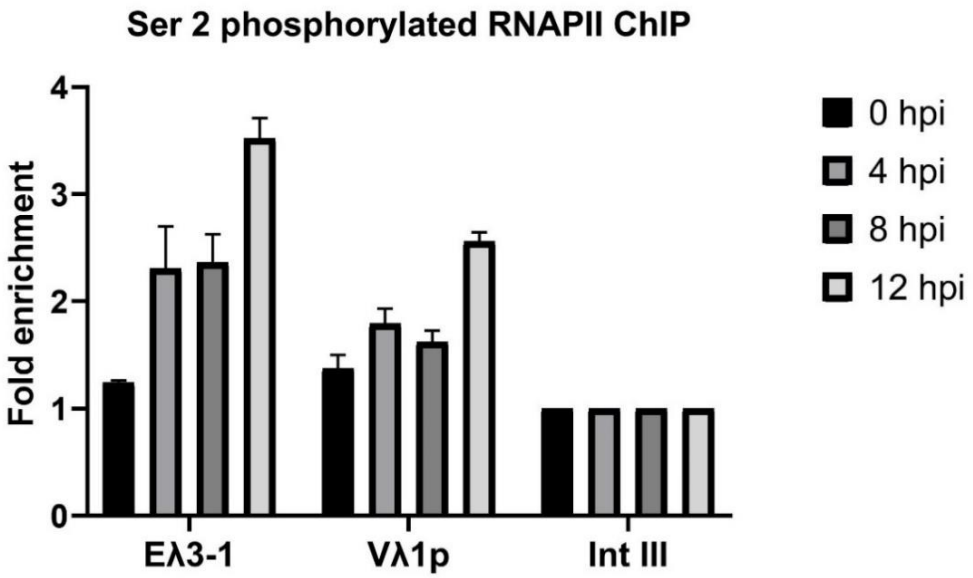

Figure 3.19 – Phosphorylation of the Ser 2 residue of CTD of RNAPII is activated at the Eλ3-1 enhancer and Vλ1 promoter in 1D1-T215 cells after 8 hpi

Ser 2 phosphorylated RNAPII binding to the Eλ3-1 enhancer and Vλ1 promoter was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at Eλ3-1 and Vλ1p and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

3.8 The architecture factor YY1 facilitates Vλ1 non-coding transcription at the late stage of activation

YY1 is a ubiquitously expressed transcription factor that belongs to the zinc finger family of DNA binding proteins. It can activate or repress transcription, depending on the context in which it binds (Sarvagalla et al., 2019). YY1 also plays an important role in mediating the chromatin folding of the IgH locus as evidenced by a YY1 conditional knock-out which led to a decrease in chromatin looping events (Liu et al., 2007). According to published ChIP-seq data from pre-B cells (Kleiman et al., 2016), YY1 binding is enriched at the Eλ3-1

enhancer. To investigate if YY1 is involved in the activation of Vλ1 non-coding transcription and the regulation of chromatin organization of the Igλ locus, knock-down of YY1 expression was performed using the shRNA lentiviral system. Western blotting analysis showed that YY1 protein levels are diminished dramatically in shYY1 1D1-T215 cells (Figure 3.20A). Analysis of Vλ1 transcript levels was subsequently performed using RT-qPCR which showed that Vλ1 non-coding transcription is significantly repressed in 1D1-T215 cells that express a specific shRNA against YY1 (Figure 3.20B). Furthermore, 3C analysis showed that interaction frequency between Eλ3-1 and Vλ1 is reduced in YY1 knock-down cells (Figure 3.20C). These data therefore indicate that YY1 is essential for the activation of Vλ1 non-coding transcription and chromatin organization of the Igλ locus.

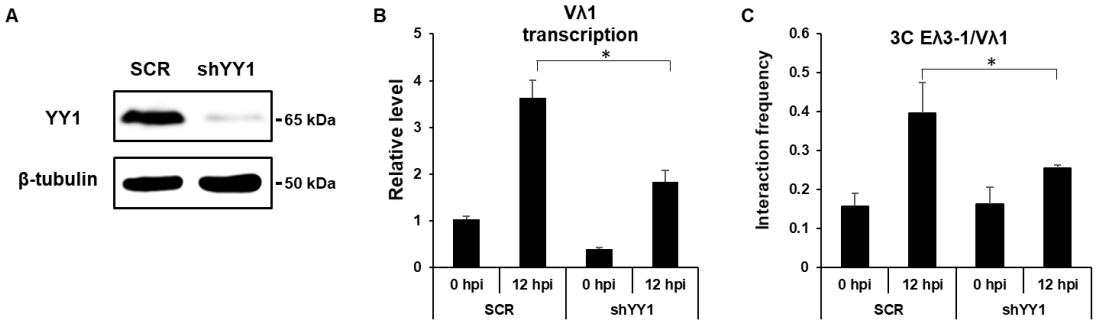

Figure 3.20 – YY1 is essential for the activation of Vλ1 non-coding transcription

A) Western blot analysis of YY1 expression levels in sh-SCR and shMed23 1D1-T215 cells. YY1 protein levels are diminished dramatically in shYY1 1D1-T215 cells. β-tubulin was used as a loading control.

B) RT-qPCR analysis of Vλ1 non-coding transcription in shSCR and shYY1 1D1-T215 cells. Vλ1 non-coding transcription is reduced in YY1 knock-down cells. Data were normalized to the expression level of the housekeeping gene, *Hprt*. Error bars show standard error of the mean (SEM) from three experimental repeats.

C) 3C analysis of interactions between Eλ3-1 and Vλ1 in shSCR and shYY1 1D1-T215 cells. The interaction frequency between Eλ3-1 and Vλ1 is decreased in YY1 knock-down cells. Data were normalized by detecting an interaction with the ERCC3 locus. Error bars show standard error of the mean (SEM) from three replicates.

* represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

To determine the time at which YY1 starts to act in the activation of Vλ1 transcription, temporal ChIP analysis was carried out in 1D1-T215 cells. Intriguingly, YY1 is enriched at both the Eλ3-1 enhancer and Vλ1 promoter and the level of binding increases from 8 to 12 hpi in both regions (Figure 3.21). To determine if the increased YY1 binding is directly caused by IRF4, co-immunoprecipitation was performed. The results showed there are no direct interactions between YY1 and IRF4 (data not shown). Previous studies have demonstrated the physical interactions between YY1 and p300 (Lee et al., 1995), and thus the increased level of YY1 may be caused by the increased recruitment of p300 to the enhancer and promoter as discussed further below. As the increase of YY1 binding to the enhancer and promoter is a late event during the activation of Vλ1 non-coding transcription and occurs in cells after the loop between Eλ3-1 and Vλ1 is already formed, as determined by temporal 3C (Figure 3.10A), a potential function of YY1 in Igλ activation may be to secure the pre-formed the enhancer-promoter loop, thus facilitating the efficient transcription of the Vλ1 gene segment.

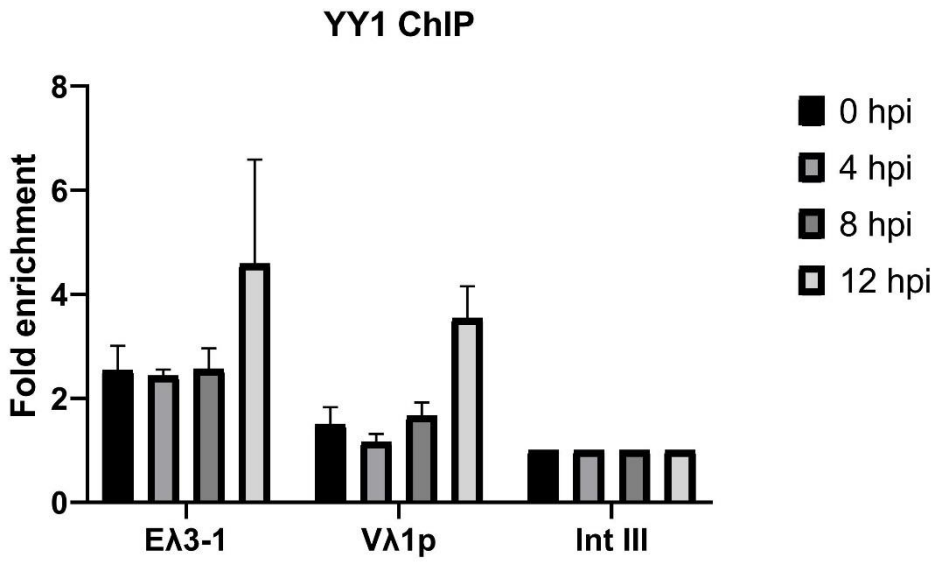

Figure 3.21 – YY1 is recruited in Eλ3-1 and Vλ1p at the late stage of Vλ1 non-coding transcription

YY1 binding to the Eλ3-1 and Vλ1 promoter was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at Eλ3-1 and Vλ1p and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

C) Discussion

This chapter aimed to examine enhancer-mediated activation of transcription using the Eλ3-1 enhancer and Vλ1 gene segment of the mouse Igλ locus as a model. To this end, I have characterised an inducible IRF4-ER pro-B cell line which allows me to induce the Eλ3-1 enhancer and follow its activation of the Vλ1 gene segment temporally. Target gene transcription, activator binding and long-range interactions between the enhancer and promoter were determined using this system. I report here IRF4 potentially increases chromatin accessibility of the Eλ3-1 enhancer and Vλ1 promoter by recruiting E2A and p300. Mediator bound to the Eλ3-1 enhancer and Vλ1 promoter, through direct interactions with IRF4, might bridge the enhancer to the promoter and facilitate the Ser 5 phosphorylation of RNAPII CTD within the enhancer-promoter loop to achieve efficient transcription. In addition, the architecture factor YY1 recruited to Eλ3-1 and the Vλ1 promoter could be involved in the stabilization of the enhancer-promoter loop.

3.9 Incomplete assembly of basal RNAPII machinery at promoters of unarranged gene segments of the Igλ locus in pro-B cells

Gene transcription is usually rate-limited at the levels of (a) preinitiation complex (PIC) assembly and/or (b) RNAPII release from promoter-proximal regions to productive elongation (Jonkers and Lis, 2015). To investigate how the Eλ3-1 enhancer activates the promoters of unarranged gene segments of the Igλ locus, it is important to know the initial state of RNAPII machinery present at the promoters of unarranged gene segments. Firstly, ChIP analysis of Rpb1 binding, the major subunit of RNAPII, was performed in primary pro-B and pre-B cells. Notably, Rpb1 binding to the Jλ1 promoter does not change from pro-B to pre-B cells, suggesting the RNAPII is already present at the Jλ1 promoter in pro-B cells. However, this does not mean that the assembly of PIC is completed in pro-B stage as the hallmark of completion of assembly of PIC is the phosphorylation of the Ser 5 residue of RNAPII CTD, which requires CDK7 to be recruited at the promoter to catalyse this phosphorylation. It was previously demonstrated that phosphorylation of the Ser 5 residue of RNAPII CTD is tightly associated with levels of H3K4me3 at promoters (Ng et al., 2003).

Furthermore, the level of H3K4me3 at the Jλ1 promoter region in pro-B cells was shown to be significantly lower than that in pre-B cells (Bevington and Boyes, 2013), implying that the Ser 5 residue of RNAPII CTD at the Jλ1 promoter is not phosphorylated in pro-B cell stage. ChIP analysis revealed that the binding of Ser 5 phosphorylated RNAPII to the Jλ1 promoter increases significantly from pro-B to pre-B cells. Thus, it is highly likely that the increased IRF4 binding to the Eλ3-1 enhancer plays an essential role in triggering the phosphorylation of the Ser 5 residue of RNAPII CTD bound to the Jλ1 promoter.

3.10 The 1D1-T215 cell line represents an ideal system to investigate enhancer-mediated promoter activation

To develop a pro-B cell line that expresses inducible IRF4, the ER ligand binding domain (Eng et al., 1997; Whitfield et al., 2015) was initially fused in frame with the N-terminus of IRF4. IRF4-ER was then cloned upstream of the IRES element in the MSCV-IRES-GFP construct for producing retrovirus. The pro-B cell line 1D1 was infected with this retrovirus for 48 hours and subsequently cells were selected by flow cytometry using the GFP reporter. The highest 10% of GFP expressing 1D1 cells were purified by flow cytometry and monoclonal cell lines were isolated using the semi-solid agar. However, IRF4-ER was found to be present in the nuclear extract of untreated cells by Western blotting (A. Smith, PhD thesis, 2018). This is likely due to the presence of contaminating estrogen and estrogen mimics in the foetal calf serum and phenol red in the cell culture media, resulting in the premature activation of IRF4-ER in 1D1 cells. In addition, previous publications demonstrated that the binding of β-estradiol to the ER can facilitate the degradation of fusion proteins by the proteasome (Alarid et al., 1999; Nawaz et al., 1999). Therefore, the different amounts of contaminating estrogen and estrogen mimics present in the culture media and serum make these cell lines an unreliable experimental system. To remove contaminating estrogen, dextran/charcoal-stripped serum and phenol red free media were used. However, this did not appear to be a viable strategy because the cells expanded poorly under these conditions. To generate a stable IRF4-ER pro-B cell line, the ER fused to IRF4 was mutated to ER_{T2}. Compared with the original ER, ER_{T2} contains four point-mutations,

namely G525R, G400V, M543A and L544A (Feil et al., 1997). ER_{T2} shows decreased sensitivity to β-estradiol and responds only to the estrogen antagonist tamoxifen or its active metabolite 4-hydroxytamoxifen (Feil et al., 1997). This successfully prevented the premature activation of IRF4 in 1D1-T215 cells as evidenced the minimal presence of IRF4-ER_{T2} in the nucleus extract of untreated samples (A. Smith, PhD thesis, 2018).

Addition of 4-hydroxytamoxifen to 1D1-T215 cells leads to the activation of Jλ1 non-coding transcription. Intriguingly, the process of Jλ1 non-coding transcription can be classified into two stages, the early and late stage. During the early stage, namely from 0 to 8 hpi, Jλ1 transcription is maintained at a low level, whereas from 8 to 12 hpi, at the late stage, Jλ1 is efficiently transcribed. To determine if the low level of Jλ1 transcription during the early stage is caused by a delay in nuclear transportation of IRF4-ER_{T2}, western blotting analysis of IRF4-ER_{T2} was performed using nuclear extracts of 1D1-T215 cells that were harvested at 1, 2, and 4 hours post induction. The results showed that the level of nuclear IRF4-ER_{T2} increases and reaches its highest level at 2 hours post induction (A. Smith, PhD thesis, 2018). IRF4 binding to Eλ3-1 was subsequently analysed in 1D1-T215 following induction. IRF4 appears to reach its highest levels at Eλ3-1 at 4 hours post induction. It is likely therefore, that only the first 2 hour delay in induction is caused by the nuclear translocation of IRF4-ER_{T2}. Furthermore, compared with primary pre-B cells, 1D1-T215 appears to have no gross alternations in chromosome number as verified by DNA content (A. Smith, PhD thesis, 2018). Taken together, these data suggest that the 1D1-T215 cell line is a highly suitable system that can be used to investigate enhancer-mediated activation of gene transcription.

Whilst the 1D1-215 cell line appears to be a viable model for investigating Eλ3-1 mediated Igλ activation, there are three caveats. Firstly, Jλ1 non-coding transcription is substantially repressed in 1D1-T215 cells compared to primary pre-B cells (Figure 3.9). This is consistent with the limited level of IRF4 (Figure 3.9) and Ser 5 phosphorylated RNAPII binding (data not shown) to the Jλ1 promoter in 1D1-T215 cells both before and after induction. According to

sequence analysis of the Jλ1 promoter, a binding site for the transcription repressor, STAT5, resides in the Jλ1 promoter. STAT5 can be activated by the tyrosine kinase v-Abl encoded by the A-MuLV, used to transform these cells, and results in reduction of chromatin accessibility of antigen receptor loci and inhibition of V(D)J recombination (Danial and Rothman, 2000; Mandal et al., 2011). Therefore, the repression of Jλ1 non-coding transcription may be caused by STAT5 binding and this could be confirmed by ChIP analysis of STAT5 binding to the Jλ1 promoter. These data therefore suggest that Jλ1 is not suitable for investigating Eλ3-1 mediated gene activation in 1D1-T215 cells. Using CRISPR-Cas9 technology, the repression of Jλ1 transcription might be alleviated by mutating the binding motif of STAT5 to abolish STAT5 binding to the Jλ1 promoter. Alternatively, Vλ1 is another essential gene segment under the control of Eλ3-1, and Vλ1 non-coding is transcribed at a similar level in 1D1-T215 cells to pre-B cells. This, combined with the significant enrichment of IRF4 and Ser 5 phosphorylated RNAPII at the Vλ1 promoter (Figure 3.10C, 3.18), meant that Vλ1 was selected instead as the Eλ3-1-activated target gene for further analysis.

It is also notable that V(D)J recombination of the Igλ locus is inhibited in the 1D1-T215 cells. V(D)J recombination requires binding of the RAG1 and RAG2 recombinase to accessible RSSs flanking target gene segments. The Jλ1 prompter contains a functional RSS but STAT5 binding to the promoter may repress the generation of open chromatin, and thereby cause the RSS to be inaccessible to RAG proteins. Moreover, RAG proteins themselves may be repressed by STAT5, which has been shown to prevent FOXO1 binding to the Erag enhancer (Amin and Schlissel, 2008; Biggs et al., 1999). Consistent with this, RAG1 expression is reduced by ~25-fold compared with wild-type pro-B cells (X. Wang, PhD thesis, 2018). It is highly likely that RAG2 expression is repressed by the same mechanism because both promoters share the Erag enhancer. Therefore, the inaccessible RSS and reduced levels of RAG expression limit the utility of this cell line somewhat as it prevents the temporal analysis of Igλ recombination. With modifications to 1D1-T215 cells, it is highly likely that the V(D)J recombination can be achieved. For example, mutating the

binding motifs of STAT5 located in the Jλ1 promoter using CRISPR-Cas9 technology as well as constitutively expressing exogenous RAGs using promoters that are insensitive to v-Abl signaling, such as EF1α, in lentiviral constructs. The final caveat of 1D1-T215 is that the retroviral integration sites of the IRF4-ER_{T2} transgene are unknown due to the random nature of virus insertion. This is could be improved by integrating IRF4-ERT2 transgene into the *Rosa26* locus using CRISPR-Cas9 mediated homology directed recombination (HDR). However, due to the low efficiency of this procedure and time constraints, it was not possible to conduct targeted insertion of the transgene. Whilst the location of integration sites is unknown, flow cytometry analysis of GFP expression shows that the IRF4-ER_{T2} transgene is stably expressed in 1D1-T215 cells for at least three months (data not shown), indicating the retroviral integration sites are not prone to silencing. Therefore, this strategy has developed a stable cell line that allows the Eλ3-1 enhancer and Vλ1 promoter interactions to be examined and has the potential to be modified for the investigation of V(D)J recombination.

3.11 Sequential order of recruitment of distinct transcription factors at enhancer-promoter loops

The inducible IRF4-ER_{T2} system that was characterised in this chapter allows temporal analysis to be performed of the non-coding transcription of unarranged gene segments of the Igλ locus as well as transcription factor binding within enhancer-promoter loops. The data acquired support a model whereby activation of Vλ1 transcription can be divided into two stages in 1D1-T215 cells, namely the early and late stages. In the early stage, Vλ1 non-coding transcription is maintained at only a low level, whereas efficient Vλ1 transcription occurs at the late stage.

During the early stage, generation of open chromatin at enhancers and promoters is essential for more transcription activators to be recruited. Accessible chromatin contains characteristic histone marks, such as H3K4me3 at active promoters and H3K27ac at active enhancers. E2A is an essential transcription factor for the progression of B cell development and knock-out of

E2A disrupts the non-coding transcription and histone modification landscape of the Igλ locus (Beck et al., 2009). As E2A was previously demonstrated to interact with IRF4 directly (Lazorchak et al., 2006), it is likely that IRF4 exhibits its function in an E2A dependent manner. Consistent with this, published ChIP-seq data from pro-B cell lines showed IRF4 (Schwickert et al., 2014) and E2A (Lin et al., 2010) are both enriched at the Eλ3-1 enhancer. This was further confirmed by ChIP-qPCR performed in primary pro-B, pre-B (J. Scott, PhD thesis, 2016) and 1D1-T215 cells (Figure 3.8B and 11). The acetyltransferase p300, which is responsible for the generation of H3K27ac at enhancers, has been demonstrated to be recruited to enhancers by E2A through direct interactions (Sakamoto et al., 2012). Re-analysis of published ChIP-seq data from a pre-B cell line, Haftl C10, showed that the majority of p300 binding is colocalised with E2A binding across the genome (van Oevelen et al., 2015). The increased chromatin accessibility across the Eλ3-1 enhancer could be directly confirmed by ChIP analysis of H3K27ac in 1D1-T215 cells, as well as by examining accessibility via DNaseI or restriction enzymes. Combined with the correlation in the temporal binding of IRF4, E2A and p300 at Eλ3-1, it is highly likely that IRF4 facilitates the generation of accessible chromatin at enhancers through recruiting E2A and p300 in the early stages of Vλ1 transcription activation.

In addition to this, the transcription activator, Mediator, seems to be recruited to enhancers and promoters at the early stage of Igλ activation (Figure 3.14C). The eukaryotic Mediator complex is comprised of approximately 30 subunits, which are classified into four modules, namely the head, middle, tail and kinase modules (Allen and Taatjes, 2015). The tail module is able to interact with different transcription cofactors at enhancer regions (Ansari and Morse, 2012). The head module is highly conserved compared with the other modules and has been shown to physically interact with RNA polymerase II and TFIIH at promoters (Esnault et al., 2008; Robinson et al., 2012). It is thus believed that the Mediator complex plays a role in bridging the enhancer to its cognate promoter. In my experiments, it was difficult initially to capture the interactions between the mediator and enhancer/promoter regions using conventional ChIP

methods, possibly because the subunits of the mediator bind DNA indirectly. Previous studies in the lab failed to detect binding of Med1 to the Eλ3-1 enhancer and Vλ1 promoter regions. To capture the interactions, two crosslinkers, formaldehyde and disuccinimidyl glutarate, were used. With these modifications, a significant enrichment of Med1 was observed at Eλ3-1 and Vλ1p in 1D1-T215 cells following induction. Temporal ChIP analysis and knock-down experiments further demonstrated the role of Mediator in bringing the enhancer and promoter into close proximity. The head module of Mediator was previously demonstrated to facilitate TFIIH binding to gene promoters to phosphorylate the Ser 5 of RNAPII CTD (Esnault et al., 2008). This is consistent with the coordinated binding of Med1 and Ser5 phosphorylated RNAPII to the Vλ1 promoter in the early stages of transcription activation. This could be further verified by ChIP-qPCR analysis of the binding of the catalytic subunit of TFIIH, CDK7, to the Vλ1 promoter. However, a ChIP-grade antibody for CDK7 is not available. Although this could be achieved by tagging the endogenous CDK7 with HA or Myc epitopes in 1D1-T215 cells using CRISPR-Cas9 mediated HDR, it was not possible to do this due to time constraints. Together, the data presented above indicate the early events in Vλ1 activation include the binding of E2A, p300, Mediator and Ser 5 phosphorylated RNAPII to the Eλ3-1 enhancer and Vλ1 promoter.

In the late stage of Vλ1 transcription activation, Vλ1 is efficiently transcribed which correlates with increased binding of Ser2 phosphorylated RNAPII to the Vλ1 promoter. Phosphorylation of the elongating form RNAPII, to generate Ser2 phosphorylated RNAPII, is catalysed by pTEFb; this facilitates RNAPII release from promoter-proximal regions (Adelman and Lis, 2012). This could be verified by ChIP-qPCR analysis of the binding of the catalytic subunit of pTEFb, CDK9. Furthermore, re-analysis of published ChIP-seq data of YY1 binding from pre-B cells showed that YY1 is present at the Eλ3-1 enhancer (Kleiman et al., 2016). YY1 belongs to the C2H2 zinc finger family of transcription factors and has been demonstrated to be essential for B cell development (Kleiman et al., 2016). Notably, one of its functions is the regulation of long-distance chromatin interactions at the IgH and Igκ loci (Atchison, 2014). Notably, from sequence

analysis, I found YY1 binding motifs are not present in the Eλ3-1 enhancer nor Vλ1 promoter and co-immunoprecipitation experiments showed that there are no direct interactions between IRF4 and YY1 (data not shown). However, temporal ChIP analysis reveals that YY1 binds to Eλ3-1 and Vλ1p and that this increases from 8 to 12 hpi (Figure 3.21), suggesting that it is a late event. Such binding may facilitate the stabilization of the Eλ3-1-Vλ1p loop to achieve the efficient transcription at the late stage of Vλ1 activation. Although YY1 is capable of binding to p300 directly (Lee et al., 1995), YY1 binding to Eλ3-1 does not correlate with p300 binding in 1D1-T215 cells following induction. This may be explained if YY1 binding to the enhancer cannot take place until the level of p300 binding meets a minimum requirement. This could be tested by over-expression of YY1 mutants that lack the p300 interacting domains in 1D1-T215 cells, followed by analysis of YY1 binding to the enhancer. The delay of YY1 binding may be also explained by YY1 binding to the enhancer-promoter loop being dependent on other transcription activators, such as enhancer RNAs (discussed in chapter 5).

Chapter 4: Dynamic activation of chromatin folding of the Igλ locus by IRF4

A) Introduction

The spatial topology of mammalian chromosomes within the nucleus has emerged as an essential player in fundamental processes such as transcription, replication, and DNA damage repair (Cavalli and Misteli, 2013). Functional enhancer-promoter communications are the determinant of tissue-specific gene transcription which are intimately associated with the way in which chromosomes are folded in three–dimensional (3D) space (Schoenfelder and Fraser, 2019). Mammalian chromatin is hierarchically folded at different levels, such as compartments, topologically associating domains (TADs) and insulated neighborhood domains (INDs), which have been postulated to represent structural and functional units of genome organization. Physical contacts between different *cis*-acting elements across the structural unit boundaries occur at relatively low levels. Efficient tissue-specific gene expression requires transcriptional enhancers to be constrained together with their cognate promoters within the same genome structural unit at the correct stage of differentiation. Therefore, to fully understand enhancer-mediated activation, it is important to unravel the mechanism by which chromatin folding facilitates enhancer-promoter interactions.

Antigen receptor loci contain a great number of gene segments and regulatory DNA elements that normally span mega-base sized chromatin regions. Establishment of the appropriate chromatin environment is a prerequisite not only for recruiting the recombination machinery (RAGs) to the correct gene segments but also for facilitating interactions between gene segments and their corresponding enhancers to activate non-coding transcription. The spatiotemporal organization of antigen receptor loci is poorly understood, primarily because of the absence of a temporal system to determine the changes of chromatin conformation and binding of transcription activators that occur. Chromatin folding of *IgH* and *Igκ* loci have been extensively investigated by DNA fluorescence *in situ* hybridization (FISH) and 3C derivative

technologies. These studies proposed that in the poised state prior to V(D)J recombination, antigen receptor loci are organized into several compartments in which multiple genomic DNA loops form rosette-like structures (Jhunjhunwala et al., 2008). Those rosette-like chromatin domains are then collapsed into a single globule as cells develop to the next stage of development (Jhunjhunwala et al., 2008). This contraction process is tightly associated with binding of architecture factors or transcriptional activators, such as CTCF, cohesin, YY1, PAX5, p300 and E2A, at the interspersed regulatory DNA elements within the locus; this also correlates with non-coding transcription of unrearranged gene segments. However, these studies cannot explore the antigen receptor locus activation and chromatin folding in fine detail. Indeed, whilst analysis of chromatin folding in B cells at different stages of development enables predictions regarding the temporal order of events, these studies cannot truly identify the temporal order of locus folding in any detail. A problem for the temporal analysis of antigen receptor locus folding is the absence of a homogenous population of lymphocytes in which antigen receptor locus activation can be induced. In Chapter 3, I described the generation of an IRF4-ER_{T2} expressing pro-B cell line, 1D1-T215. Using this system, I have demonstrated that locus folding between the Eλ3-1 enhancer and its target genes Vλ1 and Jλ1 corelates with the activation of non-coding transcription of these unrearranged gene segments after induction. This inducible Igλ cell line enables, for the first time, the analysis of the binding of transcription activators at those regulatory DNA elements and changes in chromatin structure during the activation of the locus.

In this chapter, I describe the characterization of additional *cis*-acting elements in the murine Igλ locus. I further examine the temporal binding of transcription activators, RNAPII machinery and architecture factors at these *cis*-acting elements, to decipher how the Eλ3-1 enhancer acts in concert with these regulatory DNA elements to activate the non-coding transcription of unrearranged gene segments. In addition, I determine which long-range interactions might be involved in bringing the enhancer elements and target gene segments into close spatial proximity, to build a more complete picture

how the chromatin organization and activation of non-coding transcription are coordinated.

B) Results

4.1 An IND sealing the 3’ half of the Igλ locus is already formed at the pro-B cell stage

The murine Igλ locus appears to have arisen from an evolutionary duplication event, giving rise to two recombination clusters. Each recombination cluster contains several gene segments and regulatory DNA elements with a similar organization (Figure 3.1). These two gene clusters seem to be relatively independent as V-J recombination primarily occurs between gene segments contained in the same cluster. This indicates that these two recombination clusters may reside in different chromosome environments. To investigate this, published Hi-C data from murine pro-B cells (Krijger et al., 2016) were re-analyzed. Hi-C is a powerful technique, developed in 2009, that determines the 3D architecture of the whole genome by combining proximity-based ligation and massively parallel sequencing (Lieberman-Aiden et al., 2009). As shown in Figure 4.1, chromatin interactions within the Igλ locus are separated into two INDs, sealing the 5’ half and 3’ half of the Igλ locus, respectively.

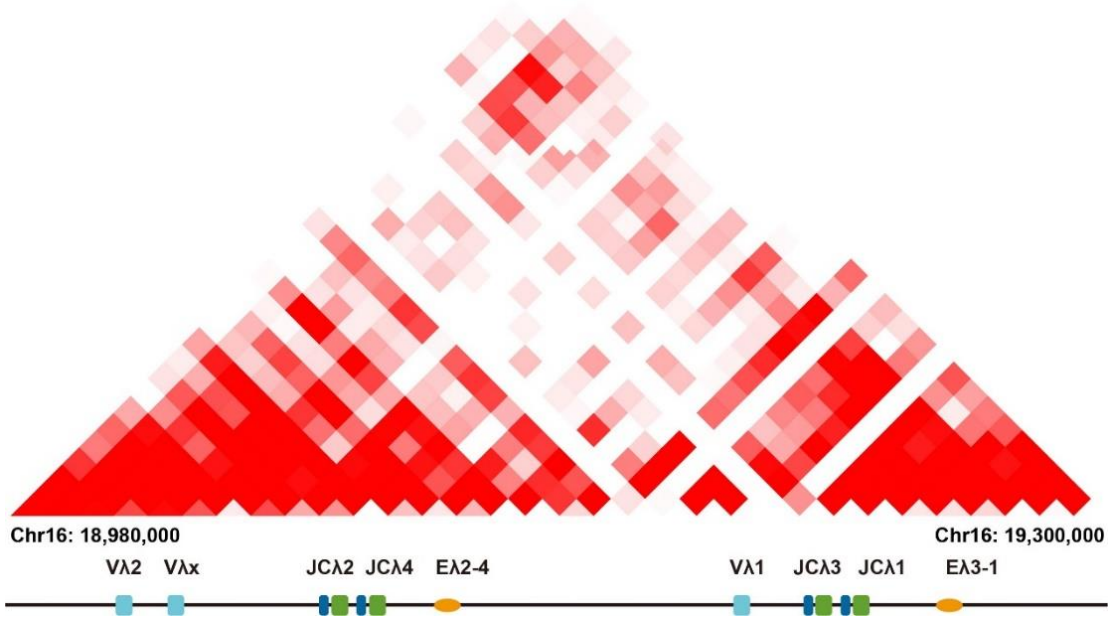

Figure 4.1 – Hi-C identifies enriched chromatin interactions within the 5' half and 3' half of the Igλ locus

Hi-C data from pro-B cells (Krijger et al., 2016) was analyzed using the Galaxy HiCExplorer web server (Wolff et al., 2020). Heatmaps showing chromatin interactions across the 260 kb of the Igλ locus are shown. Heatmap intensities indicate the interaction frequency detected in 10 kb windows. Importantly, clear interactions are constrained to the 5' and 3' halves of the Igλ locus.

CTCF and cohesin are essential architecture factors that are involved in shaping the genome into diverse chromatin domains, such as TADs and INDs. INDs are a subtype of chromatin domain and genome-wide analysis of human chromatin loops suggest that INDs vary from 25 kb to 940 kb in length and each IND contains three genes on average (Hnisz et al., 2016). To test how the Igλ locus is organized into two INDs by these architecture factors, published CTCF ChIP-seq data from *Rag2* deficient pro-B cells were processed and mapped to the Igλ locus. Analysis of CTCF binding indicated that in the 3' half of the locus, CTCF is enriched at the hypersensitive sites located approximately 24 kb downstream of Eλ3-1 referred to as HS7 and 3 kb upstream of Vλ1, referred to as HSVλ1 (Figure 4.2).

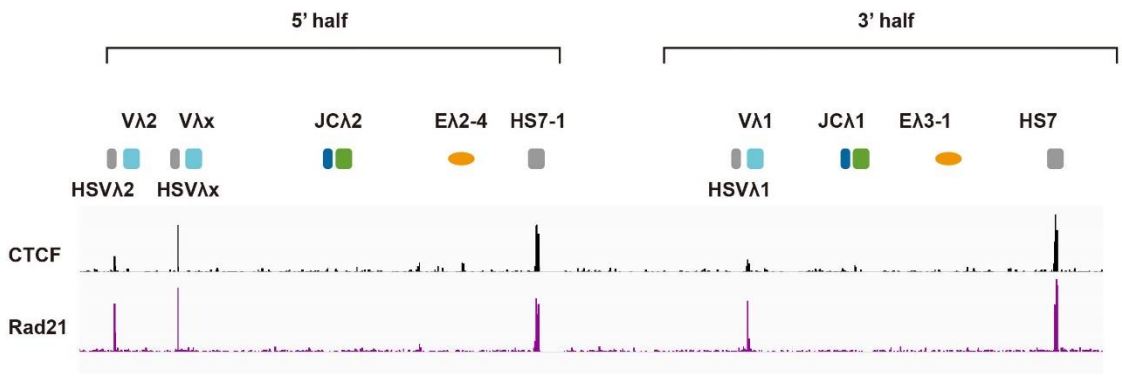

Figure 4.2 – CTCF and cohesin binding across the murine Igλ locus
CTCF and cohesin ChIP-seq data from pro-B cells (Ebert et al., 2011; McManus et al., 2011) were re-analyzed using the Galaxy web server. Rad21 is a subunit of the cohesin complex. Within the 5' half of the Igλ locus, regions downstream of Eλ2-4 (HS7-1), regions upstream of Vλ2 (HSVλ2) and Vλx (HSVλx) clearly exhibit CTCF and cohesin binding. Likewise, regions upstream of Vλ1 (HSVλ1) and downstream of Eλ3-1 (HS7) display CTCF and cohesin binding within the 3' half of the Igλ locus.

Likewise, four CTCF binding peaks were discovered in the 5' half of the Igλ locus with a small peak residing at the Eλ2-4 enhancer itself and larger peaks at hypersensitive sites downstream of the enhancer Eλ2-4 (HS7-1), upstream of Vλ2 (HSVλ2) and upstream of VλX (HSVλX, Figure 4.2). Because IND boundaries are normally co-bound by CTCF and the cohesin complex, published ChIP-seq data from pro-B cells for the cohesin component, Rad21, was reanalyzed to determine if the cohesin complex is present at any of these CTCF enriched regions within the Igλ locus. As shown in Figure 4.2, with the exception of Eλ2-4, all the previously mentioned CTCF enriched regions are bound by cohesin, indicating that the Igλ locus is organized into different INDs by CTCF/cohesin at the pro-B cell stage. Because the majority of recombinations in the Igλ locus occur between Jλ1 and Vλ1, which are located in the 3' half of the locus, I focussed on how the 3' half of the locus is organized by the CTCF/cohesin complex. To this end, ChIP-qPCR analysis of CTCF binding at the Igλ locus in 1D1-T215 cells was performed and this revealed that CTCF is indeed enriched at HS7 and HSVλ1 (Figure 4.3A).

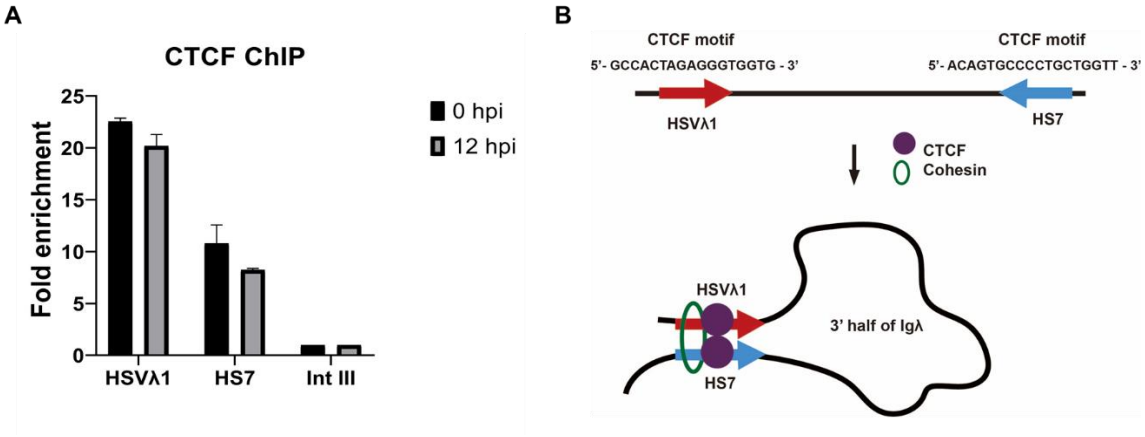

Figure 4.3 – CTCF binding to HS7 and HSVλ1 in 1D1-T215 cells

A) CTCF binding to HS7 and HSVλ1 were analysed by ChIP-qPCR in 1D1-T215 cells at 0 and 12 hpi. The fold enrichment at HS7, HSVλ1 and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

B) The binding sites at HS7 and HSVλ1 are in a convergent orientation implying physical interactions between these two genomic fragments can occur.

Notably, the majority of CTCF-mediated chromatin loops occur between CTCF binding sites that are in a convergent orientation (de Wit et al., 2015). Sequence analysis confirmed that the CTCF binding motifs discovered at HS7 and HSVλ1 are in a convergent orientation (Figure 4.3B). This implies that HS7 and HSVλ1 are likely to interact through a CTCF/cohesin loop. ChIP-qPCR analysis of cohesin binding further confirmed the enrichment of Rad21 at HS7 and HSVλ1, which adds support to the idea that these regions form chromatin contacts via a CTCF/cohesin loop (Figure 4.3B and 4.4).

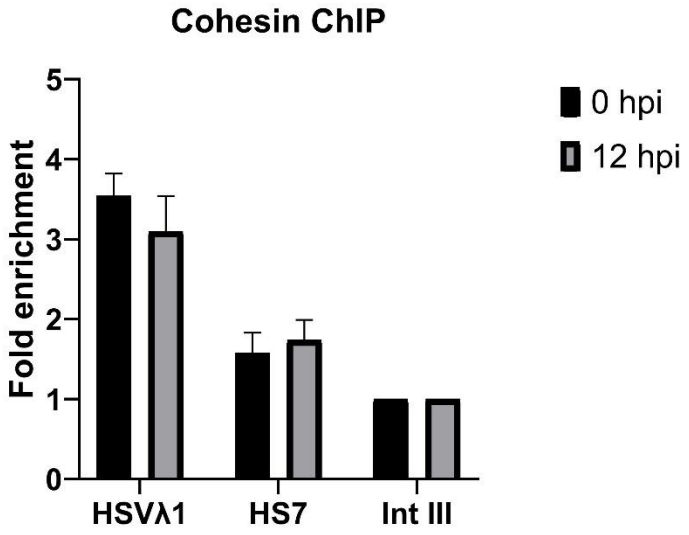

Figure 4.4 – Cohesin binding to HS7 and HSVλ1 in 1D1-T215 cells

A) Rad21 is a subunit of the cohesin complex. Rad21 binding to HS7 and HSVλ1 was analysed by ChIP-qPCR in 1D1-T215 cells at 0 and 12 hpi. The fold enrichment at HS7, HSVλ1 and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

Temporal analysis of CTCF and cohesin binding at the Igλ locus showed that there is no significant alteration in binding of these two architecture factors at HS7 and HSVλ1 in 1D1-T215 cells after induction (Figure 4.3A and 4.4). This is consistent with the CTCF/cohesin binding profiles in natural pro-B and pre-B cells (J. Scott, PhD thesis, 2016). Together, these data imply that CTCF/cohesin connects HS7 and HSVλ1 to form an 85 kb IND at the 3' half of the Igλ locus, that this is formed by the pro-B cell stage where Igλ is inactive, and that this IND is maintained during the progression to pre-B cells where Igλ is activated.

4.2 Identification of additional regulatory elements in the 3’ half of the Igλ locus

As mentioned above, CTCF/cohesin connects HS7 and HSVλ1 to form an IND, thus leading to contraction of the 3’ half of the Igλ locus and shortening the distance between the Eλ3-1 enhancer and its target genes. However, antigen receptor loci are normally in a more contracted state where they are activated (Jhunjhunwala et al., 2008). To identify additional regulatory elements which may be involved in the chromatin organization of the Igλ locus, chromatin accessibility across the 3’ half of the Igλ locus was examined as active regulatory elements are characterized by high levels of chromatin accessibility (Klemm et al., 2019). A powerful technology to probe DNA accessibility genome-wide is Assay for Transposase Accessible Chromatin using sequencing (ATAC-seq) (Buenrostro et al., 2015). Therefore, published ATAC-seq data from pro-B cells were reprocessed and mapped to the 3’ half of the Igλ locus. As shown in Figure 4.5, three ATAC signal peaks were found at Eλ3-1, HS7 and HSVλ1 and interestingly, two additional peaks were discovered approximately 27 kb upstream of Eλ3-1 and 5 kb upstream of HS7, which are referred to as HSCλ1 and HS6, respectively. Whilst examining open chromatin can identify functional DNA elements, I sought to also determine if any of these accessible DNA elements identified by ATAC-seq also display enhancer characteristics. Histone H3 lysine 4 mono-methylation (H3K4me1) is a hallmark of all transcriptional enhancers (Creyghton et al., 2010). Active enhancers are further characterized by acetylation of histone H3 lysine 27 (H3K27ac) which is catalyzed by p300 acetyltransferase (Creyghton et al., 2010). I therefore analyzed p300 ChIP-seq data from a pro-B like cell line, haftl derived C10 (van Oevelen et al., 2015), and H3K4me1 and H3K27ac ChIP-seq data from primary pro-B cells (Choukrallah et al., 2015), to locate enhancer-like elements within the 3’ half of the Igλ locus. These data reveal that, similar to the active Eλ3-1 enhancer, HSCλ1 and HS6 are both occupied by high levels of H3K27ac and p300 binding (Figure 4.5).

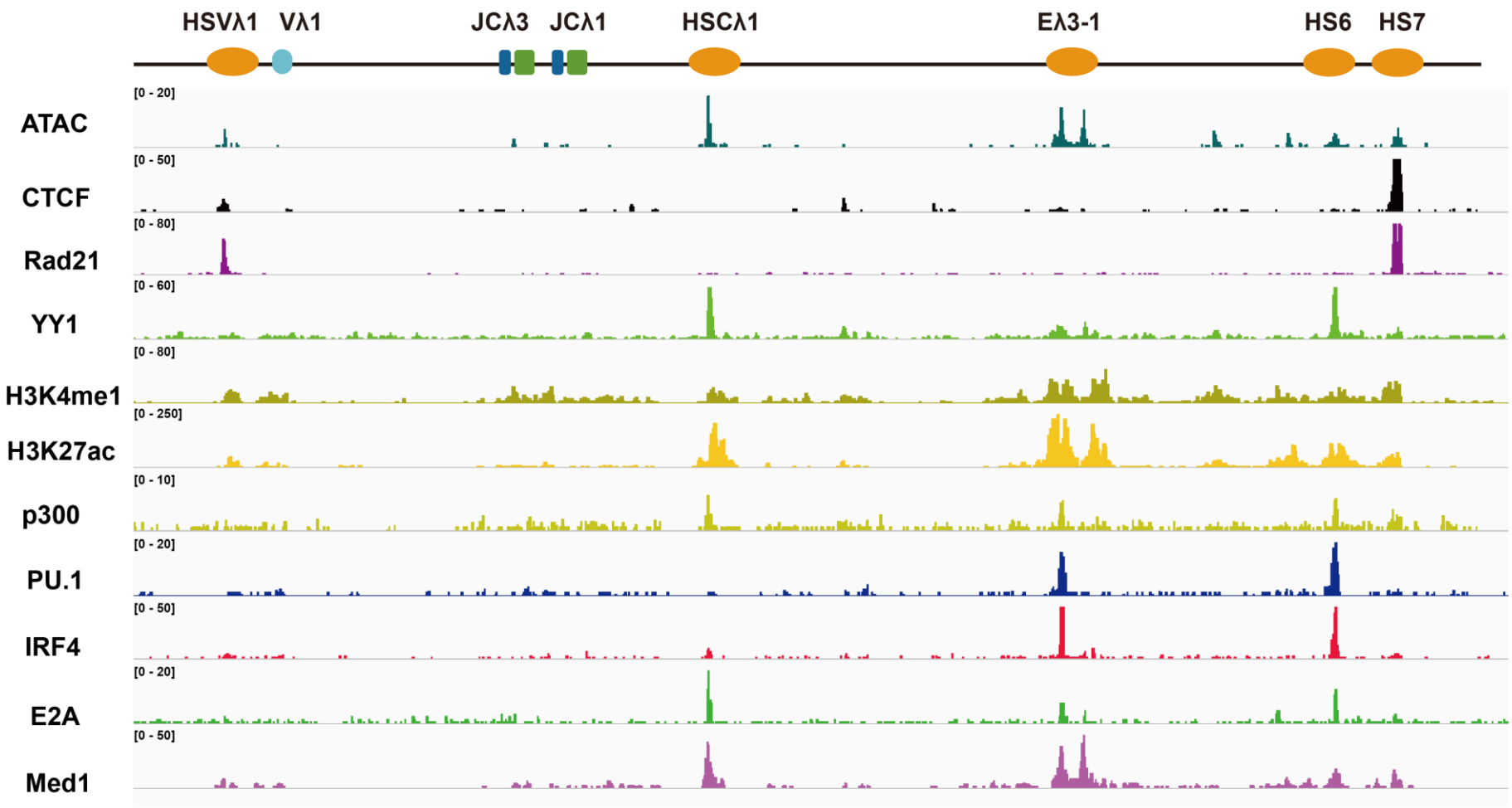

Figure 4.5 – ATAC-seq and ChIP-seq data at the 3' half of Igλ locus

ATAC -seq and ChIP-seq of architecture factors (CTCF, cohesin and YY1), enhancer hallmarks (H3K4me1, H3K27ac and p300) and transcription activators (PU.1, IRF4, E2A and Med1), mapped to the 3' half of the Igλ locus. The data sources are summarized in Table 2.2.

As HSCλ1 and HS6 display the characteristics of active enhancers, I next examined if HSCλ1 and HS6 show a similar landscape of transcription factor binding as Eλ3-1. Firstly, HSCλ1, HS6 and Eλ3-1 were subjected to sequence analysis using an integrated web tool named LASAGNA-search. This online software scans an input sequence for putative transcription factor binding sites based on built-in transcription factor binding models (Lee and Huang, 2013).

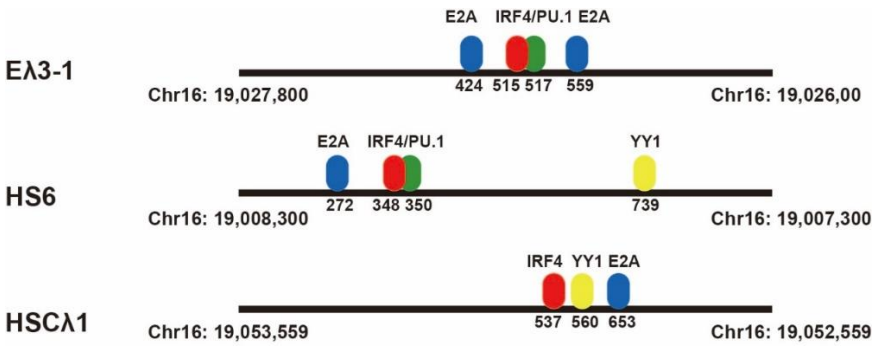

Figure 4.6 – Transcription factor motifs analysis at Eλ3-1, HS6 and HSCλ1
Transcription factor binding site analysis of IRF4 (red), PU.1 (green), E2A (blue) and YY1 (yellow) at Eλ3-1, HS6 and HSCλ1 using the LASAGNA-search tool. A 1 kb window centred on each enhancer was subject to analysis. Coordinates of identified motifs are shown relative to each enhancer and the genomic coordinates of the regions analysed are shown.

As can be seen in Figure 4.6, all three putative enhancers, Eλ3-1, HS6 and HSCλ1, contain binding motifs for similar transcription factors, including PU.1, IRF4 and E2A, with the exception of HSCλ1 which lacks a PU.1 binding motif. To unravel the transcription factor binding profile of the 3' half of the Igλ locus, published IRF4, PU.1 and E2A ChIP-seq data (Lin et al., 2010; Schwickert et al., 2014; van Oevelen et al., 2015) from pro-B cells were reanalyzed. Consistent with the predicted transcription factor binding sites, IRF4, PU.1 and E2A are significantly enriched at all enhancers except HSCλ1 (Figure 4.5). To further confirm this, ChIP-qPCR analysis of IRF4, PU.1 and E2A binding was performed in induced 1D1-T215 cells. Consistent with ChIP-seq data, E2A is present at HSCλ1, HS6 and Eλ3-1 at similar levels of enrichment in 1D1-T215 cells at 12hpi when Igλ is activated (Figure 4.7A). Similar ChIP-qPCR data

further confirmed PU.1 binding to HS6 and Eλ3-1 in induced 1D1-T215 cells (A. Smith, PhD thesis, 2018). Whilst no IRF4 signal peaks were observed at HSCλ1 in pro-B ChIP-seq data, IRF4 binding is significantly and reproducibly enriched at HSCλ1 in 1D1-T215 cells after induction (Figure 4.7B). Compared with the low levels of IRF4 binding observed at HSCλ1, IRF4 is highly enriched at Eλ3-1 and HS6 (Figure 4.7B) possibly due to its affinity being increased by pre-bound PU.1 at these two sites (Eisenbeis et al., 1995; Escalante et al., 2002). Together, these data therefore imply that the newly identified enhancer-like elements HS6 and HSCλ1 may be essential for activation of the Igλ locus.

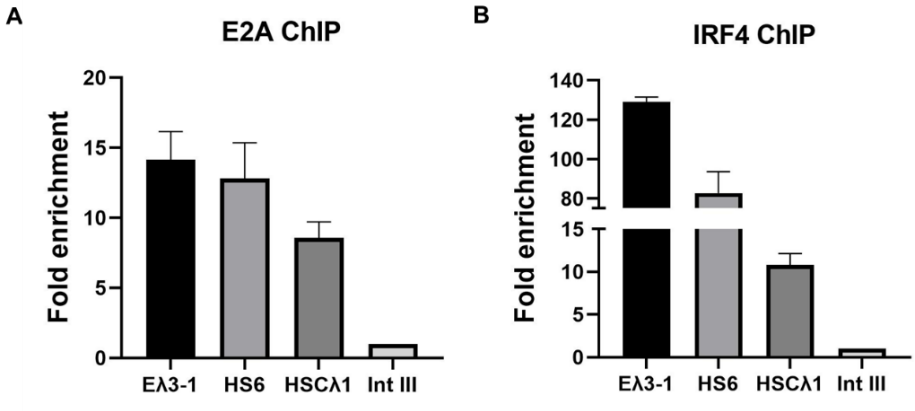

Figure 4.7 – E2A and IRF4 binding at Eλ3-1, HS6 and HSCλ1 in induced 1D1-T215 cells

A) E2A binding to Eλ3-1, HS6 and HSVλ1 was analysed by ChIP-qPCR in 1D1-T215 cells at 12 hpi. The fold enrichment at Eλ3-1, HS6, HSVλ1 and Intgene III (negative control region) is shown.

B) IRF4 binding to Eλ3-1, HS6 and HSVλ1 was analysed by ChIP-qPCR in 1D1-T215 cells at 12 hpi. The fold enrichment at Eλ3-1, HS6, HSVλ1 and Intgene III (negative control region) is shown.

All values are normalized to binding at Intgene III as a negative control. Error bars show the standard error of the mean (SEM) from three experimental repeats.

To test if these identified regulatory DNA elements can enhance gene transcription, I cloned HS6 and the promoter of Vλ1 into the pGL3-basic luciferase construct. Determination of the expression of *Firefly* and *Renilla* luciferase expression was performed as described in Chapter 2. The data shown in Figure 4.8 indicate that HS6 is indeed an enhancer of Vλ1 transcription.

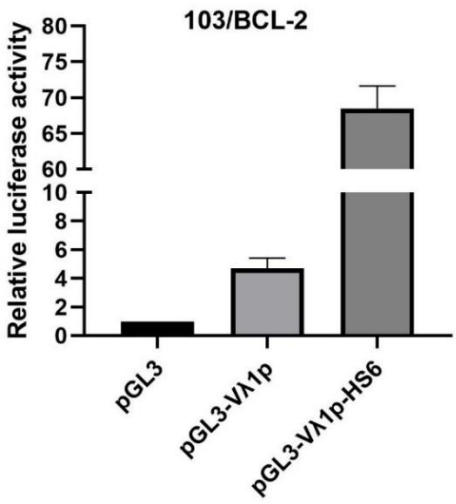

Figure 4.8 – HS6 is an enhancer of Vλ1

Luciferase activity driven by HS6 in temperature shifted 103/BCL-2 cells. The Vλ1 promoter increases luciferase activities by ~5-fold compared with the empty vector; HS6 gives a further ~15-fold increase over the Vλ1 promoter.

Similar preliminary experiments to examine if HSCλ1 has enhancer activity were, however, negative (data not shown). This may be explained by the fact that compared to HS6 and Eλ3-1, IRF4 binds to HSCλ1 at only low levels (Figure 4.7) as a binding site for its co-factor, PU.1, is absent (Figure 4.6).

4.3 Temporal analysis of chromatin interactions within the 3' half of the Igλ locus

Analysis of published Hi-C data from pro-B cells and pre-B cells suggested that a number of interactions occur during Igλ locus activation (A. Smith, PhD thesis, 2018). Whilst an increased level of chromatin interactions was observed within the 3' half of the Igλ locus in pre-B cells, the true temporal order of these interaction events cannot be deciphered from Hi-C data. The inducible 1D1-T215 cell line described in Chapter 3, however, does allow the temporal order of chromatin interactions within the Igλ locus to be traced. Considering that efficient Vλ1 and Jλ1 non-coding transcription was achieved in 1D1-T215 cells from 8 hpi to 12 hpi (Figure 3.10), I conducted 3C experiments at four, eight and twelve hours post induction using the Eλ3-1 enhancer as a viewpoint.

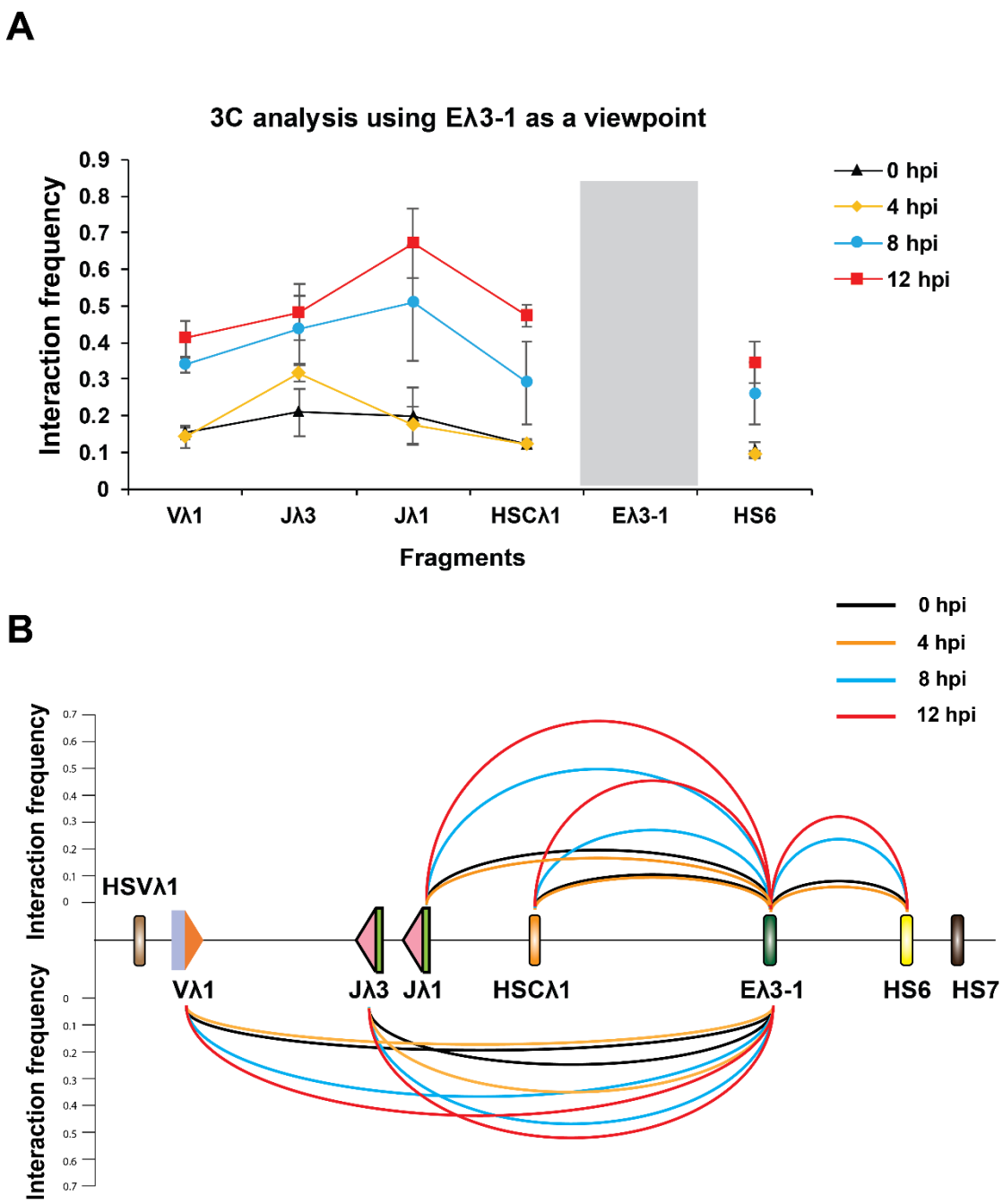

Figure 4.9 – Temporal 3C analysis of chromatin interactions formed in the 3' half of Igλ

A) Analysis of the relative interaction frequency of Dpn II fragments from the Eλ3-1 viewpoint in 1D1-T215 cells at 0, 4, 8 and 12 hpi. Error bars show the standard error of the mean (SEM) from three experimental repeats.

B) Schematic diagram of chromatin interactions within the 3' half of the Igλ locus using Eλ3-1 as the viewpoint. The height of curves between Eλ3-1 and other genomic fragments represents the average value of interaction frequency obtained from three

experimental repeats. Both plots are shown as the latter shows locus-wide interactions more clearly.

As shown in Figure 4.9, the Eλ3-1 enhancer exhibits only minimal contacts with the unrearranged gene segments (Vλ1, Jλ1 and Jλ3) as well as other enhancer-like elements (HS6 and HSCλ1) before Igλ activation. Following induction, no dramatic changes in chromatin contracts were observed within the 3' half of Igλ using Eλ3-1 as a viewpoint before 4 hpi. Remarkably, a substantial increase in chromatin interaction frequency between Eλ3-1 and the gene segments within the 3' half of the Igλ locus was observed between 4 hpi to 8 hpi, which is just prior to the substantial increase in transcription of unrearranged gene segments. These data are consistent with the idea that the establishment of chromatin environment is a prerequisite for the activation of target gene transcription. Notably, the chromatin interaction frequency between Eλ3-1 and HS6 as well as HSCλ1 appears to correlate well with changes in chromatin interactions between Eλ3-1 and the Igλ gene segments, suggesting that these two enhancer-like elements may be involved in the activation of the Igλ locus. All of these chromatin interactions increase further at the later 12 hpi time point.

4.4 HS6 and HSCλ1 are indispensable for the activation of the Igλ locus

The data presented above indicate that the physical distance between Eλ3-1 and the newly identified enhancer-like element HS6 as well as HSCλ1 is decreased in 1D1-T215 cells following induction, indicating that HS6 and HSCλ1 may generate an enhancer-hub, together with Eλ3-1. To examine if HS6 and HSCλ1 are important for maintenance of the chromatin structure organized by the putative Eλ3-1 enhancer hub, genetic mutations were separately introduced to HS6 and HSCλ1 in 1D1-T215 cells using CRISPR-Cas9 technology. To disrupt the function of these enhancer-like elements, sgRNAs were designed to knock-out (KO) binding sites for key transcription factors, such as IRF4, PU.1, E2A and YY1, within HS6 and HSCλ1. Multiple monoclonal cell lines were obtained for HS6 and HSCλ1 KO, using semi-solid agar to obtain individual clones, as described in Chapter 2. However, only one cell line of each KO was subjected to further analysis. Mutations introduced at HS6 and HSCλ1 were confirmed by Sanger sequencing. The sequencing results confirmed

deletions of binding sites for key transcription factors in HS6 (A. Smith, PhD thesis, 2018) and HSCλ1 mutant cell lines (Figure 4.10A).

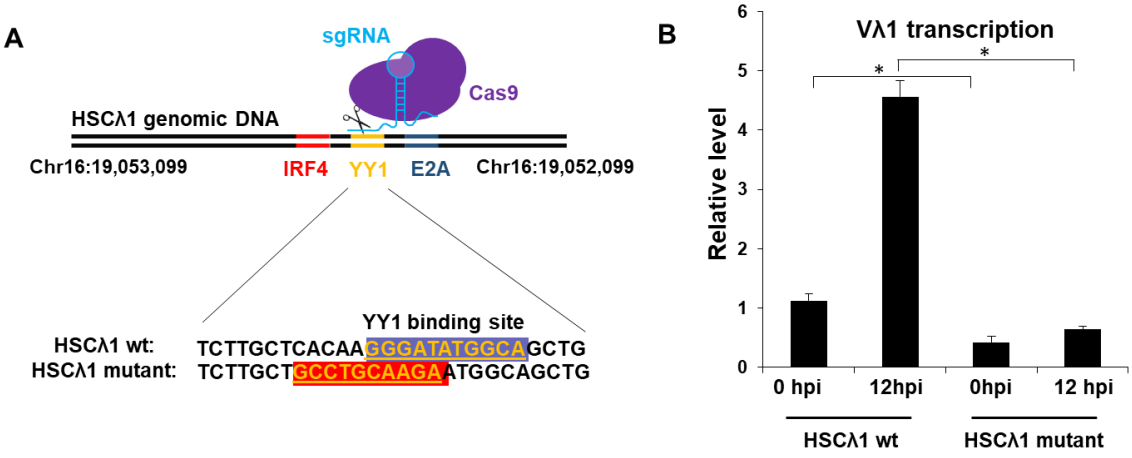

Figure 4.10 – Mutations in HSCλ1 introduced by CRISPR-Cas9 diminish the non-coding transcription of Vλ1

A) Mutations were introduced into HSCλ1 using the CRISPR-Cas9 technology. A sgRNA was designed to target a region that contains a binding motif for YY1. Monoclonal cell lines were obtained, and the mutations were confirmed by Sanger sequencing.

B) RT-qPCR analysis of Vλ1 non-coding transcription in the HSCλ1 mutant cell line. Vλ1 non-coding transcription is significantly reduced in 1D1-T215 cells which contain mutations at HSCλ1. Data were normalized to the expression level of the housekeeping gene, *Hprt*.

Error bars show standard error of the mean (SEM) from three experimental repeats.

* represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

Next, examination of the level of non-coding transcription of unrearranged gene segments showed that Vλ1 transcription is completely disrupted in both cell lines (A. Smith, PhD thesis, 2018; Figure 4.10B), indicating that both HS6 and HSCλ1 are essential for the activation of non-coding transcription. To test if the establishment of the Eλ3-1 enhancer hub and Igλ locus folding is affected by HS6 and HSCλ1 mutations, 3C experiments were performed to determine the changes in chromatin interactions between Eλ3-1 and HS6, as well as with HSCλ1, in the HS6 and HSCλ1 mutant cell lines. A substantial decrease in these chromatin interactions was observed in the HS6 mutant cell line (A. Smith,

PhD thesis, 2018). Likewise, reduced levels of chromatin interactions between Eλ3-1 and the other two enhancer-like elements were also observed in the HSCλ1 mutant cell line (Figure 4.11). These data therefore imply that the chromatin structure organized by the Eλ3-1 enhancer hub is essential for the activation of the non-coding transcription of unrearranged gene segments of the Igλ locus, as disruption of the enhancer hub by deleting of critical elements within either HS6 or HSCλ1 inhibits the Igλ locus activation triggered by IRF4.

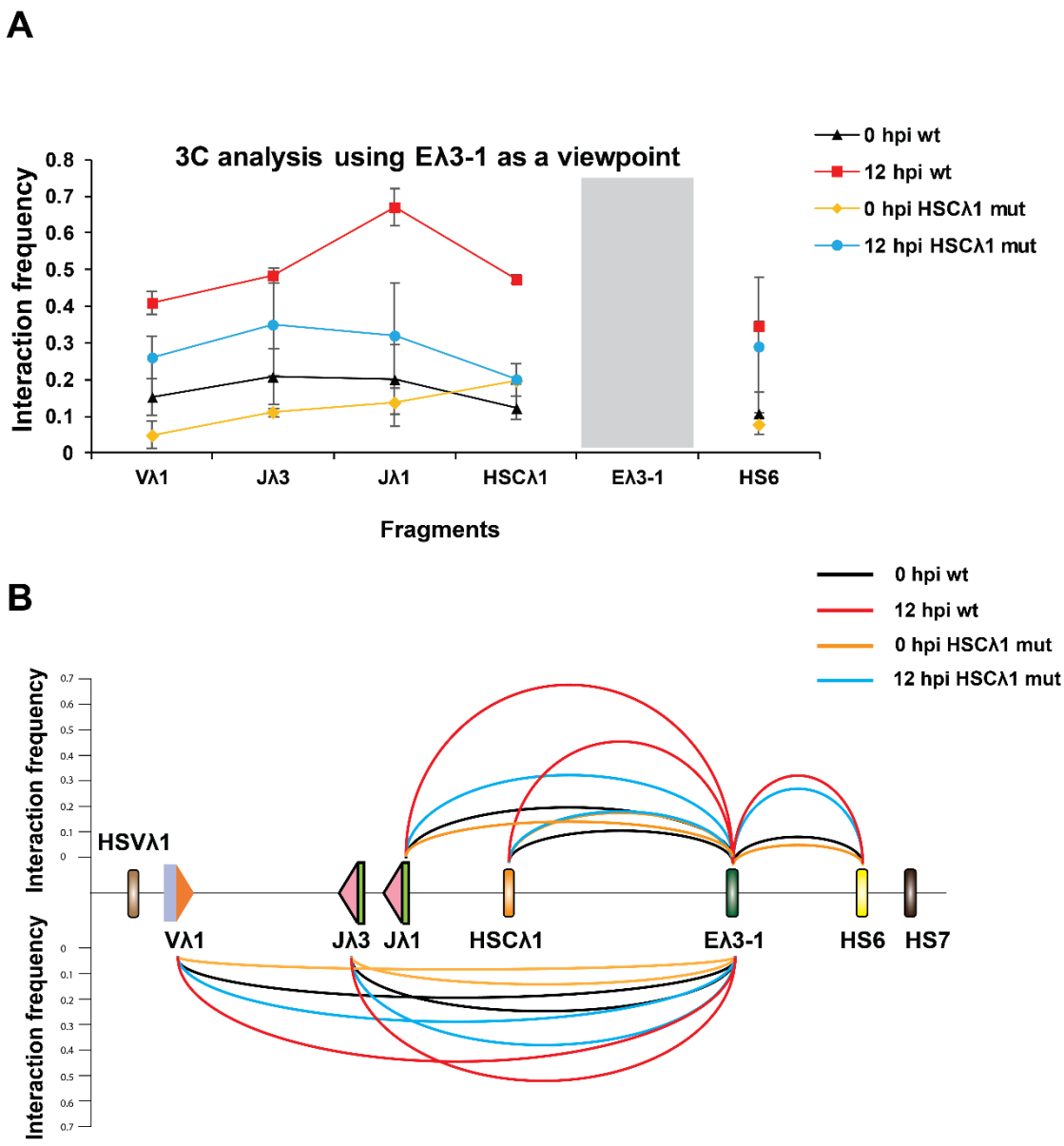

Figure 4.11 – 3C analysis of chromatin interactions formed in the 3' half of Igλ in the HSCλ1 mutant cell line

A) Analysis of the relative interaction frequency of Dpn II fragments from the Eλ3-1 viewpoint in the HSCλ1 mutant 1D1-T215 cell line. Error bars show standard error of the mean (SEM) from three experimental repeats.

B) Schematic diagram of chromatin interactions within the 3' half of the Igλ locus using Eλ3-1 as the viewpoint. The height of curves between Eλ3-1 and other genomic fragments represents the average value of interaction frequency obtained from three experimental repeats.

4.5 Eλ3-1 shares the same transcription activators with HS6 and HSCλ1

IRF4 was previously demonstrated to be the trigger for activation of the Igλ locus in pro-B cells (Bevington and Boyes, 2013). Combined published ChIP-seq data and ChIP-qPCR data (Figures 4.5 and 4.7), identified two additional IRF4 binding sites at HS6 and HSCλ1. Moreover, removal of IRF4 binding sites at Eλ3-1 as well as HS6 leads to a substantial decrease in non-coding transcription of the unrearranged gene segments the Igλ locus (A. Smith, PhD thesis, 2018). These data therefore strongly imply that elevated levels of IRF4 binding are essential for the activation of all three enhancers. Furthermore, temporal ChIP analysis of IRF4 binding in 1D1-T215 cells showed that IRF4 binding to Eλ3-1 is an early event in Igλ locus activation that has already reached its highest level by 4 hpi (Figure 3.8). To test if the HS6 and HSCλ1 elements are triggered by a similar mechanism, temporal analysis of IRF4 binding was performed in 1D1-T215 cells after induction. This shows that IRF4 binding to HS6 is also an early event that also reaches its maximal level at 4 hpi, possibly because it utilises a similar mechanism of recruitment to that at Eλ3-1, namely, via pre-bound PU.1 (Figure 4.12). Although IRF4 binding to HSCλ1 occurs at only low levels, possibly due to the absence of PU.1, it nonetheless displayed a temporal pattern of recruitment similar to that at Eλ3-1 and HS6 (Figure 4.12). Overall, this temporal analysis of IRF4 binding showed a simultaneous increase in IRF4 occupancy at Eλ3-1, HS6 and HSCλ1.

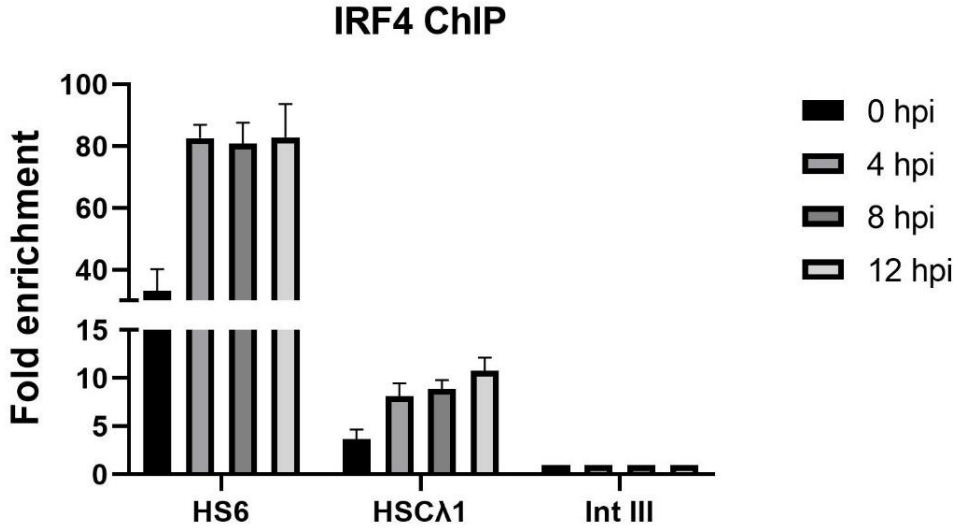

Figure 4.12 – IRF4 is recruited to HS6 and HSCλ1 in 1D1-T215 cells
IRF4 binding to HS6 and HSCλ1 was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at HS6, HSCλ1 and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

As demonstrated in the previous chapter, increased IRF4 binding to Eλ3-1 appears to allow the recruitment of E2A and p300, which could result in the increased chromatin accessibility. To examine if HS6 and HSCλ1 also facilitate increased transcription factor binding following IRF4 induction, temporal ChIP analysis of E2A and p300 binding to HS6 and HSCλ1 was carried out in induced 1D1-T215 cells. As can be seen in Figures 4.13A and B, E2A and p300 binding to HS6 and HSCλ1 is substantially increased at 8 hpi and further increases at the subsequent 12 hpi time point; this is a similar pattern of recruitment as seen at Eλ3-1. These data together suggest that similar to Eλ3-1, IRF4 interacts directly with HS6 and HSCλ1 and increased IRF4 binding results in recruitment of E2A and p300 at these two enhancers. Given that p300 is a histone acetyltransferase (Vo and Goodman, 2001) this potentially generates an open chromatin structure.

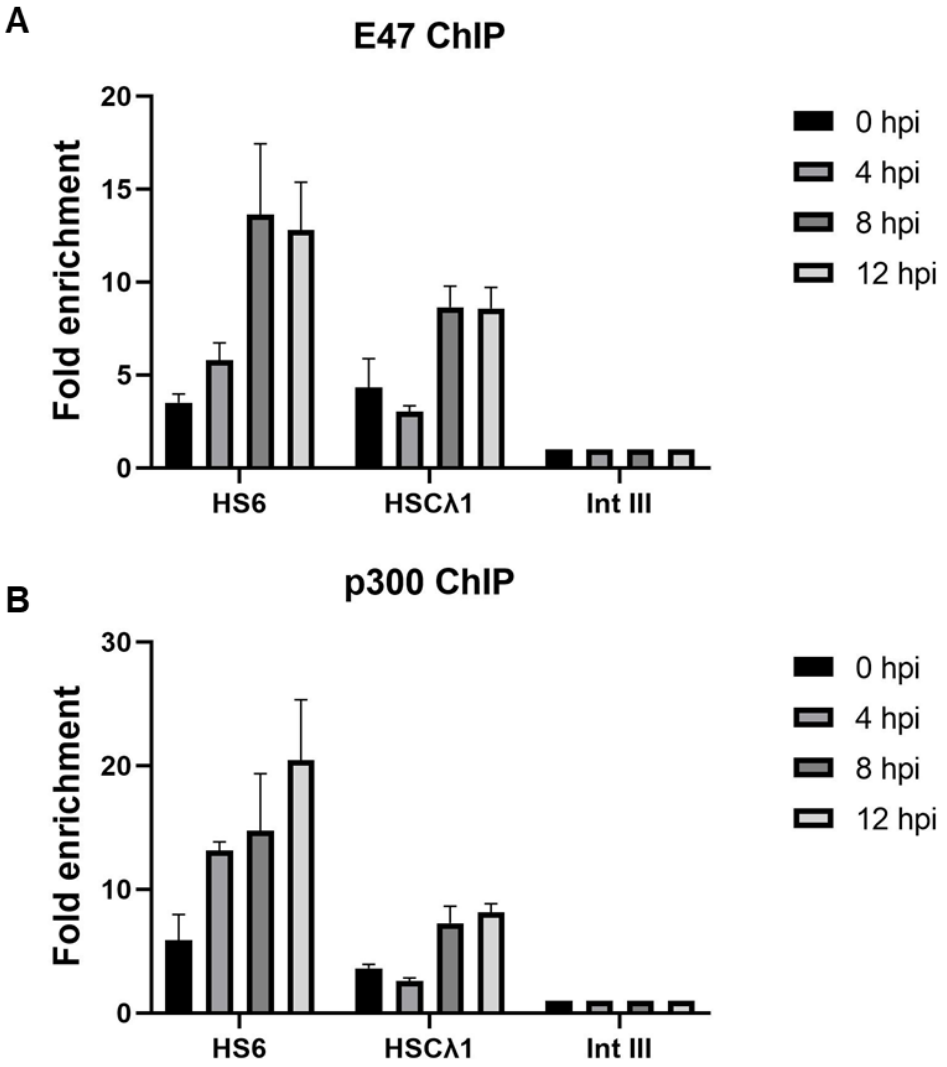

Figure 4.13 – E2A and p300 is recruited to HS6 and HSCλ1 in 1D1-T215 cells

A) E2A binding to HS6 and HSCλ1 was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at HS6, HSCλ1 and Intgene III (negative control region) is shown.

B) p300 binding to HS6 and HSCλ1 was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at HS6, HSCλ1 and Intgene III (negative control region) is shown.

All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

4.6 Mediator is essential for chromatin folding of the Igλ locus

The Mediator complex was previously shown to be involved in bridging enhancers to their cognate promoters (Malik and Roeder, 2016) and increasing evidence suggests that the Mediator complex plays a role in the regulation of long-range chromatin interactions (Chereji et al., 2017; Thomas-Claudepierre et al., 2016). Thus, it is not surprising that knock-down of Med23 in 1D1-T215 cells disrupts the links between the Eλ3-1 enhancer and its target genes (Figure 3.13). Establishment of the correct chromatin environment is a prerequisite for tissue-specific enhancer-promoter communications. Temporal 3C analysis of the Igλ locus architecture indicates that the formation of an enhancer hub by Eλ3-1, HS6 and HSCλ1 precedes the establishment of efficient transcription of target genes (Figures 3.10 and 4.9). From another perspective, the chromatin folding that is achieved by the formation of the enhancer-hub results in a shortening of the distance between Eλ3-1 and its target genes. To determine if Mediator is essential for chromatin interactions between Eλ3-1, HS6 and HSCλ1, Mediator binding was assessed by analyzing published ChIP-seq data of a core Mediator component, Med1, from pro-B cells (Whyte et al., 2013). As shown in Figure 4.5, Med1 is already present at HS6 and HSCλ1 prior to activation, as expected. Indeed, because both these elements contain IRF4 binding sites, Mediator could be loaded onto HS6 and HSCλ1 through direct interactions with IRF4. ChIP-qPCR analysis further confirmed that Med1 binding to HS6 and HSCλ1 shows a gradual increase after induction in 1D1-T215 cells, which mirrors its binding to Eλ3-1 (Figure 3.14 and 4.14).

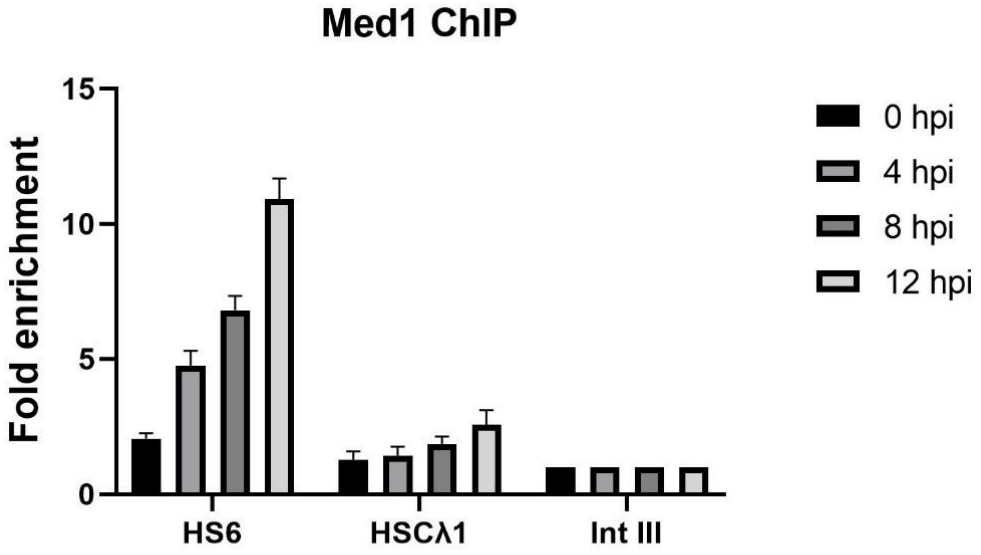

Figure 4.14 – Med1 is recruited to HS6 and HSCλ1 in 1D1-T215 cells
Med1 binding to HS6 and HSCλ1 was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at HS6, HSCλ1 and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

To determine if Mediator is responsible for the establishment of chromatin interactions between Eλ3-1 and the other enhancers, temporal 3C analysis was performed in Med23 KD 1D1-T215 cells. This shows a substantial decrease in chromatin interactions between Eλ3-1, HS6 and HSCλ1 in 1D1-T215 cells expressing an shRNA targeting Med23 (Figure 4.15). Likewise, the interaction frequency between Eλ3-1 and other gene segments, including Jλ1, Vλ1 and Jλ3, is reduced in Med23 KD cells (Figure 4.15). These data therefore indicate that Mediator is essential for the regulation of chromatin interactions during the activation of Igλ gene transcription.

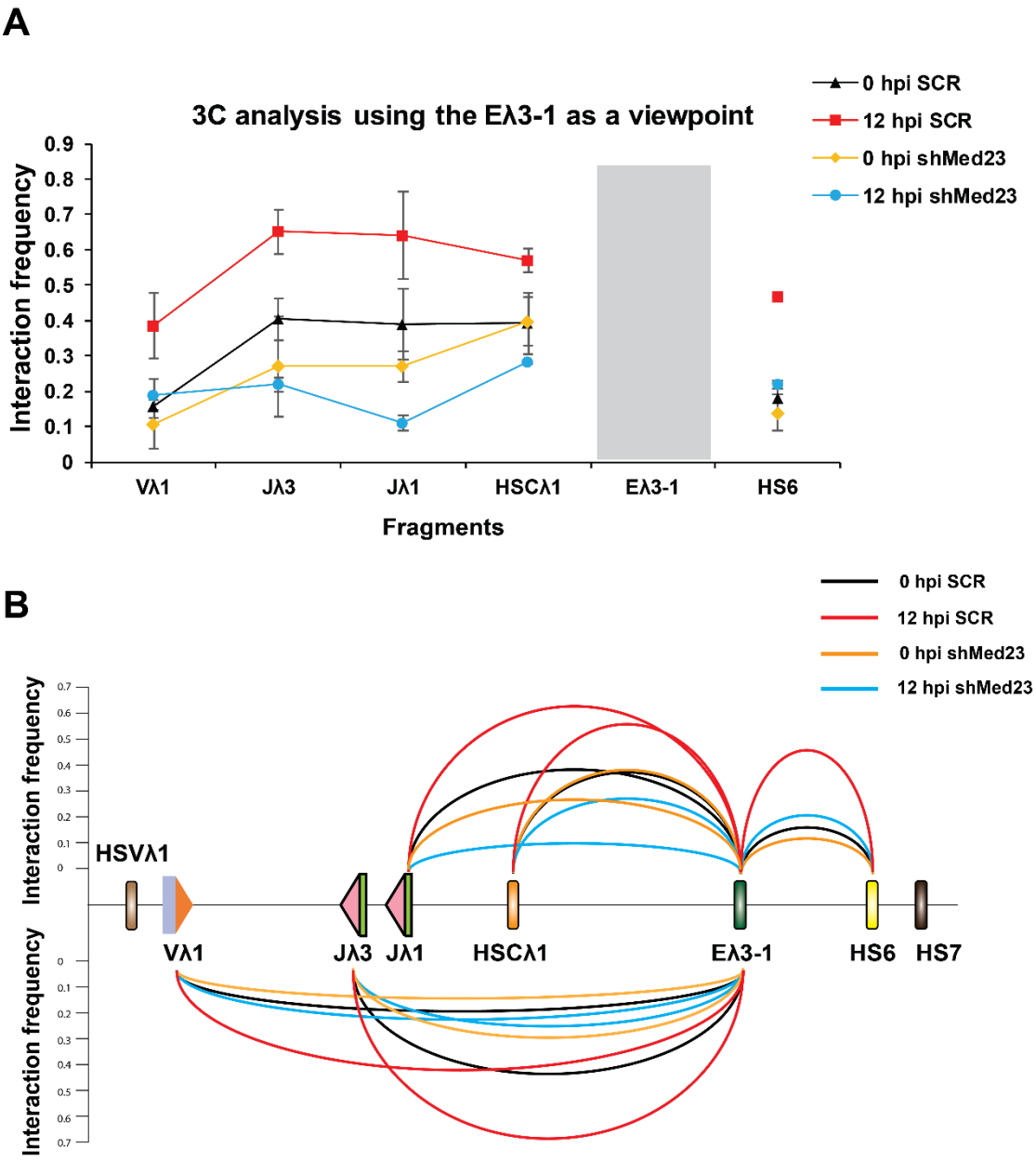

Figure 4.15 – 3C analysis of chromatin interactions formed in the 3' half of Igλ in Med23 knock down cells

A) Analysis of the relative interaction frequency of Dpn II fragments from the Eλ3-1 viewpoint in 1D1-T215 cells expressing an shRNA targeting Med23. Error bars show standard error of the mean (SEM) from three experimental repeats.

B) Schematic diagram of chromatin interactions within the 3' half of the Igλ locus using Eλ3-1 as the viewpoint. The height of curves between Eλ3-1 and other genomic fragments represents the average value of interaction frequency obtained from three experimental repeats.

4.7 YY1 binding to HS6 and HSCλ1 is essential for the stabilization of the Igλ locus chromatin folding

It has become well established that YY1 is a structural regulator of the majority of enhancer-promoter interactions (Weintraub et al., 2017). Consistent with this, YY1 has been shown to be involved in the regulation of chromatin folding of other antigen receptor loci (Liu et al., 2007). The data presented in Chapter 3 reveal that YY1 is essential for Eλ3-1-Vλ1 chromatin interactions. Re-analysis of published YY1 ChIP-seq data showed that whilst a limited level of YY1 is present at Eλ3-1, YY1 is greatly enriched at HS6 and HSCλ1 (Figure 4.5) (Kleiman et al., 2016). To determine how YY1 regulates the chromatin folding of the 3' half of the Igλ locus, temporal analysis of YY1 binding to HS6 and HSCλ1 was firstly performed in 1D1-T215 cells. As can be seen in Figure 4.16, there is a dramatic increase of YY1 binding to HS6 and HSCλ1 from 8 hpi to 12 hpi, which mirrors YY1 binding to Eλ3-1 following induction (Figure 3.21). As formation of the Eλ3-1 enhancer hub is nearly completed at 8 hpi (Figure 4.9), this may imply that YY1 is not required for the initial stage of formation of the Eλ3-1 hub.

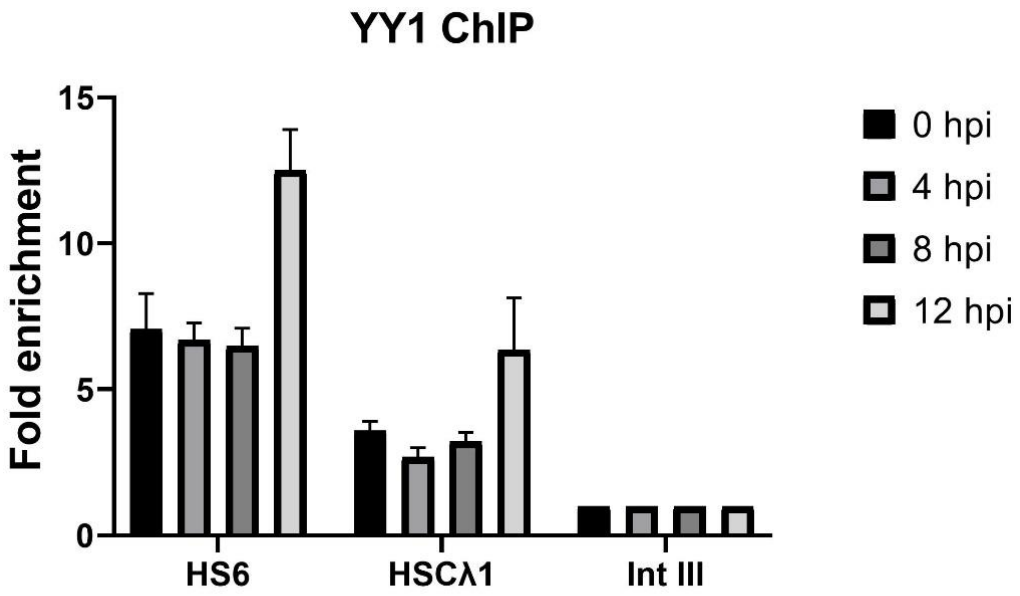

Figure 4.16 – YY1 is recruited to HS6 and HSCλ1 in 1D1-T215 cells
YY1 binding to HS6 and HSCλ1 was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at HS6, HSCλ1 and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

Next, depletion of YY1 expression was conducted in 1D1-T215 cells using the pLKO shRNA system, followed by determination of the chromatin interactions between Eλ3-1 and the gene segments. This showed that the interaction frequency between Eλ3-1 and the gene segments is disrupted completely in 1D1-T215 cells expressing an shRNA targeting YY1 (Figures 3.20 and 4.17), which correlates with the reduced levels of non-coding transcription of Igλ (Figure 3.20). Likewise, reduced levels of chromatin interactions between Eλ3-1 and HS6 as well as HSCλ1 are also observed in YY1 KD cells (Figure 4.17).

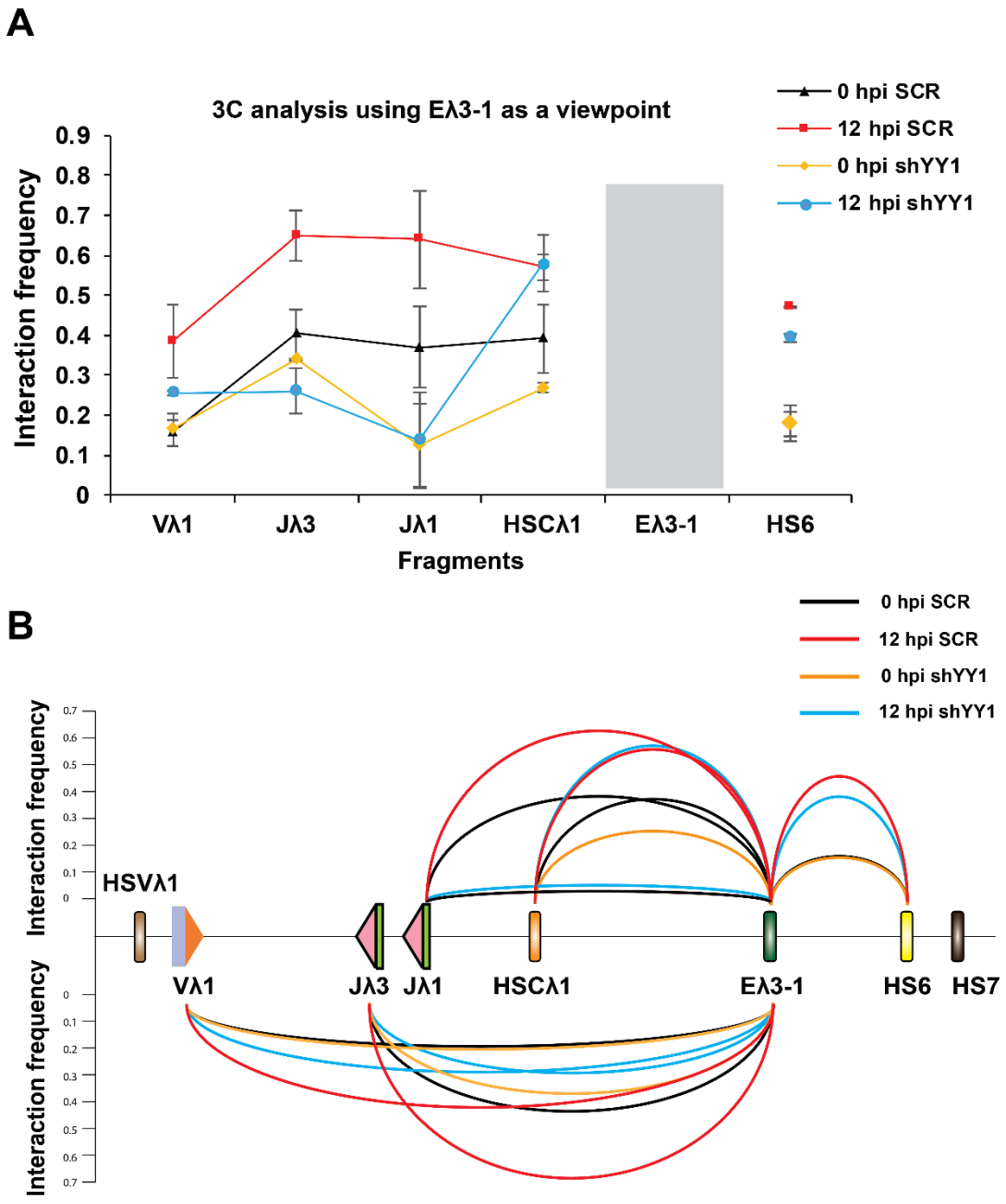

Figure 4.17 – 3C analysis of chromatin interactions formed in the 3' half of the Igλ locus in YY1 knock down cells

A) Analysis of the relative interaction frequency of Dpn II fragments from the Eλ3-1 viewpoint in 1D1-T215 cells expressing an shRNA targeting YY1. Error bars show standard error of the mean (SEM) from three experimental repeats.

B) Schematic diagram of chromatin interactions within the 3' half of the Igλ locus using Eλ3-1 as the viewpoint. The height of curves between Eλ3-1 and other genomic fragments represents the average value of interaction frequency obtained from three experimental repeats.

I previously introduced genetic mutations in the YY1 binding site located in HSCλ1 in 1D1-T215 cells using the CRISPR-Cas9 technology (Figure 4.10). Determination of the YY1 binding to the mutated HSCλ1 was subsequently performed and the results showed that although the YY1 binding motif was only partially mutated (Figure 4.10A), a significant decrease of YY1 binding was observed at HSCλ1 (Figure 4.18).

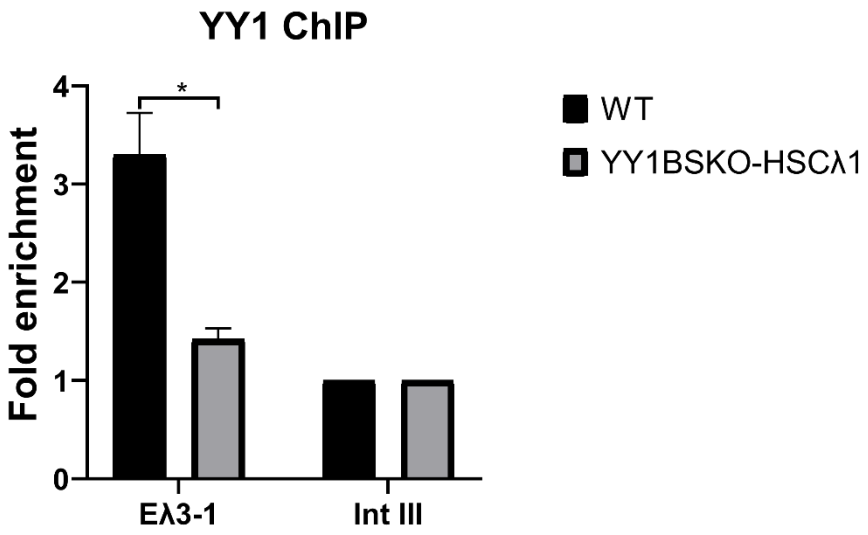

Figure 4.18 – Mutations in YY1 binding sites leads to a reduced level of YY1 enrichment at HSCλ1

YY1 binding to HSCλ1 in the HSCλ1 mutant cell line was analysed by ChIP-qPCR. The fold enrichment at HSCλ1 and Intgene III (negative control region) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats. * represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

To examine if the knock-out of the YY1 binding site within HSCλ1 influences the chromatin topology of the Eλ3-1 enhancer hub, 3C analysis was performed. This showed that chromatin interactions between Eλ3-1 and gene segments are all affected by mutation of the YY1 binding site (Figure 4.11). Consistent with this, non-coding transcription of Igλ is disrupted in the HSCλ1 mutant cell line (Figure 4.10B). Whilst the chromatin interactions between Eλ3-1 and HSCλ1 are disrupted by the mutations in the YY1 binding site within HSCλ1,

the level of chromatin interactions between Eλ3-1 and HS6 is maintained (Figure 4.11). These data therefore indicate that YY1 is essential for the chromatin structure of the Eλ3-1 enhancer hub. Whilst YY1 might not be required for the initial stage of establishment of the Eλ3-1 hub, it is likely needed to maintain the chromatin loops which are already formed.

C) Discussion

This chapter aimed to examine how chromatin structure is reorganized to facilitate the activation of target gene transcription using the mouse Igλ locus as a model. By reanalysing published ATAC-seq and ChIP-seq datasets from pro-B and pre-B cells, I identified new potential regulatory DNA elements. Temporal 3C analysis was then performed using the inducible 1D1-T215 cell line to determine the changes in long range interactions among these *cis*-acting elements. I find that an IND sealing the 3' half of the murine Igλ is already formed in the pro-B stage of B cell development where the Igλ locus is inactive. CTCF/cohesin mediates formation of the IND, resulting in chromatin folding and shortening of the distance between the Eλ3-1 enhancer and its target genes. Locus contraction of Igλ is further facilitated by two newly identified enhancer elements, HS6 and HSCλ1. Similar to Eλ3-1, both of HS6 and HSCλ1 contain binding motifs for IRF4 and are activated by increased levels of IRF4 binding. IRF4 appears to recruit E2A and p300 to HS6 and HSCλ1 to generate open chromatin. IRF4 also likely recruits the Mediator complex to HS6 and HSCλ1 to facilitate the establishment of the Eλ3-1 enhancer hub, which results in further locus contraction and shortening of the distance between the enhancers and target genes. In addition, the architecture factor YY1 is recruited to the constituent enhancers of the Eλ3-1 enhancer hub, HS6 and HSCλ1, to further cement the locus contraction.

4.8 The IND that seals the Igλ locus is conserved

Long-range interactions mediated by CTCF/cohesin are essential for chromatin contraction and rearrangement of the IgH, Igκ and T cell receptor loci by bringing regulatory DNA elements and recombining gene segments into close spatial proximity (Shih and Krangel, 2013). Previous studies showed that most of the CTCF-mediated chromatin loops occur between CTCF binding sites in a convergent orientation (Rao et al., 2014). Deletion or inversion of one of a pair of CTCF binding sites ablates the chromatin loop (de Wit et al., 2015). Using combined sequence analysis and ChIP-seq analysis, I found that the IND sealing the 3' half of the Igλ locus is formed by the CTCF binding sites within HS7 and HSVλ1 that lie in a convergent orientation. Moreover, a high level of

CTCF binding to HS7 and HSVλ1 was confirmed by ChIP-qPCR in primary pro-B (J. Scott, PhD thesis, 2016) and uninduced 1D1-T215 cells where the Igλ locus is poised (Figure 4.3A). Intriguingly, the binding levels do not change in pre-B cells (J. Scott, PhD thesis, 2016) nor in induced 1D1-T215 cells where Igλ is activated. Importantly, the cohesin component, Rad21, was confirmed to be present at HS7 and HSVλ1 in 1D1-T215 cells and its binding level also does change after induction. 3C data further demonstrate that the interactions between HS7 and HSVλ1 do not change from pro-B to pre-B cells (J. Scott, PhD thesis, 2016), suggesting that the 3' half of the Igλ locus is already sealed in a chromatin loop in pro-B cells, formed by CTCF/cohesin binding at HS7 and HSVλ1.

Recent evidence suggests that CTCF/cohesin mediated chromatin boundaries are conserved across diverse cell types (Essien et al., 2009). To examine if the IND that seals the 3' half of the Igλ locus is present in different types of cells. I re-analyzed CTCF ChIP-seq data from different mouse cells including embryonic stem cells (ESCs) and differentiated tissues such as liver, kidney, lung, spleen and heart (Shen et al., 2012).

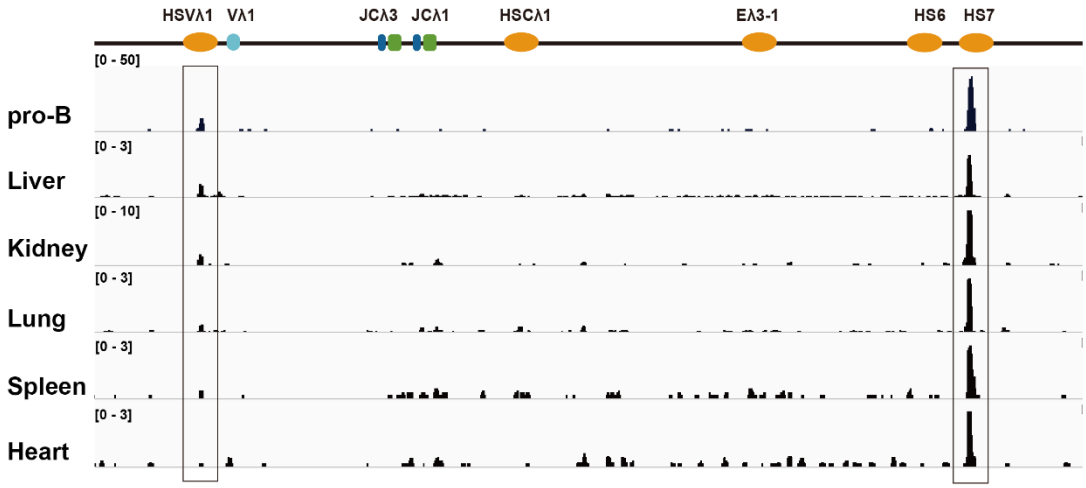

Figure 4.19 – CTCF binding to HS7 and HSVλ1 in different tissues

CTCF ChIP-seq data from different tissues (Shen et al., 2012) was re-analyzed and mapped to the 3' half of the murine Igλ locus.

Intriguingly, CTCF binding to HS7 and HSVλ1 was observed in all selected tissues (Figure 4.20), implying that the HS7-HSVλ1 IND may be formed early in development. To verify this, I analyzed the Hi-C data from different tissues using the 3D Genome database (http://promoter.bx.psu.edu/hi-c/index.html). As shown in Figure 4.19, the Igλ locus is clearly divided into two chromatin domains in different cell types including ESCs, myoblasts (the C2C12 cell line) and neural progenitor cells (NPCs). The chromatin boundaries correspond to the peaks of CTCF/cohesin enrichment at the Igλ locus. These data therefore imply that the IND that constrains the 3' half of the Igλ locus may be formed in the initial stage of development and is subsequently conserved during differentiation to various tissues.

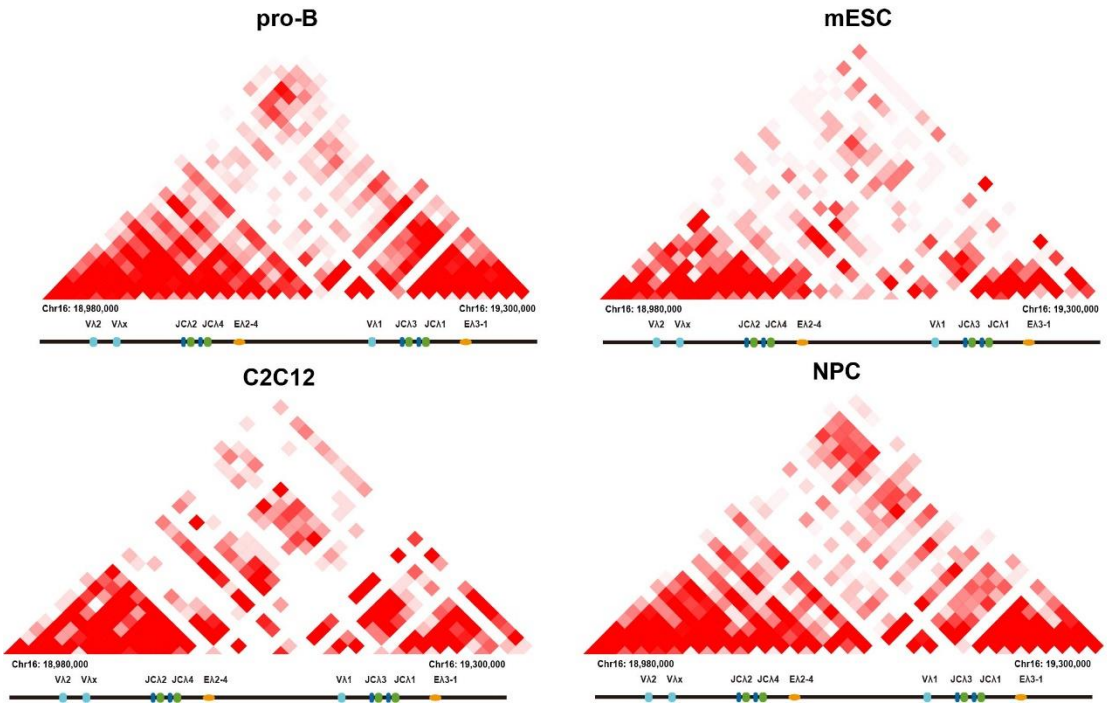

Figure 4.20 – Hi-C analysis of chromatin interactions within the Igλ locus in different cell types

Hi-C data from mouse embryonic stem cells (mESCs), myoblasts (C2C12) and neural progenitor cells (NPCs) were obtained from the 3D Genome database (Wang et al., 2018). Heatmaps of chromatin interactions across the 260 kb Igλ locus is shown. Heatmap intensities indicate the interaction frequency detected in 10 kb windows.

4.9 Eλ3-1, HS6 and HSCλ1 form a super-enhancer

By analysing ATAC-seq from pro-B cells, I discovered two additional regulatory DNA elements, named HS6 and HSCλ1, within the 3' half of the Igλ locus. Analysis of available pro-B and pre-B ChIP-seq data reveal that these two *cis*-acting elements display classical characteristics of transcriptional enhancers, such as high levels of H3K27ac, p300 binding, H3K4me1 and contain binding motifs for transcription activators, in this case, IRF4 and E2A. The binding of these factors was confirmed by ChIP-qPCR in 1D1-T215 cells. KO of binding sites for key transcription factors within HS6 and HSCλ1 confirm that these two regulatory elements are essential for Igλ activation. In addition, re-analysis of published Hi-C data demonstrated that Eλ3-1 and HS6 as well as HSCλ1 are brought into close spatial proximity in pre-B cells where Igλ is active. This was confirmed by temporal 3C analysis in 1D1-T215 cells after induction. These data therefore indicate that Eλ3-1 could interact with HS6 and HSCλ1 to form an enhancer hub during activation of the Igλ locus. This type of enhancer hub could be potentially categorized as a super-enhancer. These are characterized by a large cluster of typical enhancers that are occupied by high levels of p300, H3K27ac, H3K4me1, master transcription factors and Mediator (Hnisz et al., 2013; Pott and Lieb, 2015; Whyte et al., 2013). Genome-wide identification of super-enhancers was firstly performed by the Young lab in mouse ESCs (Whyte et al., 2013). Using a similar identification procedure, mouse B cell-specific super-enhancers were identified in pro-B cells (Qian et al., 2014). Interestingly, the genomic region covering HS6, Eλ3-1 and HSCλ1 was identified as a B cell-specific super-enhancer. Super-enhancers rely on cooperativity between transcriptional regulators bound to constituent enhancers for their function (Hnisz et al., 2017). Consistent with this, KD of Mediator and YY1 resulted in a pronounced decrease of chromatin interactions within the Eλ3-1 hub, accompanied by reduced levels of transcription of target genes. Moreover, genetic deletion of constituent enhancers within super-enhancers can disrupt the activities of other constituents within the super-enhancer (Jiang et al., 2016; Shin et al., 2016), resulting in the collapse of an entire super-enhancer (Mansour et al., 2014). Consistent with this, the establishment of the Eλ3-1 enhancer hub is disrupted in HS6 and HSCλ1

mutant 1D1-T215 cell lines. These data therefore suggest that the formation and function of the Eλ3-1 enhancer hub involves cooperative processes that bring HS6 and HSCλ1 and their bound transcription regulators into close spatial proximity.

4.10 YY1 binding to HS6 and HSCλ1 is essential to maintain the chromatin structure of the enhancer hub

YY1 is not only an architectural factor that is involved in the regulation of long-range DNA interactions (Atchison, 2014) but also an essential regulator of gene transcription (Sarvagalla et al., 2019). Sequence analysis showed that YY1 binding motifs are present only in HS6 and HSCλ1 within the Eλ3-1 enhancer hub. However, YY1 binding to Eλ3-1 was confirmed by ChIP-qPCR. This is consistent with the fact that knock-down of YY1 can disrupt chromatin interactions between Eλ3-1 and the other two enhancers. Thus, YY1 binding to Eλ3-1 appears indirect and may be mediated by interactions with other transcription factors, such as p300 (Galvin and Shi, 1997) and Mediator (Luck et al., 2020). Whilst knock-down of YY1 results in a substantial decrease of interaction frequency between Eλ3-1 and unrearranged gene segments, only a limited decrease of chromatin interactions was observed between Eλ3-1 and the other two enhancers that contain strong YY1 binding sites. In addition, knock-out of YY1 binding sites within HSCλ1 leads to a substantial decrease of chromatin interactions between Eλ3-1 and HSCλ1. By contrast, the interaction frequency between Eλ3-1 and HS6 is maintained at a high level. These data imply that YY1 is capable of binding HS6 and HSCλ1 via the strong YY1 binding motifs and this may allow YY1 binding to HS6 and HSCλ1 when its levels are only low. This could be tested by examining the level of YY1 binding to HS6 and HSCλ1 in YY1 KD cells.

As mentioned above, YY1 can directly interact with transcription regulators including p300 and Mediator (Lee et al., 1995; Luck et al., 2020). Recruitment of those transcription regulators by IRF4 appears to be an early event during Igλ activation. By contrast, YY1 binding to enhancers and promoters is a late event. There are two possible explanations: Firstly, that YY1 recruitment by

p300 and Mediator occurs in a dose-dependent manner. In the initial stage of Igλ activation, the low levels of these YY1-interacting factors bound at the enhancers are insufficient to recruit YY1 but as their levels increase, so does YY1 binding. Secondly, YY1 recruitment may be mediated by other factors, such as enhancer RNAs (Sigova et al., 2015). In chapter 5, I show evidence that eRNAs produced by Eλ3-1 are tightly associated with YY1 recruitment to Igλ.

Temporal ChIP analysis demonstrated that YY1 enrichment at the Igλ locus is a late event. This may indicate that YY1 is not required in the initial stage of Igλ activation. Consistent with this, no significant changes in chromatin interactions between Eλ3-1 and Vλ1 were observed in uninduced 1D1-T215 cells expressing an shRNA targeting YY1 (Figure 3.20C). However, reduced Vλ1 non-coding transcription was observed in uninduced YY1 KD cells, implying that YY1 is important for maintaining the basal level of non-coding transcription of Igλ. Previous studies demonstrated that YY1 can regulate gene transcription by competing and preventing the binding of transcription repressors to gene promoters (Makhlouf et al., 2014). In addition, YY1 is also capable of recruiting chromatin remodelers to gene promoters to facilitate transcription (Cai et al., 2007). Therefore, the decrease in non-coding transcription of unrearranged gene segments of Igλ may be caused by the altered transcription factor binding and chromatin accessibility at promoters due to reduced YY1 expression.

4.11 A proposed model of chromatin folding of Igλ

In conclusion, the data generated in this chapter builds a model of how the 3' half of the Igλ locus is organized to achieve the efficient transcription of target gene segments (Figure 4.21). Initially, in uninduced 1D1-T215 cells where the Igλ locus is inactive, an IND is formed by CTCF/cohesin to bring HS7 and HSVλ1 together, sealing the 3' half of the Igλ locus. Formation of such a chromatin loop is a prerequisite for the establishment of correct chromatin environment for further activation of the Igλ locus. Next, upon induction of 1D1-T215 cells with 4-hydroxytamoxifen, the level of IRF4 increases in the nucleus and this allows a higher level of IRF4 binding to all three enhancers, Eλ3-1,

HS6 and HSCλ1. Increased levels of IRF4 binding to these enhancers results in recruitment of E2A and p300, facilitating the generation of open chromatin. Simultaneously, the Mediator complex is recruited to enhancers and promoters via direct interactions with IRF4, resulting in the locus contraction. After 8 hpi, YY1 binding to enhancers and promoters was observed and this correlates with the synthesis of enhancer RNAs (discussed in Chapter 5), facilitating the stabilization of folded chromatin and leading to the efficient transcription of gene segments within the Igλ locus.

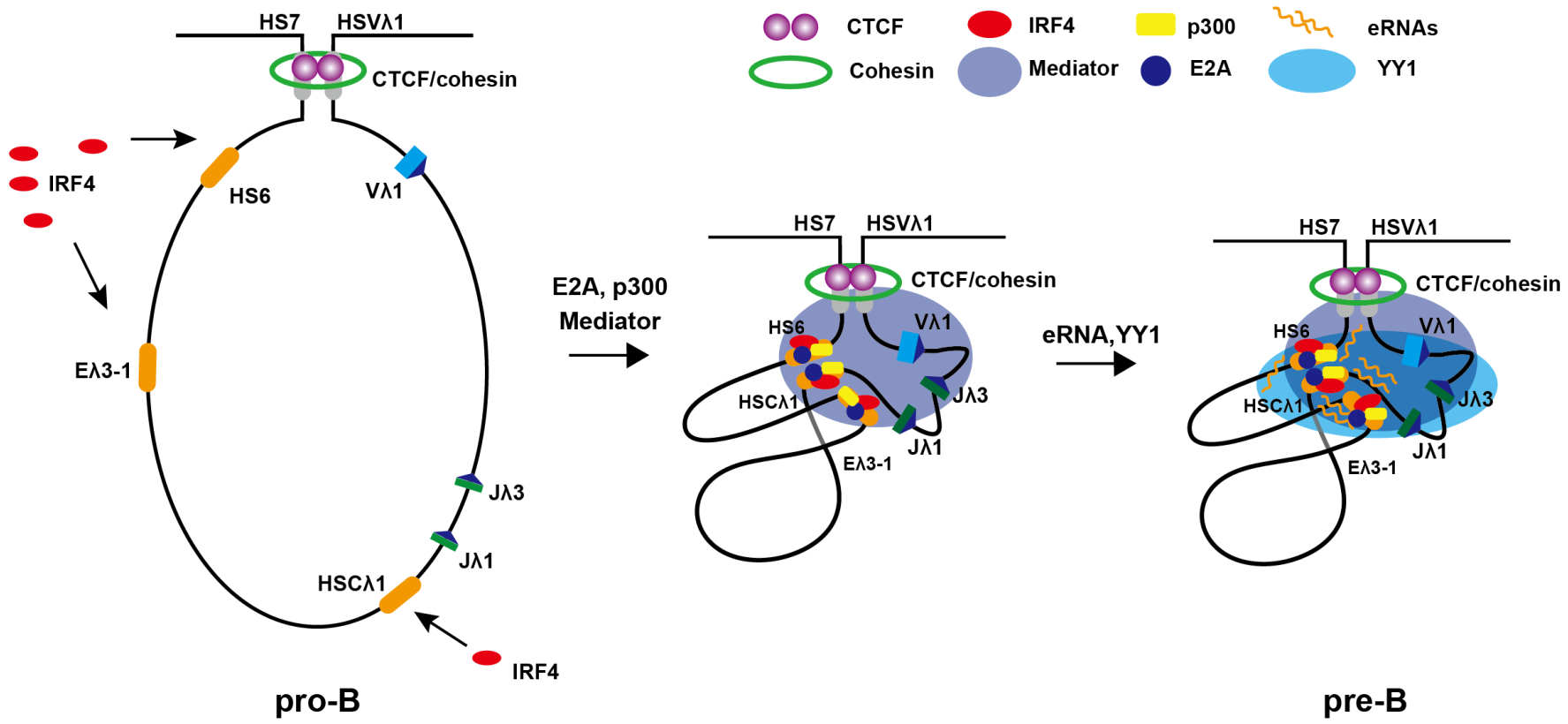

Figure 4.21 - A proposed model for enhancer-mediated Igλ activation

In pro-B cells where the Igλ locus is inactive, an IND is formed by CTCF/cohesin to bring HS7 and HSVλ1 together, sealing the 3' half of the Igλ locus. At the early stage of Igλ activation, increased levels of IRF4 binding to Eλ3-1, HS6 as well as HSCλ1 result in recruitment of E2A and p300, facilitating the generation of open chromatin. Simultaneously, the Mediator complex is recruited to enhancers and promoters via direct interactions with IRF4, resulting in locus contraction. At the late stage of Igλ activation, YY1 binds to enhancers and promoters and this correlates with the synthesis of enhancer RNAs, facilitating the stabilization of folded chromatin and leading to the efficient transcription of Igλ.

Chapter 5: The role of enhancer RNAs in the activation of the Igλ locus

A) Introduction

Transcriptional enhancers are traditionally believed to function as a transporter for delivering pre-bound transcription factors, including general transcription factors, lineage specific transcription factors, architecture factors and chromatin remodelers, to their target promoters to establish tissue-specific enhancer-promoter interplay. By using the next-generation technology, transcriptional enhancers have been demonstrated to encode enhancer RNAs in mammalian cells (Kim et al., 2010). It was initially assumed that enhancer RNAs are non-functional and are merely by-products of the RNAPII machinery. However, recent studies have demonstrated that enhancer RNAs play diverse roles in the regulation of gene transcription. For instance, enhancer RNAs can directly bind to structural factors, including Mediator (Lai et al., 2013), YY1 (Sigova et al., 2015) and cohesin (Tsai et al., 2018). Depletion of enhancer RNAs lead to the reduced enrichment of structural factors at enhancers and disrupted target gene transcription. In addition, enhancer RNAs can also facilitate the binding of transcription activators, such as c-Jun and NF-κB, to corresponding enhancers (Shii et al., 2017; Huang et al., 2018).

Enhancer RNAs can be primarily classified into two groups in terms of the transcriptional directionality. Unidirectionally transcribed enhancer RNAs are generally long (>4 kb) and polyadenylated RNAs, whereas bidirectional transcribed enhancer RNAs are relatively short (<2 kb) and non-polyadenylated (Koch et al., 2011; Natoli and Andrau, 2012). Genome-wide analysis of the expression of long non-coding RNAs suggests that the majority of enhancer RNAs belong to bidirectional enhancer RNAs (Andersson et al., 2014). Due to the lack of polyadenylation signals, the 3’ end of each bidirectional enhancer RNAs is processed by the Integrator complex instead of CPSF to complete transcription termination (Lai et al., 2015). The integrator complex was initially demonstrated to be involved in the 3’ end processing of RNAPII dependent uridylate rich small nuclear RNAs (Baillat et al., 2005). Recent studies have

shown that apart from the control of transcription termination, Integrator also plays essential roles in the regulation of transcription elongation. For instance, pause-release is a general rate-limiting step of transcription elongation (Adelman and Lis, 2012). P-TEFb is an essential regulator of the pause-release process (Peterlin, 2010) and has been demonstrated to be recruited to paused RNAPII at immediate early genes by the Integrator complex through direct interactions (Gardini et al., 2014). In addition, Integrator can activate poised enhancers by recruiting the early growth response (EGR) transcription activators, such as EGR1/2, during monocytic differentiation (Barbieri et al., 2018).

Mammalian antigen receptor loci contain multiple enhancer-like elements. Enhancer-mediated regulation of locus folding and V(D)J recombination has been extensively investigated (Proudhon et al., 2015; Schatz and Ji, 2011). However, the roles of enhancer RNAs in the regulation of activation of antigen receptor loci are unknown. Recent chromatin interaction studies revealed that enhancers that are engaged in looping with cognate promoters of protein-coding genes exhibit higher expression of enhancer RNAs and are occupied by subunits of the Integrator complex (Lai et al., 2015). Combined with the data regarding the active enhancer-promoter interactions described in Chapters 3 and 4, it is highly likely that enhancer RNAs and Integrator play roles in the regulation of activation of the Igλ locus.

In this chapter, I describe the characterization of the enhancer RNAs encoded by the Eλ3-1 enhancer within the Igλ locus in primary pro-B and pre-B cells. I further describe the temporal analysis of enhancer RNA expression and Integrator binding using the inducible IRF4 system to decipher if the alteration of levels of enhancer RNA expression and Integrator binding shows a correlation with the binding of other transcription activators. In addition, I knock-down the expression of enhancer RNAs encoded by Eλ3-1 followed by 3C analysis to build a picture of how the chromatin folding of the Igλ locus is regulated by enhancer RNAs.

B) Results

5.1 The Eλ3-1 enhancer encodes enhancer RNAs

Enhancer RNAs are a subclass of non-coding RNAs that are transcribed from active enhancers and have been demonstrated to be involved in the formation of enhancer-promoter loops and the activation of target genes (Li et al., 2016b). Active enhancers can be characterized by enhanced chromatin accessibility, enrichment of H3K27ac, binding of transcription activators, and recruitment of RNAPII (Shlyueva et al., 2014). According to analysis of published ATAC-seq and ChIP-seq data from primary pro-B cells and pro-B like cell lines, the Eλ3-1 enhancer displays the characteristics of an active enhancer at the pro-B stage of development (Figure 4.5), which implies that Eλ3-1 may encode enhancer RNAs in pro-B cells. To verify this, published RNA-seq data from pro-B cells were re-analyzed and the results show that a number of reads map to the Eλ3-1 enhancer (Figure 5.1), suggesting that Eλ3-1 is indeed capable of producing enhancer RNAs.

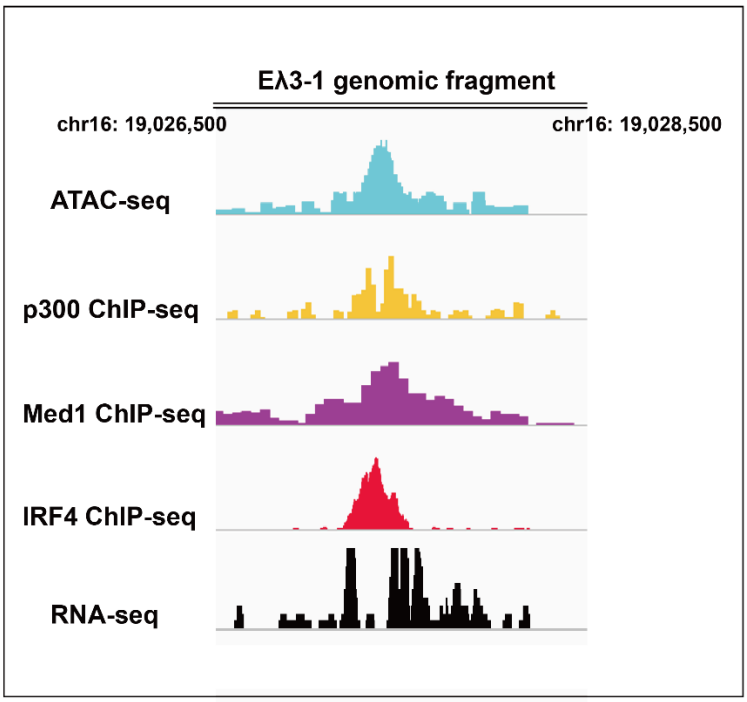

Figure 5.1 – RNA-seq analysis of the expression of Eλ3-1 enhancer RNAs

RNA-seq data from pro-B cells (Bonelt et al., 2019) was re-analyzed using the Galaxy web server. Signal peaks of ATAC-seq and ChIP-seq data from pro-B cells (adapted

from Figure 4.5) indicate the central region of the Eλ3-1 enhancer. Visualization of the mapped reads was performed in IGV. Genomic coordinates of the Eλ3-1 enhancer are shown.

To investigate if enhancer RNAs encoded by Eλ3-1 are associated with the activation of the Igλ locus, RT-qPCR was performed to examine the level of expression of Eλ3-1 enhancer RNAs in primary pro-B and pre-B cells. As shown in Figure 5.2, the level of Eλ3-1 enhancer RNAs increases significantly from pro-B to pre-B cells.

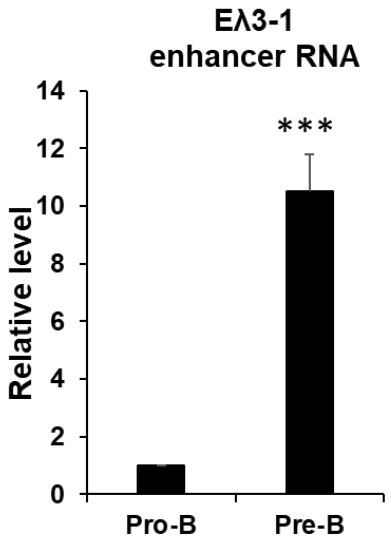

Figure 5.2 – Enhancer RNAs encoded by Eλ3-1 increase dramatically from pro-B to pre-B cells

The transcription level of the Eλ3-1 enhancer in non-transgenic pro-B and pre-B cells was analyzed by quantitative PCR. Data were normalized to the expression level of the housekeeping gene, *Hprt*. Error bars show standard error of the mean (SEM) from three experimental repeats. ***** represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

5.2 Eλ3-1 enhancer transcription correlates with YY1 binding to Eλ3-1 during the activation of the Igλ locus

The data above suggest that the Eλ3-1 enhancer RNAs are highly likely involved in the activation of the Igλ locus. Previous publications demonstrated that the enhancer RNAs can interact with diverse transcription factors, including

cohesin (Li et al., 2013), Mediator (Lai et al., 2013), YY1 (Sigova et al., 2015) and p300 (Bose et al., 2017). To determine if the change in expression of enhancer RNAs encoded by Eλ3-1 correlates with these enhancer RNA binding partners, it is important to perform the temporal analysis of the expression of enhancer RNAs encoded by Eλ3-1 during the activation of the Igλ locus. Therefore, the level of expression of enhancer RNAs encoded by Eλ3-1 was determined using RT-qPCR in the inducible Igλ system, 1D1-T215. The data shown in Figure 5.3 reveal that the total level of enhancer RNAs encoded by Eλ3-1 starts to increase from 4 hpi, just prior to the increase of YY1 binding to Eλ3-1 during Igλ locus activation (Figure 3.21). The largest increase is between 8 and 12 hours and correlates with the largest increase in YY1 binding. YY1 has been previously demonstrated to be trapped by RNAs tethered at enhancer loci (Sigova et al., 2015). These data therefore may imply that the increase of YY1 binding to Eλ3-1 is caused by enhancer RNA transcription.

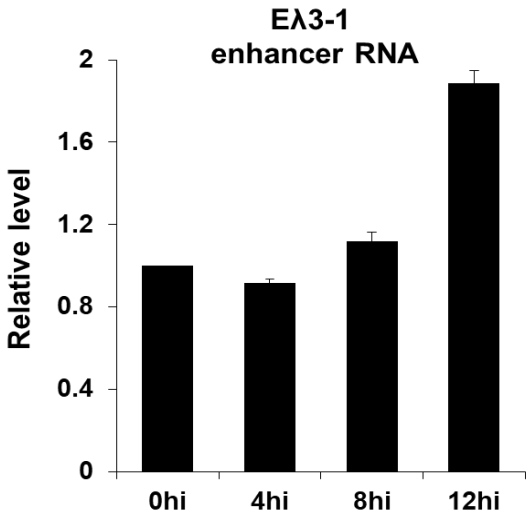

Figure 5.3 – Temporal analysis of Eλ3-1 transcription in 1D1-T215 cells

The transcription level of the Eλ3-1 enhancer was analysed by RT-qPCR in 1D1-T215 cells following induction. This shows a gradual increase from 4 to 12 hours post-induction. Data were normalized to the expression level of the housekeeping gene, *U6*. Error bars show standard error of the mean (SEM) from three experimental repeats.

5.3 Integrator is recruited to enhancers prior to enhancer transcription during Igλ locus activation

The Integrator complex has been demonstrated to be essential for biogenesis of enhancer RNAs, establishment of enhancer-promoter interactions, and for facilitating the release of paused RNAPII (Lai et al., 2015; Shii et al., 2017). To investigate if the Integrator complex is involved in the regulation of interactions between Eλ3-1 and Vλ1, temporal analysis of the binding of a core Integrator subunit, IntS11, to the Eλ3-1 enhancer and Vλ1 promoter was carried out using ChIP-qPCR in 1D1-T215 cells. As shown in Figure 5.4, the level of Integrator binding to the enhancer increase significantly from 0 to 4 hpi and reaches its highest level at 4 hpi, which is just prior to the increase of expression of Eλ3-1 enhancer RNAs. These data indicate that Integrator is an early event during the activation of non-coding transcription of Vλ1. This is consistent with the binding pattern of Integrator observed at enhancers for immediate early genes in HeLa cells induced with epidermal growth factor (Lai et al., 2015).

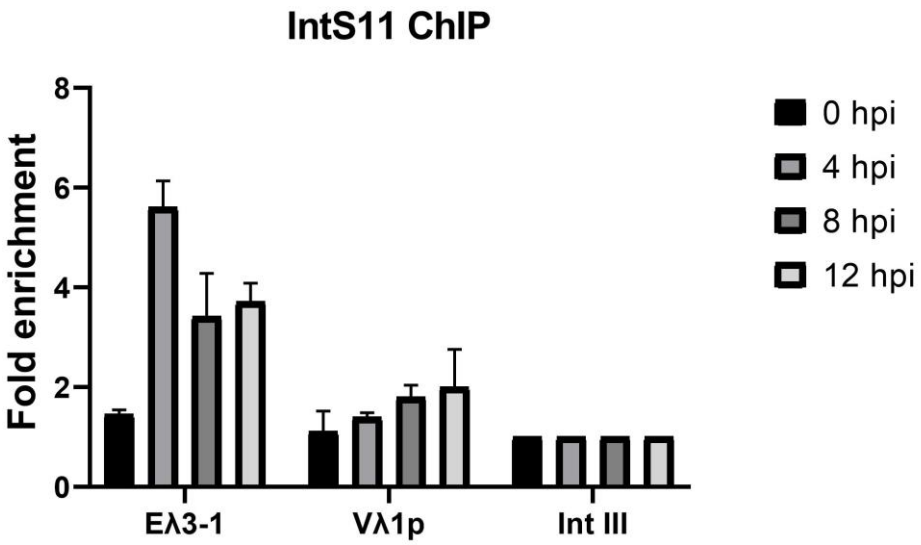

Figure 5.4 – Integrator is recruited to both Eλ3-1 and Vλ1p in 1D1-T215 cells
Integrator binding at the Eλ3-1 enhancer and Vλ1 promoter was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at Eλ3-1, Vλ1p and Intgene III is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

Rapid Integrator binding at enhancers following induction could explain the subsequent increase in the synthesis of enhancer RNAs. By contrast, Integrator binding to the Vλ1 promoter shows a gradual increase and seems to reach its highest level at 8 hpi (Figure 5.4). The difference in the Integrator binding pattern at enhancers and promoters may imply that Integrator plays a different role at gene promoters. Previous publications showed that Integrator can directly interact with regulators of RNAPII pause-release at protein-coding gene promoters, such as NELF (Stadelmayer et al., 2014) and p-TEFb (Gardini et al., 2014). These data together indicate that Integrator may be essential for the regulation of non-coding transcription of Igλ through different mechanisms.

5.4 Enhancer RNAs encoded by Eλ3-1 are bidirectional

Genome-wide analysis of transcription at enhancers suggests that the majority of enhancers are transcribed bidirectionally (Andersson et al., 2014). Global run-on sequencing (GRO-seq) is a powerful approach to identify the location and orientation of all actively transcribing RNA polymerases across the genome (Core et al., 2008). To determine if the enhancer RNAs generated by Eλ3-1 are bidirectional, published GRO-seq data from mouse pro-B cells (Bonelt et al., 2019) was re-analyzed. As shown in Figure 5.5, a number of reads were mapped to both the sense strand and anti-sense strand of the Eλ3-1 enhancer, implying that Eλ3-1 produces bidirectional enhancer RNAs in pro-B cells.

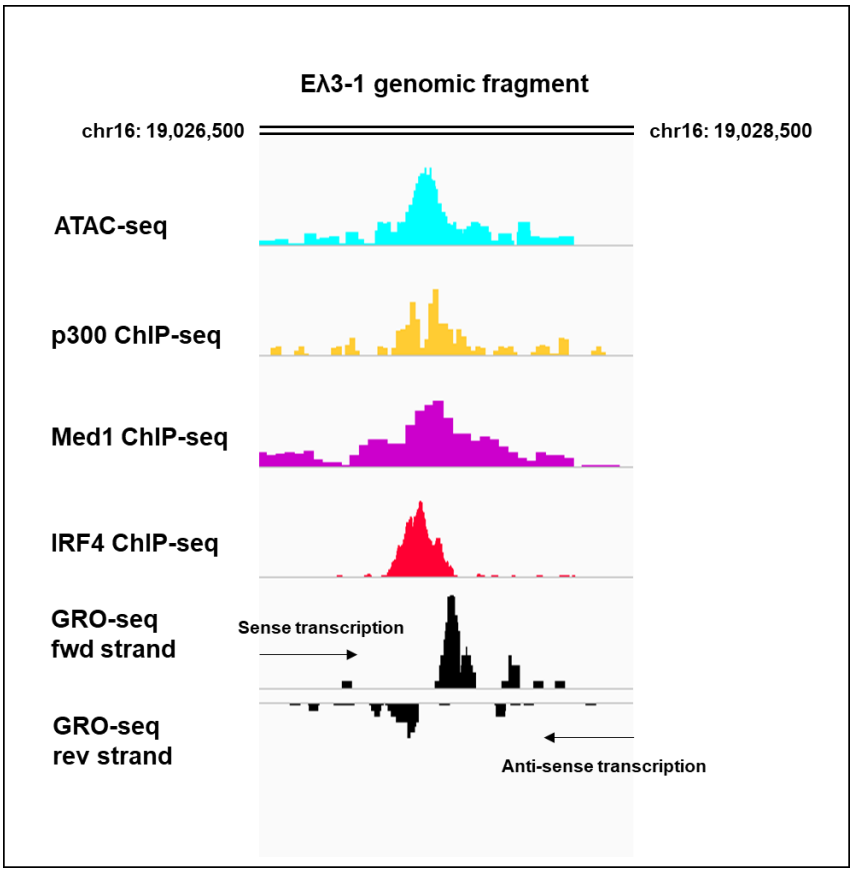

Figure 5.5 – GRO-seq analysis of the directionality of Eλ3-1 enhancer RNAs
GRO-seq data from pro-B cells (Bonelt et al., 2019) was re-analyzed using the Galaxy web server. Signal peaks of ATAC-seq and ChIP-seq data from pro-B cells (adapted from Figure 4.5) indicate the central region of the Eλ3-1 enhancer. Arrows indicate the direction of the sense and anti-sense enhancer RNA transcription. Visualization of the mapped reads was performed in IGV. Genomic coordinates of the Eλ3-1 enhancer are shown.

The data shown in Figure 5.3 suggest that the expression of the total level of Eλ3-1 enhancer RNAs starts to increase from 4 hpi when Integrator binding to the Eλ3-1 enhancer reaches its highest level following induction. However, the temporal expression pattern of the sense and anti-sense Eλ3-1 enhancer RNAs remains unclear. To investigate this, the sense and anti-sense Eλ3-1 enhancer RNAs were reverse transcribed with corresponding strand-specific primers (Figure 5.6A) and were subsequently subject to quantitative PCR analysis. The results showed that both of the sense and anti-sense Eλ3-1 transcription start to increase from 4 hpi (Figure 5.6B).

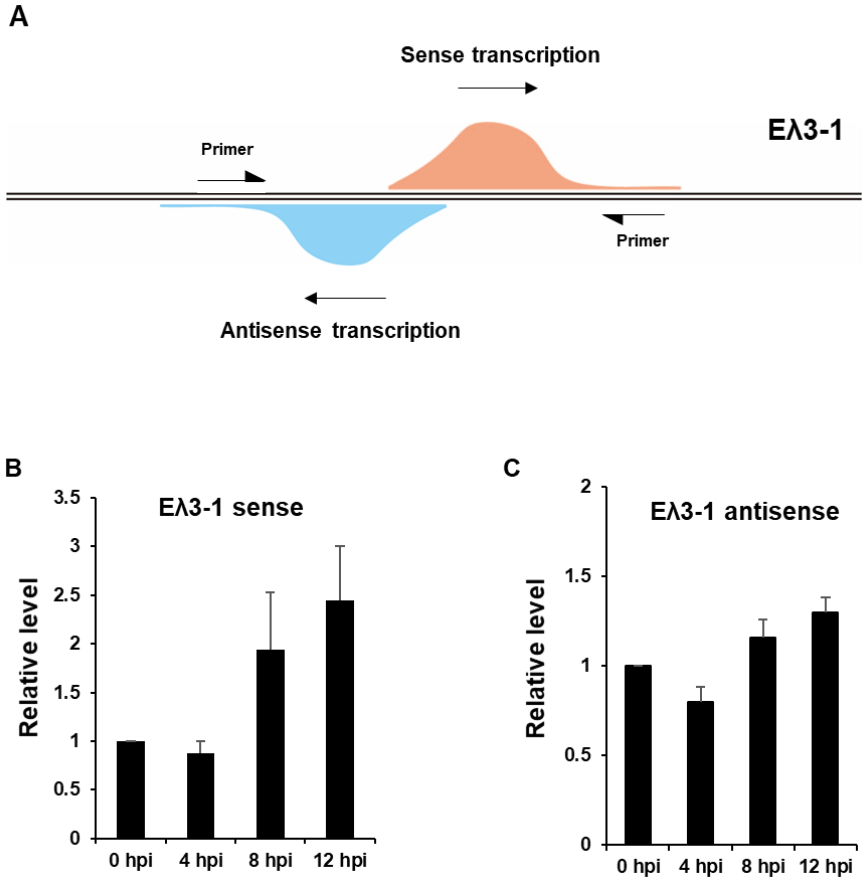

Figure 5.6 – Temporal analysis of Eλ3-1 sense and anti-sense transcription in 1D1-T215 cells

A) Schematic of reverse transcription of the Eλ3-1 enhancer RNAs using strand-specific primers

B) The expression level of the sense Eλ3-1 enhancer RNAs was analysed by RT-qPCR in 1D1-T215 cells following induction. This shows a gradual increase from 4 to 12 hours post-induction.

C) The expression level of the anti-sense Eλ3-1 enhancer was analysed by RT-qPCR in 1D1-T215 cells following induction. This also shows a gradual increase from 4 to 12 hours post-induction, albeit to a lower level than the sense RNA.

Data were normalized to the expression level of the housekeeping gene, *U6*. Error bars show standard error of the mean (SEM) from three experimental repeats.

5.5 Anti-sense enhancer RNA is intrinsically repressive to gene transcription

Previous publications demonstrated that enhancer RNAs are essential for the control of target gene transcription. To determine if enhancer RNAs encoded by the Eλ3-1 enhancer are required for the activation of Vλ1 non-coding transcription, knock-down of the sense and anti-sense Eλ3-1 enhancer RNAs were performed separately in 1D1-T215 cells using the pLKO.1 lentiviral system. RT-qPCR analysis demonstrated that, compared to 1D1-T215 cells expressing a scrambled shRNA (shSCR), the sense Eλ3-1 expression is diminished dramatically in 1D1-T215 cells expressing an shRNA targeting the sense Eλ3-1 transcripts (Figure 5.7A). Likewise, shRNA-mediated specific knock-down of Eλ3-1 anti-sense expression led to the degradation of more than 70 % of Eλ3-1 anti-sense transcripts in 1D1-T215 cells (Figure 5.7B).

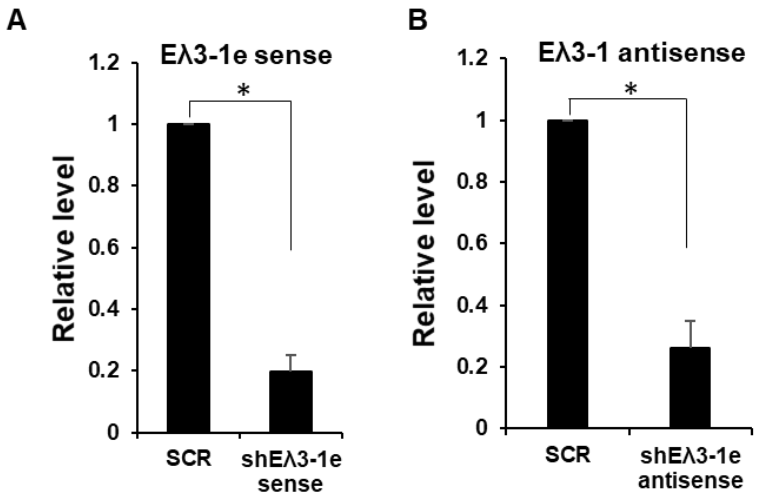

Figure 5.7 – Knock down of the sense and anti-sense Eλ3-1 enhancer RNAs using shRNA

A) RT-qPCR analysis of the level of Eλ3-1 sense transcripts in shSCR and shEλ3-1sense 1D1-T215 cells. The level of Eλ3-1 sense transcripts is diminished dramatically in cells expressing shEλ3-1sense.

B) RT-qPCR analysis of the level of Eλ3-1 anti-sense transcripts in shSCR and shEλ3-1 anti-sense 1D1-T215 cells. The level of Eλ3-1 anti-sense transcripts is diminished dramatically in shEλ3-1anti-sense cells.

Data were normalized to the expression level of the housekeeping gene, *U6*. Error bars show standard error of the mean (SEM) from three experimental repeats. * represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

Next, I sought to determine the level of Vλ1 non-coding transcription in the sense Eλ3-1 enhancer RNA knock down (shEλ3-1sense) and anti-sense Eλ3-1 enhancer RNA knock down (shEλ3-1anti-sense) 1D1-T215 cells, respectively. As can be seen in Figure 5.8A, Vλ1 non-coding transcription is reduced significantly in Eλ3-1 sense enhancer RNA knock-down 1D1-T215 cells, as might be expected. Intriguingly, compared with uninduced shSCR 1D1-T215 cells, Vλ1 non-coding transcription increases significantly in uninduced shEλ3-1anti-sense 1D1-T215 cells (Figure 5.8B). Likewise, a similar phenomenon was observed in induced shEλ3-1anti-sense and shSCR 1D1-T215 cells.

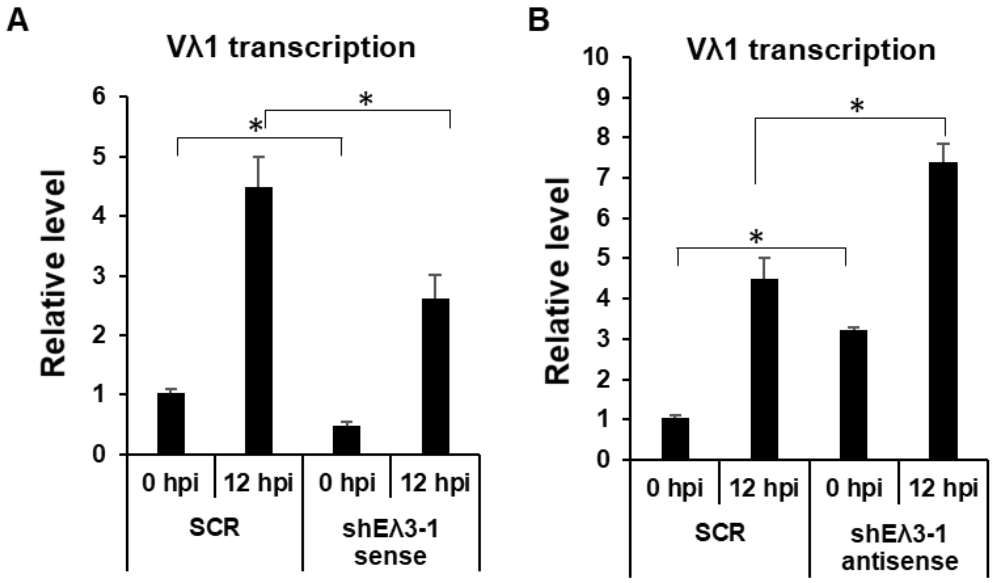

Figure 5.8 – Eλ3-1 enhancer RNAs are essential for Vλ1 non-coding transcription

A) RT-qPCR analysis of the level of Vλ1 non-coding transcription in shSCR and shEλ3-1sense 1D1-T215 cells following induction. The level of Vλ1 transcription is diminished dramatically in shEλ3-1sense cells.

B) RT-qPCR analysis of the level of Vλ1 non-coding transcription in shSCR and shEλ3-1antisense 1D1-T215 cells following induction. The level of Vλ1 transcription is increased dramatically in shEλ3-1antisense cells. Data were normalized to the expression level of the housekeeping gene, *Hprt*.

Error bars show standard error of the mean (SEM) from three experimental repeats.

* represents a p-value <0.05, ** a p-value <0.01 and *** a p-value < 0.001.

Previous publications showed that enhancer RNAs are essential for the establishment of enhancer-promoter interactions. Thus, the difference in

functionality of the sense and anti-sense Eλ3-1 enhancer RNAs may be reflected in the establishment of Eλ3-1 - Vλ1 chromatin loops. To verify this, 3C analysis of the interactions between Eλ3-1 and Vλ1 was performed in the shEλ3-1sense and shEλ3-1antisense 1D1-T215 cells, respectively. As shown in Figure 5.9A, Eλ3-1-Vλ1 interactions are completely diminished in shEλ3-1sense 1D1-T215 cells at 12 hpi, indicating that the sense Eλ3-1 enhancer RNA is vital for the establishment of enhancer-promoter chromatin loops. Surprisingly, however, an increased interaction frequency between Eλ3-1 and Vλ1 was observed in shEλ3-1anti-sense 1D1-T215 cells (Figure 5.9B). This implies that anti-sense enhancer RNAs may be intrinsically repressive to target gene transcription.

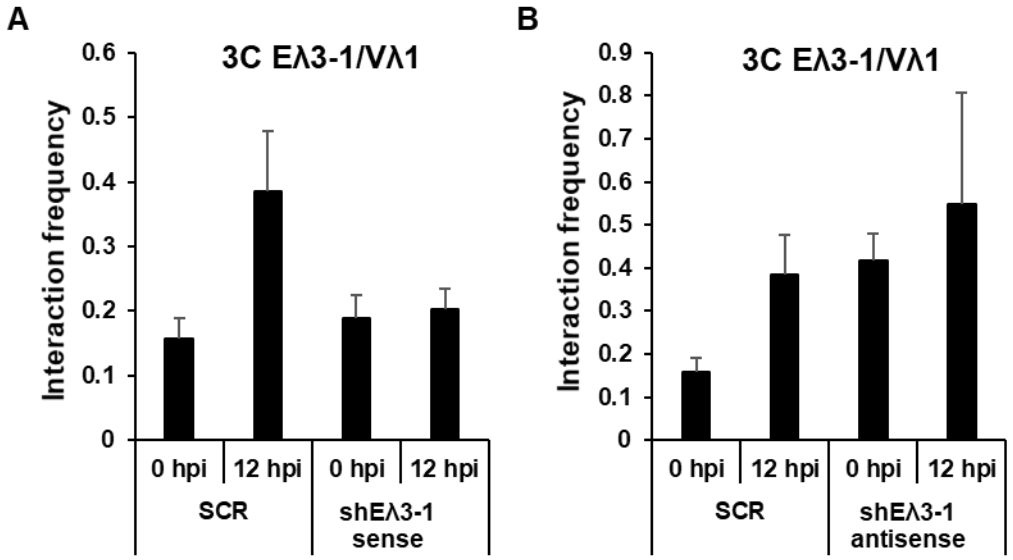

Figure 5.9 – Eλ3-1 enhancer RNAs are essential for the establishment of Eλ3-1 - Vλ1 interactions

A) 3C analysis of interactions between Eλ3-1 and Vλ1 in shSCR and shEλ3-1sense 1D1-T215 cells. The interaction frequency between Eλ3-1 and Vλ1 is decreased in sense Eλ3-1 enhancer RNA knock-down cells after induction.

B) 3C analysis of interactions between Eλ3-1 and Vλ1 in shSCR and shEλ3-1antisense 1D1-T215 cells. The interaction frequency between Eλ3-1 and Vλ1 is increased in anti-sense Eλ3-1 enhancer RNA knock-down cells after induction.

Data were normalized by detecting an interaction with the ERCC3 locus. Error bars show standard error of the mean (SEM) from three replicates.

5.6 Enhancer RNAs are essential for the establishment of the correct chromatin structure of the Igλ locus

As establishment of the correct chromatin environment is a prerequisite for the activation of non-coding transcription of unrearranged gene segments in the Igλ locus, disruption of Vλ1 non-coding transcription in enhancer RNA knock down cells may have been caused by disruption of the normal programmed change in chromatin folding during Igλ activation. To verify this, 3C analysis using the Eλ3-1 as a viewpoint was firstly performed in shEλ3-1sense 1D1-T215 cells. This showed a substantial decrease in chromatin interactions within the Eλ3-1 enhancer hub when sense Eλ3-1 enhancer RNA is knocked down (Figure 5.10). Likewise, the interaction frequency between Eλ3-1 and gene segments, including Jλ1, Vλ1 and Jλ3, is reduced in shEλ3-1sense 1D1-T215 cells. These data therefore indicate that the sense Eλ3-1 enhancer RNA is essential for the regulation of chromatin interactions during the activation of the Igλ locus.

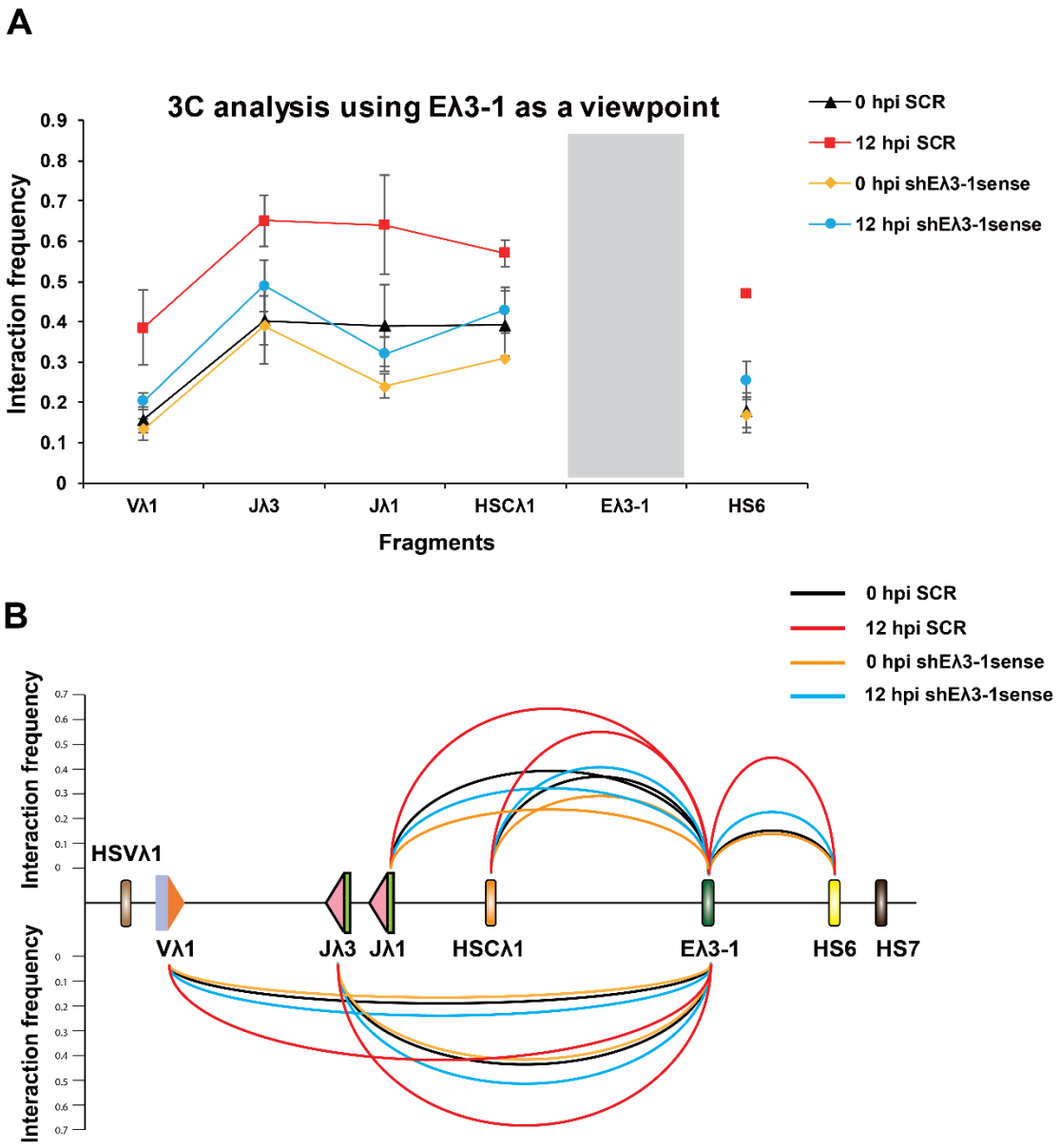

Figure 5.10 – 3C analysis of chromatin interactions formed in the 3' half of Igλ in shEλ3-1sense 1D1-T215 cells

A) Analysis of the relative interaction frequency of Dpn II fragments from the Eλ3-1 viewpoint in 1D1-T215 cells expressing an shRNA targeting the sense Eλ3-1 enhancer RNA. Error bars show standard error of the mean (SEM) from three experimental repeats.

B) Schematic diagram of chromatin interactions within the 3' half of the Igλ locus using Eλ3-1 as the viewpoint. The height of curves between Eλ3-1 and other genomic fragments represents the average value of interaction frequency obtained from three experimental repeats.

As mentioned above, the anti-sense Eλ3-1 enhancer RNA is repressive to the activation of Vλ1 non-coding transcription. To examine if the anti-sense RNAs are repressive to the chromatin organization of the whole Igλ locus, 3C experiments were next performed in shEλ3-1antisense 1D1-T215 cells. As shown in Figure 5.11, the chromatin interactions between Eλ3-1 and other functional genomic elements increase substantially in anti-sense Eλ3-1 enhancer RNA knock down cells. These data therefore suggest that the Eλ3-1 anti-sense transcripts repress the establishment of correct chromatin folding within the Igλ locus.

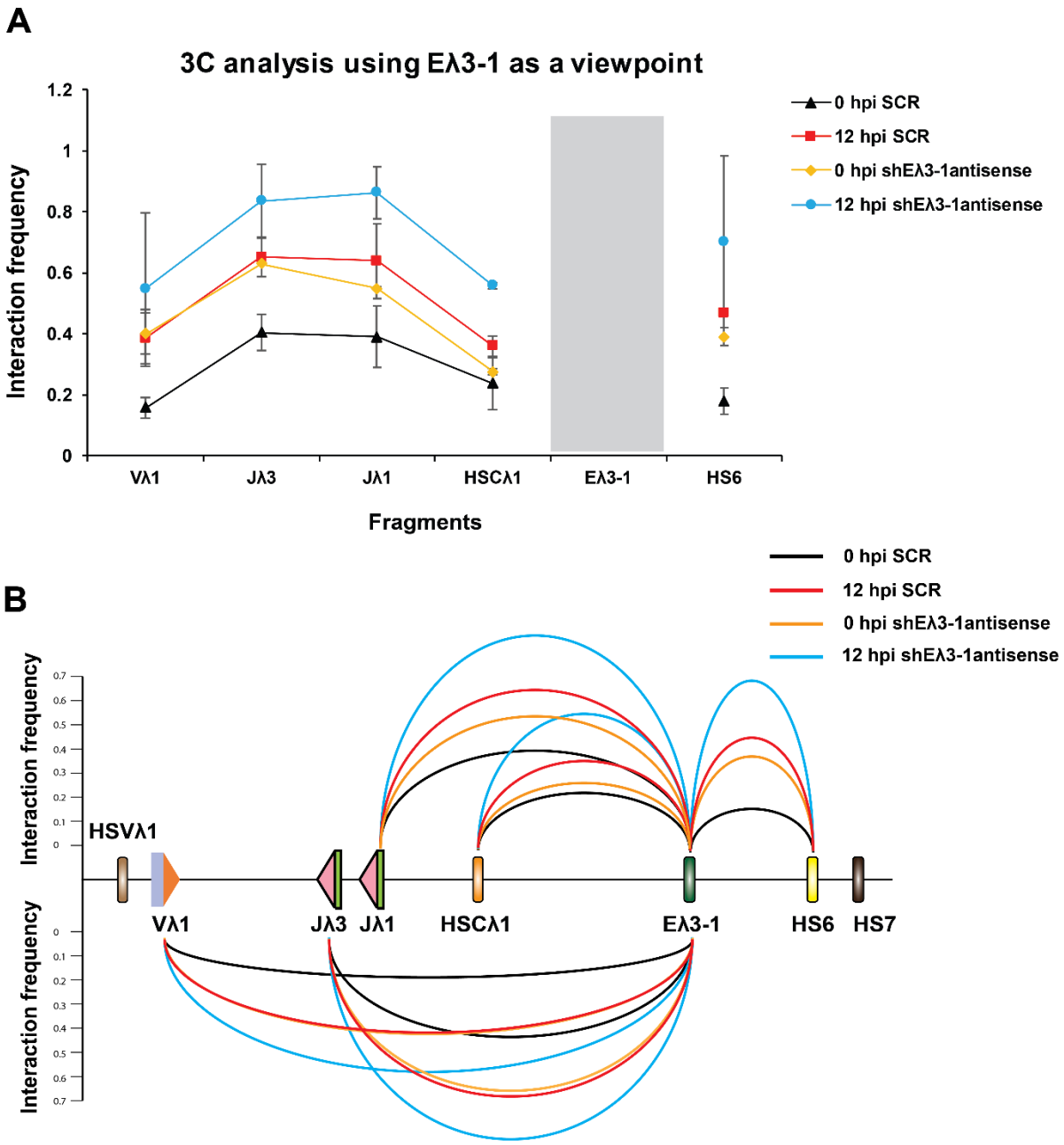

Figure 5.11 – 3C analysis of chromatin interactions formed in the 3' half of Igλ in shEλ3-1antisense 1D1-T215 cells

A) Analysis of the relative interaction frequency of Dpn II fragments from the Eλ3-1 viewpoint in 1D1-T215 cells expressing an shRNA targeting the anti-sense Eλ3-1 enhancer RNA. Error bars show standard error of the mean (SEM) from two replicates. B) Schematic diagram of chromatin interactions within the 3' half of the Igλ locus using Eλ3-1 as the viewpoint. The height of curves between Eλ3-1 and other genomic fragments represents the average value of interaction frequency obtained from two experimental repeats.

Previous publications demonstrated that enhancer RNAs exert their functions through interacting with the binding partners, such as p300, cohesion, Mediator and YY1. Temporal ChIP analysis performed in Chapter 3 and 4 showed that the increase of the level of Eλ3-1 enhancer RNAs occurs just prior to the increase of YY1 enrichment at Eλ3-1, which may suggest that expression of enhancer RNAs is a prerequisite for YY1 binding, and the disruption of the chromatin structure in enhancer RNA knock down cells may be caused by the altered YY1 binding. To test this, ChIP-qPCR analysis of YY1 binding to Eλ3-1 was performed in the shEλ3-1sense and shEλ3-1antisense 1D1-T215 cells. As shown in Figure 5.12, knock down of the Eλ3-1 sense enhancer RNA leads to the decreased YY1 binding to Eλ3-1, suggesting that enhancer RNA mediated regulation of chromatin structure is indeed associated with YY1 binding. By contrast, YY1 binding to Eλ3-1 is not affected in the Eλ3-1 anti-sense enhancer RNA knock down cells, indicating that the anti-sense enhancer RNA likely exerts its regulatory function in a YY1 independent manner.

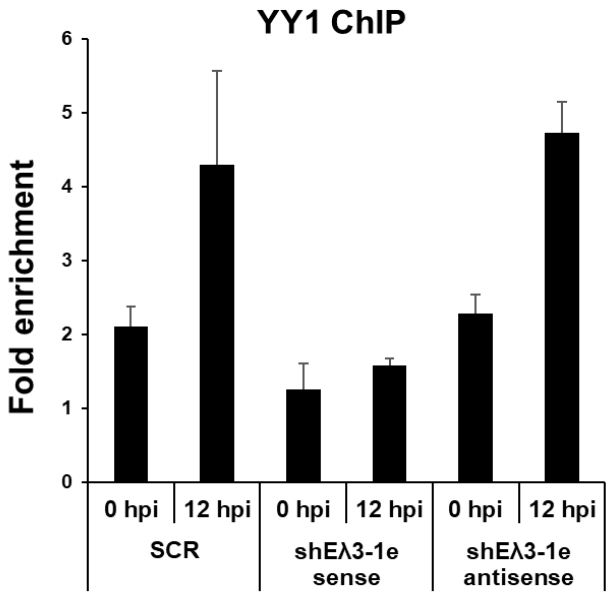

Figure 5.12 – YY1 binding to Eλ3-1 in enhancer RNA knock down cells

YY1 binding to Eλ3-1 was analysed by ChIP-qPCR in shEλ3-1sense and shEλ3-1 antisense 1D1-T215 cells. The fold enrichment at Eλ3-1 and Intgene III (negative control) is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

5.7 HS6 and HSCλ1 produce enhancer RNAs in B cells

The data shown in Chapter 4 reveal that Eλ3-1 shows a similar transcription factor binding pattern to two newly identified enhancer-like elements, HS6 and HSCλ1. Eλ3-1 interacts with these two putative enhancers to form a super-enhancer during the activation of the Igλ locus. This may imply that HS6 and HSCλ1 produce enhancer RNAs in pro-B and pre-B cells, like Eλ3-1. To verify this, GRO-seq data from pro-B cells was mapped to the HS6 and HSCλ1 enhancers. As shown in Figure 5.13, a number of reads were mapped to both of the sense and anti-sense strands of HS6 and HSCλ1, suggesting that both HS6 and HSCλ1 produce bidirectional enhancers in pro-B cells.

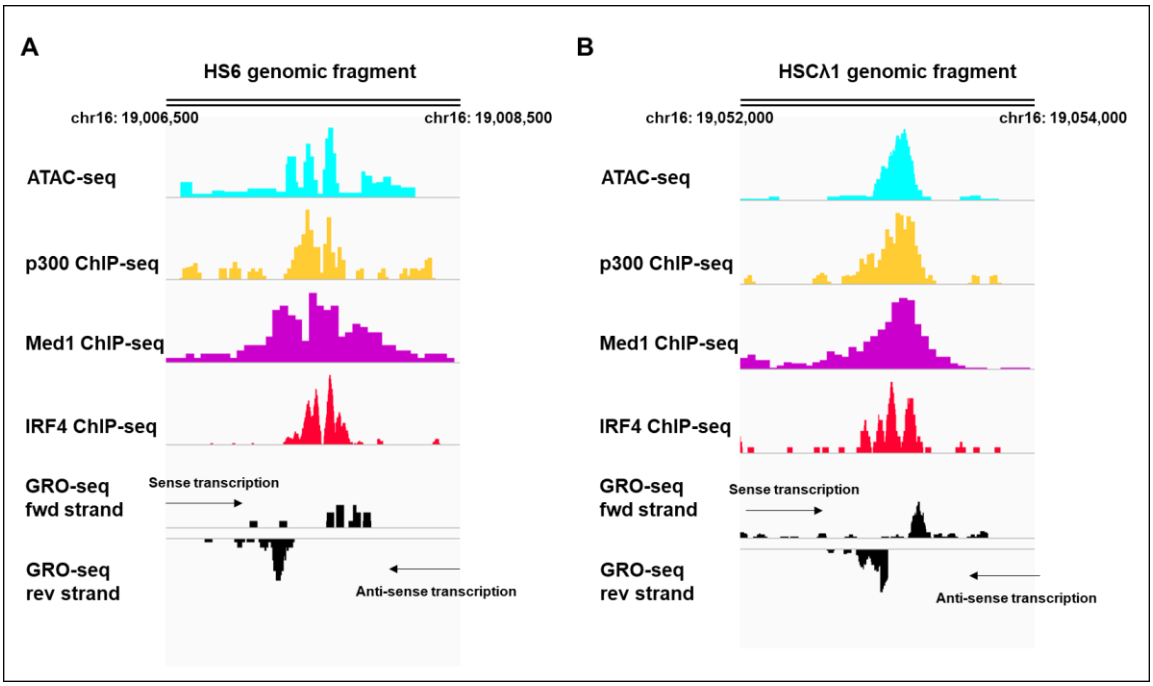

Figure 5.13 – GRO-seq analysis of HS6 and HSCλ1 enhancer RNAs

GRO-seq data from pro-B cells (Bonelt et al., 2019) was re-analyzed using the Galaxy web server. Signal peaks of ATAC-seq and ChIP-seq data from pro-B cells (adapted from Figure 4.5) indicate the central region of the HS6 (A) and HSCλ1 (B) enhancers. Arrows indicate the direction of the sense and anti-sense enhancer RNA transcription. Visualization of the mapped reads was performed in IGV. Genomic coordinates of the HS6 and HSCλ1 enhancers are shown.

In addition, ChIP-qPCR analysis of Integrator binding to HS6 and HSCλ1 was performed in 1D1-T215 cells following induction. The results show that Integrator binding to HS6 and HSCλ1 reaches its highest level at 4 hpi, which is very similar to the binding pattern observed at Eλ3-1. These data therefore indicate that HS6 and HSCλ1 may share the same mechanism as Eλ3-1 to produce bidirectional enhancer RNAs and that these enhancer RNAs function in a similar way at all three enhancers.

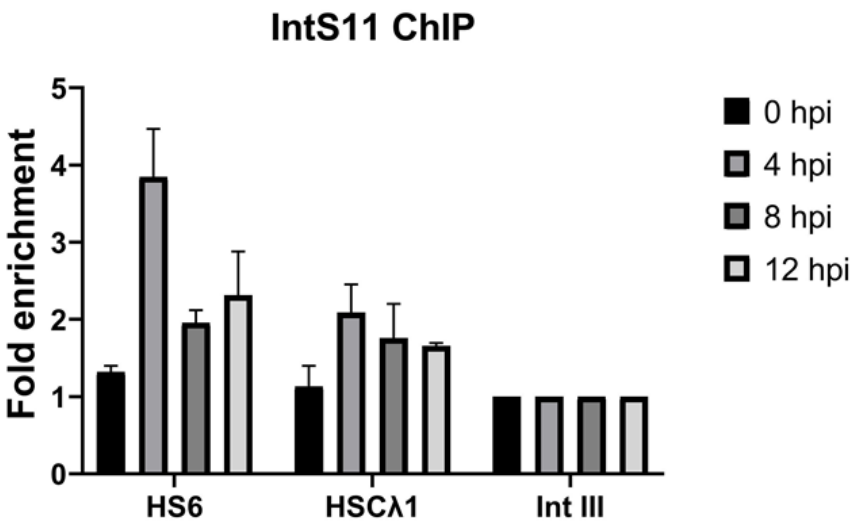

Figure 5.14 – Integrator is recruited to HS6 and HSCλ1 in 1D1-T215 cells
Integrator binding at the HS6 and HSCλ1 enhancer was analysed by ChIP-qPCR in 1D1-T215 cells following induction. The fold enrichment at HS6 and HSCλ1 and Intgene III is shown. All values are normalized to binding at Intgene III as a negative control. Error bars show standard error of the mean (SEM) from three experimental repeats.

C) Discussion

This chapter aimed to examine how enhancer RNAs mediate the regulation of target gene transcription using the mouse Igλ locus as a model. By re-analyzing published RNA-seq and GRO-seq datasets, bidirectional enhancer RNAs encoded by the B specific enhancer, Eλ3-1, were found to be present in pro-B cells. Temporal analysis of the expression of these enhancer RNAs in the inducible 1D1-T215 cell line showed that the levels of both sense and anti-sense Eλ3-1 enhancer RNAs start to increase at 4hpi, just after the increase in Integrator binding to Eλ3-1. Knock down of the sense Eλ3-1 enhancer RNA leads to the disruption of chromatin folding, accompanied by reduced non-coding transcription of the Vλ1 promoter. Surprisingly, knock down of the anti-sense Eλ3-1 enhancer RNAs results in the chromatin contraction and increased levels of non-coding transcription of the Vλ1 promoter. In addition, the constituent enhancers of the Eλ3-1 super-enhancer, HS6 and HSCλ1, also produce enhancer RNAs in pro-B cells.

5.8 Eλ3-1 enhancer RNAs increase substantially from pro-B to pre-B cells

The Eλ3-1 enhancer has been shown to be essential for the regulation of V(D)J recombination of the Igλ locus during the differentiation from pro-B to pre-B cells (Haque et al., 2013). Enhancer RNAs are transcribed from active enhancers which are characterized by high levels of transcription activator binding and specific histone modifications, such as H3K27ac and H3K4me1. Published ChIP-seq, RNA-seq and GRO-seq datasets confirm that Eλ3-1 is an active enhancer in pro-B cells and can generate bidirectional enhancer RNAs. RT-qPCR analysis of the total level of Eλ3-1 enhancer RNAs shows that the expression of Eλ3-1 enhancer RNAs increases more than 10-fold from pro-B to pre-B cells (Figure 5.2). The expression of the Eλ3-1 enhancer RNAs also increases in 1D1-T215 cells following induction. However, only a 2-fold increase in these levels was observed at 12 hpi. As mentioned in Chapter 3, the 1D1-T215 cell line is a derivative of the 1D1 cell line which is immortalized by A-MuLV. It was previously demonstrated that the v-Abl tyrosine kinase

encoded by A-MuLV can repress antigen receptor loci via STAT5 signalling. Sequence analysis shows that a STAT5 binding motif is present in Eλ3-1 when LASAGNA-search is used with a relatively low stringency cut-off value. This may explain the reduced increase of Eλ3-1 enhancer RNAs observed in 1D1-T215 cells. However, previous ChIP-qPCR data suggest that STAT5 is not bound to Eλ3-1 in pro-B cells. These data together indicate that the limited increase of Eλ3-1 enhancer RNAs may be not caused by STAT5 signalling. Consistent with this, an approximately 7-fold increase of Eλ3-1 enhancer RNAs was reproducibly observed in temperature shifted 103/BCL-2 cells (data not shown). These cells are immortalized by a temperature sensitive mutant of A-MuLV and the v-Abl kinase is inactivated following temperature shift. The altered increase in enhancer RNA expression may alternatively be explained by the requirement for other changes in 1D1-T215 cells that occur much later following induction. Indeed, the Eλ3-1 enhancer RNAs increase by nearly 6-fold in 1D1-T215 cells at 24 hpi (data not shown).

The total level of Eλ3-1 enhancer RNAs increases 2-fold in 1D1-T215 cells at 12 hpi. To determine the expression pattern of sense and anti-sense enhancer RNAs, RT-qPCR analysis was performed, which showed that the sense Eλ3-1 enhancer RNA is increased to a greater extent than the anti-sense Eλ3-1 enhancer RNA. As the initial ratio of sense and anti-sense enhancer RNAs are unknown, how much each enhancer RNA contributes to the activation is currently unclear. This could be addressed using absolute quantitative qPCR, using standard curves with known copy numbers, to determine the copies of each enhancer RNA present.

5.9 Integrator binding is essential for the activation of the Igλ locus

The Integrator complex directs 3’ end processing of enhancer RNAs and facilitates their transcription termination. Temporal ChIP analysis of Integrator binding to Eλ3-1 shows that Integrator enrichment at Eλ3-1 reaches its highest level in 1D1-T215 cells at 4 hpi, indicating that Integrator recruitment is an early event during the activation of the Igλ locus. However, how Integrator is recruited to the Eλ3-1 is unknown. There is no evidence that IRF4 can physically interact

with the Integrator complex. However, among IRF4 binding partners, the Mediator complex has been shown to interact with Integrator directly. Thus, Integrator may be recruited indirectly by IRF4 via interacting with Mediator. This can be verified by examining the level of Integrator binding to Eλ3-1 in Mediator KD 1D1-T215 cells after induction with 4-hydroxytamoxifen.

Integrator reaches its highest level at 4 hpi, which is just prior to the increase of the expression of enhancer RNAs. This is consistent with the role of Integrator in enhancer RNA end processing and transcription termination. However, Integrator binding to Vλ1 shows a gradual increase from 0 to 8 hpi and seems to reach its highest level at 8 hpi. The difference in the Integrator binding pattern may suggest Integrator exerts different functions at enhancers and promoters. Recent publications showed that Integrator can interact with the transcription elongation factor, p-TEFb, to facilitate the transition from paused Pol II to elongating Pol II, resulting in increased transcription efficiency at protein-coding gene promoters. Thus, enrichment of Integrator at the Vλ1 promoter may activate the p-TEFb to achieve efficient transcription. This could be verified by performing the temporal ChIP analysis of p-TEFb binding at the Vλ1 promoter to see if the p-TEFb binding to the Vλ1 promoter is a late event during the activation of the Igλ locus. If this is the case, further experiments can be conducted to knock down of Integrator followed by examination of the p-TEFb binding to the Vλ1 promoter.

5.10 Enhancer RNAs are essential for the activation of the Igλ locus

GRO-seq from pro-B cells suggests enhancer RNAs encoded by Eλ3-1 are bidirectional (Figure 5.5). Temporal analysis of the expression of Eλ3-1 sense and anti-sense enhancer RNAs showed that they both start to increase at 4 hpi (Figure 5.6), which is similar to the increase in total enhancer RNA levels. These data suggest that the enhancer RNAs play roles in the later stages of Igλ activation. However, knock down of Eλ3-1 sense enhancer RNA in uninduced 1D1-T215 cells results in a decrease in Vλ1 non-coding transcription (Figure 5.8A), indicating that basal enhancer RNA levels are still essential for the target gene transcription. This raises the question of how the basal

enhancer RNA levels contribute to target gene activation. Previous publications demonstrated that enhancer RNAs can facilitate enhancer-promoter interactions through interacting with architecture factors, such as cohesin (Tsai et al., 2018), Mediator (Lai et al., 2013) and YY1 (Sigova et al., 2015). However, 3C analysis of the chromatin interaction frequency between the Eλ3-1 enhancer and Vλ1 promoter does not change in uninduced Eλ3-1 sense enhancer RNA knock down cells (Figure 5.9), implying that the reduced Vλ1 transcription in enhancer RNA knock down cells may not involve the architecture factors. Enhancer RNAs are also thought to regulate gene transcription via stimulating the acetyltransferase p300 (Bose et al., 2017). This could be verified by determining the chromatin accessibility and level of H3K27ac at Eλ3-1 in the Eλ3-1 sense enhancer RNA knock down cells compared to cells expressing the scrambled shRNA.

Regulation of enhancer-promoter interactions is mediated by architectural factors, such as CTCF, cohesion, YY1 and Mediator (Schoenfelder and Fraser, 2019). These architectural proteins have been demonstrated to bind to large numbers of endogenous RNAs (Lai et al., 2013; Pan et al., 2020; Saldana-Meyer et al., 2019; Wai et al., 2016). Enhancer RNAs have been shown to be involved in the regulation of enhancer-promoter loops via facilitating recruitment of cohesion and Mediator (Lai et al., 2013; Tsai et al., 2018). Knock down of enhancer RNAs can lead to reduced enrichment of Mediator and cohesion at the corresponding enhancers and promoters, accompanied by disrupted enhancer-promoter interactions (Lai et al., 2013; Tsai et al., 2018). However, previous studies found that enhancer RNA knock down has no effect on loading of the cohesion complex at the corresponding enhancers (Hah et al., 2013; Mousavi et al., 2013; Schaukowitch et al., 2014). This may suggest that cohesion only contributes to the enhancer-promoter interactions mediated by enhancer RNAs at some loci. Consistent with this, no cohesion enrichment was detected at the Eλ3-1 enhancer and Vλ1 promoter (data not shown). It seems possible that enhancer RNAs which have no effects on cohesion loading may lack a specific binding domain for cohesion. Thus, enhancer RNA may interact with other architecture factors to facilitate the establishment of enhancer-

promoter interactions, such as Mediator (Lai et al., 2013) and YY1 (Sigova et al., 2015). Temporal ChIP analysis of Mediator and YY1 binding to the Igλ locus reveal that only YY1 binding to the Igλ locus follows the synthesis of enhancer RNA (Figures 3.14 and 3.21), suggesting that enhancer RNA synthesis may be a prerequisite for YY1 recruitment to target enhancers and promoters within the Igλ locus. This can be verified by determining the level of YY1 binding to enhancers and promoters within the Igλ locus in Eλ3-1 enhancer RNA knock down 1D1-T215 cells following induction.

The data presented in this chapter demonstrate that enhancer RNAs encoded by Eλ3-1 are essential for the chromatin organization of the Igλ locus. However, previous publications showed that inhibition of enhancer RNA production by flavopiridol, an inhibitor of CDK9, does not change the normal chromatin landscape nor inhibit looping to target gene promoters (Hah et al., 2013), suggesting enhancer RNA transcripts are not required in this case. Together, these data indicate that enhancer RNAs work in different ways at different enhancers.

5.11 Anti-sense enhancer RNAs are repressive to target gene transcription

Previous publications show that enhancer RNAs are essential for the establishment of enhancer-promoter interactions (Arnold et al., 2019). Consistent with this, knock down of the Eλ3-1 sense enhancer RNA disrupts the chromatin interactions between the Eλ3-1 enhancer and its target promoters, accompanied by reduced target gene transcription (Figure 5.10). However, increased Vλ1 non-coding transcription was observed in the Eλ3-1 anti-sense enhancer RNA knock down cells (Figure 5.11), indicating that anti-sense enhancer RNAs may be repressive to target gene transcription. This raises a question of how anti-sense enhancer RNAs work to repress target gene transcription. Single-stranded non-coding RNA molecules do not work in isolation in the complex nuclear environment. These molecules can form sophisticated functional domains to physically interact with diverse RNA binding proteins to exert their functions in different biological processes (Kung et al.,

2013; Rinn and Chang, 2012). As the increased target gene transcription is accompanied by increased chromatin interactions within the Igλ locus in the anti-sense enhancer RNA knock down cells, the anti-sense enhancer RNA may play roles in the repressing the recruitment of architecture factors to the Igλ locus. This is could be verified by determining the level of architecture factors, such as Mediator and YY1, at Eλ3-1 in the Eλ3-1 anti-sense enhancer RNA knock down cells. In turn, this may raise the question of how the anti-sense enhancer RNA suppresses the recruitment of architecture factors. One possibility is this could be due to reduced chromatin accessibility. Previous publications show that numerous long non-coding RNAs are capable of recruiting transcription repressors, such as the heterogeneous nuclear ribonucleoproteins (Carpenter et al., 2013) and polycomb repressive complex (Brockdorff, 2013), thereby decreasing the chromatin accessibility at enhancers and promoters. Moreover, as the sense enhancer RNA facilitates recruiting architecture factors to enhancers via its functional domains, the anti-sense enhancer RNA may interact with the sense enhancer RNA to disrupt the interactions between the sense enhancer RNA and architecture factors. Consistent with this, computation analysis showed that the Eλ3-1 anti-sense enhancer RNA can hybridize with the sense enhancer RNA to form a stable structure (data not shown).

Notably, however, knock-down of the anti-sense enhancer RNA makes a negligible difference to YY1 binding to Eλ3-1, which suggests that the anti-sense enhancer RNA works via a mechanism independent of YY1 recruitment. It may also work independently of counteracting this aspect of sense enhancer RNA function. Nonetheless, these data indicate that the anti-sense enhancer RNA is essential for transcriptional regulation and that the interplay between the sense enhancer RNAs and anti-sense enhancer RNAs may represent a novel mechanism utilized by cells to achieve precise regulation of gene transcription.

Chapter 6: Discussion

In this thesis, I characterized an inducible pro-B cell line, 1D1-T215, to investigate the dynamics of long-range enhancer-promoter contacts. This system led to three key discoveries, that enhance our understanding of the activation of enhancer-promoter interactions. Specifically, 1) I uncovered the temporal order of events leading to enhancer-promoter interactions; 2) I uncovered the temporal order of events that lead to chromatin folding of the Igλ locus; 3) I showed that antisense enhancer RNAs are repressive to enhancer-promoter interactions and chromatin organization.

6.1 Characterisation of an inducible system to investigate long-range enhancer-promoter contacts

The gene model used to investigate long-range enhancer-promoter interactions in this study is the murine Igλ locus. The non-coding transcription and V(D)J recombination that occur within this locus are tightly regulated by the Eλ3-1 enhancer (Haque et al., 2013). Previous publications from our lab showed that the Eλ3-1 enhancer contains binding motifs for diverse transcription factors, such as IRF4, PU.1 and E2A, and furthermore, equipping pro-B cells with a pre-B level of a single transcription factor, IRF4, is sufficient to activate non-coding transcription of the Igλ locus and all associated chromatin changes (Bevington and Boyes, 2013). To temporally dissect the critical events in the activation of the murine Igλ locus, I describe the characterization of an inducible pro-B cell line, 1D1-T215, in Chapter 3, which is capable of activating non-coding transcription of Igλ gene segments upon induction with 4-hydroxytamoxifen.

To generate immortalized pro-B cell lines, there are two well-established methods, including long-term pro-B cell growth in presence of IL-7 (Corfe et al., 2007) and infection of pro-B cells with A-MuLV (Rosenberg and Baltimore, 1976). These methods constitutively activate the signalling pathways that regulate pro-B cell proliferation and that are orchestrated by IL-7 signalling (Banerjee and Rothman, 1998; Clark et al., 2014; Shore et al., 2002). Previous data from our lab showed that immortalization of pro-B cells from transgenic

mice that express an IRF4-ER transgene was successfully achieved by long-term cell culture in the presence of IL-7 (A. Smith, PhD thesis, 2018). However, the expression of the IRF4-ER transgene, which is driven by a λ5/VpreB promoter was not detected, even though the IRF4-ER transgene is integrated into the B cell genome (A. Smith, PhD thesis, 2018). The inability to detect the expression of IRF4-ER may be caused by transgene silencing. For instance, transgene silencing can be attributed to a repressive chromatin environment caused by nearby genomic sequences (Matzke and Matzke, 1998). Infection of pro-B cells with A-MuLV is an alternative, effective strategy for immortalization. The v-Abl tyrosine kinase encoded by A-MuLV leads to the constitutive activation of JAK1/3 which in turn mediates phosphorylation of STAT5 that is activated by IL-7, ultimately resulting in pro-B cells that can proliferate in the absence of IL-7 (Danial et al., 1995). Pro-B cells were highly efficiently immortalized by A-MuLV, resulting in a number of pro-B cell lines (A. Smith, 2018, PhD thesis). These cell lines were then subject to analysis of the expression of endogenous IRF4 and PU.1. The cell line, 1D1, which displays a similar expression pattern of selected transcription factors to primary pro-B cells (A. Smith, PhD thesis, 2018), was used for further experiments. The IRF4-ER_{T2} transgene was introduced into 1D1 cells and integrated into the cell genome via retroviral transduction, as described in Chapter 3. To select cells with transgene expression, two types of selection markers were used: fluorescence-based (EGFP) and antibiotic-based (puromycin) markers. The advantage of using EGFP for cell selection is that the transgene expression can be monitored by checking the intensity of EGFP fluorescence. As cells may lose transgene expression during long-term cell culture due to epigenetic modifications, leading to transgene silencing (Jahner et al., 1982), flow cytometry can be used to re-select those cells that still express high levels of transgene. The advantage of using antibiotic selection is that the selection antibiotic (e.g. puromycin) can be added to culture media to kill those cells which lose transgene expression during long-term cell culture to thus generate a population with a high level of transgene expression. After three months of cell culture, there was no significant change in transgene expression in these two types of cell lines (data not shown). This means that both of these two types of

cell lines can be used for further experiments. The 1D1-IRF4-ER$_{T2}$ pro-B cell clone 15 (referred to as 1D1-T215) shows the most substantial increase of Igλ transcription following induction (Figure 3.7) and was therefore chosen for further analysis.

The major advantage of using the Igλ locus as a model to examine the activation of Igλ non-coding transcription is that this activation depends solely on the presence of pre-B cell levels of IRF4. This can be achieved in the 1D1-T215 cell line via induction with 4-hydroxytamoxifen. This implies that the transcriptome and genome organizational changes in 1D1-T215 cells after induction may be a representative of the events that occurred in wide-type pre-B cells. This is different from the inhibition of v-Abl in A-MuLV-transformed pro-B cells using the ABL tyrosine kinase inhibitor, STI-571, or the inactivation of a temperature-sensitive v-Abl mutant by temperature shift, as these lead to a number of non-physiological alterations in gene expression (Chen et al., 1994; Muljo and Schlissel, 2003). In addition, since the IRF4-ER activity is tightly regulated, the exact effects of IRF4 on the establishment of enhancer-promoter interactions and chromatin organization within the Igλ locus can be examined with a high degree of detail. This is facilitated by the rapid expansion of the cell line and low variability between 1D1-T215 cells. Whilst there are many advantages to using the 1D1-T215 cell line to investigate enhancer-promoter interactions and chromatin folding, there are also several caveats. Immortalization with A-MuLV, as used to generate the parent cell line of 1D1-T215 cells, represses Jλ1 non-coding transcription compared to primary pro-B cells. This may be mediated by direct binding of the transcription repressor, STAT5, to the Jλ1 promoter. Reduced expression of RAG proteins was also observed in 1D1-T215 cells; this is also likely caused by constitutively activated v-Abl signalling which prevents FOXO1 binding to the Erag enhancer (Amin and Schlissel, 2008; Biggs et al., 1999). Nonetheless, it seems possible that the V(D)J recombination could be achieved at the Igλ locus by mutating the STAT5 binding motifs located in the promoters of Igλ gene segments and by overexpressing exogenous RAGs using promoters that are insensitive to v-Abl signalling.

6.2 What do enhancers deliver to their promoters to activate transcription?

Efficient gene transcription depends on the establishment of tissue-specific enhancer-promoter contacts (Matharu and Ahituv, 2015; Schoenfelder and Fraser, 2019). Establishment of enhancer-promoter interactions raises two essential questions. Firstly, what do enhancers deliver to activate RNAPII machinery bound at their cognate promoters, and secondly, what is the structural basis for the formation of enhancer-promoter loops?

To investigate how the Eλ3-1 enhancer activates non-coding transcription of the Igλ locus, I began to examine what Eλ3-1 delivers to the promoters of target gene segments. To address this, I examined the basal RNAPII machinery bound at the target gene promoters in primary pro-B and pre-B cells. The elevated levels of Ser 5 phosphorylated RNAPII observed at the Jλ1 promoter in pre-B cells compared to pro-B cells indicates that phosphorylation of the Ser 5 residues of CTD of RNAPII pre-bound at Jλ1 promoter is an important part of the activation of this promoter (Figure 3.5). It has been already demonstrated that the Eλ3-1 enhancer recruits IRF4 during the differentiation of pro-B to pre-B cells and just increasing IRF4 levels is sufficient to trigger the activation of the Igλ locus in pro-B cells (Bevington and Boyes, 2013). Therefore, the first question regarding enhancer-promoter interactions can be transformed to how IRF4 activates the phosphorylation of Ser 5 residues of RNAPII that is pre-bound to the promoters of target gene segments of the Igλ locus.

The IRF4 inducible pro-B cell line, 1D1-T215, enables the temporal analysis of critical events during the activation of the Igλ locus to be performed to address this question. Published ChIP-seq data regarding transcription regulators provides clues about what the Eλ3-1 enhancer may deliver to target gene promoters. Those candidate transcription regulators can be divided into three groups: lineage-specific transcription factors, histone modifiers and general transcription activators. IRF4, E2A and PU.1 are lymphocyte-specific transcription factors that play essential roles in the regulation of lymphocyte proliferation and differentiation (Bain et al., 1994; Lu et al., 2003; Mittrucker et

al., 1997; Scott et al., 1994). Temporal ChIP analysis confirmed the IRF4 binding to the Eλ3-1 enhancer, and demonstrated that the increase of IRF4 binding to Eλ3-1 is an early event during the activation of the Igλ locus (Figures 3.8). p300 is an acetyltransferase that is responsible for the acetylation of H3K27, as well as other lysine targets (Ogryzko et al., 1996). Acetylated chromatin displays enhanced chromatin accessibility, allowing more transcription factors to be recruited to the regulatory regions (Li et al., 2007; Vo and Goodman, 2001; Watanabe et al., 2013). Specific interactions between lineage-specific transcription factors and histone modifiers may lead to histone modifications at *cis*-acting elements during cell development (Li et al., 2007). p300 has been demonstrated to interact with E2A directly (Qiu et al., 1998). Temporal ChIP analysis confirmed similar increases in the binding of these factors to Eλ3-1 during the activation of the Igλ locus (Figure 3.12), suggesting their co-recruitment. Mediator is a multiprotein complex that is vital for the regulation of gene transcription and is particularly important in mediating enhancer/promoter interactions (Allen and Taatjes, 2015). Whilst the head module of Mediator can directly activate the RNAPII machinery bound to gene promoters (Esnault et al., 2008; Robinson et al., 2012), the tail module interacts with lineage-specific transcription factors bound at enhancers (Ansari and Morse, 2012). The functionality of Mediator allows a link between enhancers and promoters to be established. Co-IP experiments confirmed such a direct interaction between IRF4 and Med23 (Figure 3.13A). Temporal ChIP analysis revealed further that Mediator recruitment at the Eλ3-1 is an early event during the activation of the Igλ locus (Figure 3.14). Notably, phosphorylation of the Ser 5 residues of CTD of RNAPII is catalyzed by CDK7 which has been shown to be recruited to gene promoters by Mediator (Esnault et al., 2008; Valay et al., 1995). These data therefore support a model of enhancer-mediated activation of target gene transcription: lineage-specific transcription factors IRF4, PU.1 and E2A are firstly recruited to the Eλ3-1 enhancer during B cell differentiation, Eλ3-1 bound lineage-specific transcription factors subsequently interact with the histone modifier p300 and the transcription activator Mediator to establish accessible chromatin structure and activate the RNAPII machinery bound at target gene promoters.

Next, to investigate the structural basis of enhancer-promoter loops, published ChIP-seq data regarding architectural factors including CTCF, cohesin and YY1 were analyzed. CTCF is an essential regulator of genome structure via binding to itself to form homodimers, which can cause the bound genomic DNA to form chromatin loops (Phillips and Corces, 2009). The cohesin complex plays an essential role in holding sister chromatids together during meiosis and mitosis (Klein et al., 1999; Nasmyth and Haering, 2009) and is capable of bringing two distant genomic fragments into close proximity (Nasmyth and Haering, 2009). Cohesin can directly interact with CTCF and usually colocalises with CTCF to form chromatin loops (Hansen et al., 2017; Merkenschlager and Odom, 2013; Wendt and Peters, 2009). However, published ChIP-seq data from pro-B cells show that there is no significant enrichment of CTCF and cohesin at the Eλ3-1 enhancer nor at the promoters of Igλ gene segments (Figure 4.2). This was further confirmed in 1D1-T215 cells by ChIP-qPCR both before and after induction (Figure 4.3). YY1 is another structural regulator that has been demonstrated to interact with CTCF (Donohoe et al., 2007; Schwalie et al., 2013) and cohesin (Pan et al., 2013) to establish chromatin loops. YY1 can also contribute to the establishment of chromatin loops independently, especially enhancer-promoter loops (Weintraub et al., 2017). Published YY1 ChIP-seq data from pro-B cells show that YY1 is enriched at the Eλ3-1 enhancer (Figure 4.5). Temporal ChIP analysis demonstrates that YY1 binding to Eλ3-1 is a late event during the activation of the Igλ locus (Figure 3.21). Furthermore, knock down of YY1 disrupts Eλ3-1 and Vλ1 interactions completely, accompanied by reduced Vλ1 non-coding transcription (Figure 3.20). This suggests that YY1 is indispensable for maintaining the interactions between Eλ3-1 and Vλ1. YY1 has been previously demonstrated to directly interact with Mediator (Luck et al., 2020) and p300 (Lee et al., 1995). However, Mediator and p300 binding to Eλ3-1 are early events during the activation of the Igλ locus (Figures 3.12 and 3.14), in contrast to the binding pattern of YY1 (Figure 3.20). This may suggest YY1 enrichment at Eλ3-1 is caused by other factors. Enhancer RNAs are a subclass of long non-coding RNAs that are involved in the regulation of enhancer-promoter interactions (Arnold et al., 2019). It has been demonstrated that YY1

contains RNA binding domains and its binding to enhancers relies on enhancer RNAs tethered at enhancer regions (Sigova et al., 2015). Temporal analysis of the expression of the enhancer RNAs encoded by Eλ3-1 shows that the level of enhancer RNAs start to increase from 4 hpi (Figure 5.3), just prior to the increase of YY1 binding to Eλ3-1. Similar to mRNAs, enhancer RNAs are also transcribed by the RNAPII machinery. However, the 3' end of enhancer RNA is processed by the Integrator complex instead of the cleavage and polyadenylation specificity factors, facilitating the maturation of enhancer RNAs and further release of enhancer RNAs from transcribing RNAPII (Lai et al., 2015). Integrator binding to Eλ3-1 reaches its highest level at 4 hpi (Figure 5.4), which is just before the increase of enhancer RNAs. Integrator can interact with Mediator directly and its binding to Eλ3-1 is an early event, which may suggest that Integrator is recruited by Mediator during the activation of the Igλ locus. These data support the following model by which the enhancer-promoter loops are established: the lineage-specific transcription factor IRF4 facilitates the recruitment of Integrator to the Eλ3-1 enhancer, leading to the synthesis of enhancer RNAs. Enhancer RNAs encoded by the Eλ3-1 enhancer in turn contribute to the recruitment of the structural regulator YY1 to enhancers and promoters to stabilize the long-range interactions.

6.3 Chromatin organization of the Igλ locus is triggered by IRF4

Activation of the Igλ locus requires the establishment of the correct chromatin environment which culminates in bringing enhancers and promoters into close proximity. To determine how the chromatin structure of the Igλ locus is activated by a single transcription factor, IRF4, I reanalyzed published ATAC-seq and ChIP-seq data regarding architecture factors from pro-B cells. This led to the identification of two genomic elements, HS7 and HSVλ1 that display high levels of CTCF and cohesin binding in pro-B cells (Figure 4.5). Further ChIP-qPCR data confirmed that CTCF and cohesin are co-bound to HS7 and HSVλ1 and that the level of binding of both factors to HS7 and HSVλ1 does not change from pro-B cells to pre-B cells (J. Scott, PhD thesis, 2016). This was confirmed by temporal ChIP analysis in 1D1-T215 cells (Figures 4.3 and 4.4). 3C experiments further reveal that HS7 can interact with HSVλ1 and the interaction

frequency between these two elements does change from pro-B to pre-B cells (J. Scott, 2016, PhD thesis). These data suggest that HS7 and HSVλ1 may already establish a chromatin domain that seals the 3' end of the Igλ locus in pro-B cells. According to the analysis of CTCF and cohesin ChIP-seq data from different tissues, I found that CTCF/cohesin mediated chromatin boundaries at HS7 and HSVλ1 are conserved across tissue types (Figure 4.19). These data reveal that the HS7-HSVλ1 chromatin loop may be formed early in development. Formation of the HS7-HSVλ1 chromatin loop results in the locus contraction, thus shortening the distance between the Eλ3-1 enhancer and unrearranged gene segments of Igλ. Similar chromatin loops are also formed in the 5' half of the duplicated Igλ locus (data not shown). The formation of separate 5' and 3' chromatin domains within the Igλ locus seems to provide an explanation for the V(D)J recombination rarely occurs between gene segments located in the 5' and 3' halves of the Igλ locus (Sanchez et al., 1991).

CTCF/cohesin mediated locus folding seems insufficient for the establishment of enhancer-promoter interactions within the Igλ locus. It has been demonstrated that equipping pro-B cells with a pre-B level of IRF4 can activate the Igλ locus completely (Bevington and Boyes, 2013). To investigate how IRF4 further facilitate the chromatin organization of the Igλ locus, published IRF4 ChIP-seq data from pro-B cells were reanalyzed. The results show that IRF4 is enriched at not only the Eλ3-1 enhancer but also an enhancer-like element, HS6 (Figure 4.5). Combined with IRF4 ChIP-qPCR data from 1D1-T215 cells, another IRF4 binding region was discovered within HSCλ1 (Figure 4.7). This raises the question of how IRF4 activates the Igλ locus via binding to Eλ3-1, HS6 and HSCλ1. E2A is essential for lymphocyte development and has been shown to increase the chromatin accessibility through interacting with the histone acetyltransferase, p300 (Qiu et al., 1998). Therefore physical interactions between E2A and IRF4 (Lazorchak et al., 2006) may facilitate p300 binding to IRF4 bound enhancers, leading to enhancers that are more accessible to other transcription regulators. Temporal ChIP analysis of p300 and E2A confirm that, similar to IRF4 binding, E2A and p300 binding to the Igλ locus are both early events during locus activation (Figures 3.11, 3.12 and 4.13).

Temporal 3C analysis of interactions between Eλ3-1 and HS6 as well as HSCλ1 indicates that Eλ3-1, HS6 and HSCλ1 may form an enhancer hub during the activation of the Igλ locus (Figure 4.9). Mediator has been shown to be enriched at the constituent enhancers of enhancer hubs (Whyte et al., 2013). Published ChIP-seq data show that Mediator is present at Eλ3-1, HS6 and HSCλ1 in pro-B cells (Figure 4.5), which was further confirmed by the ChIP-qPCR in 1D1-T215 cells (Figure 4.14). Mediator has also been demonstrated to be involved in the regulation of chromatin structure especially enhancer-promoter interactions (Allen and Taatjes, 2015). The correlation between Mediator binding to the Igλ locus and 3C interactions within the Igλ locus implies that the locus contraction during the early stage of Igλ activation may be caused by Mediator which brings enhancers and promoters into close proximity via increased binding to these regulatory elements.

As mentioned above, YY1 is essential for stabilizing the interactions between the Eλ3-1 enhancer and its cognate promoters at the late stages of activation of the Igλ locus. Published YY1 ChIP-seq data show that YY1 is present at HS6 and HSCλ1 (Figure 4.5). Temporal ChIP analysis reveals that YY1 binding to all three constituent enhancers of the Eλ3-1 enhancer hub starts to increase in 1D1-T215 cells from 8 hpi (Figures 3.21 and 4.16). Increased YY1 binding to enhancers can be attributed to the synthesis of enhancer RNAs (Lai et al., 2015). Published GRO-seq data from pro-B cells reveal that all three constituent enhancers of the Eλ3-1 enhancer hub likely encode enhancer RNAs (Figures 5.5 and 5.13). Consistent with this, increased YY1 binding to Eλ3-1, HS6 and HSCλ1 is observed just after the increase of Eλ3-1 enhancer RNA expression. These data suggest that YY1 binding to *cis*-acting elements within the Igλ locus at the late stage of Igλ locus activation may be responsible for maintaining the chromatin structure formed at the early stage. Together, these data therefore support a three-step model to explain the chromatin structure formed during the activation of the Igλ locus: Step 1: Formation of the CTCF/cohesin mediated chromatin loop between HS7 and HSVλ1; Step 2: IRF4 facilitates locus contraction through interacting with Mediator; Step 3: This chromatin structure is maintained by YY1/eRNAs.

This model of the Igλ locus organization implies that the unrearranged gene segments are recruited to the Eλ3-1 enhancer hub during the activation of the Igλ locus. This provides an explanation for the coordinate upregulation of Vλ1 and Jλ1 non-coding transcription. Previous publications showed that the majority of Igλ recombination events occur between the Vλ1 and Jλ1 gene segments (Boudinot et al., 1994). However, it is still unknown how the bias in recombination between Vλ1 and Jλ1 is achieved as Jλ3 is also recruited to the enhancer hub. 3C analysis of chromatin interactions using Vλ1 as the viewpoint was not performed and it is possible that the chromatin folding triggered by IRF4 leads to Vλ1 being in closer proximity to Jλ1, which may explain the increased recombination. Another explanation for the biased recombination may be caused by the difference in RSS structure at Jλ1 and Jλ3 promoters. The recombination efficiency of RSSs can be evaluated by the recombination information content (RIC) score (Cowell et al., 2002). The RSS at the Jλ1 promoter is scored as functional (pass) with RIC > -19.03, while RSS at Jλ3 promoter pass with RIC > -19.55, suggesting RSS at the Jλ1 promoter is marginally more likely to be recognized by RAG machinery.

6.4 Interplay between sense and anti-sense enhancer RNAs and transcription factor trapping

Enhancer RNAs are a subclass of nuclear-localized long non-coding RNAs which are synthesized by tissue-specific enhancers during cell development (Arnold et al., 2019). Whilst an increasing number of publications demonstrate that enhancer RNAs are functional biomolecules, the precise mechanism of how enhancer RNAs are involved in transcriptional control remains enigmatic. Enhancer RNAs are generally believed to function via direct interactions with diverse transcription factors to contribute to the transcriptional control. These enhancer RNA binding partners include transcription activators, transcription repressors, histone modifiers and architectural factors (Arnold et al., 2019; Guttman and Rinn, 2012; Lam et al., 2014). To investigate the roles of the enhancer RNAs encoded by Eλ3-1 in the regulation of non-coding transcription of the Igλ locus, temporal analysis of the expression of enhancer RNAs was

performed. This showed that the expression of Eλ3-1 enhancer RNAs is a relatively late event (Figures 5.3 and 5.6). Enhancer RNAs have been shown to bind to the acetyltransferase p300 directly to stimulate histone acetylation, leading to the activation of gene transcription (Bose et al., 2017). Consistent with this, Vλ1 non-coding transcription was diminished in Eλ3-1 sense enhancer RNA knock down 1D1-T215 cells (Figure 5.8). However, p300 binding to Eλ3-1 is an early event which does not correlate with the enhancer RNA expression. This suggests that interactions between enhancer RNAs and p300 may not be the main driver for target gene activation. Moreover, enhancer RNAs can facilitate the binding of transcription activators to target genes to activate transcription. The Mediator complex contains multiple subunits that are capable of binding enhancer RNAs, such as Med1 and Med12 (Lai et al., 2013). Depletion of enhancer RNAs results in a reduced level of Mediator binding to target genes, accompanied by a disrupted transcription of target genes (Lai et al., 2013). This may also explain the decrease in Vλ1 non-coding transcription in Eλ3-1 sense RNA knock down 1D1-T215 cells (Figure 5.8). Mediator has also been shown to be involved in bringing enhancers and promoters into close proximity (Malik and Roeder, 2016). Consistent with this, depletion of Med23 disrupts the interactions between the Eλ3-1 enhancer and other *cis*-acting elements within the Igλ locus (Figure 4.15). However, the chromatin contacts between Eλ3-1 and other regulatory elements are also diminished in the Eλ3-1 sense enhancer RNA knock down cells (Figure 5.10). It is unclear whether these disrupted chromatin interactions within the Igλ locus in Eλ3-1 sense enhancer RNA knock down cells are mediated by Mediator. To verify this, ChIP analysis of Mediator binding to the Igλ locus needs to be performed in the Eλ3-1 enhancer RNA knock down 1D1-T215 cells. Enhancer RNAs have been demonstrated to also recruit architectural factors to enhancers and promoters to facilitate the establishment of chromatin interactions. Published ChIP-seq analysis shows that the structural factor YY1 is present at the Igλ locus (Figure 4.5). Similar to the Eλ3-1 sense enhancer RNA, depletion of YY1 also disrupts the chromatin interactions formed within the Igλ locus (Figure 4.11). The close correlation found between YY1 binding to the Igλ locus and expression of the Eλ3-1 enhancer RNAs strongly implies that increased YY1 binding is caused

by the expression of enhancer RNAs. Consistent with this, the level of YY1 binding to Eλ3-1 decreases in the Eλ3-1 sense enhancer RNA knock-down cells (Figure 5. 12).

As mentioned previously, enhancer RNAs encoded by the Eλ3-1 are bidirectional. To explore if the anti-sense enhancer RNA plays a similar role in the regulation of the target gene transcription, knock down of the anti-sense enhancer RNA was conducted. Intriguingly, target gene transcription is increased in the Eλ3-1 antisense enhancer RNA knock down 1D1-T215 cells (Figure 5.8). This implies that the antisense enhancer RNA is repressive to target gene transcription. Consistent with this, chromatin interactions within the Igλ locus are increased in the Eλ3-1 antisense enhancer RNA knock down cells (Figure 5.11). This raises the question of how antisense enhancer RNAs repress target gene transcription and enhancer-promoter interactions. Previous publications show that RNAs are capable of recruiting transcription repressors. For instance, the long non-coding RNA RepA and HOTAIR can interact with the polycomb repressive complex directly to establish a chromatin structure that is repressive to transcription (Rinn et al., 2007; Yuan et al., 2012; Zhao et al., 2008). In addition, anti-sense enhancer RNAs may hybridise with sense enhancer RNAs. Computation analysis of RNA-RNA interactions showed that the Eλ3-1 sense enhancer RNA can hybridise with the anti-sense enhancer RNA (data not shown). Previous publications demonstrated that sense enhancer RNAs can physically interact with YY1 and knock down of sense enhancer RNAs leads to reduced binding of YY1 at enhancers (Sigova, A., et al, 2015), implying that the sense enhancer RNA may form unique secondary or tertiary structures recognized by YY1. This is consistent with the data shown in Figure 5.12. However, knock-down of the Eλ3-1 anti-sense enhancer RNA does not affect YY1 binding to Eλ3-1. This indicates that the Eλ3-1 anti-sense enhancer RNA may repress target gene transcription in a YY1 independent manner. Together, these data support the following model of how the anti-sense enhancer RNAs repress target gene transcription: The anti-sense enhancer RNA directly interacts with transcription repressors to inhibit target

gene transcription; it can also indirectly repress gene transcription through hybridizing with the sense enhancer RNA.

Conclusions and further directions

In this thesis, I describe the characterization of an inducible system to determine the temporal order of events of the activation of enhancer-promoter interactions and chromatin organization. Based on data generated in this thesis, I present a model by which the enhancer-mediated activation of gene transcription and chromatin folding: the initiation of the activation of the Igλ locus is achieved by increased binding of a single transcription factor IRF4 at three enhancer-like elements, Eλ3-1, HS6 and HSCλ1; increased binding of IRF4 facilitates the recruitment of E2A, p300, Mediator and Integrator to enhancers and promoters at the early stage of activation of the Igλ locus, accompanied by an increased chromatin interaction frequency between the *cis*-acting regulatory elements; enhancer RNAs and YY1 binding are increased to stabilize the chromatin structure of Igλ at the late stage of the activation of the Igλ locus (Figure 4.21).

Whilst the inducible pro-B cell line 1D1-T215 led to key discoveries regarding enhancer-promoter interactions and chromatin organization of the Igλ locus, genetic manipulation of STAT5 binding sites within promoters of Igλ gene segments and overexpression of RAG proteins need to be performed to determine the V(D)J recombination. Moreover, a more sensitive technique, such as Capture-C (Davies et al., 2016), should be used to detect the chromatin interactions within the Igλ locus. Because only ~5% of recombinations of mouse light chain antigen receptor genes occurs at the Igλ locus, combined with the low amount of amplifiable ligation products recovered by 3C, the assay has poor sensitivity. In addition, the interplay between sense enhancer RNA and anti-sense enhancer RNA needs to be investigated. For instance, investigation if the protein binding partners differ between the sense and antisense enhancer RNA; prediction and characterization of functional domains formed by the sense and anti-sense enhancer RNAs that may alter their binding to essential transcription regulators or their hybridization to each other.

References

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

Abarrategui, I., and Krangel, M.S. (2009). Germline transcription: a key regulator of accessibility and recombination. Adv Exp Med Biol *650*, 93-102.

Acker, J., de Graaff, M., Cheynel, I., Khazak, V., Kedinger, C., and Vigneron, M. (1997). Interactions between the human RNA polymerase II subunits. J Biol Chem *272*, 16815-16821.

Acquaviva, J., Chen, X., and Ren, R. (2008). IRF-4 functions as a tumor suppressor in early B-cell development. Blood *112*, 3798-3806.

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet *13*, 720-731.

Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. Mol Cell 34, 387-393.

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.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat Rev Mol Cell Biol *16*, 155-166.

Allison, L.A., Moyle, M., Shales, M., and Ingles, C.J. (1985). Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell *42*, 599-610.

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.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T.*, et al.* (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455-461.

Ansari, S.A., and Morse, R.H. (2012). Selective role of Mediator tail module in the transcription of highly regulated genes in yeast. Transcription *3*, 110-114.

Ardehali, M.B., Mei, A., Zobeck, K.L., Caron, M., Lis, J.T., and Kusch, T. (2011). Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. EMBO J *30*, 2817-2828.

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

Arnold, P.R., Wells, A.D., and Li, X.C. (2019). Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. Front Cell Dev Biol *7*, 377.

Atchison, M.L. (2014). Function of YY1 in Long-Distance DNA Interactions. Front Immunol *5*, 45.

Baillat, D., and Wagner, E.J. (2015). Integrator: surprisingly diverse functions in gene expression. Trends Biochem Sci 40, 257-264.

Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell 123, 265-276.

Bain, G., Maandag, E.C., Izon, D.J., Amsen, D., Kruisbeek, A.M., Weintraub, B.C., Krop, I., Schlissel, M.S., Feeney, A.J., van Roon, M., et al. (1994). E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. Cell 79, 885-892.

Bajpai, R., Chen, D.A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C.P., Zhao, Y., Swigut, T., and Wysocka, J. (2010). CHD7 cooperates with PBAF to control multipotent neural crest formation. Nature 463, 958-962.

Balvay, L., Soto Rifo, R., Ricci, E.P., Decimo, D., and Ohlmann, T. (2009). Structural and functional diversity of viral IRESes. Biochim Biophys Acta 1789, 542-557.

Barolo, S. (2012). Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. Bioessays 34, 135-141.

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

Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell 33, 729-740.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell 27, 299-308.

Bankovich, A.J., Raunser, S., Juo, Z.S., Walz, T., Davis, M.M., and Garcia, K.C. (2007). Structural insight into pre-B cell receptor function. Science 316, 291-294.

Bannister, A.J., and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. Nature 384, 641-643.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res 21, 381-395.

Barbieri, E., Trizzino, M., Welsh, S.A., Owens, T.A., Calabretta, B., Carroll, M., Sarma, K., and Gardini, A. (2018). Targeted Enhancer Activation by a Subunit of the Integrator Complex. Mol Cell 71, 103-116 e107.

Batista, C.R., Li, S.K., Xu, L.S., Solomon, L.A., and DeKoter, R.P. (2017). PU.1 Regulates Ig Light Chain Transcription and Rearrangement in Pre-B Cells during B Cell Development. J Immunol 198, 1565-1574.

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.

Bernecky, C., Plitzko, J.M., and Cramer, P. (2017). Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. Nat Struct Mol Biol 24, 809-815.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 120, 169-181.

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.

Bettridge, J., Na, C.H., Pandey, A., and Desiderio, S. (2017). H3K4me3 induces allosteric conformational changes in the DNA-binding and catalytic regions of the V(D)J recombinase. Proc Natl Acad Sci U S A 114, 1904-1909.

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

Biggs, W.H., 3rd, Meisenhelder, J., Hunter, T., Cavenee, W.K., and Arden, K.C. (1999). Protein kinase B/Akt-mediated phosphorylation promotes nuclear exclusion of the winged helix transcription factor FKHR1. Proc Natl Acad Sci U S A 96, 7421-7426.

Black, B.L., and Olson, E.N. (1998). Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. Annu Rev Cell Dev Biol 14, 167-196.

Bonelt, P., Wohner, M., Minnich, M., Tagoh, H., Fischer, M., Jaritz, M., Kavirayani, A., Garimella, M., Karlsson, M.C., and Busslinger, M. (2019). Precocious expression of Blimp1 in B cells causes autoimmune disease with increased self-reactive plasma cells. EMBO J 38, 1-19.

Bonifer, C., Vidal, M., Grosveld, F., and Sippel, A.E. (1990). Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. EMBO J 9, 2843-2848.

Bose, D.A., Donahue, G., Reinberg, D., Shiekhattar, R., Bonasio, R., and Berger, S.L. (2017). RNA Binding to CBP Stimulates Histone Acetylation and Transcription. Cell 168, 135-149 e122.

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.

Boyd, K.E., and Farnham, P.J. (1999). Coexamination of site-specific transcription factor binding and promoter activity in living cells. Mol Cell Biol 19, 8393-8399.

Bradney, C., Hjelmeland, M., Komatsu, Y., Yoshida, M., Yao, T.P., and Zhuang, Y. (2003). Regulation of E2A activities by histone acetyltransferases in B lymphocyte development. J Biol Chem 278, 2370-2376.

Brandle, D., Muller, C., Rulicke, T., Hengartner, H., and Pircher, H. (1992). Engagement of the T-cell receptor during positive selection in the thymus down-regulates RAG-1 expression. Proc Natl Acad Sci U S A 89, 9529-9533.

Breen, J.J., Agulnick, A.D., Westphal, H., and Dawid, I.B. (1998). Interactions between LIM domains and the LIM domain-binding protein Ldb1. J Biol Chem 273, 4712-4717.

Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. RNA 19, 429-442.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol *109*, 21.29.1-21.29.9.

Bulger, M., van Doorninck, J.H., Saitoh, N., Telling, A., Farrell, C., Bender, M.A., Felsenfeld, G., Axel, R., and Groudine, M. (1999). Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. Proc Natl Acad Sci U S A 96, 5129-5134.

Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. Cell 56, 549-561.

Cai, Y., Jin, J., Yao, T., Gottschalk, A.J., Swanson, S.K., Wu, S., Shi, Y., Washburn, M.P., Florens, L., Conaway, R.C., et al. (2007). YY1 functions with INO80 to activate transcription. Nat Struct Mol Biol 14, 872-874.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol Cell 49, 825-837.

Carpenter, S., Aiello, D., Atianand, M.K., Ricci, E.P., Gandhi, P., Hall, L.L., Byron, M., Monks, B., Henry-Bezy, M., Lawrence, J.B., et al. (2013). A long noncoding RNA mediates both activation and repression of immune response genes. Science 341, 789-792.

Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.F., and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. Nat Genet 32, 623-626.

Cavalli, G., and Misteli, T. (2013). Functional implications of genome topology. Nat Struct Mol Biol 20, 290-299.

Chen, J., Ezzeddine, N., Waltenspiel, B., Albrecht, T.R., Warren, W.D., Marzluff, W.F., and Wagner, E.J. (2012). An RNAi screen identifies additional members of the Drosophila Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation. RNA 18, 2148-2156.

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, X., Zhao, J., Gu, C., Cui, Y., Dai, Y., Song, G., Liu, H., Shen, H., Liu, Y., Wang, Y., et al. (2018). Med23 serves as a gatekeeper of the myeloid potential of hematopoietic stem cells. Nat Commun 9, 3746.

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.

Chereji, R.V., Bharatula, V., Elfving, N., Blomberg, J., Larsson, M., Morozov, A.V., Broach, J.R., and Bjorklund, S. (2017). Mediator binds to boundaries of chromosomal interaction domains and to proteins involved in DNA looping, RNA metabolism, chromatin remodeling, and actin assembly. Nucleic Acids Res 45, 8806-8821.

Choukrallah, M.A., Song, S., Rolink, A.G., Burger, L., and Matthias, P. (2015). Enhancer repertoires are reshaped independently of early priming and heterochromatin dynamics during B cell differentiation. Nat Commun 6, 8324.

Chowdhury, D., and Sen, R. (2001). Stepwise activation of the immunoglobulin mu heavy chain gene locus. EMBO J 20, 6394-6403.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell 9, 279-289.

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.

Clayton, A.L., Hazzalin, C.A., and Mahadevan, L.C. (2006). Enhanced histone acetylation and transcription: a dynamic perspective. Mol Cell 23, 289-296.

Cocea, L., De Smet, A., Saghatchian, M., Fillatreau, S., Ferradini, L., Schurmans, S., Weill, J.C., and Reynaud, C.A. (1999). A targeted deletion of a region upstream from the Jkappa cluster impairs kappa chain rearrangement in cis in mice and in the 103/bcl2 cell line. J Exp Med 189, 1443-1450.

Cogne, 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.

Corden, J.L., Cadena, D.L., Ahearn, J.M., Jr., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc Natl Acad Sci U S A 82, 7934-7938.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322, 1845-1848.

Ong, C.T., and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet 15, 234-246.

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.

Cowell, L.G., Davila, M., Kepler, T.B., and Kelsoe, G. (2002). Identification and utilization of arbitrary correlations in models of recombination signal sequences. Genome Biol *3*, research0072.1–0072.20.

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.

Curry, J.D., Geier, J.K., and Schlissel, M.S. (2005). Single-strand recombination signal sequence nicks in vivo: evidence for a capture model of synapsis. Nature Immunol 6, 1272-1279.

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

Danial, N.N., and Rothman, P. (2000). JAK-STAT signaling activated by Abl oncogenes. Oncogene *19*, 2523-2531.

Davies, J.O., 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 Laat, W., and Grosveld, F. (2003). Spatial organization of gene expression: the active chromatin hub. Chromosome Res 11, 447-459.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol 8, e1000384.

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

Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A., et al. (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. Cell 158, 849-860.

Degner, S.C., Wong, T.P., Jankevicius, G., and Feeney, A.J. (2009). Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. J Immunol *182*, 44-48.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell *149*, 1233-1244.

Di Croce, L., and Helin, K. (2013). Transcriptional regulation by Polycomb group proteins. Nat Struct Mol Biol 20, 1147-1155.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376-380.

Donohoe, M.E., Zhang, L.F., Xu, N., Shi, Y., and Lee, J.T. (2007). Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. Mol Cell *25*, 43-56.

Dou, Y., Milne, T.A., Ruthenburg, A.J., Lee, S., Lee, J.W., Verdine, G.L., Allis, C.D., and Roeder, R.G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat Struct Mol Biol *13*, 713-719.

Douziech, M., Coin, F., Chipoulet, J.M., Arai, Y., Ohkuma, Y., Egly, J.M., and Coulombe, B. (2000). Mechanism of promoter melting by the xeroderma pigmentosum complementation group B helicase of transcription factor IIH revealed by protein-DNA photo-cross-linking. Mol Cell Biol *20*, 8168-8177.

Dynlacht, B.D., Hoey, T., and Tjian, R. (1991). Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell *66*, 563-576.

Ebert, A., McManus, S., Tagoh, H., Medvedovic, J., Salvagiotto, G., Novatchkova, M., Tamir, I., Sommer, A., Jaritz, M., and Busslinger, M. (2011). The distal V(H) gene cluster of the Igh locus contains distinct regulatory elements with Pax5 transcription factor-dependent activity in pro-B cells. Immunity *34*, 175-187.

Ebmeier, C.C., Erickson, B., Allen, B.L., Allen, M.A., Kim, H., Fong, N., Jacobsen, J.R., Liang, K., Shilatifard, A., Dowell, R.D.*, et al.* (2017). Human TFIIH Kinase CDK7 Regulates Transcription-Associated Chromatin Modifications. Cell Rep *20*, 1173-1186.

Ebright, R.H., and Busby, S. (1995). The Escherichia coli RNA polymerase alpha subunit: structure and function. Curr Opin Genet Dev 5, 197-203.

Eckner, R., Yao, T.P., Oldread, E., and Livingston, D.M. (1996). Interaction and functional collaboration of p300/CBP and bHLH proteins in muscle and B-cell differentiation. Genes Dev 10, 2478-2490.

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.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A.*, et al.* (2019). The Pfam protein families database in 2019. Nucleic Acids Res *47*, D427-D432.

Elkin, S.K., Ivanov, D., Ewalt, M., Ferguson, C.G., Hyberts, S.G., Sun, Z.Y., Prestwich, G.D., Yuan, J., Wagner, G., Oettinger, M.A.*, et al.* (2005). A PHD

finger motif in the C terminus of RAG2 modulates recombination activity. J Biol Chem *280*, 28701-28710.

Eng, F.C., 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. Mol Cell Biol *17*, 4644-4653.

Engel, H., Ruhl, H., Benham, C.J., Bode, J., and Weiss, S. (2001). Germ-line transcripts of the immunoglobulin lambda J-C clusters in the mouse: characterization of the initiation sites and regulatory elements. Mol Immunol *38*, 289-302.

Escalante, C.R., Brass, A.L., Pongubala, J.M., Shatova, E., Shen, L., Singh, H., and Aggarwal, A.K. (2002). Crystal structure of PU.1/IRF-4/DNA ternary complex. Mol Cell *10*, 1097-1105.

Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., Holstege, F., and Werner, M. (2008). Mediator-dependent recruitment of TFIIH modules in preinitiation complex. Mol Cell *31*, 337-346.

Essien, K., Vigneau, S., Apreleva, S., Singh, L.N., Bartolomei, M.S., and Hannenhalli, S. (2009). CTCF binding site classes exhibit distinct evolutionary, genomic, epigenomic and transcriptomic features. Genome Biol *10*, R131.

Farley, E., and Levine, M. (2012). HOT DNAs: a novel class of developmental enhancers. Genes Dev 26, 873-876.

Featherstone, K., Wood, A.L., Bowen, A.J., and Corcoran, A.E. (2010). The mouse immunoglobulin heavy chain V-D intergenic sequence contains insulators that may regulate ordered V(D)J recombination. J Biol Chem *285*, 9327-9338.

Feaver, W.J., Svejstrup, J.Q., Henry, N.L., and Kornberg, R.D. (1994). Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell *79*, 1103-1109.

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.

Fernex, C., Capone, M., and Ferrier, P. (1995). The V(D)J recombinational and transcriptional activities of the immunoglobulin heavy-chain intronic enhancer can be mediated through distinct protein-binding sites in a transgenic substrate. Mol Cell Biol 15, 3217-3226.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and Lobanenkov, V.V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol 16, 2802-2813.

Finch, J.T., Lutter, L.C., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M., and Klug, A. (1977). Structure of nucleosome core particles of chromatin. Nature *269*, 29-36.

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.

Forget, D., Robert, F., Grondin, G., Burton, Z.F., Greenblatt, J., and Coulombe, B. (1997). RAP74 induces promoter contacts by RNA polymerase II upstream and downstream of a DNA bend centered on the TATA box. Proc Natl Acad Sci U S A *94*, 7150-7155.

Foxler, D.E., James, V., Shelton, S.J., Vallim, T.Q., Shaw, P.E., and Sharp, T.V. (2011). PU.1 is a major transcriptional activator of the tumour suppressor gene LIMD1. FEBS Lett *585*, 1089-1096.

Fujinaga, K., Irwin, D., Huang, Y., Taube, R., Kurosu, T., and Peterlin, B.M. (2004). Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. Mol Cell Biol *24*, 787-795.

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.

Fraser, P., and Grosveld, F. (1998). Locus control regions, chromatin activation and transcription. Curr Opin Cell Biol 10, 361-365.

Galvin, K.M., and Shi, Y. (1997). Multiple mechanisms of transcriptional repression by YY1. Mol Cell Biol *17*, 3723-3732.

Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J.M., Wagner, E.J., Lazar, M.A., Shilatifard, A., and Shiekhattar, R. (2014). Integrator regulates transcriptional initiation and pause release following activation. Mol Cell *56*, 128-139.

Geier, J.K., and Schlissel, M.S. (2006). Pre-BCR signals and the control of Ig gene rearrangements. Semin Immunol *18*, 31-39.

Geng, Y., Yu, Q., Sicinska, E., Das, M., Schneider, J.E., Bhattacharya, S., Rideout, W.M., Bronson, R.T., Gardner, H., and Sicinski, P. (2003). Cyclin E ablation in the mouse. Cell *114*, 431-443.

Gerdes, T., and Wabl, M. (2002). Physical map of the mouse lambda light chain and related loci. Immunogenetics *54*, 62-65.

Giles, K.E., Gowher, H., Ghirlando, R., Jin, C., and Felsenfeld, G. (2010). Chromatin boundaries, insulators, and long-range interactions in the nucleus. Cold Spring Harb Symp Quant Biol 75, 79-85.

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.

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., Ji, Y., Skok, J., Roldan, E., Jung, S., Cedar, H., and Bergman, Y. (2005). Epigenetic ontogeny of the Igk locus during B cell development. Nat Immunol *6*, 198-203.

Gorman, J.R., and Alt, F.W. (1998). Regulation of immunoglobulin light chain isotype expression. Adv Immunol *69*, 113-181.

Gottschling, D.E. (2000). Gene silencing: two faces of SIR2. Curr Biol 10, R708-711.

Grana, X., De Luca, A., Sang, N., Fu, Y., Claudio, P.P., Rosenblatt, J., Morgan, D.O., and Giordano, A. (1994). PITALRE, a nuclear CDC2-related protein kinase that phosphorylates the retinoblastoma protein in vitro. Proc Natl Acad Sci U S A *91*, 3834-3838.

Grawunder, U., Leu, T.M., 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.

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.

Grimley, P.M., Dong, F., and Rui, H. (1999). Stat5a and Stat5b: fraternal twins of signal transduction and transcriptional activation. Cytokine Growth Factor Rev *10*, 131-157.

Gruber, T.M., and Gross, C.A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol 57, 441-466.

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

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

Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. Nature *482*, 339-346.

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.

Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome Res *23*, 1210-1223.

Hampsey, M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol Mol Biol Rev *62*, 465-503.

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., Bevington, S.L., and Boyes, J. (2013). The Elambda(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.

Hardison, R., Slightom, J.L., Gumucio, D.L., Goodman, M., Stojanovic, N., and Miller, W. (1997). Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. Gene 205, 73-94.

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.

Harrison, M.M., Li, X.Y., Kaplan, T., Botchan, M.R., and Eisen, M.B. (2011). Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. PLoS Genet *7*, e1002266.

Hartzog, G.A., Wada, T., Handa, H., and Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. Genes Dev 12, 357-369.

Hawley, D.K., and McClure, W.R. (1983). Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Res 11, 2237-2255.

He, Y., Fang, J., Taatjes, D.J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. Nature 495, 481-486.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39, 311-318.

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.

Herz, H.M., Mohan, M., Garruss, A.S., Liang, K., Takahashi, Y.H., Mickey, K., Voets, O., Verrijzer, C.P., and Shilatifard, A. (2012). Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev 26, 2604-2620.

Hill, L., Ebert, A., Jaritz, M., Wutz, G., Nagasaka, K., Tagoh, H., Kostanova-Poliakova, D., Schindler, K., Sun, Q., Bonelt, P.*, et al.* (2020). Wapl repression by Pax5 promotes V gene recombination by Igh loop extrusion. Nature *584*, 142-147.

Hizume, K., Araki, S., Hata, K., Prieto, E., Kundu, T.K., Yoshikawa, K., and Takeyasu, K. (2010). Nano-scale analyses of the chromatin decompaction induced by histone acetylation. Arch Histol Cytol *73*, 149-163.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, 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.

Hnisz, D., Day, D.S., and Young, R.A. (2016a). Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. Cell 167, 1188-1200.

Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., et al. (2016b). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science 351, 1454-1458.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13-23.

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 kappa E3' enhancer. J Immunol *188*, 2276-2284.

Hong, J.W., Hendrix, D.A., and Levine, M.S. (2008). Shadow enhancers as a source of evolutionary novelty. Science 321, 1314.

Hsu, A.P., Johnson, K.D., Falcone, E.L., Sanalkumar, R., Sanchez, L., Hickstein, D.D., Cuellar-Rodriguez, J., Lemieux, J.E., Zerbe, C.S., Bresnick, E.H.*, et al.* (2013). GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. Blood *121*, 3830-3837, S3831-3837.

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.

Huang, Z., Du, G., Huang, X., Han, L., Han, X., Xu, B., Zhang, Y., Yu, M., Qin, Y., Xia, Y.*, et al.* (2018). The enhancer RNA lnc-SLC4A1-1 epigenetically regulates unexplained recurrent pregnancy loss (URPL) by activating CXCL8 and NF-kB pathway. EBioMedicine *38*, 162-170.

Hunt, T.P., and Magasanik, B. (1985). Transcription of glnA by purified Escherichia coli components: core RNA polymerase and the products of glnF, glnG, and glnL. Proc Natl Acad Sci U S A 82, 8453-8457.

Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. Exp Mol Med *49*, e324.

Igarashi, K., Fujita, N., and Ishihama, A. (1991). Identification of a subunit assembly domain in the alpha subunit of Escherichia coli RNA polymerase. J Mol Biol 218, 1-6.

Imamura, Y., Oda, A., Katahira, T., Bundo, K., Pike, K.A., Ratcliffe, M.J., 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 kappa light chain intronic enhancer and 3' enhancer in kappa 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.

Isel, C., and Karn, J. (1999). Direct evidence that HIV-1 Tat stimulates RNA polymerase II carboxyl-terminal domain hyperphosphorylation during transcriptional elongation. J Mol Biol *290*, 929-941.

Ishihama, A. (1981). Subunit of assembly of Escherichia coli RNA polymerase. Adv Biophys *14*, 1-35.

Ishihama, A., Fujita, N., and Glass, R.E. (1987). Subunit assembly and metabolic stability of E. coli RNA polymerase. Proteins *2*, 42-53.

Ishihama, A. (1992). Role of the RNA polymerase alpha subunit in transcription activation. Mol Microbiol 6, 3283-3288.

Ivanov, D., Kwak, Y.T., Guo, J., and Gaynor, R.B. (2000). Domains in the SPT5 protein that modulate its transcriptional regulatory properties. Mol Cell Biol *20*, 2970-2983.

Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science 288, 1422-1425.

Jahner, D., Stuhlmann, H., Stewart, C.L., Harbers, K., Lohler, J., Simon, I., and Jaenisch, R. (1982). De novo methylation and expression of retroviral genomes during mouse embryogenesis. Nature *298*, 623-628.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jhunjhunwala, S., van Zelm, M.C., Peak, M.M., Cutchin, S., Riblet, R., van Dongen, J.J., Grosveld, F.G., Knoch, T.A., and Murre, C. (2008). The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. Cell *133*, 265-279.

Jhunjhunwala, S., van Zelm, M.C., Peak, M.M., and Murre, C. (2009). Chromatin architecture and the generation of antigen receptor diversity. Cell *138*, 435-448.

Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R., and Schatz, D.G. (2010). The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. Cell *141*, 419-431.

Jiang, T., Raviram, R., Snetkova, V., Rocha, P.P., Proudhon, C., Badri, S., Bonneau, R., Skok, J.A., and Kluger, Y. (2016). Identification of multi-loci hubs

from 4C-seq demonstrates the functional importance of simultaneous interactions. Nucleic Acids Res *44*, 8714-8725.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503, 290-294.

Jin, J., Cai, Y., Li, B., Conaway, R.C., Workman, J.L., Conaway, J.W., and Kusch, T. (2005). In and out: histone variant exchange in chromatin. Trends Biochem Sci *30*, 680-687.

Johnson, K., Hashimshony, T., Sawai, C.M., Pongubala, J.M., 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.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 167-177.

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.

Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C., et al. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol Cell 51, 310-325.

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, A., and Dean, A. (2012). Chromatin loop formation in the beta-globin locus and its role in globin gene transcription. Mol Cells *34*, 1-5.

Kim, T.K., Ebright, R.H., and Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIH. Science *288*, 1418-1422.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S.*, et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

Kim, T.K., Lagrange, T., Wang, Y.H., Griffith, J.D., Reinberg, D., and Ebright, R.H. (1997). Trajectory of DNA in the RNA polymerase II transcription preinitiation complex. Proc Natl Acad Sci U S A *94*, 12268-12273.

Kim, T.K., and Shiekhattar, R. (2015). Architectural and Functional Commonalities between Enhancers and Promoters. Cell 162, 948-959.

Kimura, M., Ishiguro, A., and Ishihama, A. (1997). RNA polymerase II subunits 2, 3, and 11 form a core subassembly with DNA binding activity. J Biol Chem *272*, 25851-25855.

Kimura, M., and Ishihama, A. (1995). Functional map of the alpha subunit of Escherichia coli RNA polymerase: amino acid substitution within the amino-terminal assembly domain. J Mol Biol 254, 342-349.

Kimura, M., and Ishihama, A. (1996). Subunit assembly in vivo of Escherichia coli RNA polymerase: role of the amino-terminal assembly domain of alpha subunit. Genes Cells 1, 517-528.

Kleiman, E., Jia, H., Loguercio, S., Su, A.I., and Feeney, A.J. (2016). YY1 plays an essential role at all stages of B-cell differentiation. Proc Natl Acad Sci U S A *113*, E3911-3920.

Klein, F., Mahr, P., Galova, M., Buonomo, S.B., Michaelis, C., Nairz, K., and Nasmyth, K. (1999). A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. Cell *98*, 91-103.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat Rev Genet *20*, 207-220.

Klenova, E.M., Nicolas, R.H., Paterson, H.F., Carne, A.F., Heath, C.M., Goodwin, G.H., Neiman, P.E., and Lobanenkov, V.V. (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. Mol Cell Biol 13, 7612-7624.

Klose, R.J., Kallin, E.M., and Zhang, Y. (2006). JmjC-domain-containing proteins and histone demethylation. Nat Rev Genet 7, 715-727.

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C.*, et al.* (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nat Struct Mol Biol *18*, 956-963.

Kolodziej, P.A., and Young, R.A. (1991). Mutations in the three largest subunits of yeast RNA polymerase II that affect enzyme assembly. Mol Cell Biol *11*, 4669-4678.

Kondo, M., Weissman, I.L., and Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell *91*, 661-672.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.

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 *296*, 158-162.

Krajewski, W.A., and Becker, P.B. (1998). Reconstitution of hyperacetylated, DNase I-sensitive chromatin characterized by high conformational flexibility of nucleosomal DNA. Proc Natl Acad Sci U S A *95*, 1540-1545.

Krangel, M.S. (2003). Gene segment selection in V(D)J recombination: accessibility and beyond. Nat Immunol *4*, 624-630.

Krangel, M.S. (2009). Mechanics of T cell receptor gene rearrangement. Curr Opin Immunol *21*, 133-139.

Krijger, P.H., Di Stefano, B., de Wit, E., Limone, F., van Oevelen, C., de Laat, W., and Graf, T. (2016). Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming. Cell Stem Cell *18*, 597-610.

Kumari, G., and Sen, R. (2015). Chromatin Interactions in the Control of Immunoglobulin Heavy Chain Gene Assembly. Adv Immunol *128*, 41-92.

Kung, J.T., Colognori, D., and Lee, J.T. (2013). Long noncoding RNAs: past, present, and future. Genetics *193*, 651-669.

Kuo, T.C., and Schlissel, M.S. (2009). Mechanisms controlling expression of the RAG locus during lymphocyte development. Curr Opin Immunol *21*, 173-178.

Kwiatkowski, N., Zhang, T., Rahl, P.B., Abraham, B.J., Reddy, J., Ficarro, S.B., Dastur, A., Amzallag, A., Ramaswamy, S., Tesar, B.*, et al.* (2014). Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. Nature *511*, 616-620.

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.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev *12*, 34-44.

Lai, F., Gardini, A., Zhang, A., and Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. Nature *525*, 399-403.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497-501.

Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. Trends Biochem Sci *39*, 170-182.

Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. Cell *152*, 1021-1036.

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 Borukhov, S. (2016). Bacterial RNA Polymerase-DNA Interaction-The Driving Force of Gene Expression and the Target for Drug Action. Front Mol Biosci 3, 73.

Lee, C., and Huang, C.H. (2013). LASAGNA-Search: an integrated web tool for transcription factor binding site search and visualization. Biotechniques *54*, 141-153.

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, J.E., Wang, C., Xu, S., Cho, Y.W., Wang, L., Feng, X., Baldridge, A., Sartorelli, V., Zhuang, L., Peng, W.*, et al.* (2013). H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. Elife *2*, e01503.

Lee, J.S., Galvin, K.M., See, R.H., Eckner, R., Livingston, D., Moran, E., and Shi, Y. (1995). Relief of YY1 transcriptional repression by adenovirus E1A is mediated by E1A-associated protein p300. Genes Dev *9*, 1188-1198.

Lee, M.G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature 437, 432-435.

Lettice, L.A., Williamson, I., Devenney, P.S., Kilanowski, F., Dorin, J., and Hill, R.E. (2014). Development of five digits is controlled by a bipartite long-range cis-regulator. Development 141, 1715-1725.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707-719.

Li, N., Johnson, D.C., Weinhold, N., Studd, J.B., Orlando, G., Mirabella, F., Mitchell, J.S., Meissner, T., Kaiser, M., Goldschmidt, H.*, et al.* (2016a). Multiple myeloma risk variant at 7p15.3 creates an IRF4-binding site and interferes with CDCA7L expression. Nat Commun *7*, 13656.

Li, Q., Barkess, G., and Qian, H. (2006). Chromatin looping and the probability of transcription. Trends Genet 22, 197-202.

Li, Q., Peterson, K.R., Fang, X., and Stamatoyannopoulos, G. (2002). Locus control regions. Blood 100, 3077-3086.

Li, Q., Zhang, M., Duan, Z., and Stamatoyannopoulos, G. (1999). Structural analysis and mapping of DNase I hypersensitivity of HS5 of the beta-globin locus control region. Genomics 61, 183-193.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X.*, et al.* (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516-520.

Li, W., Notani, D., and Rosenfeld, M.G. (2016b). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat Rev Genet *17*, 207-223.

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.L., Nien, C.Y., Liu, H.Y., Metzstein, M.M., Kirov, N., and Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature 456, 400-403.

Liang, Z., and Biggin, M.D. (1998). Eve and ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in Drosophila. Development 125, 4471-4482.

Liao, S.M., Zhang, J., Jeffery, D.A., Koleske, A.J., Thompson, C.M., Chao, D.M., Viljoen, M., van Vuuren, H.J., and Young, R.A. (1995). A kinase-cyclin pair in the RNA polymerase II holoenzyme. Nature *374*, 193-196.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.*, et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Lin, J.J., Lehmann, L.W., Bonora, G., Sridharan, R., Vashisht, A.A., Tran, N., Plath, K., Wohlschlegel, J.A., and Carey, M. (2011). Mediator coordinates PIC assembly with recruitment of CHD1. Genes Dev *25*, 2198-2209.

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

Liu, H., Schmidt-Supprian, M., Shi, 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.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell *167*, 1170-1187.

Losada, A., Hirano, M., and Hirano, T. (1998). Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. Genes Dev *12*, 1986-1997.

Löffert, D., Schaal, S., Ehlich, A., Hardy, R.R., Zou, Y.R., Müller, W. & Rajewsky, K. (1994). Early B-cell development in the mouse: insights from mutations introduced by gene targeting. Immunological Reviews *137*, 135–153.

Lu, C., Ward, A., Bettridge, J., Liu, Y., and Desiderio, S. (2015). An autoregulatory mechanism imposes allosteric control on the V(D)J recombinase by histone H3 methylation. Cell Rep *10*, 29-38.

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.

Luck, K., Kim, D.K., Lambourne, L., Spirohn, K., Begg, B.E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F.J., Charloteaux, B.*, et al.* (2020). A reference map of the human binary protein interactome. Nature *580*, 402-408.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Makhlouf, M., Ouimette, J.F., Oldfield, A., Navarro, P., Neuillet, D., and Rougeulle, C. (2014). A prominent and conserved role for YY1 in Xist transcriptional activation. Nat Commun *5*, 4878.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet 11, 761-772.

Malik, S., and Roeder, R.G. (2016). Mediator: A Drawbridge across the Enhancer-Promoter Divide. Mol Cell *64*, 433-434.

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.

Malu, S., Malshetty, V., Francis, D., and Cortes, P. (2012). Role of non-homologous end joining in V(D)J recombination. Immunologic research *54*, 233-246.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature 444, 953-956.

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.

Mansour, M.R., Abraham, B.J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A.D., Etchin, J., Lawton, L., Sallan, S.E., Silverman, L.B.*, et al.* (2014). Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science *346*, 1373-1377.

Martensson, I.L., and Ceredig, R. (2000). Review article: role of the surrogate light chain and the pre-B-cell receptor in mouse B-cell development. Immunology *101*, 435-441.

Martinez-Salas, E. (1999). Internal ribosome entry site biology and its use in expression vectors. Curr Opin Biotechnol *10*, 458-464.

Matharu, N., and Ahituv, N. (2015). Minor Loops in Major Folds: Enhancer-Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. PLoS Genet *11*, e1005640.

Mathelier, A., Shi, W., and Wasserman, W.W. (2015). Identification of altered cis-regulatory elements in human disease. Trends Genet *31*, 67-76.

Mathew, R., and Chatterji, D. (2006). The evolving story of the omega subunit of bacterial RNA polymerase. Trends Microbiol 14, 450-455.

Matthews, A.G., Kuo, A.J., Ramon-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.

Matzke, A.J., and Matzke, M.A. (1998). Position effects and epigenetic silencing of plant transgenes. Curr Opin Plant Biol *1*, 142-148.

McBlane, F., and Boyes, J. (2000). Stimulation of V(D)J recombination by histone acetylation. Curr Biol *10*, 483-486.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D.L. (1997). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev *11*, 3306-3318.

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.

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

McMurry, M.T., and Krangel, M.S. (2000). A role for histone acetylation in the developmental regulation of VDJ recombination. Science *287*, 495-498.

Merkenschlager, M., and Odom, D.T. (2013). CTCF and cohesin: linking gene regulatory elements with their targets. Cell *152*, 1285-1297.

Millevoi, S., and Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. Nucleic Acids Res 38, 2757-2774.

Milne, C.D., and Paige, C.J. (2006). IL-7: a key regulator of B lymphopoiesis. Semin Immunol *18*, 20-30.

Min, J., Feng, Q., Li, Z., Zhang, Y., and Xu, R.M. (2003). Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. Cell 112, 711-723

Mittrucker, H.W., Matsuyama, T., Grossman, A., Kundig, T.M., Potter, J., Shahinian, A., Wakeham, A., Patterson, B., Ohashi, P.S., and Mak, T.W. (1997). Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function. Science *275*, 540-543.

Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K.*, et al.* (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell *124*, 1283-1298.

Mohan, M., Herz, H.M., Smith, E.R., Zhang, Y., Jackson, J., Washburn, M.P., Florens, L., Eissenberg, J.C., and Shilatifard, A. (2011). The COMPASS family of H3K4 methylases in Drosophila. Mol Cell Biol *31*, 4310-4318.

Morcillo, P., Rosen, C., Baylies, M.K., and Dorsett, D. (1997). Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila. Genes Dev 11, 2729-2740.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P., and Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res 9, 6047-6068.

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.

Moss, T., and Stefanovsky, V.Y. (2002). At the center of eukaryotic life. Cell *109*, 545-548.

Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell *51*, 606-617.

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

Murthy, K.G., and Manley, J.L. (1995). The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. Genes Dev 9, 2672-2683.

Myer, V.E., and Young, R.A. (1998). RNA polymerase II holoenzymes and subcomplexes. J Biol Chem *273*, 27757-27760.

Nagulapalli, M., Maji, S., Dwivedi, N., Dahiya, P., and Thakur, J.K. (2016). Evolution of disorder in Mediator complex and its functional relevance. Nucleic Acids Res 44, 1591-1612.

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.

Nakayama, J., Yamamoto, M., Hayashi, K., Satoh, H., Bundo, K., Kubo, M., Goitsuka, R., Farrar, M.A., and Kitamura, D. (2009). BLNK suppresses pre-B-cell leukemogenesis through inhibition of JAK3. Blood *113*, 1483-1492.

Nasmyth, K., and Haering, C.H. (2009). Cohesin: its roles and mechanisms. Annu Rev Genet *43*, 525-558.

Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. Annu Rev Genet *46*, 1-19.

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. Proc Natl Acad Sci U S A *96*, 1858-1862.

Nelsen, B., Tian, G., Erman, B., Gregoire, J., Maki, R., Graves, B., and Sen, R. (1993). Regulation of lymphoid-specific immunoglobulin mu heavy chain gene enhancer by ETS-domain proteins. Science 261, 82-86.

Nemazee, D. (2006). Receptor editing in lymphocyte development and central tolerance. Nat Rev Immunol *6*, 728-740.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell *11*, 709-719.

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.

Nonet, M.L., and Young, R.A. (1989). Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of Saccharomyces cerevisiae RNA polymerase II. Genetics 123, 715-724.

Nudler, E. (2009). RNA polymerase active center: the molecular engine of transcription. Annu Rev Biochem 78, 335-361.

Nutt, S.L., Heavey, B., Rolink, A.G., and Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature *401*, 556-562.

O'Shea, J.J., and Plenge, R. (2012). JAK and STAT signaling molecules in immunoregulation and immune-mediated disease. Immunity *36*, 542-550.

Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell *87*, 953-959.

Ohnishi, K., and Melchers, F. (2003). The nonimmunoglobulin portion of lambda5 mediates cell-autonomous pre-B cell receptor signaling. Nat Immunol *4*, 849-856.

Oldfield, A.J., Yang, P., Conway, A.E., Cinghu, S., Freudenberg, J.M., Yellaboina, S., and Jothi, R. (2014). Histone-fold domain protein NF-Y

promotes chromatin accessibility for cell type-specific master transcription factors. Mol Cell 55, 708-722.

Palstra, R.J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet *35*, 190-194.

Pan, H., Jin, M., Ghadiyaram, A., Kaur, P., Miller, H.E., Ta, H.M., Liu, M., Fan, Y., Mahn, C., Gorthi, A.*, et al.* (2020). Cohesin SA1 and SA2 are RNA binding proteins that localize to RNA containing regions on DNA. Nucleic Acids Res *48*, 5639-5655.

Pan, X., Papasani, M., Hao, Y., Calamito, M., Wei, F., Quinn Iii, W.J., Basu, A., Wang, J., Hodawadekar, S., Zaprazna, K.*, et al.* (2013). YY1 controls Igkappa repertoire and B-cell development, and localizes with condensin on the Igkappa locus. EMBO J *32*, 1168-1182.

Parker, S.C., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Program, N.C.S., et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc Natl Acad Sci U S A 110, 17921-17926.

Peng, J., Zhu, Y., Milton, J.T., and Price, D.H. (1998). Identification of multiple cyclin subunits of human P-TEFb. Genes Dev *12*, 755-762.

Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A., and Bejerano, G. (2013). Enhancers: five essential questions. Nat Rev Genet *14*, 288-295.

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.

Perry, M.W., Boettiger, A.N., Bothma, J.P., and Levine, M. (2010). Shadow enhancers foster robustness of Drosophila gastrulation. Curr Biol 20, 1562-1567.

Peterlin, B.M. (2010). Transcription elongation takes central stage: the P-TEFb connection. Cell Cycle *9*, 2933-2934.

Peterson, M.G., Tanese, N., Pugh, B.F., and Tjian, R. (1990). Functional domains and upstream activation properties of cloned human TATA binding protein. Science *248*, 1625-1630.

Petrenko, N., Jin, Y., Wong, K.H., and Struhl, K. (2016). Mediator Undergoes a Compositional Change during Transcriptional Activation. Mol Cell 64, 443-454.

Pettersson, S., Cook, G.P., Bruggemann, 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.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev *20*, 2922-2936.

Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194-1211.

Plaschka, C., Lariviere, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F., et al. (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. Nature 518, 376-380.

Plaschka, C., Nozawa, K., and Cramer, P. (2016). Mediator Architecture and RNA Polymerase II Interaction. J Mol Biol 428, 2569-2574.

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.

Poss, Z.C., Ebmeier, C.C., and Taatjes, D.J. (2013). The Mediator complex and transcription regulation. Crit Rev Biochem Mol Biol 48, 575-608.

Pott, S., and Lieb, J.D. (2015). What are super-enhancers? Nat Genet *47*, 8-12.

Price, D.H. (2000). P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. Mol Cell Biol *20*, 2629-2634.

Proudhon, C., Hao, B., Raviram, R., Chaumeil, J., and Skok, J.A. (2015). Long-Range Regulation of V(D)J Recombination. Adv Immunol *128*, 123-182.

Pugh, B.F., and Tjian, R. (1991). Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev *5*, 1935-1945.

Pui, J.C., Allman, D., Xu, L., DeRocco, S., Karnell, F.G., Bakkour, S., Lee, J.Y., Kadesch, T., Hardy, R.R., Aster, J.C.*, et al.* (1999). Notch1 expression in early lymphopoiesis influences B versus T lineage determination. Immunity *11*, 299-308.

Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K.R., Resch, W., Liang, G., Tang, Z., Mathe, E., Benner, C.*, et al.* (2014). B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. Cell *159*, 1524-1537.

Qiu, Y., Sharma, A., and Stein, R. (1998). p300 mediates transcriptional stimulation by the basic helix-loop-helix activators of the insulin gene. Mol Cell Biol *18*, 2957-2964.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Ramon-Maiques, S., Kuo, A.J., Carney, D., Matthews, A.G., Oettinger, M.A., Gozani, O., and Yang, W. (2007). The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. Proc Natl Acad Sci U S A *104*, 18993-18998.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S.*, et al.*

(2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.

Rasmussen, E.B., and Lis, J.T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc Natl Acad Sci U S A *90*, 7923-7927.

Rienzo, M., and Casamassimi, A. (2016). Integrator complex and transcription regulation: Recent findings and pathophysiology. Biochim Biophys Acta 1859, 1269-1280.

Rickert, R.C. (2013). New insights into pre-BCR and BCR signalling with relevance to B cell malignancies. Nat Rev Immunol *13*, 578-591.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu Rev Biochem *81*, 145-166.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E.*, et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311-1323.

Riquet, F.B., Tan, L., Choy, B.K., Osaki, M., Karsenty, G., Osborne, T.F., Auron, P.E., and Goldring, M.B. (2001). YY1 is a positive regulator of transcription of the Col1a1 gene. J Biol Chem 276, 38665-38672.

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.

Robinson, P.J., Bushnell, D.A., Trnka, M.J., Burlingame, A.L., and Kornberg, R.D. (2012). Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II. Proc Natl Acad Sci U S A *109*, 17931-17935.

Robinson, P.J., Trnka, M.J., Bushnell, D.A., Davis, R.E., Mattei, P.J., Burlingame, A.L., and Kornberg, R.D. (2016). Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. Cell 166, 1411-1422 e1416.

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.

Romeo, V., and Schumperli, D. (2016). Cycling in the nucleus: regulation of RNA 3' processing and nuclear organization of replication-dependent histone genes. Curr Opin Cell Biol 40, 23-31.

Romier, C., Cocchiarella, F., Mantovani, R., and Moras, D. (2003). The NF-YB/NF-YC structure gives insight into DNA binding and transcription regulation by CCAAT factor NF-Y. J Biol Chem 278, 1336-1345.

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.

Rosenberg, N., and Baltimore, D. (1976). A quantitative assay for transformation of bone marrow cells by Abelson murine leukemia virus. J Exp Med *143*, 1453-1463.

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R.L. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science 262, 1407-1413.

Roth, J.F., Shikama, N., Henzen, C., Desbaillets, I., Lutz, W., Marino, S., Wittwer, J., Schorle, H., Gassmann, M., and Eckner, R. (2003). Differential role of p300 and CBP acetyltransferase during myogenesis: p300 acts upstream of MyoD and Myf5. EMBO J 22, 5186-5196.

Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, C.D. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nat Rev Mol Cell Biol 8, 983-994.

Rouaud, P., Vincent-Fabert, C., Saintamand, A., Fiancette, R., Marquet, M., Robert, I., Reina-San-Martin, B., Pinaud, E., Cogne, M., and Denizot, Y. (2013). The IgH 3' regulatory region controls somatic hypermutation in germinal center B cells. J Exp Med *210*, 1501-1507.

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.

Rzymski, T., Mikula, M., Zylkiewicz, E., Dreas, A., Wiklik, K., Golas, A., Wojcik, K., Masiejczyk, M., Wrobel, A., Dolata, I.*, et al.* (2017). SEL120-34A is a novel CDK8 inhibitor active in AML cells with high levels of serine phosphorylation of STAT1 and STAT5 transactivation domains. Oncotarget *8*, 33779-33795.

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.

Saldana-Meyer, R., Rodriguez-Hernaez, J., Escobar, T., Nishana, M., Jacome-Lopez, K., Nora, E.P., Bruneau, B.G., Tsirigos, A., Furlan-Magaril, M., Skok, J.*, et al.* (2019). RNA Interactions Are Essential for CTCF-Mediated Genome Organization. Mol Cell *76*, 412-422.

Sanchez, P., Nadel, B., and Cazenave, P.A. (1991). V lambda-J lambda rearrangements are restricted within a V-J-C recombination unit in the mouse. Eur J Immunol *21*, 907-911.

Sarvagalla, S., Kolapalli, S.P., and Vallabhapurapu, S. (2019). The Two Sides of YY1 in Cancer: A Friend and a Foe. Front Oncol *9*, 1230.

Satija, R., and Bradley, R.K. (2012). The TAGteam motif facilitates binding of 21 sequence-specific transcription factors in the Drosophila embryo. Genome Res *22*, 656-665.

Saunders, A., Core, L.J., and Lis, J.T. (2006). Breaking barriers to transcription elongation. Nat Rev Mol Cell Biol *7*, 557-567.

Schatz, D.G., and Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. Nat Rev Immunol *11*, 251-263.

Schaukowitch, K., Joo, J.Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. Mol Cell *56*, 29-42.

Schlissel, M., Voronova, A., and Baltimore, D. (1991). Helix-loop-helix transcription factor E47 activates germ-line immunoglobulin heavy-chain gene transcription and rearrangement in a pre-T-cell line. Genes Dev 5, 1367-1376.

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.

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet *20*, 437-455.

Schwalie, P.C., Ward, M.C., Cain, C.E., Faure, A.J., Gilad, Y., Odom, D.T., and Flicek, P. (2013). Co-binding by YY1 identifies the transcriptionally active, highly conserved set of CTCF-bound regions in primate genomes. Genome Biol *14*, R148.

Schwickert, T.A., Tagoh, H., Gultekin, 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.

Seroz, T., Hwang, J.R., Moncollin, V., and Egly, J.M. (1995). TFIIH: a link between transcription, DNA repair and cell cycle regulation. Curr Opin Genet Dev *5*, 217-221.

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.

Shah, J.N., Shao, G., Hei, T.K., and Zhao, Y. (2008). Methylation screening of the TGFBI promoter in human lung and prostate cancer by methylation-specific PCR. BMC Cancer *8*, 284.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V.*, et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119, 941-953.

Shih, H.Y., and Krangel, M.S. (2013). Chromatin architecture, CCCTC-binding factor, and V(D)J recombination: managing long-distance relationships at antigen receptor loci. J Immunol *190*, 4915-4921.

Shii, L., Song, L., Maurer, K., Zhang, Z., and Sullivan, K.E. (2017). SERPINB2 is regulated by dynamic interactions with pause-release proteins and enhancer RNAs. Mol Immunol *88*, 20-31.

Shin, M.K., and Koshland, M.E. (1993). Ets-related protein PU.1 regulates expression of the immunoglobulin J-chain gene through a novel Ets-binding element. Genes Dev 7, 2006-2015.

Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu Rev Biochem *81*, 65-95.

Shin, H.Y., Willi, M., HyunYoo, K., Zeng, X., Wang, C., Metser, G., and Hennighausen, L. (2016). Hierarchy within the mammary STAT5-driven Wap super-enhancer. Nat Genet *48*, 904-911.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet *15*, 272-286.

Shore, S.K., Tantravahi, R.V., and Reddy, E.P. (2002). Transforming pathways activated by the v-Abl tyrosine kinase. Oncogene *21*, 8568-8576.

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 *350*, 978-981.

Skaar, J.R., Ferris, A.L., Wu, X., Saraf, A., Khanna, K.K., Florens, L., Washburn, M.P., Hughes, S.H., and Pagano, M. (2015). The Integrator complex controls the termination of transcription at diverse classes of gene targets. Cell Res 25, 288-305.

Song, S.H., Hou, C., and Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. Mol Cell 28, 810-822.

Spangler, L., Wang, X., Conaway, J.W., Conaway, R.C., and Dvir, A. (2001). TFIIH action in transcription initiation and promoter escape requires distinct regions of downstream promoter DNA. Proc Natl Acad Sci U S A *98*, 5544-5549.

Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., Raffel, R., Sobhian, B., Severac, D., Rialle, S.*, et al.* (2014). Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. Nat Commun *5*, 5531.

Stadhouders, R., de Bruijn, M.J., Rother, M.B., Yuvaraj, S., Ribeiro de Almeida, C., Kolovos, P., Van Zelm, M.C., van Ijcken, W., Grosveld, F., Soler, E.*, et al.* (2014). Pre-B cell receptor signaling induces immunoglobulin kappa locus

accessibility by functional redistribution of enhancer-mediated chromatin interactions. PLoS Biol *12*, e1001791.

Stanhope-Baker, P., Hudson, K.M., Shaffer, A.L., Constantinescu, A., and Schlissel, M.S. (1996). Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity in vitro. Cell *85*, 887-897.

Sterner, D.E., and Berger, S.L. (2000). Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev 64, 435-459.

Steward, M.M., Lee, J.S., O'Donovan, A., Wyatt, M., Bernstein, B.E., and Shilatifard, A. (2006). Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. Nat Struct Mol Biol *13*, 852-854.

Stopka, T., Amanatullah, D.F., Papetti, M., and Skoultchi, A.I. (2005). PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. EMBO J 24, 3712-3723.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature 403, 41-45.

Su, W., Jackson, S., Tjian, R., and Echols, H. (1991). DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes Dev 5, 820-826.

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.

Sun, X., Zhang, Y., Cho, H., Rickert, P., Lees, E., Lane, W., and Reinberg, D. (1998). NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. Mol Cell *2*, 213-222.

Sutherland, C., and Murakami, K.S. (2018). An Introduction to the Structure and Function of the Catalytic Core Enzyme of Escherichia coli RNA Polymerase. EcoSal Plus 2018.

Tedder, T.F. (2009). CD19: a promising B cell target for rheumatoid arthritis. Nat Rev Rheumatol 5, 572-577.

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.

Thomas-Claudepierre, A.S., Robert, I., Rocha, P.P., Raviram, R., Schiavo, E., Heyer, V., Bonneau, R., Luo, V.M., Reddy, J.K., Borggrefe, T.*, et al.* (2016). Mediator facilitates transcriptional activation and dynamic long-range contacts at the IgH locus during class switch recombination. J Exp Med *213*, 303-312.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575-581.

Tsai, K.L., Tomomori-Sato, C., Sato, S., Conaway, R.C., Conaway, J.W., and Asturias, F.J. (2014). Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. Cell *157*, 1430-1444.

Tsai, P.F., Dell'Orso, S., Rodriguez, J., Vivanco, K.O., Ko, K.D., Jiang, K., Juan, A.H., Sarshad, A.A., Vian, L., Tran, M.*, et al.* (2018). A Muscle-Specific Enhancer RNA Mediates Cohesin Recruitment and Regulates Transcription In trans. Mol Cell *71*, 129-141 e128.

Turowski, T.W., and Tollervey, D. (2016). Transcription by RNA polymerase III: insights into mechanism and regulation. Biochem Soc Trans *44*, 1367-1375.

Valay, J.G., Simon, M., Dubois, M.F., Bensaude, O., Facca, C., and Faye, G. (1995). The KIN28 gene is required both for RNA polymerase II mediated transcription and phosphorylation of the Rpb1p CTD. J Mol Biol *249*, 535-544.

van Oevelen, C., Collombet, S., Vicent, G., Hoogenkamp, M., Lepoivre, C., Badeaux, A., Bussmann, L., Sardina, J.L., Thieffry, D., Beato, M.*, et al.* (2015). C/EBPalpha Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. Stem Cell Reports *5*, 232-247.

Vannini, A., and Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Mol Cell *45*, 439-446.

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.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., et al. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell 142, 967-980.

Vernimmen, D., and Bickmore, W.A. (2015). The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. Trends Genet *31*, 696-708.

Vo, N., and Goodman, R.H. (2001). CREB-binding protein and p300 in transcriptional regulation. J Biol Chem *276*, 13505-13508.

Volpi, S.A., Verma-Gaur, J., Hassan, R., Ju, Z., Roa, S., Chatterjee, S., Werling, U., Hou, H., Jr., 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.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., et al. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev *12*, 343-356.

Wadman, I.A., Osada, H., Grutz, G.G., Agulnick, A.D., Westphal, H., Forster, A., and Rabbitts, T.H. (1997). The LIM-only protein Lmo2 is a bridging molecule

assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. EMBO J *16*, 3145-3157.

Wai, D.C., Shihab, M., Low, J.K., and Mackay, J.P. (2016). The zinc fingers of YY1 bind single-stranded RNA with low sequence specificity. Nucleic Acids Res *44*, 9153-9165.

Walter, J., and Biggin, M.D. (1996). DNA binding specificity of two homeodomain proteins in vitro and in Drosophila embryos. Proc Natl Acad Sci U S A 93, 2680-2685.

Walter, J., Dever, C.A., and Biggin, M.D. (1994). Two homeo domain proteins bind with similar specificity to a wide range of DNA sites in Drosophila embryos. Genes Dev 8, 1678-1692.

Wang, C., Lee, J.E., Lai, B., Macfarlan, T.S., Xu, S., Zhuang, L., Liu, C., Peng, W., and Ge, K. (2016). Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. Proc Natl Acad Sci U S A *113*, 11871-11876.

Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res 22, 1680-1688.

Wang, Q., Carroll, J.S., and Brown, M. (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. Mol Cell 19, 631-642.

Wang, S., Linde, M.H., Munde, M., Carvalho, V.D., Wilson, W.D., and Poon, G.M. (2014). Mechanistic heterogeneity in site recognition by the structurally homologous DNA-binding domains of the ETS family transcription factors Ets-1 and PU.1. J Biol Chem 289, 21605-21616.

Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., et al. (2018). The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol. *19*, 151.

Watanabe, S., Radman-Livaja, M., Rando, O.J., and Peterson, C.L. (2013). A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. Science *340*, 195-199.

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 e1528.

Wendt, K.S., and Peters, J.M. (2009). How cohesin and CTCF cooperate in regulating gene expression. Chromosome Res *17*, 201-214.

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*, 227-234.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307-319.

Wilson, A., Held, W., and MacDonald, H.R. (1994). Two waves of recombinase gene expression in developing thymocytes. J Exp Med *179*, 1355-1360.

Witte, S., O'Shea, J.J., and Vahedi, G. (2015). Super-enhancers: Asset management in immune cell genomes. Trends Immunol 36, 519-526.

Wittschieben, B.O., Otero, G., de Bizemont, T., Fellows, J., Erdjument-Bromage, H., Ohba, R., Li, Y., Allis, C.D., Tempst, P., and Svejstrup, J.Q. (1999). A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. Mol Cell *4*, 123-128.

Wolff, J., Rabbani, L., Gilsbach, R., Richard, G., Manke, T., Backofen, R., and Gruning, B.A. (2020). Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. Nucleic Acids Res 46, W11-W16.

Woychik, N.A. (1998). Fractions to functions: RNA polymerase II thirty years later. Cold Spring Harb Symp Quant Biol *63*, 311-317.

Woychik, N.A., and Hampsey, M. (2002). The RNA polymerase II machinery: structure illuminates function. Cell *108*, 453-463.

Wu, J., and Grunstein, M. (2000). 25 years after the nucleosome model: chromatin modifications. Trends Biochem Sci 25, 619-623.

Wu, M., Wang, P.F., Lee, J.S., Martin-Brown, S., Florens, L., Washburn, M., and Shilatifard, A. (2008). Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Mol Cell Biol *28*, 7337-7344.

Wu, Y., Albrecht, T.R., Baillat, D., Wagner, E.J., and Tong, L. (2017). Molecular basis for the interaction between Integrator subunits IntS9 and IntS11 and its functional importance. Proc Natl Acad Sci U S A 114, 4394-4399.

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). Vkappa gene repertoire and locus contraction are specified by critical DNase I hypersensitive sites within the Vkappa-Jkappa intervening region. J Immunol *190*, 1819-1826.

Xiang, Y., Park, S.K., and Garrard, W.T. (2014). A major deletion in the Vkappa-Jkappa intervening region results in hyperelevated transcription of proximal Vkappa genes and a severely restricted repertoire. J Immunol *193*, 3746-3754.

Xiang, Y., Zhou, X., Hewitt, S.L., Skok, J.A., and Garrard, W.T. (2011). A multifunctional element in the mouse Igkappa locus that specifies repertoire and Ig loci subnuclear location. J Immunol *186*, 5356-5366.

Xiao, B., Wilson, J.R., and Gamblin, S.J. (2003). SET domains and histone methylation. Curr Opin Struct Biol 13, 699-705.

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, F., Zhang, K., and Grunstein, M. (2005). Acetylation in histone H3 globular domain regulates gene expression in yeast. Cell 121, 375-385.

Yamaguchi, Y., Inukai, N., Narita, T., Wada, T., and Handa, H. (2002). Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA. Mol Cell Biol *22*, 2918-2927.

Yamaguchi, Y., Shibata, H., and Handa, H. (2013). Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. Biochim Biophys Acta *1829*, 98-104.

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.

Yokoyama, A. (2018). RNA Polymerase II-Dependent Transcription Initiated by Selectivity Factor 1: A Central Mechanism Used by MLL Fusion Proteins in Leukemic Transformation. Front Genet *9*, 722.

Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T.*, et al.* (2012). Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. Science *337*, 971-975.

Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. Cell Res 21, 564-578.

Zabidi, M.A., and Stark, A. (2016). Regulatory Enhancer-Core-Promoter Communication via Transcription Factors and Cofactors. Trends Genet *32*, 801-814.

Zeitlinger, J., Zinzen, R.P., Stark, A., Kellis, M., Zhang, H., Young, R.A., and Levine, M. (2007). Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. Genes Dev 21, 385-390.

Zhang, G., and Darst, S.A. (1998). Structure of the Escherichia coli RNA polymerase alpha subunit amino-terminal domain. Science *281*, 262-266.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev *15*, 2343-2360.

Zhang, Y., Zhang, X., Ba, Z., Liang, Z., Dring, E.W., Hu, H., Lou, J., Kyritsis, N., Zurita, J., Shamim, M.S.*, et al.* (2019). The fundamental role of chromatin loop extrusion in physiological V(D)J recombination. Nature *573*, 600-604.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750-756.

Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M., and Tuan, D. (2007). A facilitated tracking and transcription mechanism of long-range enhancer function. Nucleic Acids Res 35, 5532-5544.

Zhuang, Y., Soriano, P., and Weintraub, H. (1994). The helix-loop-helix gene E2A is required for B cell formation. Cell 79, 875-884.

Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., and Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. Cell 138, 1122-1136.