



universität
wien

DISSERTATION

Titel der Dissertation

Investigation of protein and DNA interactions of Pax5
in B-cell lineage commitment

Verfasserin

Dipl.Ing.Mol.Biol./Ökol.

Jasna Medvedović

angestrebter akademischer Grad

Doktorin der Naturwissenschaften (Dr. re.nat.)

Wien, 2011

Studienkennzahl lt. Studienblatt:

A > 091 490 <

Dissertationsgebiet lt. Studienblatt:

Molekulare Biologie

Betreuer:

Prof. Dr. Meinrad Busslinger

• Zusammenfassung	1
• Summary	3
• INTRODUCTION	
• Mechanisms of gene regulation	5
○ Chromatin marks	7
○ Chromatin-regulatory complexes	8
○ Posttranslational regulation of non-histone proteins	12
• B cell development	15
○ Transcription factors	15
○ Pax5 in B-cell lineage commitment and maintenance	18
○ Molecular mechanism of B-lineage commitment	20
○ Pax protein family	21
• V(D)J recombination	25
○ The structure of the immunoglobulin heavy chain locus	25
○ Rag1/2, “the V(D)J endonuclease”	26
○ Accessibility control	27
○ Chromatin modifications	28
○ Germline transcripts	29
○ <i>Igh</i> locus control through <i>cis</i> -regulatory elements	30
○ Nuclear positioning and topology of the <i>Igh</i> locus	32
○ Factors involved in <i>Igh</i> locus activation and looping	33
○ CTCF and looping of complex loci	37
• Aim of the thesis	40
• RESULTS I	
• Identification of posttranslational modifications by mass spectrometry	41
• Sumoylation of Pax5	46
○ Recombinant Pax5 can be sumoylated <i>in vitro</i>	47
○ Pax5 can be sumoylated in HEK293T cells	50
• Interaction partners of Pax5	
○ Pax5 recruits chromatin-remodelling and modifying-complexes	54
○ Interaction of Pax5 with Nono	62
○ Interaction of Pax5 with CTCF	64
• RESULTS II	
• CTCF, Pax5 and E2A bind to conserved PAIR sequences	71

• Quality controls of the circular chromosome conformation capture (4C) method	73
• PCR amplification of the 4C library	76
• Genome distribution and normalisation of deep-sequenced 4C reads in <i>Rag2</i> ^{-/-} pro B cells, <i>Rag2</i> ^{-/-} <i>Pax5</i> ^{-/-} pro-B cells and MEFs	79
• 4C interactions of the 3' regulatory elements in the <i>Igh</i> locus	84
• 4C interactions of PAIR elements 4, 5 and 8	91
• Interactions of the pseudogene PG.4.28 from the proximal V _H region	96
• DISCUSSION I	99
• DISCUSSION II	109
• EXPERIMENTAL PROCEDURES	113
• Isolation and immortalisation of bone marrow pro-B cells	113
• Nuclear extract preparation and streptavidin affinity pulldown for mass spectrometry analysis	113
• Isolation of SDS-PAGE fractioned proteins and tryptic digestion for mass spectrometry analysis	114
• Mass spectrometry data analysis	115
• Analysis of proteins identified both in Pax5-Bio and CTCF-Bio streptavidin pulldowns	116
• Nuclear extract preparation with benzonase and validation of Pax5-interactions	116
• Immunoprecipitation and streptavidin pulldown for Western blot analysis	117
• Co-immunoprecipitation of Pax5 with PTIP and CTCF in HEK293T cells	118
• <i>In vitro</i> sumoylation assay	119
• <i>In vivo</i> sumoylation assay in transiently transfected HEK293T cells	120
• Antibodies	121
• Electrophoretic mobility shift assays (EMSA)	122
• Isolation of pro-B cells for 4C analysis	123
• <u>Circular Chromosome Conformations Capture</u> (4C) method	123
• Primer design for 4C in the mouse <i>Igh</i> locus	124
• Solexa sequencing	125
• Bioinformatic analysis	125
• REFERENCES	127
• CREDITS FOR CONTRIBUTION TO THIS WORK	144
• ACKNOWLEDGMENTS	145
• Curriculum vitae	146

Zusammenfassung

Der Determinierungsprozess einer Zelle zu einer bestimmten Zelllinie wird durch das Zusammenspiel von Transkriptionsfaktoren, Chromatin-Regulatoren und Signal-Transduktionswegen reguliert. Der Transkriptionsfaktor Pax5 ist der B-Zell-spezifische Determinierungsfaktor. In seiner dualen Funktion reprimiert Pax5 die Expression von Genen anderer Zelllinien einerseits und aktiviert B-Zell-spezifische Gene andererseits. Dabei interagiert Pax5 mit Proteinkomplexen, die eine aktivierende oder reprimierende Funktion haben. In früheren Studien konnten einige Interaktionspartner von Pax5 identifiziert werden wie z.B. TBP, eine Komponente des basalen Transkriptionsfaktors TFIID oder aber Brg1, eine Untereinheit des BAF-Chromatin Remodelungskomplexes sowie verschiedene Groucho Corepressoren.

Bisher wurden diese Interaktionsstudien in transient transfizierten Zelllinien untersucht. Ein Ziel dieser Arbeit bestand darin unter physiologischen Bedingungen in einem pro-B Zell System mit endogener Pax5 Expression einerseits die bisher bekannten Pax5 Interaktionspartner zu bestätigen und andererseits neue Interaktionspartner zu identifizieren. Hierzu wurde die *Pax5^{Bio/Bio}* knock-in Maus verwendet, die ein Pax5 Protein mit einer Biotinylierungssequenz (Pax5-Bio) generiert, das bei Coexpression der *E.coli* Biotin Ligase BirA biotinyliert wird. Das biotinylierte Pax5 konnte in Streptavidin-Pulldown Experimenten erfolgreich aufgereinigt werden. Interaktionspartner wurden durch Massenspektrometrie und Western Blot Experimente analysiert. Im Zuge dessen konnte gezeigt werden, dass Pax5 in pro-B Zellen den BAF-Chromatin Remodelungskomplex rekrutiert, sowie TFIID und Proteine mit Histon-Acetyltransferase Aktivität (CBP). Weiterhin konnten Komponenten (PTIP, RbBP5) des Histon-Methyltransferase Komplexes MLL mit aufgereinigt werden. Der MLL Komplex führt zur Methylierung von H3K4 - einer Markierung aktiven Chromatins. Der NCoR Komplex, welcher hingegen eine Histon-Deacetylase Funktion besitzt, interagiert ebenfalls mit Pax5. Zudem konnte gezeigt werden, dass dementsprechende Änderungen der Chromatinmodifikationen an Pax5 regulierten Genen nach Rekrutierung dieser Komplexe stattfinden und von diesen abhängig sind. Daher lässt sich zusammenfassen, dass Pax5 in der frühen B-Zell Entwicklung Genexpression auf epigenetischer und transkriptioneller Ebene reguliert.

Eine wichtige Eigenschaft von B-Zellen ist die Produktion eines vielfältigen Antigenrezeptor Repertoires, das das Überleben von Säugetieren in einer Pathogen-reichen Umwelt ermöglicht. Diese Vielfalt wird durch den Prozess der sog. V(D)J Rekombination erreicht. Dabei werden diskontinuierliche variable (V), diversity (D) und joining (J) Segmente der Immunoglobulin (Ig)

Gene während der B-Zell Entwicklung rearrangiert. Um eine gleichmäßige Repräsentation aller Gensegmente zu gewährleisten, kontrahiert der Immunoglobulinschwerekette (*Igh*) Gen Locus im pro-B Zell Stadium. Dadurch werden die 3Mb voneinander entfernte V_H Gene und D_H Segmente nah zusammen gebracht. Diese Kontraktion ist abhängig von Pax5. PAIR Elemente der distalen V_H Regionen des *Igh* Locus enthalten Pax5- CTCF- sowie E2A Bindestellen und zeigen eine Pax5 abhängige Expression von antisense Transkripten während des pro-B Zellstadiums. Es ist bekannt, dass CTCF während bestimmter Entwicklungsstadien zusammen mit anderen spezifischen Transkriptionsfaktoren zur Schleifenbildung komplexer Regionen im Genom führen kann.

Daher sollte untersucht werden, ob PAIR Elemente die Kontraktion des *Igh* Locus regulieren. Diese Hypothese konnte durch die circular chromosom conformation capture (4C) Methode in Kombination mit deep sequencing bestätigt werden. Obgleich 2.5 Mb voneinander entfernt, interagieren PAIR Elemente mit Enhancern am 3'-Ende des *Igh* Locus. Diese Interaktion konnte nur in Pax5 exprimierenden pro-B Zellen festgestellt werden. Dieser Ansatz trägt einen weiteren Teil zu dem Verständnis des Mechanismus bei, wie Pax5 die Kontraktion des *Igh* Locus und sowie die V(D)J Rekombination reguliert.

Summary

Lineage commitment is a result of the interplay of transcription factors, chromatin regulators and cell signalling. Pax5 is a B-cell commitment factor with a dual role of activating B-cell specific genes and simultaneously repressing genes important for alternative lineages. Pax5 acts by interacting with protein complexes that mediate its activating or repressive function. Some interaction partners of Pax5 have been previously identified, such as the TBP component of the basal transcription factor TFIID, the Brg1 subunit of the BAF-chromatin remodelling complex and Groucho corepressors. However, most of these studies were done by using cell lines and transient transfection assays. We wanted to confirm the known Pax5-partner proteins, and to identify novel Pax5-interacting proteins in a physiological pro-B cell system where Pax5 is expressed at an endogenous level. Therefore, I took advantage of the *Pax5^{Bio/Bio}* knock-in mouse, which expresses a biotin tagged Pax5 protein (Pax5-Bio) together with its modifying *E. coli* biotin ligase BirA. From these mice I isolated the biotinylated Pax5 protein by streptavidin pulldown. We analysed the copurified proteins by mass spectrometry and Western blotting. We could confirm that Pax5 recruits the chromatin-remodelling BAF complex, the basal transcription complex TFIID and proteins with histone-acetyltransferase activity (CBP) in pro-B cells. Furthermore, we found that Pax5 associates with the components (PTIP, RbBP5) of the histone methyltransferase MLL complex, establishing the active H3K4 methylation mark, as well as with the NCoR complex and the associated histone deacetylase activity. Moreover, the corresponding changes in chromatin marks at Pax5-regulated genes are accompanied by, and dependent on, the recruitment of these complexes. In summary, Pax5 regulates gene transcription in early B-cell development by orchestrating epigenetic and transcriptional changes on its target genes.

A hallmark of the B cell lineage is the expression of a diverse antigen receptor repertoire, which allows the survival of mammals in a pathogen-rich environment. This diversity arises by the process of V(D)J recombination, which assembles the variable regions of immunoglobulin (Ig) genes from discontinuous variable (V), diversity (D) and joining (J) gene segments during B cell development. To allow equal representation of all gene segments, the immunoglobulin heavy-chain (*Igh*) locus undergoes contraction at the pro-B cell stage, which juxtaposes distal V_H genes next to proximal D_H segments, although they may be separated by genomic distance of up to 3 Mb. The contraction critically depends on Pax5. PAIR elements in the distal V_H region of the *Igh* locus contain Pax5-, CTCF- and E2A-binding sites and give rise to Pax5-dependent antisense transcription at the pro-B cell stage. Since CTCF, in collaboration with specific

transcription factors, mediates developmental stage-specific looping of other complex loci, we hypothesised that the PAIR elements could regulate *Igh* locus contraction. I confirmed this by using the circular chromosome conformation capture (4C) method coupled with deep sequencing. Although they are separated by more than 2.5 Mb, the PAIR elements interacted with the enhancers at the 3' end of the *Igh* locus only in pro-B cells that were expressing Pax5. With this study we have come closer to understanding the mechanism by which Pax5 controls *Igh* locus contraction and V(D)J recombination.

INTRODUCTION

Mechanisms of gene regulation

Genomic DNA is the template of our heredity. By comparing changes in the DNA sequence we can reconstruct the time and sequence of events which led to the evolution of more complex organisms. However, the drastic increase in organism complexity does not correlate with an increase in the number of protein-coding genes. Another layer of complexity is added by the structure of nucleosomes that form fundamental units of chromatin. A nucleosome consists of 147 bp of DNA wrapped around dimers of histones H2A/H2B and dimers of histones H3/H4 (Smith and Shilatifard 2010). Nucleosomes form arrays of “beads on a string” which contract the long DNA fibre in the restricted space of a nucleus (Smith and Shilatifard 2010). The chromatin structure presents a constraint when cells need to transcribe a gene or replicate the DNA (Smith and Shilatifard 2010). In response to this challenge many powerful enzymes evolved that physically or chemically modulate the structure of chromatin in response to various developmental or environmental cues (Smith and Shilatifard 2010). “Chromatin signalling” (Smith and Shilatifard 2010) acts primarily on N-terminal “tails” of histones (Figure 1). Unstructured and flexible, they can easily interact with regulators and undergo different posttranslational modifications. The most versatile residue in their sequences are lysines. A lysine located in the histone tail can exist in five different states unmodified, methylated, acetylated, ubiquitinated and sumoylated, in varying degrees and combinatorial combinations (Figure 1), together comprising “the histone code” (Jenuwein and Allis, 2001). Various effector-molecules can read the code, and translate it to different transcriptional outcomes (Jenuwein and Allis, 2001). Patterns of histone modifications are linked with functional genomic elements and they undergo dynamic changes throughout differentiation and development (Sims and Reinberg, 2008). Methylation of lysines on histone H3 has distinct functional consequences depending on the residue and the degree of modification (mono, di, tri) (Zhou *et al.*, 2011). The functional significance of acetylation depends less on the specific residue, but the accessibility of chromatin remains positively correlated with the degree of modification (Zhou *et al.*, 2011).



(B) The functional interplay between methylation and acetylation that occurs in p53 and histones. Methylation of p53 K372 (p53K372me1) by SET-domain-containing protein-9 (SET9) facilitates the recruitment of the TIP60 histone acetyltransferase (HAT) complex and the subsequent acetylation of p53K382. Similarly, the SET1 histone methyltransferase methylates H3K4 (H3K4me1), thereby facilitating the recruitment of the NuA3 HAT complex and the acetylation of H3K14. Thus, signalling pathways that involve Lys methylation and acetylation occur on both histone and non-histone proteins in a similar fashion. (From: Sims and Reinberg, 2007)

6

Chromatin marks

Over the last decade extensive work has been done in defining chromatin landscapes of gene-regulatory elements. In this section, I will describe some of the conclusions drawn from these studies as reviewed by Zhou *et al.*, (2011). Most mammalian promoters are in regions of high CpG content, “high CpG content promoters” (HCPs) in contrast to “low CpG content promoters” (LCPs). The two categories have distinct modes of regulation (Zhou *et al.*, 2011). In embryonic stem (ES) cells, HCPs are active “by default”. Histone H3 lysine 4 trimethylation (H3K4me3) is enriched and associated with other features of accessible chromatin such as histone acetylation, presence of H3.3 histone variant and hypersensitivity to DNase I digestion (Figure 2). RNA polymerase II (RNAPII) binds promoters marked with H3K4me3. The initiating RNAPII contributes to the chromatin accessibility through interactions with chromatin modifiers (Zhou *et al.*, 2011). Although transcription is initiated, functional, full-length transcripts might be missing. Additional regulation of HCP-genes may exist at the transition from initiation to elongation (Zhou *et al.*, 2011). Inactive HCPs carry the repressive modification H3K27me3 and are inaccessible to RNAPII. Poised HCPs are marked with both H3K4me3 and H3K27me3. Due to the repressive mark they do not elongate and make productive mRNAs (Zhou *et al.*, 2011). Such bivalent modifications are often present in haematopoietic progenitors at gene-promoters of developmental regulators (Zhou *et al.*, 2011). Bivalency allows rapid activation of transcription at the next developmental stage (Zhou *et al.*, 2011). Inactive LCPs lack chromatin marks and the DNA can be methylated (Zhou *et al.*, 2011) at cytosine nucleotides, which provides binding platforms for the gene-silencing complexes (Zhou *et al.*, 2011). The chromatin is poised by dimethylation of H3K4 and selectively activated with H3K4me3 after binding of specific transcription factors. In haematopoietic progenitors such promoters are often associated with cell-type specific genes (for example, structural proteins) that become induced during differentiation (Zhou *et al.*, 2011). In summary, the signature of active promoter contains H3K4me2, H3K4me3, high levels of acetylation (H3K27 and H3K9) and they usually overlap with DNase I hypersensitive regions (Figure 2A; Zhou *et al.*, 2011).

Enhancer DNA elements are located distantly from the transcription start site. Through long-range interactions enhancers positively influence transcription at promoters by recruiting transcription factors, co-activators and RNAPII (Zhou *et al.*, 2011). Previously, the presence of H3K4me2 and H3K9ac marks and the lack of H3K4me3 were used to define enhancers but recent work shows that their chromatin signature is often identical to promoters (H3K4me2, H3K4me3, H3K27ac and H3K9ac) (Ernst *et al.*, 2010). Therefore defining features of enhancers are that unlike promoters, their genomic location does not overlap with the transcription start site

and their activity is cell type specific with the purpose of fine-tuning gene expression in response to specific developmental and environmental needs.

Upon differentiation of a cell to a committed state many genes become stably repressed thus they are silenced within chromatin decorated by repressive histone marks. H3K9me2 is enriched in large heterochromatin domains bound to the nuclear lamina (LADs) (Zhou *et al.*, 2011). Blocks of H3K27me3 extend across silent genes and intergenic regions. While lamina interactions have not been observed in ES, H3K27me3 are often seeded at the ES cell stage and further expanded in differentiated cells (Zhou *et al.*, 2011).

In summary, gene activation is an ordered interplay of DNA elements, transcription factors and complexes that modify the structure of chromatin. Transcription factors recognize the DNA regulatory elements and activate them by recruiting histone methyl transferases and acetylases. The active regions spread and activate enhancer elements that in turn form stable complexes with promoters and focus gene activation to genes fulfilling the requirements a specific cell type is faced with.

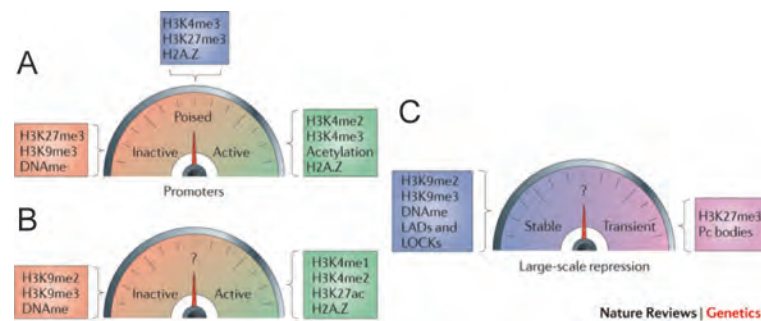


Figure 2. Histone modifications can result in active, inactive or poised state of genes

(A) At promoters, they can contribute to fine-tuning of expression levels, from active to poised to inactive or even intermediate levels. (B) At distal sites, histone marks correlate with levels of enhancer activity. (C) On a global scale, they may confer repression of varying stabilities and be associated with different genomic features. For example, lamina-associated domains (LADs) in the case of stable repression and Polycomb (Pc) bodies in the case of context-specific repression. DNAm, DNA methylation; LOCK, large organized chromatin K modification. (From: Zhou *et al.*, 2011)

Chromatin-regulatory complexes

Chromatin (de)methylation and (de)acetylation is mediated by large enzymatic complexes consisting of several subunits. The core components have homologues from yeast to mammals. However, the complexity of these complexes increases with the complexity of organisms. Although at first glance some enzymes with the same function may seem redundant, deeper inspection of their post-translational modifications (PTMs), or modulator-subunits, reveals specific roles. Another important group of enzymes are the chromatin remodellers. They use

acetylated and methylated residues of histone-tails, as well as various structural motifs in transcription factors, as binding platforms. From there they actively reposition nucleosomes thereby enabling or disabling the transcriptional machinery to access the genes.

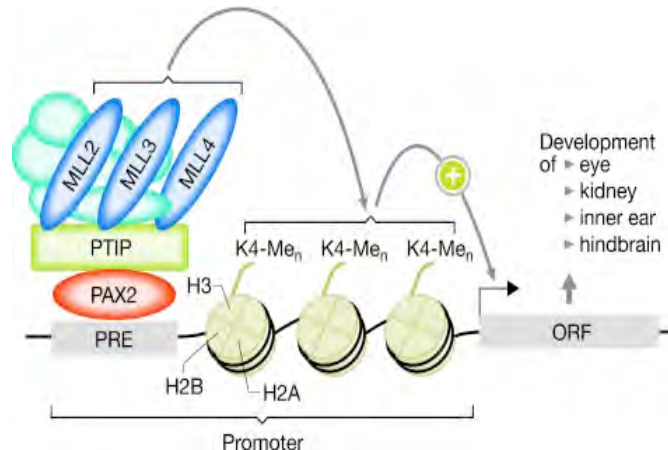


Figure 3. Diagram representing the composition of the MLL complex.

PTIP targets HMTase complexes to gene promoters. PAX2 binds to PREs (Pax2-response-elements) in the promoters of a range of genes that are required for the development of various organs. PTIP is brought to promoters by PAX2, and recruits SET1-like lysine methyltransferase complexes that methylate H3K4, thereby activating gene transcription. PTIP was found to be an integral component of several SET domain-containing HMTase complexes that differ in their catalytic subunits, which can be MLL2, MLL3 plus MLL4 or MLL2 plus MLL3 (Patel *et al.*, 2007). H3K4, lysine 4 of histone H3; HMTase, histone methyltransferase; Me, methyl; MLL, mixed-lineage leukaemia; ORF, open reading frame; PAX2, paired box gene 2; PREs, PAX2-response elements; PTIP, PAX-transactivation domain-interacting protein; SET, Su(Var)3-9, E(z), Trx. (From: Muñoz and Rouse, 2009)

In progenitor and differentiated cells transcriptional decisions are maintained throughout cell divisions, as “cellular memory”. In *Drosophila* this is accomplished by antagonistic chromatin modifying complexes of Polycomb and Trithorax groups (Ringrose and Paro, 2004). Polycomb group proteins recognise Polycomb response elements (PRE) (Ringrose and Paro, 2004) in the genome of *Drosophila*. Polycomb repressive complex 2 contains Enhancer of zeste (EZH1 and EZH2) histone methyltransferases, which methylate H3K27 creating a binding site for Polycomb protein (Pc), a component of the PRC1 that has ubiquitin ligase activity (Ringrose and Paro, 2004). The first vertebrate PRE was identified in mammals at the *MafB* gene (Sing *et al.*, 2009), while a potential PRE has been identified between the *HOXD11* and *HOXD12* loci in human ES cells, a region that contains YY1 binding sites and recruits PRC1 and PRC2 (Woo *et al.*, 2010). In mammals YY1 is one of the DNA-binding components of the mammalian PRC2 (Atchison *et al.*, 2003), which recruits EZH2 to muscle specific genes (Caretti *et al.*, 2004). Trithorax, the counterpart of Polycomb in *Drosophila*, activates the genes by methylating H3K4 (Smith and Shilatifard 2010). The mammalian homologue is MLL (Figure 3) that has a conserved SET-

domain with lysine methyltransferase activity (Smith and Shilatifard, 2010). Mammals have at least six H3K4 methyltransferases Set1A/B, MLL1, MLL2/ALR, MLL3/HLR and MLL4-5. They all have common core subunits Ash2, Wdr5 and RbBP5 (Figure 3). MLL1 and MLL2 interact with the tumor suppressor Menin, which directs them to *Hox* genes and other targets, whereas MLL3 and MLL4 contain NCOA6, PTIP, PA-1 and histone H3K27 demethylase UTX (Smith and Shilatifard 2010). PTIP interacts with Pax2, the transcription factor specifying kidney development (Figure 3; Muñoz and Rouse, 2009). Through interactions with Pax2, PTIP participates in gene activation serving as a scaffold that recruits MLL2 to Pax2-response elements (PRE; Patel *et al.*, 2007). H3K4 methylation by MLL complexes creates a binding platform for proteins with Phd finger domains (Smith and Shilatifard 2010). Some examples are Taf3 subunit of the basal transcription factor TFIID, histone acetyltransferases or chromatin remodelling complexes (CRCs), which in turn use energy derived from ATP hydrolysis to disrupt histone-DNA contacts. The best-defined mammalian CRCs are BAF (Brahma-related gene (BRG)/Brahma (BRM)-associated factor) complexes (Ho and Crabtree, 2010; Figure 4A). They dissociate patches of DNA from the surface of the histone octamer (nucleosome disruption) by moving the nucleosomes to neighbouring DNA segments *in cis* (nucleosome sliding) or *in trans* (nucleosome transfer) (Chi, 2004). CRCs cooperate with histone-modifying enzymes. Subunits of BAF have bromodomains that mediate interaction with acetylated histones (Figure 4B; Chi, 2004). In a positive feedback loop, CRCs in turn facilitate acetylation by disrupting nucleosomes (Chi, 2004). They can also aid repression of transcription as components of the nucleosome remodelling and deacetylation (NuRD) complex with associated HDAC1 and HDAC2 members (Chi, 2004). Signalling pathways control the location and/or activities of BAF complexes through induction of transcriptional activators or post-translational modifications (Chi, 2004). In response to such stimuli BAF can be transiently recruited to promoters and mediate gene activation (“hit and run”), or poise a promoter for rapid gene induction in response to amplifying signalling cascades (Chi, 2004; Figure 4B).

Histone acetyl transferase (HAT) complexes (Gcn5, P300/CBP, PCAF) are recruited to regions of transcribed genes where they act on H4K16 and H3K9 (Chi, 2004). The negative charge of acetyl groups has a dual role; it decreases the interaction of histones with DNA and creates a binding surface for bromodomain-containing complexes, for example BAF or TFIID (Chi, 2004). Counteracting complexes have histone deacetylase activity (HDAC). Two models of action have been proposed for HATs and HDAC. Studies in yeast suggested that they are recruited to various genes by a transcription regulator (Peserico and Simone, 2009). Current models propose that HATs and HDACs are simultaneously present on target genes, whereas

activation of transcription depends on other factors or complexes that stabilise one or the other activity (Peserico and Simone, 2009). For example CBP and p300 coregulators participate in many physiological processes by acetylating histone and non-histone proteins (Peserico and Simone, 2009). Their structure consisting of a bromodomain, an autoregulatory domain and domains for interactions with different proteins, provides them a scaffolding role between transcription factors, co-activators and components of the basal transcriptional complex (Peserico and Simone, 2009). They have overlapping, but each of them, also unique functions. For example, CBP but not p300 is required for normal hematopoietic development (Peserico and Simone, 2009).

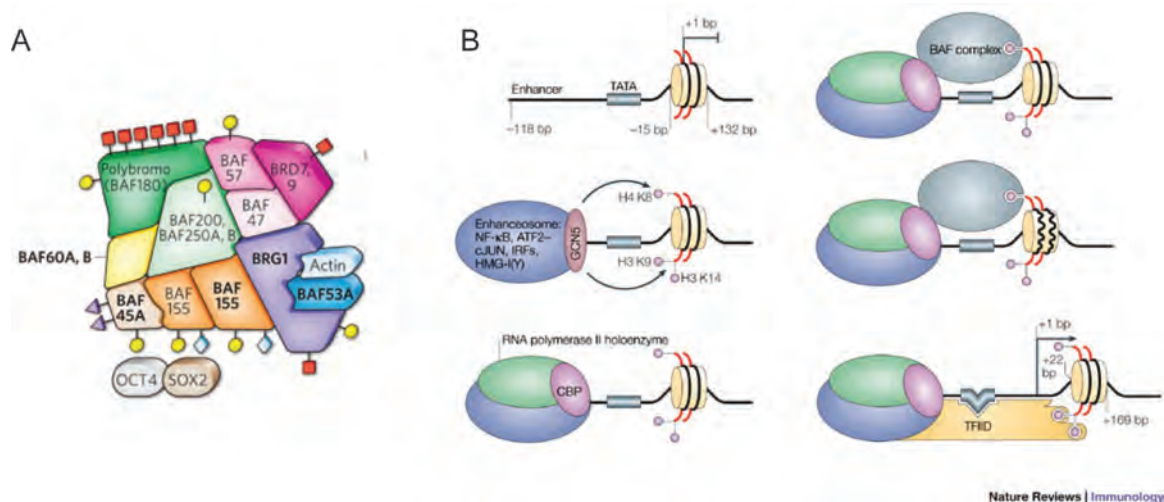


Figure 4. Diagrams showing the structure and function of the BAF-chromatin remodelling complex.

(A) The diagram depicts the composition of BAF complexes in embryonic stem cells (ESC). Some subunits of the BAF complex are stable independent of the cell type whereas others are variable, highlighted in bold. In some cases, key transcription factors that work in cooperation with BAF complexes (such as OCT4 and SOX2 in ESCs). These transiently associated transcription factors are not shown in contact with the main complex to distinguish them from the subunits of the complex. (From: Ho and Crabtree, 2010)

(B) Important events at the INF-β promoter following viral infection that involve the BAF complex. The transcription start site (located at +1 base pairs, bp) of the 250-bp promoter of the gene encoding interferon-beta (IFN-beta) is covered by a positioned nucleosome. The promoter also contains an enhancer that, following viral infection, directs the assembly of the nucleoprotein complex known as the enhanceosome, which contains nuclear factor-kappaB (NF-kappaB), activating transcription factor 2 (ATF2) in a heterodimer with cJUN, IFN-regulatory factors (IRFs) and the high-mobility group protein HMG-I(Y). The enhanceosome recruits the histone acetyltransferase (HAT) GCN5 (general control of amino-acid synthesis 5), which acetylates a subset of lysine residues in the nucleosome: lysine residue 9 in the tail of histone H3 (H3 K9), H3 K14 and H4 K8. Acetylation at different lysine residues specifies distinct outcomes, consistent with the "histone-code hypothesis". GCN5 then leaves the promoter, and the RNA polymerase II holoenzyme (lacking the transcription factor TFIID) is recruited together with CBP (cyclic AMP-responsive-element-binding protein (CREB)-binding protein), which is another HAT. BAF (Brahma-related gene (BRG)/Brahma (BRM)-associated factor) complexes are then recruited, in part by contacting the acetylated H4 K8 through the bromodomain of BRG and disrupt the nucleosome (wavy lines), making it accessible to TFIID. Nucleosomal disruption probably facilitates the interaction between the double bromodomain in the TAF250 (TATA-box-binding-protein (TBP)-associated factor 250) subunit of TFIID and acetylated H3 K9 and H3 K14. The TBP subunit of TFIID bends the DNA and forces the disrupted nucleosome to slide to a position 36 bp downstream, thereby stabilizing TFIID binding, a prerequisite for the induction of IFN-beta expression.

Posttranslational regulation of non-histone proteins

Signalling pathways that involve posttranslational modification of lysines can also occur on non-histone proteins (Sims and Reinberg, 2007). Methylation of p53 by SET-domain-containing proteins (SET9) facilitates recruitment of histone acetyltransferase complexes and subsequent p53 acetylation in response to DNA damage (Ivanov *et al.*, 2007). Methylated arginines can modulate the transactivation potential of transcription factors (Kowenz-Leutz *et al.*, 2010). The dynamics of phosphorylation on serine, threonine and tyrosine residues is a long known mechanism of transmitting and amplifying various signals within and between cells. Phosphorylation of transcription factors can generate binding sites for cofactors. For example, phosphorylation of CREB creates a binding site for CBP, which then additionally modifies CREB by acetylation (Figure 5; reviewed by Yang, 2005).

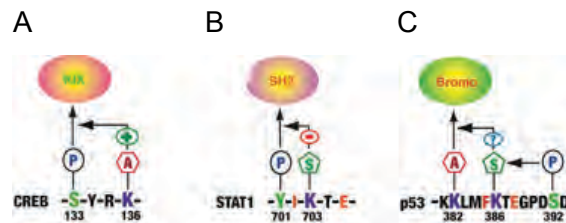


Figure 5. Agonistic and antagonistic effects of neighbouring modifications.

(A) Ser133 phosphorylation of CREB generates a binding site for the KIX domain of CBP or p300, which acetylates Lys136 and strengthens KIX binding.

(B) Tyr701 phosphorylation of STAT1 promotes dimerization through its SH2 domain. Lys703 resides within a consensus sumoylation site and its sumoylation may affect Tyr701 phosphorylation and/or SH2 binding.

(C) p53 acetylation at Lys382 generates a binding site for the bromodomain of CBP. It remains to be determined whether Ser392 phosphorylation stimulates Lys386 sumoylation and if the sumoylation facilitates Lys382 acetylation and subsequent association with CBP. (From: Yang, 2005)

Considering the size of acetyl-, methyl- and phospho-groups that posttranslationally modify proteins, we may wonder how these small chemical groups can induce such dramatic changes on structure and function of large macromolecules as proteins. Part of the reason is that combinations of these modifications, at multiple sites, can act in synergism or antagonism towards a specific function (Figure 5; Yang, 2005). On the other hand, some modifications consist of small proteins, which can modulate the function of a protein even more than addition of small chemical groups. An example of such protein groups are small ubiquitin like modifiers (SUMO), which regulate diverse cellular processes including transcription, chromosome replication and DNA repair (Figure 6; reviewed by Geiss-Friedlander and Melchior, 2007).

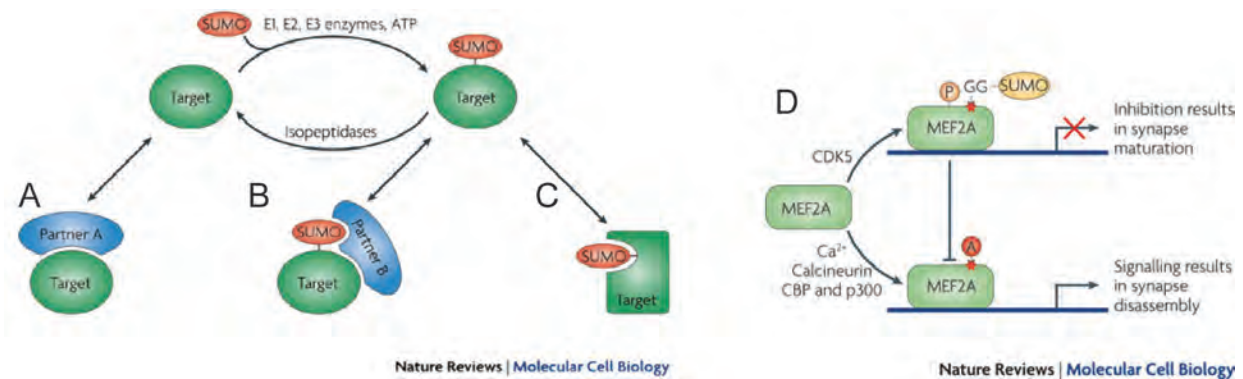


Figure 6. Consequences sumoylation can have for a modified protein.

(A) Sumoylation can interfere with the interaction between the target and its partner, in which case the interaction can only occur in the absence of sumoylation.

(B) Sumoylation can provide a binding site for an interacting partner, for example via a so-called non-covalent SUMO (small ubiquitin-related modifier)-interaction/ binding motif (SIM/SBM).

(C) Sumoylation can result in a conformational change of the modified target. If the modified target also contains a SIM/SBM, intramolecular interaction between SUMO and the SIM/SBM can lead to a conformational change. So far, this has only been reported for one target, thymine DNA glycosylase. (From: Geiss-Friedlander and Melchior, 2007)

(D) Phosphorylation of the transcription factor myocyte-specific enhancer factor 2A (MEF2A) by cyclin-dependent kinase 5 (CDK5) within a PDSM enhances its modification by SUMO, inhibiting transcription and resulting in synapse maturation. Calcium signalling results in dephosphorylation by the calcium-dependent phosphatase calcineurin and acetylation of MEF2A by the histone acetyl transferases CREB-binding protein (CBP) and p300, converting it from a transcriptional repressor to a transcriptional activator and causing synapse disassembly. (From: Gareau and Lima, 2010)

Vertebrates have four SUMO isoforms, SUMO 1-4. SUMO2 and SUMO3 share 97% sequence identity, thus they are referred to as SUMO2/3 (Geiss-Friedlander and Melchior, 2007). SUMO1 is only 50% identical in sequence to SUMO2/3 and, unlike the latter, it cannot form chains (Geiss-Friedlander and Melchior, 2007). In contrast to other ubiquitous family-members, SUMO4 is expressed only in the kidney, lymph nodes and the spleen (Geiss-Friedlander and Melchior, 2007). SUMO is conjugated to the target protein through a sequential enzymatic cascade by an E1 activating enzyme, E2 (Ubc9) conjugating enzyme and an E3 protein ligase (Geiss-Friedlander and Melchior, 2007). The same E2 enzyme processes all SUMO-target proteins, whereas several E3-ligase families have been identified, specific for different substrates (Geiss-Friedlander and Melchior, 2007). The SUMO acceptor lysine is present within a consensus SUMO motif $\Psi\text{KX(D/E)}$, where Ψ is a large hydrophobic and x any residue (Geiss-Friedlander and Melchior, 2007). This motif interacts directly with Ubc9. Longer motifs that include additional sequences have been identified in some substrates. Since phosphorylation enhances SUMOylation *in vivo* and *in vitro*, the phosphorylation dependant SUMO motif (PDSM) $\Psi\text{K(E/D)XXSP}$ was identified in proteins modified by Proline-directed kinases (Hietakangas, 2006). In proteins with negatively charged amino acid-dependent SUMO motif (NDSM) $\Psi\text{K(E/D)XXEEEE}$ (Yang *et al.*, 2006). The negative charge probably acts analogously upon phosphorylation and in residues nearby the SUMO-acceptor lysine (Geiss-

Friedlander and Melchior, 2007). Evidence is emerging how sumoylation can regulate the activity of some transcription factors. Many reports exist describing how SUMO-conjugation creates a surface with increased affinity for various histone deacetylases (HDACs), thereby converting dual transcription factors into repressors or enhancing the repressive potential of canonical co-repressors (Geiss-Friedlander and Melchior, 2007). Elk-1-SUMO interacts with HDAC2 (Yang and Sharrocks, 2004). Recruitment of HDAC6 is promoted after sumoylation of p300 coactivator (Girdwood *et al.*, 2003). Groucho corepressor is sumoylated on multiple lysine residues, which enhances its activity by creating an interaction surface for HDAC1 (Ahn *et al.*, 2009). Myocyte-specific enhancer factor 2A is phosphorylated by cyclin-dependent kinase 5 (CDK5) within the PSDM (Shalizi *et al.* 2006). Subsequent sumoylation turns Mef2A into a repressor resulting in synapse maturation (Shalizi *et al.* 2006). Dephosphorylation and acetylation of Mef2A by CBP/p300 convert it back to a transcriptional activator and cause synapse disassembly (Shalizi *et al.* 2006; Figure 6D). On the contrary, sumoylation of Ikaros transcription factor in thymocytes prevents its interaction with HDAC-dependent and HDAC-independent co-repressors and inhibits the ability of Ikaros to repress transcription (Gomez-del Arco *et al.*, 2004). Sumoylation of GATA-1, a regulator of haematopoiesis, aids sequestering of differentially regulated genes to distinct subnuclear compartments (Lee *et al.*, 2009). During erythroid maturation, sumoylation promoted the genes that are activated in collaboration with Friend of GATA (FOG-1) to migrate from the nuclear periphery (Lee *et al.*, 2009). FOG-1- and SUMO1-independent genes remained at the periphery (Lee *et al.*, 2009). Members of the polycomb repressive complex SUZ12 and EZH2 can also be modified by SUMO (Riising *et al.*, 2008). Sumoylation of EZH2 *in vitro* is required for the assembly of the PRC2 complex (Riising *et al.*, 2008). However the physiological role of PRC2-sumoylation has not been identified (Riising *et al.*, 2008).

There are many reports how sumoylation affects functions of proteins other than transcription factors and regulators (Geiss-Friedlander and Melchior, 2007). However, a general drawback of this emerging field is that the physiological role often remains enigmatic or inconclusive, shedding doubts that sumoylation might be redundant to other modifications and regulatory mechanisms. Evidence against such doubts comes from disruption of SUMO1 in mice, which caused embryonic lethality, while haploinsufficiency induced developmental defects (split lip and palate) (Alkuraya *et al.*, 2006). On the contrary, another study reported that *Sumo-1^{-/-}* mice displayed normal development while the previously observed phenotype was lacking (Zhang *et al.*, 2008). Therefore, SUMO2 and SUMO3 might compensate SUMO1 functions *in vivo* (Zhang

et al., 2008). In conclusion, further investigation is required to deepen the understanding of regulatory trajectories of sumoylation and the interplay of modification by SUMO1, 2, 3 or 4.

B-cell development

Transcription factors

The flexibility of transcriptional decisions results in a potential for differentiation to various cell types, each with specific role vital for some function in complex organisms. Lymphocytes originate from pluripotent haematopoietic stem cells (HSCs). The pluripotency derives from low-level expression of key regulatory genes of different hematopoietic lineages (Akashi, 2003). Some genes in ES cells and progenitors are embedded in “bivalent” chromatin (Bernstein *et al.*, 2006). Such regions allow rapid gene activation in response to appropriate stimuli provided by the interplay of transcription factors, chromatin regulators and cell signalling (Wang *et al.*, 2009). Gradually, dominant gene expression programs emerge restricting the differentiation potential (Ramirez *et al.*, 2011; Figure 7). HSCs develop into B cells through sequential differentiation via progenitor stages known as multipotent progenitor (MPP), lymphoid-primed multipotent progenitor (LMPP), all-lymphoid progenitor (ALP), B-cell biased lymphoid progenitor (BLP) and pre-pro-B cells (Inlay *et al.*, 2009). B lymphopoiesis occurs in the foetal liver and post-natal bone marrow (Ramirez *et al.*, 2011).

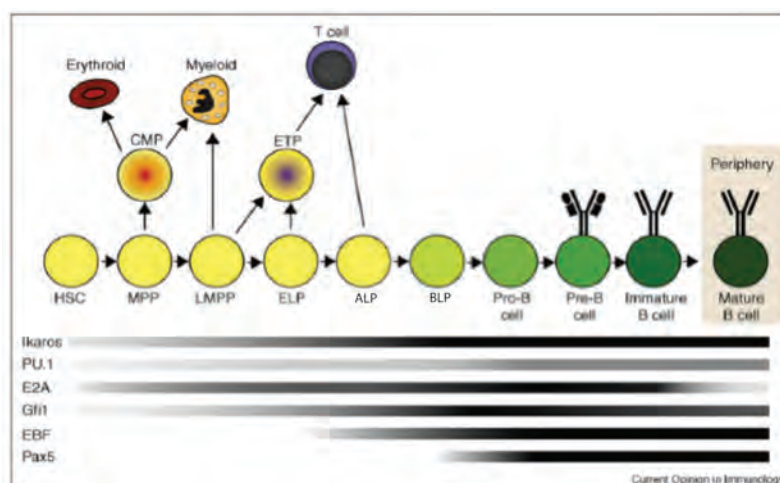


Figure 7. Transcription factor expression during B lymphopoiesis.

The progression of cells from hematopoietic stem cells through stages of B lymphopoiesis is shown. Shaded bars represent the levels of gene expression of transcription factors that are important in B cell development during the course of differentiation. HSCs (hematopoietic stem cells), MPPs (multipotent progenitors), LMPPs (lymphoid primed MPPs), ELPs (early lymphoid progenitors), ALPs (all lymphoid progenitors), BLPs (B-cell biased lymphoid progenitors), ETPs (early T lineage progenitors). Darker shading indicates increased gene expression. Important branch points during B lymphopoiesis are shown with arrows indicating alternative developmental pathways (Modified from: Ramirez *et al.*, Current Opinion in Immunology 2010, 22:177–1842010).

Differentiation towards the lymphoid lineage critically depends on several transcription factors and signalling receptors (Ramirez *et al.*, 2011; Figure 7). Ikaros (*Ikzf1*) is a Krüppel-like zinc finger transcription factor whose deficiency causes cell developmental arrest at the LMPP stage (Georgopoulos *et al.*, 1992). It can act as a transcriptional activator and repressor by recruiting the BAF chromatin remodelling or Mi-2/NuRD histone deacetylase complexes (Kim *et al.*, 1999; Sridharan and Smale, 2007). Ikaros is believed to be important for the expression of the Flt3 and IL-7 cytokine receptors at the MPP stage (Kirstetter *et al.*, 2002). Signalling from these receptors directs the gene expression program towards the lymphoid lineage (Ramirez *et al.*, 2011). Another important transcription factor at the LMPP stage is PU.1 (*Sfpi1*), an Ets domain transcription factor (Ramirez *et al.*, 2011). Low dosage of PU.1 favours B lymphoid differentiation, whereas high dosage the myeloid lineage (DeKoter and Singh 2000). Growth factor independent 1 (Gfi1) transcriptional repressor might mediate these opposing outcomes (Ramirez *et al.*, 2011). Gfi1 regulates the expression of PU.1 by displacing it from its autoregulatory element (Spooner *et al.*, 2009). Ikaros can upregulate Gfi in a subset of MPPs, thereby contributing to B cell differentiation at the expense of granulocytes (Spooner *et al.*, 2009).

E2A (*Tcf2a*) is a basic helix-loop-helix transcription factor occurring in two splice isoforms E12 and E47 (Bain *et al.*, 1994; Zhuang *et al.*, 2004). The homodimer of E47 is predominant in B cells (Beck *et al.*, 2009). E2A is important for the expression of lymphoid-lineage specific genes in MPPs (Semerad *et al.*, 2009). In LMPP E2A activates genes synergistically with Ikaros, PU.1, Gfi1 and interleukin-7 signalling (via Stat5) (Ramirez *et al.*, 2011). Most importantly, E2A induces the expression of early B cell factor 1 (*Ebf1*) leading to B-cell lineage specification (Kwon *et al.*, 2008).

Ebf1 is a helix-loop-helix transcription factor (Hagman *et al.*, 1993). E2A and Ebf1 bind in a coordinated way to a large set of loci with critical roles in early B cell development, including *Vpreb1*, *Vpreb3*, *Cd19*, *Bst1*, *Foxo1*, *Pou2f1*, *Cd79a* and *Pax5* (Lin *et al.*, 2010) thereby specifying the B-cell lineage (Figure 8). Subsequently Pax5 controls B-cell commitment at the transition to the pro-B cell stage by restricting the developmental potential of lymphoid progenitors to the B cell pathway (Nutt *et al.*, 1999).

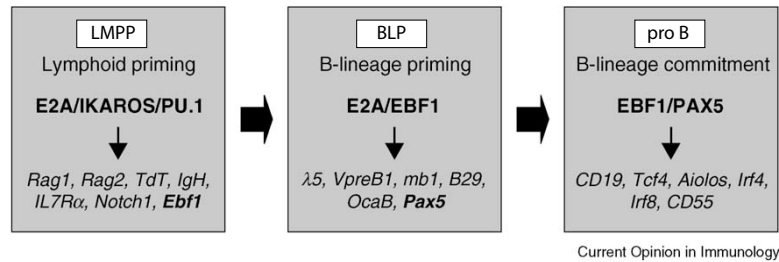


Figure 8. The stage specific crucial events involved in B lymphoid commitment. Crucial transcription factors as well as a set of relevant target genes are indicated in each of the boxes. LMPP: lymphoid myeloid primed progenitor. (From: Bryder and Sigvardsson, Current Opinion in Immunology 2010, 22:148–153).

The action of stage-specific transcription factors results in the expression of cell surface markers, according to growth factor requirements and sequential rearrangement of the immunoglobulin heavy chain locus (*Igh*) (Figure 9; Kondo *et al.*, 1997; Hardy *et al.*, 1991). The changes occurring during B-cell development are described as reviewed by Holmes *et al* (2007) and Monroe and Dorshkind, (2007). The earliest committed B-cell progenitors are large cycling cells with their *Igh* locus in either germ line or D_H-J_H configuration. These pro-B cells can be cultured *in vitro* on stromal cells in the presence of interleukin-7 (IL-7). They express $\lambda 5$, *VpreB*, *Igα* and *Igβ* (Monroe and Dorshkind, 2007). In frame V_H-D_H-J_H recombination leads to expression of a rearranged Igμ protein. Two Igμ chains pair with $\lambda 5$ and VpreB surrogate light chains forming the pre-B cell receptor (pre-BCR). Signalling through the pre-BCR induces proliferation and initiates recombination of the immunoglobulin light chain (*Igl*). Cells that have successfully rearranged *Igl*, express the B-cell receptor (BCR) on their surface. These immature B cells migrate from the bone marrow to the periphery. After several transitional stages, they become quiescent mature B cells that circulate through the blood and peripheral lymphoid organs. Mature B cells can be directly stimulated by a T-cell independent antigen and rapidly develop into antibody-secreting plasmablasts. In response to T-cell dependent antigens, B cells form germinal centres (GC). Provided with T-cell help, B cells in GCs proliferate, undergo somatic hypermutation and class switch recombination (CSR). Positively selected B cells, with high affinity antigen receptors, become terminally differentiated plasma cells, or memory cells (Holmes *et al*, 2007).

Up to the stage of terminal differentiation into antibody producing cells, the maintenance of B cell identity critically depends on the transcription factor Pax5 (Lin *et al.*, 2002; Mikkola *et al.*, 2002; Cobaleda *et al*, 2007; Figure 9).

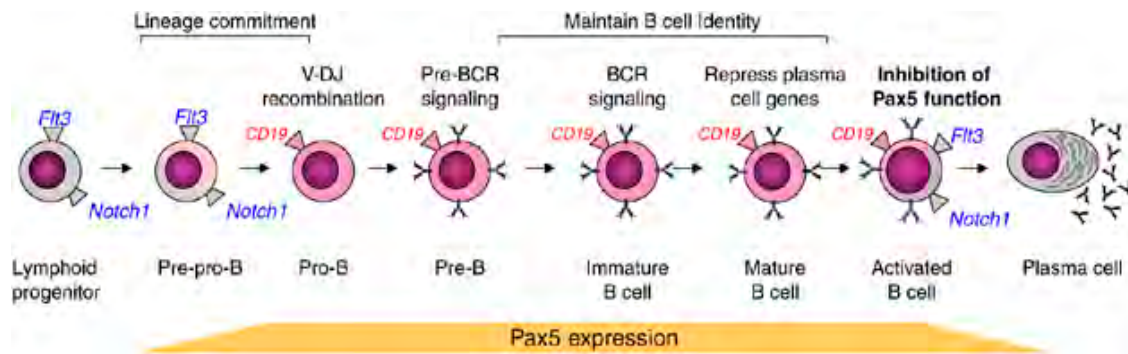


Figure 9. The expression and function of paired box gene 5 (Pax5) during B-cell development.

Simplified model of the stages of B-cell development. Pax5 expression is initiated at the pre-pro-B cell stage and maintained at a stable level throughout B-cell ontogeny before being downregulated during plasma cell development. Major functions of Pax5 are indicated on the upper portion of the figure. The cell surface expression of the protein products of two Pax5-repressed genes, Flt3 and Notch1, and one activated gene, CD19, is indicated. Pax5 function in activated B cells is inhibited on a post-translational level, allowing for the re-expression of Flt3 and Notch. Pax5 is then transcriptionally silenced in plasma cells. Pre-BCR, pre-B cell receptor; BCR, B cell receptor. (From: Holmes *et al.*, 2007)

Pax5 in B-cell lineage commitment and maintenance

B cell development critically depends on three transcription factors, E2A, Ebf1 and Pax5. In the absence of E2A and Ebf1, B cell developmental arrests at the earliest stage, before D_H-J_H rearrangement of the *Igh* locus (Zhuang *et al.*, 1994; Bain *et al.*, 1994). In Pax5 deficient embryos B-cell development is blocked before the appearance of B220⁺ progenitors in the fetal liver, whereas it continues to the c-Kit⁺B220⁺ progenitor stage in the bone marrow of adult mice (Urbanek *et al.*, 1994; Nutt *et al.*, 1997). The c-Kit⁺B220⁺ Pax5^{-/-} progenitors express E2A, Ebf1 and several other B-cell specific genes (Nutt *et al.*, 1997). They resemble early B-cell precursors can be propagated at this stage in culture with stromal ST2 cells and IL-7 (Rolink *et al.*, 1991). However, they are still not committed to the B-cell lineage, as they survive even upon depletion of IL-7 (Nutt *et al.*, 1999). Furthermore, under stimulation with appropriate cytokines they differentiate into functional macrophages, osteoclasts, dendritic cells, granulocytes and natural killer cells (Nutt *et al.*, 1999). During culture with OP9 stromal cells expressing Delta-like 1(DL1) Pax5^{-/-} cells differentiate into T cells (Höflinger *et al.*, 2004). Upon transplantation into lethally irradiated recipient mice, Pax5^{-/-} pro-B cells home to the bone marrow, undergo self-renewal (Rolink *et al.*, 1999) and give rise to all the hematopoietic cell types mentioned above (Schaniel *et al.*, 2002). Similarly to HSCs they retain their self-renewal and long-term reconstitution potential even after more than 150 cell divisions (Schaniel *et al.*, 2002). Unlike HSCs Pax5^{-/-} pro-B cells differentiate more efficiently into the lymphoid lineage than the myeloid lineage. The latter lineage can be strongly induced by ectopic expression of a myeloid transcription factor (Haevey *et al.*, 2003). Restoration of Pax5 expression suppresses the multilineage potential of Pax5^{-/-}

progenitors and directs them to become mature B cells (Nutt *et al.*, 1999). In conclusion, Pax5 is the B cell lineage commitment factor needed to restrict alternative developmental options of lymphoid progenitors (Mikkola *et al.*, 2002; Nutt *et al.*, 1999). Furthermore, Pax5 is continuously required to maintain the B lymphoid transcription program from early stages of B cell development until the onset of plasma cells differentiation upon antigen stimulation of the BCR (Mikkola *et al.*, 2002). Conditional *Pax5* inactivation enables committed pro-B cells to differentiate into dendritic cells, macrophages or osteoclasts, in response to appropriate cytokines (Mikkola *et al.*, 2002). Even mature peripheral B cells, upon conditional loss of Pax5, dedifferentiate to uncommitted progenitors in the bone marrow and can rescue T lymphopoiesis in the thymus of T-cell deficient mice (Cobaleda *et al.*, 2007a). Despite their advanced differentiation stage, mature B cells retain an extraordinary plasticity that is actively kept under control by Pax5 (Cobaleda *et al.*, 2007a). This finding renders Pax5 not only the commitment factor, but also the guardian of B cell identity and function (Cobaleda *et al.*, 2007b). Only at the final transition stage in B cell development, from mature B cells to antibody-producing plasma cells, does Pax5 become inactivated and the transcriptional program radically changes (Lin *et al.*, 2002). Many Pax5-repressed target genes becoming activated (Lin *et al.*, 2002). The switch between the two transcriptional programs is still not completely understood. Blimp1 (*Prdm1*) orchestrates the transcription program in plasma cells, by binding and repressing *Pax5*, while the reverse is true in mature B cells (Lin *et al.*, 2002). However many Pax5 repressed target genes become activated, prior to Blimp1 expression, when Pax5 protein can still be detected (Kallies *et al.*, 2007). The function of Pax5 might initially be inactivated by a post-translational modification or an interaction with an unknown protein (Kallies *et al.*, 2007). Subsequently Pax5-mediated repression of *Prdm1* is released, which leads to the stabilisation of the plasma-cell differentiation program (Kallies *et al.*, 2007).

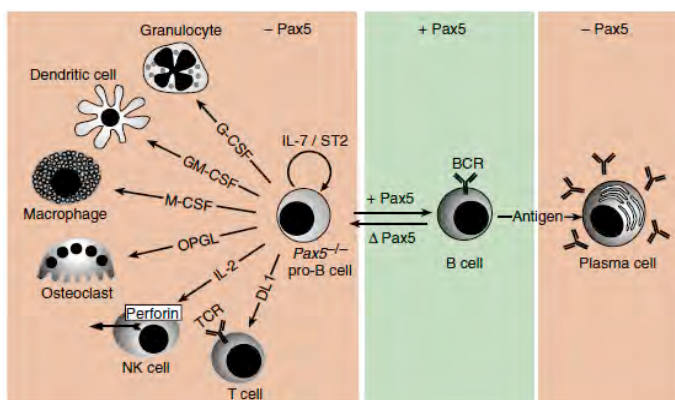


Figure 10. B cell lineage commitment by Pax5. The uncommitted *Pax5*^{-/-} pro-B cells are able to differentiate into several hematopoietic cell types either in vitro in the presence of the indicated cytokines or in vivo after transplantation into recipient mice. Conditional Pax5 deletion (Δ Pax5) results in retrodifferentiation of B lymphocytes to an uncommitted progenitor cell stage. OPGL, osteoprotegerin ligand (also known as RANKL or TRANCE); ST2, stromal ST2 cells; TCR, T cell receptor. (From: Cobaleda *et al.*, 2007)

Molecular mechanism of B-lineage commitment

At B cell commitment Pax5 downregulates classes of genes that code for proteins involved in cell-cell communication, cell adhesion, migration, cellular metabolism and nuclear processes (Delogu *et al.*, 2006). Several genes code for cell surface receptors needed for differentiation or survival of hematopoietic progenitors such as Flt3 and Ly6a (Sca-1) (Delogu *et al.*, 2006). Other repressed genes code for receptors that signal during differentiation of alternative hematopoietic lineages. The macrophage colony stimulating factor (M-CSF) receptor (*Csr1r*) is required for myeloid and Notch for T lineage differentiation, respectively (Nutt *et al.*, 1999; Souabni *et al.*, 2002); CD28 is a costimulatory molecule on the surface of T cells; FcR γ is a signal-transducing component of the Fc receptors of myeloid cells; Gp49b is an inhibitory receptor expressed in natural killer and mast cells. Intracellular signalling molecules include the adaptor molecule Grap2 of the (pre)T-cell receptor and M-CSF receptor signalling pathways and the transmembrane adaptor NTSL of the Fc γ and Fc ϵ receptor signalling pathways (Delogu *et al.*, 2006). Evidently, by downregulating multiple signal transduction pathways, Pax5 prevents the responsiveness of committed B lymphocytes to lineage inappropriate signals (Delogu *et al.*, 2006). Additionally, Pax5 mediates adherence of B cells to bone marrow niches consisting of IL-7-expressing CXCL12⁺ stromal cells (Tokoyoda *et al.*, 2004) by downregulating multiple components of chemokine and integrin signalling pathways (Delogu *et al.*, 2006).

B-cell development and function result from a complex regulatory network of proteins involved in B cell signalling, adhesion, migration, antigen presentation and germinal-center B cell formation (Schebesta *et al.*, 2007). Pax5 controls the pre(BCR) signalling by activating genes coding for cell surface co-stimulatory molecules, transmembrane components, cytoplasmic signal transducers to downstream transcription factors (Schebesta *et al.*, 2007). Furthermore, Pax5 maintains and activates genes encoding other transcription factors Ebf1 (Reossler *et al.*, 2007), Lef1 (Nutt *et al.*, 1998), CIITA (Horcher *et al.*, 2001). Direct Pax5-activated target genes display active histone marks on promoters (H3K9ac⁺, H3K4me3⁺, H3K4me2⁺) and enhancers (H3K9ac⁺, H3K4me3⁺, H3K4me3⁻), which are absent in *Pax5*^{-/-} cells (Schebesta *et al.*, 2007). Pax5 induces active chromatin marks at its target genes by recruiting chromatin remodelling and modifying complexes (McManus *et al.*, 2011).

How can Pax5 play a dual role of repressing lineage-inappropriate genes and simultaneously activating B-cell specific genes (Nutt *et al.*, 1999)? Firstly, in order to regulate the majority of the B cell transcriptome, the structure of Pax5 must favour binding to a large set of target gene. Secondly, recruitment of different interacting proteins probably determines whether the target gene will be set “on” or “off”.

Pax protein family

The evolution and function of Pax protein family was reviewed by Bouchard *et al.* (2003). The Pax proteins constitute a family of transcription factors essential for early development from invertebrates to humans. All family members contain a highly conserved DNA-binding paired domain. The mammalian gene family consists of nine members which can be classified into four groups based on sequence similarity and presence of different structural motifs: Pax1/9, Pax3/7, Pax4/6 and Pax2/5/8 (Figure 11). Pax3/7 and Pax4/6 homologues contain a full homeodomain, as the second DNA-binding motif. Pax2/5/8 family contains a partial homeodomain, which serves as a protein-protein interaction motif. Pax1/9 group completely lacks the homeodomain. Other conserved sequences of Pax proteins are the octapeptide motif and the C-terminal proline-serine-threonine (PST)-rich region.

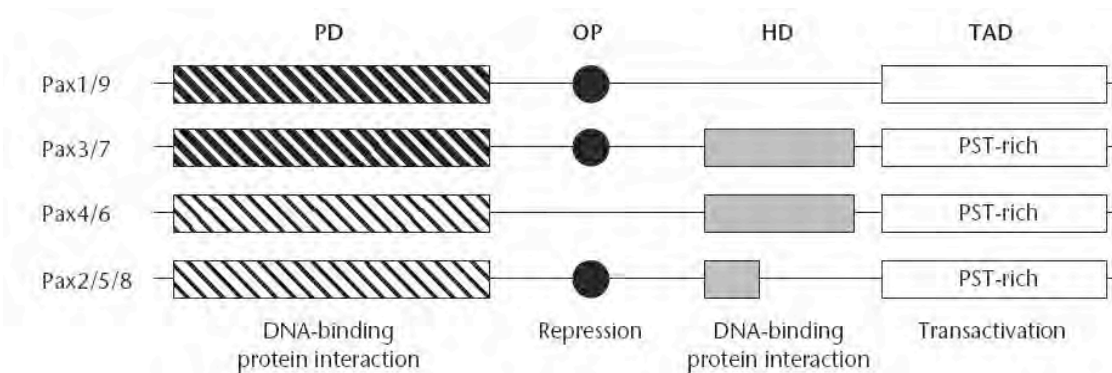


Figure 11. Structure and classification of mammalian Pax proteins.

The different domains of Pax proteins, and their function, are indicated. The consensus sequence of the conserved octapeptide (OP) is (Y/H)SI(N/D/S)GILG. PD, paired domain; HD, homeodomain, TAD, transactivation domain; PST, proline-serine-threonine rich region (From: Bouchard *et al.*, 2008)

The paired domain consists of 128 amino acids structured in N- and C-terminal subdomains. Each subdomain contains a homeodomain-like helix-turn-helix (HLTH) motif and contacts nucleotides in adjacent major grooves of the DNA helix (Xu *et al.*, 1999). Flexible linker sequences connect both HLTH motifs and interact with the minor DNA groove. As a consequence of this structure, the consensus DNA-recognition motif is 17 nucleotides long, and contains two half-sites. Each half-site independently contributes to the overall binding affinity. However naturally occurring Pax5-binding sites considerably deviate from the consensus sequence (Czerny *et al.*, 1993). The paired domain is highly conserved in Pax homologues throughout animal kingdom, from lower eukaryotes to humans. Although generally considered to

be a DNA-interaction domain, the paired domain can by itself recruit different Ets family proteins and cooperatively increase their DNA-binding affinity (Garvie *et al.*, 2001).

At the onset of vertebrate evolution, four *Pax* subgroups have diversified by gene duplications to contain multiple members. During development, different members of the same subgroup are often expressed hierarchically in spatially and temporally overlapping patterns. *Pax2/5/8* genes form an organizing centre at the midbrain-hindbrain boundary (MHB) which gives rise to the midbrain and the cerebellum. *Pax2* is the first to be expressed and regulates transcription of *Pax5* and *Pax8* (Ye *et al.*, 2001). Apart from overlapping expression, members of the same subgroup have acquired unique expression patterns and roles, which I will discuss further by using *Pax5* as an example.

In addition to the developing midbrain, *Pax5* is expressed in the adult testis and in all stages of B-cell development except in terminally differentiated plasma cells (Barberis *et al.*, 1990). Human and mouse *Pax5* differ in only 3 amino acids, suggesting that *Pax5* proteins are under high evolutionary constraints even outside of their paired domain (Adams *et al.*, 1992). Whereas bipartite nature of the paired domain allows binding to a large variety of DNA sequences, other domains and sequence motif are responsible for recruitment of partner proteins, which can result in positive or negative regulation of target genes.

The partial homeodomain of *Pax5* is involved in interaction with the TATA-box binding protein (TBP) of the TFIID basal transcription initiation complex (Eberhard and Busslinger, 1999). Discovery of *Pax5*-mediated recruitment of this complex provided a direct link to the mechanism how *Pax5* can initiate the B-cell specific transcription of *CD19* gene, which instead of a TATA box contains a high affinity *Pax5*-binding site. *Pax5* can also recruit the chromatin remodelling BAF complex by interacting with its core component Brg1 (Barlev *et al.*, 2003; McManus *et al.*, 2011). Transcriptional regulator Daxx interacts with *Pax5* either as a co-repressor or, can mediate the recruitment of the co-activator CBP (Emelyanov *et al.*, 2002), a member of a histone acetyltransferase complex. The octapeptide motif of *Pax5* can recruit Grg4 (TLE4), a member of the Groucho family (Eberhard *et al.*, 2000). Groucho proteins localise in the nucleus, although they lack a DNA binding domain and are recruited by DNA-binding transcription factors to specific control regions, where they bind to N-terminal tails of histone H3 (Palaparti *et al.*, 1997). They interact with histone deacetylases (Chen *et al.*, 1999) leading to the removal of active chromatin marks. The C-terminal domain of *Pax5* contains 88 amino acids of which 11 are prolines while more than 25 are hydroxylated (serine, threonine and tyrosine [PST]) (Adams *et al.*, 1992). Consequently, the domain is probably exposed on the surface of the protein as an unstructured hydrophilic tail (Dörfler and Busslinger, 1996). The 55 amino acids

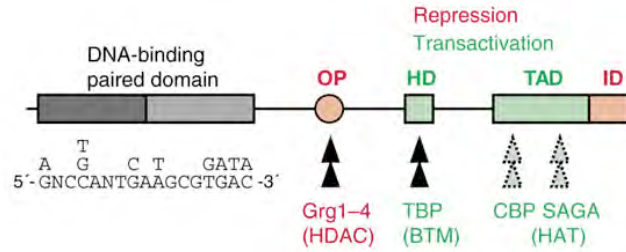


Figure 12. The functional domains and consensus recognition sequence of Pax5 with known interacting proteins. The consensus sequence is based on the alignment of 30 high-affinity Pax5-binding sites. N denotes a position where each of the four nucleotides is acceptable. Dashed gray arrows indicate a likely but not yet proven interaction between the Pax5 transactivation domain and the CREB-binding protein (CBP) or SAGA complex. Grg1-4 refers to the four full-length members of the mammalian Groucho protein family. BTM, basal transcription machinery; HAT, histone acetyltransferase; HD, partial homeodomain; HDAC, histone deacetylase; ID, inhibitory domain; OP, conserved octapeptide; TAD, transactivation domain; TBP, TATA-binding protein. (From: Cobaleda *et al.*, 2007)

of these PST-rich sequences act as a potent transactivation domain (TAD). The C-terminal amino acids of the PST-rich region exert a strong negative regulation on the upstream transactivation domain (Dörfler and Busslinger, 1996). If the paired domain is fused directly to the PST-rich region, the mutant has a transactivation potential equal to the full-length protein (Dörfler and Busslinger, 1996). Upon deletion of the C-terminal inhibitory sequence, the transactivation potential increases by 30 fold (Dörfler and Busslinger, 1996). Furthermore, the transactivation and the adjacent inhibitory domains are structurally and functionally conserved in all members of the Pax2/5/8 protein subfamily, the zebrafish Pax-b and the distantly related sea urchin Pax-258 (Dörfler and Busslinger, 1996). The latter suggests that the PST-rich region is an important regulatory module determining the transcriptional activity of Pax2/5/8 family members (Dörfler and Busslinger, 1996). Similar mechanism of auto regulation exists in C/EBP β , where the regulatory domain lowers the transcription-activation potential of the N-terminally positioned transactivation domain (Kowenz-Leutz, 1994). Recent studies have unravelled that the mechanism of C/EBP β -autoregulation depends on posttranslational modifications (Kowenz-Leutz *et al.*, 2010; Figure 13). Methylation of conserved arginines by PRMT4/CARM1 interferes with recruitment of the BAF-chromatin remodelling complex and the transcription regulatory Mediator complex (Kowenz-Leutz *et al.*, 2010). Phosphorylation downstream ras/MAPkinase signalling abolishes the interaction between PRMT4/CARM1 and C/EBP β and consequently methylation (Kowenz-Leutz *et al.*, 2010). According to these findings a model of negative crosstalk between phosphorylation and methylation has been proposed: in the absence of

activating signals, C/EBP β becomes methylated and remains transcriptionally inactive or displays repressor functions (Kowenz-Leutz *et al.*, 2010). Phosphorylation downstream MAPkinase signalling transiently prevents methylation of C/EBP β enabling co-activator recruitment and gene activating functions (Kowenz-Leutz *et al.*, 2010).

Flexibility of the consensus-binding motif recognised by the paired domain enables regulation of a large number of genes. The complexity increases if we imagine that other context-dependant transcription factors can influence the DNA-binding affinity of Pax5 and *vice versa*. Upon binding to target genes, Pax5 can recruit co-repressor or co-activator complexes (McManus *et al.*, 2011). These two structural properties endow Pax5 a dual ability to elicit opposing transcriptional outcomes on different genes in the same cell type.

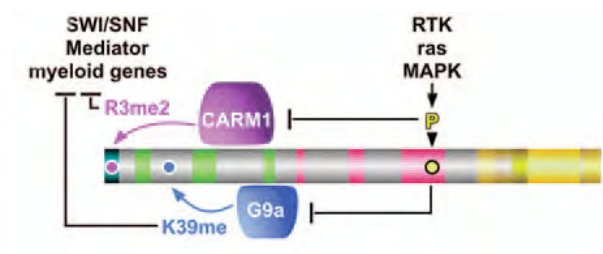


Figure 13. Scheme of the structure and crosstalk between phosphorylation and arginine/lysine methylation in C/EBP β . Green region represents the transactivation domain, pink the regulatory domain of C/EBPs. Towards the C-terminus follow: basic/acid motif, the DNA binding region, the fork and the leucine zipper regions. Model of signaling through receptor tyrosine kinase (RTK) / ras-MAPK pathways causes phosphorylation within the regulatory domain. Phosphorylation abrogates interaction with G9a and PRMT4 that methylate C/EBP β K39 and R3, respectively. Methylation at K39 interferes with activation of myeloid genes and methylation of R3 interferes with recruitment of BAF (SWI/SNF) that is required for the activation of a subset of myeloid genes. (From: Leutz *et al.*, 2011)

V(D)J recombination

The biological purpose of lymphocytes is protection against the continuously changing world of pathogens. Many of them are simple organisms that can rapidly mutate, creating more serious challenges. Flexibility of conforming to these changes stems from the enormous diversity of antigen receptor repertoire generated by B and T lymphocytes.

Antigen receptors consist of two identical immunoglobulin (Ig) heavy chains (HC) and two identical immunoglobulin light chains (LC). The C-terminal portion of Ig chains is called the constant region (C_H). Upon antigen stimulation, the constant region determines the effector function of antibodies, the secreted forms of Ig molecules. The N-terminal region of Igs is highly variable, in every B cell clone specific for only one epitope of a foreign antigen.

The structure of the immunoglobulin heavy chain locus

Exons of the immunoglobulin variable region are assembled from discontinues V (variable), D (diversity) and J (joining) gene segments of the murine *Igh* locus, which spreads over about 3 Mb at the long arm of chromosome 12 (Perlot and Alt, 2008). From the 3' to the 5' end the *Igh* locus consists of 8 constant region genes (C_H), 4 J_H gene segments, 10-13 D_H gene segments and around 150 V_H gene segments spread over 2.5 Mb (Figure 14). The V_H genes are classified into 16 partially interspersed gene families. Depending on their position in the *Igh* locus they can be divided into proximal (3' end of the V_H cluster; V_H 7183 and V_H Q52), intermediate (e.g. V_H S107) and distal (5' end of the V_H cluster; V_H J558) V_H genes (Figure 14). The V_H region contains many pseudogenes and a large proportion (39%) of interspersed, mostly LINE, repeats (Johnston *et al.*, 2006).

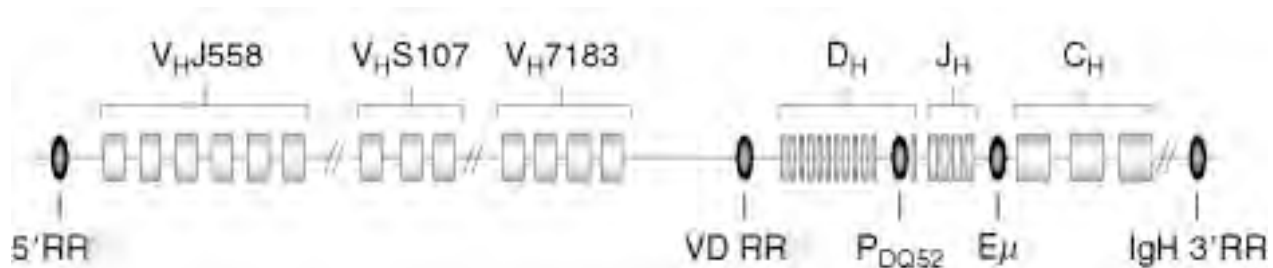


Figure 14. Schematic depiction of the murine IgH locus.

V_H , D_H , J_H gene segments and C_H exons are shown as rectangles, known and potential regulatory elements as ovals. The V_H families V_H J558, V_H S107, and V_H 7183 are depicted as examples for distal, intermediate, and proximal V_H families, respectively. The cis-regulatory elements PDQ52 (promoter of DQ52), $E\mu$ (intronic enhancer), and IgH 3'RR (IgH 3' regulatory region) are depicted. The potential regulatory element VD RR (V_H - D_H intergenic regulatory region) is depicted. Drawing not to scale. (From Alt and Perlot, 2008)

Several cis-regulatory elements have been defined. At the very 3' end, the 3'regulatory region (3'RR) is composed of seven DNase hypersensitive sites (Figure 14, 15; Perlot and Alt, 2008). The intronic enhancer, E_μ, is located in the intron between the J_H4 and C_H exons (Perlot and Alt, 2008). Flanking sequences upstream V_H gene segments contain conventional TATA-box (J558 and SM7 families) or TATA-related sequences and ATGCAAAT octamer-sequence motif for binding of POU- (Pou-domain containing octamer binding protein) family transcription factors Oct1 and Oct2 (Johnston *et al.*, 2006). Related V_HJ558 and V_HSM7 gene families contain, at conserved position downstream from the octamer, a downstream Ig control element (DICE), (Johnston *et al.*, 2006). Many antisense transcripts, identified throughout the locus, arise from less-well defined promoters (Perlot and Alt, 2008).

Rag1/2, “the V(D)J endonuclease”

V(D)J recombination is a process of assembling antigen receptor variable region genes from component V_H, D_H and J_H gene segments (Perlot and Alt, 2008). It proceeds in an ordered and differentiation-stage specific manner, with D_H-J_H joining preceding V_H-DJ_H (Perlot and Alt, 2008). Enzymes with two types of activities mediate this process: lymphocyte specific (Rag1/Rag2 endonuclease, TdT) and generally expressed (nonhomologous end joining factors (NHEJ)). Together these factors comprise the “V(D)J recombinase” (Alt, Immunology Summer School, 2009). All V_H, D_H and J_H gene segments are flanked by recombination signal sequences (RSSs), which consist of a conserved palindromic heptamer and a conserved AT-rich nonamer separated by a 12 or 23 bp long spacer (Perlot and Alt, 2008). D_H gene segments are flanked by 12 bp RSS on both sides, whereas V_H and J_H are flanked by 23 bp RSS (Perlot and Alt, 2008). The dimer of Rag1 and Rag2 proteins (Rag1/2), recognizes and binds a pair of RSS with different spacer lengths, which allows for efficient V(D)J recombination only between gene segments flanked by 12 bp and 23 bp RSS (Perlot and Alt, 2008). This prevents direct V_H-J_H joining in the context of the 12/23 rule (Perlot and Alt, 2008). Binding to a compatible pair of RSS activates the endonuclease activity of Rag1/2, which cuts precisely at the border of coding and recognition sequences (Perlot and Alt, 2008). The RSS remain blunt-ended, whereas the coding ends (CEs) of V_H, D_H or J_H segments form closed hairpins that are recognized and joined by members of the NHEJ pathway (Perlot and Alt, 2008).

Rag-mediated breaks are potentially dangerous for the integrity of the genome and viability of the cells since erroneous V(D)J recombination can cause chromosomal translocations contributing to the development of leukemias and lymphomas (Perlot and Alt, 2008). To prevent this from happening, the process is tightly regulated on several levels as reviewed by Perlot and

Alt (2008): (1) lymphoid restricted expression of *Rag1/2*, (2) complex regulation of chromatin accessibility of *Igh* gene segments.

Rag1/2 is expressed only in B and T cells and can be detected already at the CLP stage (later subdivided to ALP and BLP, Inlay *et al.*, 2009), therefore D-J_H recombined *Igh* can be found outside of the B-cell lineage (Perlot and Alt, 2008). On the contrary, V_H to DJ_H joining occurs only in pro-B cells. In the case of nonfunctional rearrangement, the second *Igh* allele is recombined. Successful recombination is followed by expression of the μ HC protein which pairs with VpreB and λ 5 surrogate light chains forming a pre-B cell receptor (pre BCR) (Perlot and Alt, 2008). Signaling from the pre-BCR through Syk (spleen tyrosine kinase) activates the phosphoinositide-3-kinase (PI3K) pathway (Herzog *et al.*, 2009). This leads to phosphorylation of Foxo transcription factor and targets it for nuclear export and degradation (Herzog *et al.*, 2009). Since Foxo activates the *Rag* genes, its inactivation decreases *Rag1/2* transcription (Herzog *et al.*, 2009). At the same time phosphorylation reduces the half-life of Rag2 (Herzog *et al.*, 2009). This sequence of events stimulates proliferation, prevents the rearrangement of the second *Igh* allele and blocks *Igk* rearrangement (Herzog *et al.*, 2009). In the meantime, a negative feedback loop that is generated through BLNK (SLP65), the signalling adaptor of Syk, leads to reactivation of Rag2, *Igk* light chain rearrangement and differentiation to mature B-cell stage (Herzog *et al.*, 2009).

Accessibility control

B and T lymphocytes share a common Rag1/2 endonuclease although the T cell receptor gene segments are recombined only in thymocytes, whereas the B cell receptor only in B cells (Osipovich and Oltz, 2010). Furthermore, the *Igh* rearrangement occurs monoallelically despite Rag1/2 activity (Osipovich and Oltz, 2010). The “accessibility model” for control of V(D)J recombination predicts that changes in chromatin landscape, within *Igh* and *Tcr* gene segments, determine their relative accessibility to the V(D)J recombinase complex (Osipovich and Oltz, 2010). These changes are orchestrated through the interplay of transcription factors, epigenetic regulators, and genetic *cis* elements of *Igh* DNA sequence (Osipovich and Oltz, 2010).

The recruitment of Rag1/2 endonuclease is controlled by cell type specific accessibility of genes, which can be regulated at the level of chromatin modification and localization of the *Igh* locus to different nuclear compartments (Perlot and Alt, 2008; Osipovich and Oltz, 2010).

Chromatin modifications

Active chromatin marks correlate with V(D)J recombination. Hyperacetylated histone 4, H3K9ac and H3K4me2 are present in the DJ_H region in early pro-B cells poised to undergo D_H to J_H rearrangement, but they are almost absent in thymocytes (Chakraborty *et al.*, 2007). Histone modification state in the D_H-C_μ region is heterogeneous (Chakraborty *et al.*, 2007). Only the 5'DFL16.1 and the 3' DQ52 D_H segments possess active H3 and H4 acetylation and H3K4me2 marks, (Chakraborty *et al.*, 2007; Malin *et al.*, 2010) thus they are recombined with the highest frequency (Chakraborty *et al.*, 2007). The remaining D_H gene segments are in the middle of a repeated sequence approximately 4 kb long (Chakraborty *et al.*, 2007), they display lower levels of antisense transcripts and their hypoacetylated heterochromatic state is constantly maintained by action of histone deacetylases (Chakraborty *et al.*, 2007). J_H segments 3' from the DQ52 contain the highest levels of acetylated and H3K4 tri-methylated histones in the D_H-C_μ domain (Chakraborty *et al.*, 2007; Malin *et al.* 2010). Since active promoters (PQ52 and iE_μ) flank the J_H region, the transcription factors bound to these promoters might be the cause of hyperacetylation (Chakraborty *et al.*, 2007). Brg1 (the ATPase subunit of BAF complex) associates with the hyperacetylated chromatin in immunoglobulin loci (Morshead *et al.*, 2003) and is required for germline *Igh* transcription (Osipovich *et al.*, 2010). Active transcription in this region probably recruits RNA polymerase-associated histone methyltransferases (HMTs), leading to enrichment of H3K4me3 (Osipovich *et al.*, 2010). Rag2 protein contains a C-terminal PhD finger domain with high binding affinity for H3K4me3 histone modification enriched at promoters of actively transcribed genes (Matthews *et al.*, 2007). Mutation of a conserved tryptophan (W453) in the PhD domain impairs V(D)J recombination (Liu *et al.*, 2007). ChIP analysis demonstrated that Rag2 binds active promoters throughout the genome (Ji *et al.*, 2010). Therefore, targeting of Rag1/2 endonuclease activity likely depends on specific binding of Rag1 to accessible RSSs enriched for H3K4me3 (Ji *et al.*, 2010).

Although it was previously considered that chromatin activation of the V_H genes precedes and is necessary for V_H-DJ_H recombination, a recent extensive ChIP-on-Chip analysis showed no, or only very small amounts of active histone modifications H3K4me2 and H3K9ac at most V_H genes in pro-B cells (Malin *et al.*, 2009). Active chromatin was consistently present only at V_H3609 gene segments, which are interspersed within the distal V_HJ558 gene cluster, and V_HSM7 genes (Malin *et al.*, 2009). Therefore, germline transcription and V_H-DJ_H recombination can proceed in the absence of overt chromatin activation of V_H genes (Malin *et al.*, 2009).

Repressive histone mark H3K9me2 is absent in the D-J_H region of pro-B cells, but present in thymocytes (Chakraborty *et al.*, 2007). In agreement with the observation that distal V_H gene

families cannot be rearranged upon loss of Pax5, although the acetylation pattern of the locus remains unchanged, removal of H3K9me2 marks in the V_H region depends on Pax5 (Hesslein *et al.*, 2003; Nutt *et al.*, 1999). Interestingly, the same phenotype emerges if a Polycomb group (PcG) protein Ezh2 is deleted. With its H3K27 methyltransferase activity, PcG propagates the silenced chromatin state (Ringrose and Paro, 2004), therefore the cause of the knock-out phenotype remains enigmatic.

Cytosines in the mammalian DNA can be methylated in CpG dinucleotides and thereby inhibit binding of transcription factors, or recruit methyl-CpG binding proteins (Zhou *et al.*, 2011). They enforce a silent chromatin state by recruiting complexes with histone deacetylase activity (HDACs) (Jaenisch and Bird, 2003). Methylated RSSs can abolish Rag1/2-mediated cleavage and V(D)J recombination (Whitehurst *et al.*, 2000).

In conclusion, during *Igh* locus recombination the following sequence of epigenetic events is probable: the V(D)J endonuclease Rag1/2 is tethered to the J_H elements through specific binding of H3K4me3 by the PhD finger of Rag2 (Matthews *et al.*, 2007). From this proximal location Rag1/2 may act to form a synapse with D_H and V_H elements (Malin *et al.*, 2009). After D_H-J_H recombination in B lymphoid progenitors, the rearranged D_H element is incorporated into the active chromatin domain at the J_H-E μ region, which facilitates subsequent V_H-DJ_H recombination (Malin *et al.*, 2009).

Germline transcripts

Germline transcription precedes rearrangement of V_H, D_H and J_H gene segments. Sense germline transcripts originate at promoters upstream of the *Igh* gene segments and their expression pattern correlates with the accessibility of chromatin, as reviewed by Perlot and Alt (2008). Germline *Igh* is transcribed from the PDQ52 promoter of the 3' most D_H segment towards the C μ . After D-J_H rearrangement the recombined DJ_H element is transcribed. Unrearranged germline transcripts arise from individual V_H promoters and get polyadenylated. Antisense germline transcription also promotes active chromatin (Bolland *et al.*, 2004), increases the accessibility of the locus and correlates with active V_H to DJ_H recombination. Antisense D_H transcription and D_H to J_H recombination are reduced in mice lacking the intronic enhancer E μ (Afshar *et al.*, 2006; Perlot *et al.*, 2005).

E μ -dependent antisense transcription before D_H to J_H recombination extends into the V_H-D_H intergenic region before and after V(D)J recombination but terminates 40 kb from the first V_H gene (Featherstone *et al.*, 2010). It is suggested that subsequent V_H antisense transcription before V_H to DJ_H recombination must be actively prevented (Featherstone *et al.*, 2010). The V_H-

D_H region contains six hypersensitive sites ([HS]; Featherstone *et al.*, 2010). One conserved HS upstream the first D_H gene locally regulates D_H genes, whereas two other HSs mark a sharp decrease in antisense transcription, bind CTCF *in vivo* and have enhancer-blocking activity (Featherstone *et al.*, 2010). These HSs may form an insulator that prevents spreading of E_μ-dependent chromatin opening into the V_H region. Thus they may control ordered V(D)J recombination (Featherstone *et al.*, 2010). Antisense V_H transcripts comprise one or multiple V_H segments. Their starting and ending sites are unknown (Perlot and Alt, 2008).

In conclusion, both sense and antisense transcription presumably increase the accessibility of the *Igh* locus and poise it for V(D)J recombination.

***Igh* locus control through *cis*-regulatory elements**

The two most important *cis*-regulatory elements in the *Igh* locus are the E_μ enhancer and the 3' regulatory region, which are initial binding sites for transcription factors and origins of active chromatin marks. Other *cis*-regulatory elements upstream of V_H gene segments become active prior to V_H to D-J_H recombination. (Perlot and Alt, 2008)

The E_μ enhancer spans over 700 bp and can be subdivided to 220 bp long core sequence and the flanking matrix attachment regions (MARs) (Subrahmanyam and Sen 2010). Deletion of the E_μ in B cells and in the germline of mice reduced D-J_H rearrangement and severely impaired V_H to DJ_H rearrangement (Afshar *et al.*, 2006; Perlot *et al.*, 2005). Although it was suggested that PDQ52 promoter might compensate the role of E_μ, the double mutant with deleted E_μ and PDQ52 did not display a more severe V(D)J phenotype (Afshar *et al.*, 2006). The knockout mice have reduced H3/H4 histone acetylation throughout the D_H-C_μ including the DFL16.1, DQ52 and J_H regions (Chakraborty *et al.*, 2009). Therefore E_μ positively regulates histone acetylation at distantly located elements. Interestingly the long distance effects are asymmetric within the D_H-C_μ region; in the 3' direction the acetylation drops to background within 15kb, while in the 5' end acetylation extends more than 50 kb away from the E_μ to DFL16.1 (Chakraborty *et al.*, 2009). The deletion reduced germline transcription of proximal V_H genes located between 150 and 450 kb upstream. This finding showed that the long-range influence of E_μ extends even further. The E_μ seems to act in patches: it activates DFL16.1 while skipping the intervening D segments and the transcription of V_H7183 genes without any influence on the *Adam6* genes, located in between DFL1.16 and proximal V_H genes (Chakraborty *et al.*, 2009).

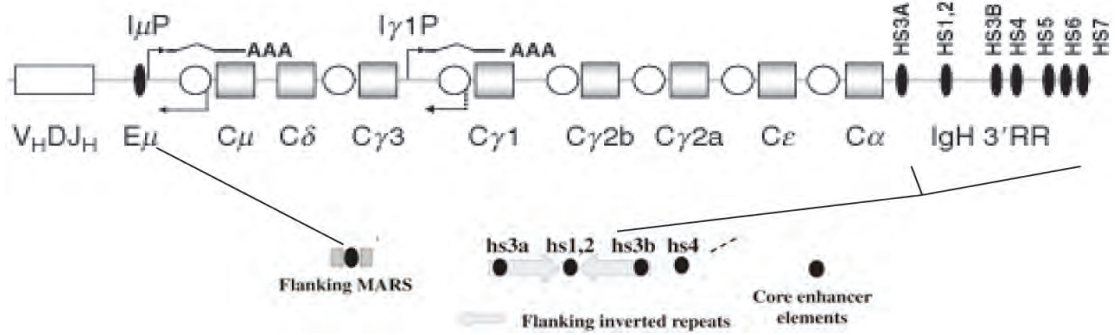


Figure 15. Schematic depiction of the the 3' part of the IgH locus.

An assembled V_HDJ_H exon is shown as a white rectangle, C_H genes as squares, E_μ and individual DNaseI hypersensitive sites within the IgH 3'RR are depicted as black ovals, switch regions as white circles. I promoters are located upstream of every switch region, only μ and $\gamma 1$ I promoters ($I_{\mu P}$, $I_{\gamma 1 P}$) are depicted. Transcripts from I promoters get spliced and polyadenylated. Switch regions also get transcribed in the antisense orientation. Concomitant transcription from $I_{\mu P}$ and, for example, $I_{\gamma 1 P}$ can target AID to μ and $\gamma 1$ switch regions and thereby initiate CSR to $C\gamma 1$. (From Alt and Perlot, 2008). Locations of E_μ (with its flanking matrix attachment regions, MAR) and 3'RR (encompassing four transcriptional enhancers with flanking inverted repeats) are highlighted.

The 3' regulatory region (3'RR) consists of hypersensitive sites HS4, HS3b, HS1,2, and HS3a scattered over 35 kb at the very 3' end of the locus, downstream of $C\alpha$. HS4 is active throughout B cell development (Michaelson *et al.*, 1995), whereas HS3a, HS1-2 and HS3b mostly during late stages (Michaelson *et al.*, 1995). These enhancers have synergistic activities. HS3A and HS3B are arranged as an inverted repeat centered at HS1-2, with virtually identical sequences (Saleque *et al.*, 1997). Deletion of HS1-2 or HS3a had no effect on IgH expression or class switch recombination (CSR) (Manis *et al.*, 1998). On the contrary, the deletion of HS3b and HS4 decreases IgH expression in resting B cells, germline transcription and class switch recombination (CSR) after stimulation (Vincet-Fabert *et al.*, 2010). Mice with a deletion of the entire 3'RR had normal B-cell development and B-cell numbers in bone marrow, spleen and blood. These B cells could differentiate into normal plasma cells but they secreted less antibodies and had no class switch recombination (CSR) (Vincet-Fabert *et al.*, 2010). Only $C\gamma 1$ transcripts and IgG1 CSR were partially independent from the 3'RR, thus they might be under control of other elements that cooperate with the 3'RR (Vincet-Fabert *et al.*, 2010). The iE_μ enhancer is suggested as an interaction-candidate because its deletion slightly impacts CSR and IgH synthesis (Vincet-Fabert *et al.*, 2010). Chromosome conformation capture studies have shown that the 3'RR stimulates transcription by physically interacting with J_H and V_H (Ju *et al.*, 2007) genes. Although the interaction included the E_μ , activation of IgH expression and the interaction with the V_H region were normal in cells with the E_μ -deletion (Ju *et al.*, 2007). These

long-range interactions are B cell specific; they were observed in plasma cells and splenic B cells, but absent in splenic T cell (Ju *et al.*, 2007).

The HS3B and E μ contain LPS-stimulation-inducible binding sites of Yy1 (Gordon *et al.*, 2003). These sites are important in class switch recombination (Gordon *et al.*, 2003). Yy1 interacts with CTCF (Donohoe *et al.*, 2009) a transcriptional regulator involved in loop formation within complex loci throughout the genome. Therefore, Yy1 binding in the 3' RR might specify long range interactions that promote efficient IgH secretion and CSR.

The 3' boundary of the *Igh* locus consists of hypersensitive sites HS5-7 that lack enhancer activity but bind CTCF (Garett *et al.*, 2004). HS5 and HS6 have insulator activity *in vitro* (Garett *et al.*, 2004) and may form a boundary between the *Igh* and other downstream genes, thereby protecting the locus from inappropriate signals.

Similarly, the 5' end of the *Igh* locus contains a cluster of DNase hypersensitive sites (Pawlitzky *et al.*, 2006). HS1 is pro-B cell specific and it contains binding sites for transcription factors PU.1, E2A and Pax5 (Pawlitzky *et al.*, 2006). Therefore a regulatory function in respect to the *Igh* locus was suggested. However, B lymphocytes isolated from mice with deletion of HS1, or the entire 5' HS cluster, undergo normal development without any defects in V(D)J recombination, allelic exclusion, or class switch recombination (Perlot *et al.*, 2010).

Nuclear positioning and topology of the *Igh* locus

The nuclear periphery acts as a repressive compartment where the chromatin is enriched with H3K9me. Fluorescence in situ hybridization (FISH) analysis has demonstrated that the *Igh* and *Igk* loci are sequestered to the nuclear periphery in multipotent progenitors and T cells through interactions with the nuclear lamina, whereas the *Igl* locus is not compartmentalized (Figure 16; Kosak *et al.*, 2002). Similar recruitment of *Igh* and *Igk* loci is probably a consequence of similar organization of their V_H regions, composed of 200 and 140 segments, respectively (Kosak *et al.*, 2002). Peripheral positioning sequesters *Igh* and *Igk* from transcription and recombination machineries (Kosak *et al.*, 2002). At the pro-B cell stage these loci are relocated to the centre of the nucleus where the *Igh* locus undergoes large scale contraction which facilitates V(D)J rearrangement of distal V_H genes (Fuxa *et al.*, 2004). This mechanism ensures equal usage of all V_H gene segments, although they are dispersed over 2.5 Mb in the *Igh* locus, and enables high diversity of variable antigen receptor regions that are essential for the defense against pathogens (Fuxa *et al.*, 2004).

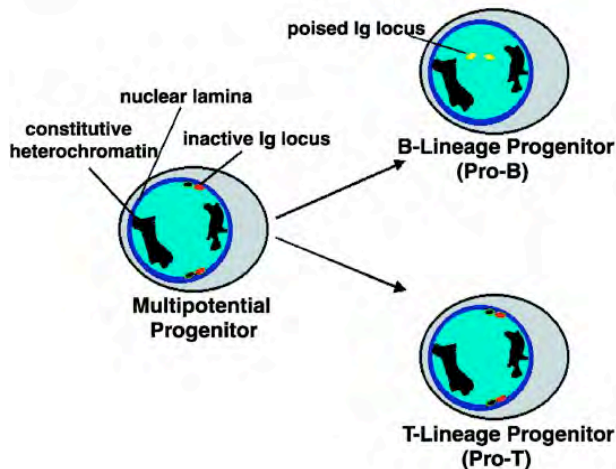


Figure 16. A model depicting the developmentally regulated subnuclear compartmentalization of Ig loci.

In their default state, the IgH and Igk loci are sequestered at the periphery through a proposed interaction with the nuclear lamina. During early B cell development, this interaction is disrupted, allowing the two loci to be poised for activation of transcription and V(D)J recombination, which is accompanied by a contraction of the locus. (From: Kosak *et al.*, 2002)

Factors involved in *Igh* locus activation and looping

Joining of D_H and J_H segments initiates at the pre-pro B cell stage. E2A transcription factor binds to E-box elements in the intronic enhancers κE2 and E5 in the *Igk* locus, and to the Eμ in the *Igh* locus (Massari and Murre, 2000). E2A also regulates Rag1/2 expression by binding an enhancer upstream of *Rag2* (Hsu *et al.* 2008). Although E2A has an essential role for the recombination of the *Igk* locus, it may be redundant in the *Igh* locus since forced expression of Ebf1 in *E2A*^{-/-} multipotent progenitors is sufficient to induce D_H to J_H and V_H to DJ_H recombination (Seet *et al.* 2000).

The first factor reported to be essential for V_H-DJ_H recombination is Pax5. The *Igh* locus of *Pax5*^{-/-} pro-B cells is in an extended configuration that physically separates the distal from the proximal V_H genes (Figure 17; Fuxa *et al.*, 2004). V_H-DJ_H recombination of the distal J558 V_H genes is 50 fold reduced, although D_H-J_H recombination occurs at normal frequency (Nutt *et al.*, 1997). With decreasing distance of V_H gene segments from the D-J_H region the efficiency of V_H to DJ_H joining progressively increases (Hesslein *et al.*, 2003). This demonstrates that Pax5 is essential for recombination of distal, but not proximal V_H genes (Hesslein *et al.*, 2003). Ectopic expression of Pax5 in thymocytes leads to same V_H-DJ_H phenotype as loss of Pax5 in pro-B cells (Fuxa *et al.*, 2004), extended *Igh* configuration and V_H to D-J_H recombination restricted to proximal V_H genes. These data indicate that efficient rearrangement of distal V_H genes requires cooperation of Pax5 with an unknown second factor that is absent in T cells (Fuxa *et al.*, 2004) or not recruited to the *Igh*.

A phenotype similar to Pax5 loss was discovered in mice deficient for Ezh2, a component of the Polycomb (PcG) complex, which harbours a conserved SET domain with H3K9- and H3K27-

specific methyltransferase activity (Su *et al.*, 2003). Mice deficient for Ezh2 have reduced numbers of pre-B and immature B cells in the bone marrow (Su *et al.*, 2003). Inactivation of Ezh2 after pro-B cell stage has no effect on maturation and B-cell activation, showing that Ezh2 has a development-stage specific role (Su *et al.*, 2003). Ezh2 deficient pro-B cells have reduced rearrangement of the distal J558V_H gene segments, and fewer cells express intracellular μ -chain (Su *et al.*, 2003). ChIP analysis of histone modifications in pro-B cells revealed that H3K27me2 is localised exclusively on the proximal V_H7183/ V_H Q52 portion of the Igh locus, but it was absent from the distal V_H genes where active histone marks H3K36me2 and H3K36me3 were detected (Xu *et al.*, 2008). Pax5^{-/-} cells lacked H3K27me2 modifications in the *Igh* locus (Xu *et al.*, 2008). These findings and the deficiency of distal V_H gene rearrangement in Pax5^{-/-} and Ezh2^{-/-} cells together imply that Pax5 might recruit Ezh2 to proximal V_H7183/ V_H Q52 genes (Xu *et al.*, 2008). This would lower their availability to the recombinase complex and bias recombination towards the distal V_H genes (Xu *et al.*, 2008). As support for this hypothesis, no H3K27me2 was found in fetal B cells, which preferentially rearrange the V_H7183/ V_H Q52 genes (Xu *et al.*, 2008). On the contrary, a study from our own laboratory (Ebert *et al.*, 2011) found no H3K27me3 or H3K36me3 modifications above background along the entire V_H gene cluster. Therefore the mechanism how Ezh2 might contribute to *Igh* locus contraction remains enigmatic and debatable.

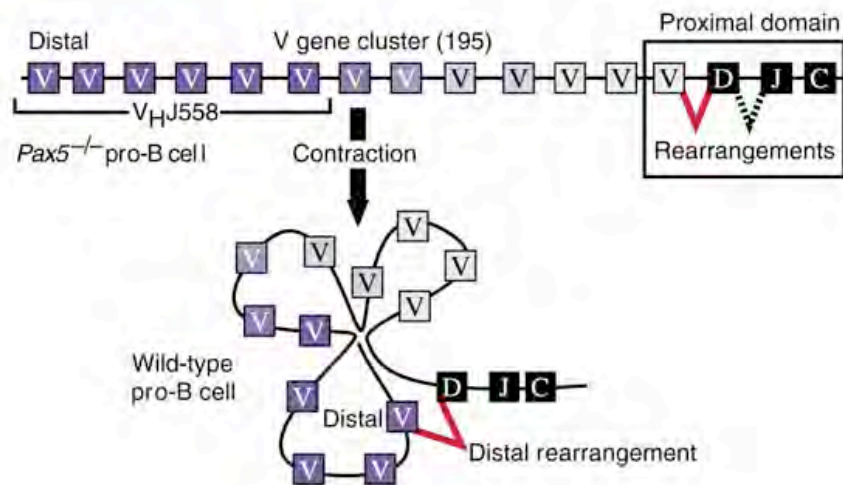


Figure 17. Contraction of the *Igh* locus by looping depends on the expression of Pax5.

The *Igh* locus is in an extended configuration in Pax5^{-/-} (the same phenotype occurs in Yy1^{-/-}, Ezh2^{-/-}) pro-B cells, which allows V(D)J recombination to take place only in the proximal domain. In wild-type pro-B cells, all V_H genes participate in V_H-DJ_H rearrangements owing to contraction of the *Igh* locus by looping

Another transcription factor causing a defect in distal V_H to DJ_H recombination is Yy1 (Ying yang 1) is a zinc finger protein that can act as a transcriptional activator, repressor or bind an initiator element, depending on the promoter context (Liu *et al.*, 2007). During development Yy1 can recruit the Polycomb complex (Atchison *et al.*, 2003). Conditional deletion of Yy1 caused a block in pro-B to pre-B transition (Liu *et al.*, 2007). Similarly to Pax5 loss, D- J_H recombination was normal, whereas the rearrangement frequency of V_H gene segments was decreasing proportionally to their distance from the D- J_H region (Liu *et al.*, 2007). In agreement to this, the *Igh* locus in *Yy1*^{-/-} cells was in an extended configuration (Liu *et al.*, 2007). Cells expressed normal levels of Pax5, Ezh2, other transcription factors, Rag1/2 and signalling molecules involved in V(D)J recombination and pro-pre-B cell transition (Liu *et al.*, 2007). This suggested that Yy1 has a direct, unique role in locus contraction (Liu *et al.*, 2007). Since Ezh2 and Yy1 interact in the Polycomb complex, where Yy1 is required to recruit Ezh2 to DNA during muscle differentiation (Satijn *et al.*, 2001; Caretti *et al.*, 2004), it is possible that a similar mechanism exists in the *Igh* locus. Lastly, Yy1 is important in later stages of B cell development according to several lines of evidence: Yy1 knockout B cells that express μ -chain (10%) fail to differentiate into pre-B, immature and mature B cells and a prerrearranged IgH transgene does not completely rescue the Yy1 knockout phenotype (Liu *et al.*, 2007).

Ikaros plays a crucial role prior to B cell lineage specification with the loss of it leading to an early and complete block in B cell development (Kirstetter *et al.* 2002). The molecular mechanism of the block has not been explained in detail so far, however one study suggested that Ikaros might play a role in V(D)J recombination. Hematopoietic progenitors isolated from *Ikzf1*^{-/-} mice were transduced with Ebf1, which led to differentiation into pro-B cells (Reynaud *et al.* 2008). Despite normal levels of Ebf1 and Pax5, these cells did not express *Rag1/2* and *Dntt* (Reynaud *et al.* 2008). They underwent D- J_H rearrangement, but were unable to recombine V_H gene segments (Reynaud *et al.* 2008). Germline transcripts from the promoters of J558 gene segments were reduced in *Ikzf1*^{-/-} pro-B cells (Reynaud *et al.* 2008). Although located centrally in the nucleus, the *Igh* locus was in an extended configuration (Reynaud *et al.* 2008). These findings suggested that Ikaros might have an important role in V(D)J recombination. However, some contradictions in the system used for the study should be noted. A main discrepancy is how the cells could undergo D- J_H recombination if they did not express *Rag1/2* genes. Furthermore, unlike the studies describing the phenotypes of *Pax5*^{-/-}, *Ezh2*^{-/-} and *Yy1*^{-/-} mice, this one used retroviral transduction, which might have elicited other Ikaros-independent changes that allowed recombination in the absence of Rag1/2. Therefore, the contribution of Ikaros to V(D)J recombination awaits deeper investigation.

In addition to *in vivo* studies in mutant mice and FISH analysis of the *Igh* locus contraction (Fuxa *et al.*, 2004), Jhujhunwala and co-workers (2008) used statistical models and *in situ* hybridization to investigate the 3D architecture of the *Igh* locus and their results are described below. They used *E2A* deficient pre-pro-B cells, and *Rag2*^{-/-} cells that are committed to the B cell lineage, but the *Igh* locus is in germline configuration because Rag2 protein is not expressed. In non-lymphoid cells the *Igh* locus is in an extended configuration, but in pre-pro B cells it becomes more contracted. During the transition to the pro-B cell stage, the entire set of V_H genes is positioned to similar distances from the D_HJ_H elements. The chromatin fiber containing V_H, D_H and J_H gene segments is more flexible in pro-B cells, adopting more configurations than in pre-pro-B cells. Whereas the probabilities of V_H gene segments to encounter D_H or J_H elements in pre-pro B cells correlated with genomic distance of gene segments, in pro-B cells the probabilities were similar for all V_H genes, regardless of their genomic separation. Computer simulations of potential chromatin configurations suggested that the *Igh* locus is organized according to the Multi-Loop-Subcompartment Model (MLS) in pre-pro B cells. The MLS topology consists of multi-loop subcompartments (1Mb) connected by 63-126 kb linkers. In pro-B cells the locus condenses more than predicted by the MLS model. Consistent with this, Jhujhunwala *et al.* (2008) showed that the D_H segments were positioned far away from the majority of V_H regions in pre-pro B cells. Furthermore, even the proximal and distal V_H regions were spatially separated. On the contrary, in pro-B cells, proximal and distal V_H genes merged and the entire V_H repertoire was in close proximity to the D_H-J_H elements.

The described changes underpin a physical mechanism by which the diverse antibody repertoire is established (Jhujhunwala *et al.* 2008). The most proximal V_H genes V_H81X and V_HQ52 are most frequently recombined due to small genomic distance from the D_H genes. However, little correlation was observed between the genomic location of other V_H gene segments and their usage (Jhujhunwala *et al.* 2008). The explanation for this phenomenon is that the *Igh* locus folds as a bundle of loops allowing the entire V_H repertoire similar access to the D_H-J_H elements (Jhujhunwala *et al.* 2008).

CTCF and looping of complex loci

CTCF is an 11-zinc finger nuclear protein found on all vertebrate insulators and boundary elements (Wallace and Felsenfeld, 2007). CTCF binding to insulator sequences prevents the interactions between promoters and enhancers, or the spread of repressive chromatin modifications into neighbouring active domains (Xie *et al.*, 2007). CTCF interacts with a number of proteins involved in transcription regulation and chromatin modification, for example Yy1 (Donohoe *et al.*, 2009) and the chromodomain helicase (CHD8) of the MLL complex (Ishihara *et al.*, 2006). Recently, genome wide ChIP mapping experiments revealed that 60-67% CTCF-binding sites are shared with the cohesin (Smc1, Smc3, Rad21 and Scc1) complex (Parelho *et al.*, 2008; Wendt *et al.*, 2008). Cohesin facilitates sister-chromatid cohesion during cell division (Wendt and Peters, 2009). However, in most cases the cohesin/CTCF co-localisation serves to restructure complex genomic loci to enhance gene expression specifically for a developmental stage and cell-type (Majumder and Boss 2010). CTCF and cohesin form loops that bring gene-regulatory elements in a position poised for activation or isolate inappropriate genes out of the activation hub (Wendt and Peters, 2009). The *Igh* locus has a complex 3D architecture whose topology drastically changes during B cell development (Jhujhunwala *et al.*, 2008), which most likely involves CTCF.

Conserved binding sites of cohesin and CTCF in intergenic regions assist in forming different spatial conformations of complex loci mediated through the interaction with a specific transcription factor (Majumder *et al.*, 2008; Hadjur *et al.*, 2009; Sekimata *et al.*, 2009). During differentiation of naïve CD4 T cells into T_H1 cells, the abundance of CTCF and cohesin on conserved sequences around the *INFG* locus increases (Hadjur *et al.*, 2009). Differentiation specific DNA interactions are established in collaboration with T-bet, T_H1 cell specifying factor (Hadjur *et al.*, 2009; Sekimata *et al.*, 2009). These long-range interactions depend on cohesin, since they were reduced to background levels upon siRNA knockdown of Rad21, although normal CTCF binding was detected (Hadjur *et al.*, 2009). These results suggest that CTCF recruits cohesin to specific sites where it can mediate chromosomal *cis*-interactions (Hadjur *et al.*, 2009). In this way cohesin may affect the probability with which gene regulatory elements interact with each other during development (Hadjur *et al.*, 2009).

Major histocompatibility class II (MHC-II) locus contains a cluster of highly polymorphic genes that encode components of MHC-II molecules, pseudogenes and genes that are structurally unrelated to MHC-II molecules, but are involved in the process of antigen presentation. MHC-II genes are expressed in B cells, macrophages and dendritic cells and can be induced by gamma interferon (INF- γ) (Majumder *et al.*, 2008). Class II transactivator (CIITA) is a cell-type specific

and INF- γ -inducible transcription factor that interacts with other ubiquitous transcription regulatory- and chromatin-modifying complexes to induce the expression of the MHC-II locus (Majumder *et al.*, 2008). CTCF-binding sites in intergenic regions throughout the MHC-II locus (Majumder *et al.*, 2008) interacted with binding sites in gene promoters and thereby induced maximal levels of transcription (Choi *et al.*, 2010). By comparing MHC-II/CIITA expressing (Raji) and non-MHC-II/CIITA expressing cells (RJ2.2.5), two three dimensional structures of the human MHC-II were defined (Figure 18; Choi *et al.*, 2010). When MHC-II genes are not expressed the region consists of smaller loops formed by nearby CTCF sites that dimerise (Yusufzai *et al.*, 2004) with each other (Choi *et al.*, 2010). The interaction frequency is proportional to CTCF occupancy and more favoured if the distance between the sites is smaller (Choi *et al.*, 2010). When CIITA is expressed, it binds promoters of MHC-II genes and recruits CTCF (Majumder *et al.*, 2008) enabling it to dimerise at distant binding sites, which increases the 3D-complexity of the locus. The architecture formed by multiple CTCF binding sites brings promoters in close proximity to each other in a domain of high transcriptional activity (Choi *et al.*, 2010). Formation of CTCF-CTCF-CIITA loops coordinates expression of these antigen presentation and selection genes by bringing all the regulatory components in the same 3D space (Majumder and Boss 2010), the “transcription factory” (Schoenfelder *et al.*, 2010) or a “chromatin hub” (Palstra *et al.*, 2003).

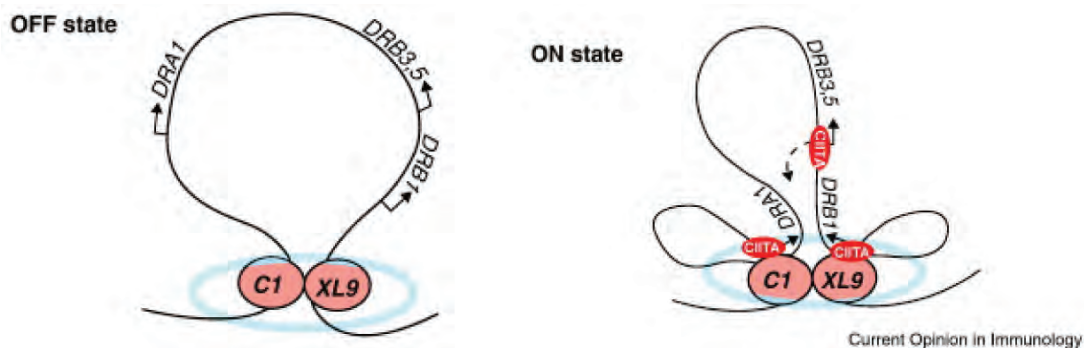


Figure 18. MHC-II insulator long-range looping model.

In the “OFF state”, MHC-insulators bound by CTCF (orange), C1 and XL9, interact to form a chromatin loop and interaction focus. Cohesin (blue) likely encircles the interacting CTCF foci maintaining the integrity/stability of the structure or its function. When MHC-II genes are induced (ON), CIITA present at the promoters interacts with CTCF bound insulators forming a second series of interactions and sub loops. While this cartoon represents the HLA-DR subregion, additional interactions are possible with these elements. (From: Choi *et al.*, 2010)

ChIP-Chip studies identified many CTCF binding sites in the *Igh* locus: in the 3'RR, DFL16.1 (the most 5' D_H gene) and within the V_H region (Degner *et al.*, 2009). The sites in the V_H 7183/Q52 region are close to the RSS, whereas in the distal V_H J558 region CTCF binds upstream of the coding V_H genes in intergenic regions (Degner *et al.*, 2009). CTCF-binding pattern was similar in B lineage progenitors and T cells, suggesting that CTCF is not the cell type specific factor that promotes drastic changes in *Igh* contraction at the pro-B cell stage (Degner *et al.*, 2009). However, Rad21 cooccupies CTCF binding sites throughout the *Igh* locus with much lower enrichment in pre-B cells and thymocytes than in pro-B cells (Degner *et al.*, 2009). Although CTCF and Rad21 might have the capacity to organize the locus into a rosette-like structure that would provide all V_H gene segments with similar rearrangement probability, they most likely, depend on a B cell specific factor to do so (Degner *et al.*, 2009). Implications for the latter lie in the V(D)J phenotypes described upon the loss of Pax5, Yy1, Ezh2 or Ikaros. Whether similar phenotypes arise after deletion of CTCF and cohesin has been investigated in our and other laboratories. However, since CTCF has important functions genome-wide, it was not possible to obtain viable CTCF^{-/-} cells in numbers that allow analysis.

Aim of the thesis

Pax5 has critical roles in the most important aspects of B cell development. It was the first identified transcription factor responsible for committing progenitor cells down a “one way road” of a definitive lineage. Moreover, Pax5 was the first discovered regulatory factor necessary for *Igh* locus contraction.

At commitment to the B cell lineage, Pax5 uses the properties of its degenerate sequence recognition paired domain to bind to a vast array of genes. Other domains collaborate in *cis* (together) and in *trans* with other transcription factors and chromatin regulators to activate B-cell specific genes and to keep regulators of other lineages under control. Some partner proteins of Pax5 have been discovered, but surely they do not comprise the full Pax5-interactome. The sequence of Pax5 protein contains motifs for different posttranslational modifications (PTMs). The ability of PTMs to alter the affinity for interaction partners or DNA, and the localization of proteins within a cell, highlights the importance they could have in the modulation of functions of Pax5. These “post-translational isoforms” of Pax5 could be correlated with the biochemical and physical changes they induce in the chromatin structure during gene regulation. Because of these reasons, I focused on expanding the know interactome and regulome of Pax5.

Interestingly, the changes in chromatin that occur prior to V(D)J recombination are reminiscent of gene regulation, although they appear to be even more complex. The regulation of locus contraction prior to V(D)J recombination is still enigmatic. At the 3' end of the locus two regulatory elements, the intronic E μ enhancer, and the 3' RR, can mediate long-range interactions. Until recently, no regulatory elements with active chromatin marks located in the V_H region were known. Ebert *et al.* (2011) have discovered intergenic elements, which are located upstream of the distal V_H3609 genes. They contain binding sites for CTCF, Rad21, E2A and Pax5. Whereas CTCF and E2A occupy these elements already at the pre-pro-B cell stage and remain there in mature B cells, Pax5 binds only at the pro-B cell stage. Moreover, the induction of active chromatin marks depends on Pax5, and results in spliced, non-coding pro-B cell-specific antisense transcripts. Therefore, they were named Pax5-activated intergenic repet (PAIR) elements. We propose that PAIRs are V_H regulatory elements that induce pro-B cell specific contraction of the V_H region. In lymphoid progenitors, the neighbouring CTCF-binding sites might form localised loops within the V_H region. In pro-B cells, Pax5-dependent chromatin activation of the PAIR elements might promote long-range interactions of CTCF bound in PAIR and in other elements throughout the locus. The cohesin ring would stabilise the formed loops. This would together serve to reposition the distal V_H genes in proximity to the DJ_H region, enriched for H3K4me3 and Rag1/2 proteins. We will use the circular chromosome conformation capture (4C) technology, combined with the next generation sequencing, to test the described model.

RESULTS I

Posttranslational modifications change the charge or the conformation of a protein allowing it to act in different cellular contexts or compartments. Phosphorylation is a key modification, through which cells propagate signals to adjust gene expression for a specific need. Evidence is accumulating that acetylation, methylation and sumoylation can serve as switches between activating or inhibitory roles of transcription factors. The mechanism of these changes is common for most proteins, it can alter the binding of interacting partners that can synergise or antagonise certain activities. This highlights the importance of posttranslational modifications for regulatory functions of Pax5. Hence, I focused on identifying posttranslational modification and interaction partners of Pax5.

Identification of posttranslational modifications by mass spectrometry

To investigate posttranslational modifications (PTMs) of Pax5 by mass spectrometry, I took advantage of a knock-in mouse carrying the biotin acceptor sequence at the C-terminus of Pax5 (Pax5-Bio) that was generated in our group (Figure 19A; McManus *et al.*, 2011). This biotin-tag can be biotinylated *in vivo* upon co-expression of *Escherichia coli* biotin ligase BirA (de Boer *et al.*, 2003). The interaction between biotin and streptavidin is the strongest non-covalent interaction known in nature (10^{15} M) thereby allowing the purification of Pax5 in large quantity and high purity. To map PTMs I needed to identify more than 90% of the Pax5 protein sequence by mass spectrometry. For this purpose I had to start with 6×10^9 pro-B cells for the preparation of nuclear extracts and subsequent purification of this large quantity of Pax5. To avoid the need of killing many mice to isolate these cells, we generated a pro-B cell line by transforming bone marrow cells with Abelson murine leukemia virus (Ab-MuLV) (Rosenberg *et al.*, 1975). I tested whether Pax5-Bio protein was fully biotinylated in these cells in two ways. Firstly, I added recombinant streptavidin before fractionating the nuclear extract by SDS-PAGE, resulting in a 20 kDa shift of the Pax5-Bio/streptavidin complex (Figure 19B). Secondly, I performed Pax5-Bio precipitation with streptavidin magnetic beads leading to complete depletion of Pax5-Bio from the nuclear extract (Figure 19C). I controlled the specificity of both methods by using nuclear extracts isolated from cells that express only BirA (Ab-BirA cells).

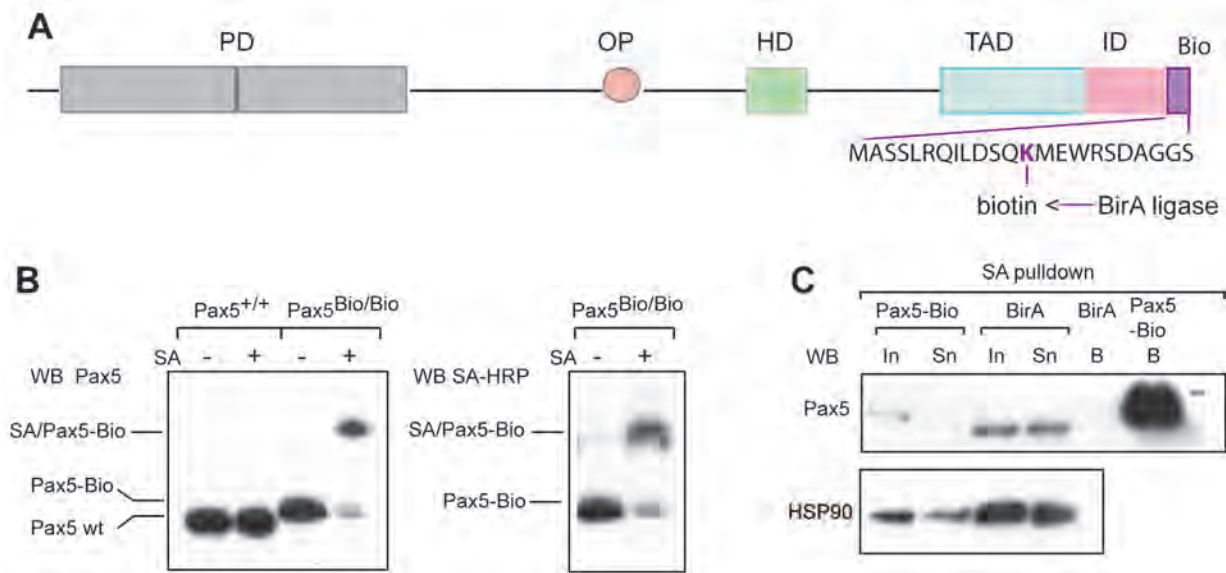


Figure 19. Biotinylation efficiency of Pax5 in Ab-MuLV transformed pro-B cells.

(A) Schematic diagram of Pax5-Bio protein with its C-terminal biotin acceptor sequence. OP, octapeptide; HD, partial homeodomain; TAD, transactivation domain; ID, inhibitory domain. (From: McManus *et al.*, 2011)

(B) Size fractionation of Pax5 and Pax5-Bio proteins, with or without incubating the nuclear extract with streptavidin.

Nuclear extracts (20 µg) were fractionated on SDS-PAGE and analysed by Western blot with anti-Pax5 antibody (left). Tagged Pax5 protein migrates slightly higher than wild type (compare 1st and 2nd lane with 3rd). An aliquot of each nuclear extract (Pax5-Bio or Pax5-wt) was incubated for 5 minutes with 5 µg of recombinant streptavidin. The majority of biotin-tagged Pax5 migrates around 20 kDa higher than in the absence of streptavidin demonstrating that the tag is biotinylated *in vivo* (lane 4, left). The same nuclear extract was reprobed with streptavidin (SA-HRP) horseradish-peroxidase (right). Residual Pax5 band that was not shifted is detected both with the Pax5 antibody and SA-HRP demonstrating that the added amount of streptavidin was not sufficient for a complete shift. This suggests that ~100% of Pax5 protein is biotinylated.

(C) *In vivo* biotinylated Pax5-Bio protein is efficiently precipitated from the nuclear extract in the affinity purification with streptavidin-coated magnetic beads. Input (In; 1/100 of nuclear extract), supernatant (Sn; 1/100) separated from the streptavidin-coated magnetic beads with a magnet and material bound to beads (B) were analysed with anti-Pax5 antibody.

By using approximately 50 mg of total nuclear protein for streptavidin (SA) pulldown, I could detect the Pax5-Bio band on a SDS-PAGE Coomassie-stained gel (Figure 20A). To increase the coverage of the protein's primary structure, important for the identification of posttranslational modification, the Pax5-containing gel band was digested with three tryptic enzymes trypsin, chymotrypsin and subtilisin and the resulting peptides were extracted by sonication.

Mass spectrometric analysis revealed nine phosphorylation sites (Figure 20B), most of them localized between the paired and the transactivation domain, and only one in the transactivation domain (Figure 20C). Three phospho-serines (S189, S206, S283) were detected in every experiment. Phosphorylations of another serine (S344) and a threonine (T285) were detected twice. Phosphorylated tyrosines (Y179, Y388) were detected in only one experiment. Interestingly, the phospho-serine and phospho-threonine residues were within the motifs (Figure 20C, larger underlined font in *italics*) recognized by proline-directed kinases (for example MAP

kinase (MAPK)) (<http://elm.eu.org/>, The Eukaryotic Linear Motif resource for Functional Sites in Proteins). Moreover the phospho-serines are within motifs engaged in phosphorylation dependant interaction with Class IV WW domains. The proteins containing the latter are involved in different processes such as ubiquitin-mediated protein degradation, mitotic regulation (http://elm.eu.org/elmPages/LIG_WW_4.html), transcription and RNA processing (Ingham *et al.*, 2005). The phospho-serine residues (S171, S174), detected only once, are within a recognition motif for casein kinase 1 (CK1), which has been reported to have a role in cell division, DNA repair and glycogen metabolism. Clustal W alignment of human and amphioxus Pax protein homologues showed that the residues that I detected of being phosphorylated, do not have a high degree of evolutionary conservation, even between members of the same protein family (alignment not shown). This does not diminish their significance, but rather highlights them as versatile regions where evolutionary processes can act in order to specify a protein for a specific developmental role.

In summary, Pax5 is constitutively phosphorylated at the pro-B cell stage by a member of a MAPK family suggesting that this signalling pathway might play an important role in maintaining the function of Pax5 in pro-B cells. Furthermore, phosphorylated motifs might be controlling the interactions with proteins containing WW class IV domains. Phospho-residues that were detected in only one experiment might be related with more transient functions and could act as switches that can be activated by specific stimuli or may be more important at a different stage of B cell development.

Pax5 was methylated on two aspartate (D53, D125) and a lysine (K98) residues in the paired domain, an aspartate (D125) in the transactivation domain and two arginines (R359, R377) in the inhibitory domain (Figure 20B and C). The methylation of the latter arginines may be important for the function of the inhibitory domain as it was detected in all three experimental replicas. Moreover, Clustal W alignment showed that these arginines are conserved in Pax homologues (Figure 20D). Upstream of the inhibitory domain, there is an amino acid sequence known as the transactivation domain, whose potential to activate transcription is controlled by the inhibitory domain (Dörfler and Busslinger, 1996). A similar juxtaposition of the transactivation and inhibitory domains exists in C/EBP β . The function of the modular inhibitory domain strongly depends on the methylation of arginines in the inhibitory domain, which interferes with the recruitment of the BAF and Mediator complexes by the transactivation domain (Kowenz-Leutz *et al.*, 2010). Furthermore, phosphorylation of the transactivation domain in C/EBP β , downstream of the Ras/MAPkinase signalling, prevents the interaction with the arginine methyltransferase (Kowenz-Leutz *et al.*, 2010). A similar crosstalk between phosphorylation and methylation might

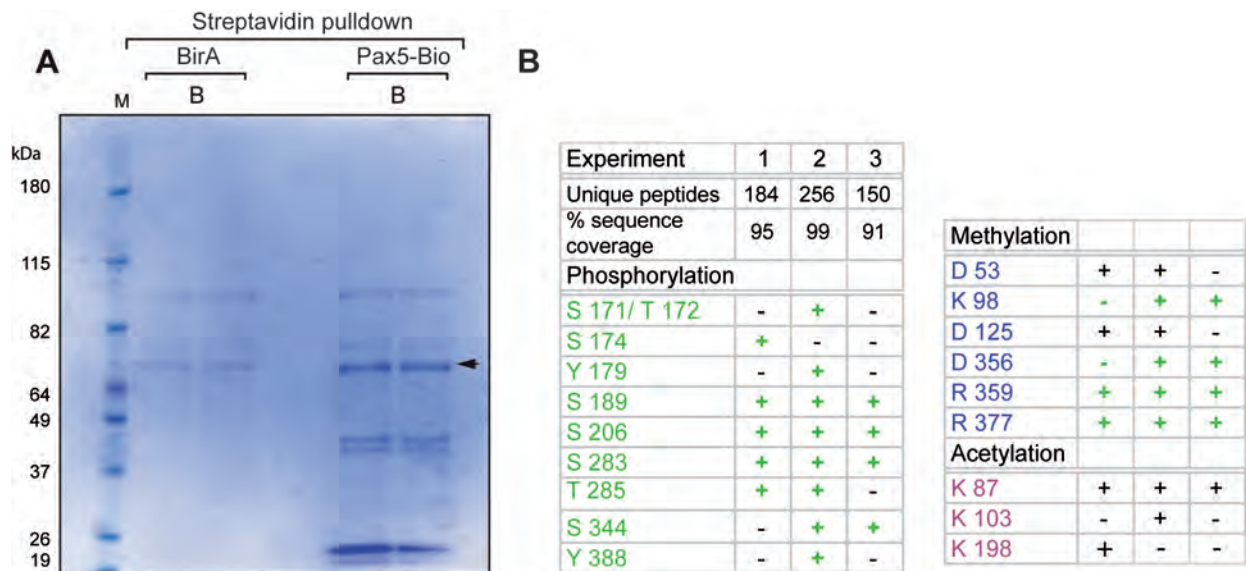
exist in the regulation of Pax5, since a MAPkinase target serine (S344) was determined in the transactivation domain.

The methylation of aspartate residues (D3, D125, D356) was an unexpected finding (Figure 20B and C). However, a study in HeLa cells has estimated that around 2% of human proteins might contain aspartate methylations (Sprung *et al.*, 2010). Methylation neutralizes the negatively charged aspartate side chains possibly remodelling the protein's structure and interactions, thereby providing an extra layer in regulation of its function. Nevertheless, data confirming the biological significance of an aspartate methylation

is still lacking.

Three lysines within Pax5 are acetylated, two sites in the paired domain (K87, K103) and one site (K198) in the nuclear localisation signal (NLS) sequence (Figure 20C). Only K87 could be considered of functional importance since it was detected in all three experiments, whereas others were found only once. The methylation and acetylation of lysines in the paired domain might regulate the interaction of Pax5 with the DNA.

In summary, we have found that Pax5 can be phosphorylated, methylated and acetylated. Each type of modification can occur on more than one residue, but for now we do not know the degree of methylation (mono-, di-, tri-) of the residues. Our results suggest that Pax5 can exist in pro-B cells in several posttranslational isoforms, which might have distinct functions.



C

1 MDLEKNYPTPRTSRTGHGGVNQLGGVFNVRPLPDVVRQRIVELAHQGVR
51 PC^DISRQLRVSHGCVSKILGRYYETGSIKPGVIGGS^KPKVATPKVVE^KIAEY
103 ^KRQNPTMFAWEIRDRLAERVC^DNDTVPSVSSINRIIRTKVQQPPNQVPV
153 ASSHSIVSTGSVTQVSSV^GTD^SSAGSS^SISGIL^GIT^SPSA^DTNKR^KRDE^G
203 IQ^ESPVPNGHSLPGRDFLRKQMRGDLFTQQQLEVLDRVFERQHYS^DIFT
252 TTEP^KPEQ^ITEYSAMASLAGGLDDMKANL^TSPT^TPADIGSSVPGPQSYPI
302 VTGRDLASTTLPGYPPHVPPAGQGSYSAPTLTGMVPGSE^FSG^SSPYSH^P
350 QYSSYN^DSW^RFPNPGLLGSPYYYSAA^RGAAPPAAATAY^DDRH

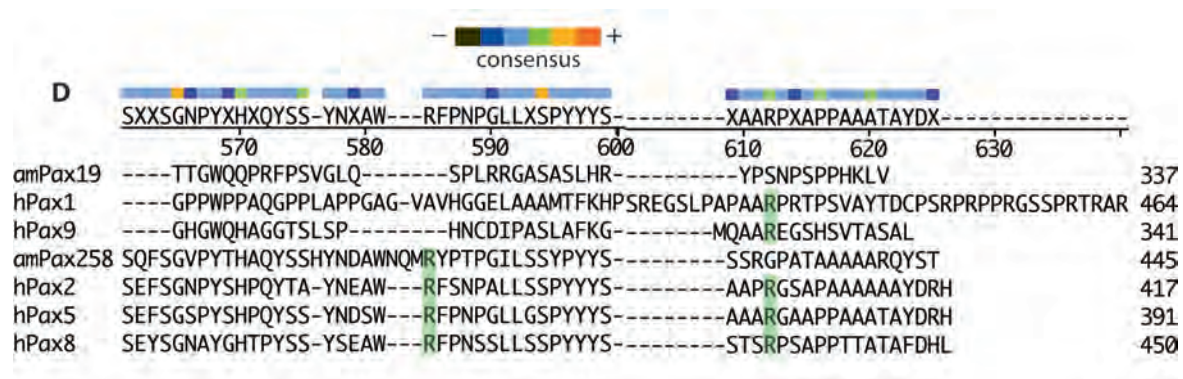


Figure 20. Identification of posttranslational modifications (PTMs) of Pax5.

Abelson murine leukemia virus transformed pro-B cells of the *Pax5^{Bio/Bio}* (Pax5-Bio) or control *Rosa26^{BirA/BirA}* (BirA) genotype were used for the streptavidin (SA) pulldown from 50 mg of total nuclear extract. The proteins bound to the streptavidin-coated magnetic beads were eluted by boiling in SDS-buffer and fractioned on SDS-PAGE. The gels were stained with Colloidal Coomassie blue buffer and the Pax5-Bio gel-band was excised, cleaved with trypsin, chymotrypsin and subtilisin, extracted and analysed by mass spectrometry in 3 replicas.

(A) An example of the size fractionation of proteins bound to streptavidin beads (B) after one of the three replicas used for PTM analysis. The arrowhead indicates the band containing purified biotinylated Pax5. The first lane contains the markers (M) with size in kilo Daltons (kDa) indicated in the left side of the gel.

(B) Table showing the reproducibility in detection of posttranslational modifications.

Unique peptides refer to the number of peptide sequences belonging to the Pax5 protein that were identified in each experiment after digestion with trypsin, chymotrypsin and subtilisin. Percentage (%) of sequence coverage indicates how much of Pax5-amino acid sequence was identified by the unique peptides in each experiment. The numbers describe amino acids' position in the Pax5 sequence.

(C) Pax5 domains are shaded as follows: grey – paired domain, red – octapeptide motif and the inhibitory domain, green – partial homeodomain and transactivation domain. Primary sequence of Pax5 with colour-coded posttranslationally modified amino acids: methylated, blue; acetylated, violet; phosphorylated, green. Phosphorylated residues S189, S206, S283 and T285, S344 are within motifs *GITSPS*, *GIQESPV*, *NLTSTPT* and *FSGSPY* respectively, which could engage in MAPK-phosphorylation dependant interaction with Class IV WW domains. These motifs are highlighted in italics and underlined in the sequence shown. Boxed **IKPEQ** motif is a consensus SUMO-acceptor motif (discussed later). Motifs in dashed boxes contain alternative, putative SUMO-acceptor lysines (discussed later).

(D) Pax-homologues sequence alignment of the inhibitory domain. Conservation of methylated arginines in human (hPax) or amphioxus (amPax) homologues. The sequences were aligned with Clustal W alignment using the Gonnet algorithm. Conserved arginines are boxed in green. The ancestor amphioxus Pax258 and the human protein homologues Pax2, Pax5 and Pax8 have conserved R359, whereas R377 is, in addition to the human Pax2, 5, and 8 homologues, conserved in Pax1 and Pax9. The colour-code for the relative level of conservation is indicated at the top.

Sumoylation of Pax5

Certain posttranslational modifications involve the conjugation of small proteins, such as ubiquitin or Small Ubiquitin related MOdifiers (SUMO1, 2 or 3) to lysines of the recipient proteins. Sumoylation regulates diverse processes (Geiss-Friedlander and Melchior, 2007) including transcriptional control and the localization of proteins in different cellular compartments and (Gill *et al.*, 2004). Isopeptide bond formation between SUMO and a specific lysine depends on an ATP-dependant enzymatic cascade consisting of an E1-activating enzyme heterodimer of Aos1 and Uba2, E2 conjugating enzyme Ubc9 and an E3 ligase (Geiss-Friedlander and Melchior, 2007). The E1 heterodimer AOS1–UBA2 activates SUMO in an ATP-dependent reaction, preparing it for the transfer to the catalytic E2 enzyme UBC9 (Geiss-Friedlander and Melchior, 2007). Then, an isopeptide bond formed between the C-terminal Gly residue of SUMO and a Lys residue in the substrate is mediated by an E3 ligase (Geiss-Friedlander and Melchior, 2007; Figure 22A). During this process, Ubc9 directly interacts with the substrates via their SUMO consensus motif **ΨKxE/D**, where Ψ represents a hydrophobic and x any amino acid (Rodriguez *et al.*, 2000) whereas enhanced target modification depends on the E3 ligase (Hay *et al.*, 2005).

The amino acid sequence of Pax5 contains the consensus SUMO motif, IKPE (Figure 20C, boxed). Furthermore, the motif is conserved at the same position in the sequence of human paralogues of the Pax2/5/8 protein family Pax2 (IKSE) and Pax8 (TKGE) (Figure 21). Considering these points, I set out to investigate the sumoylation of Pax5 *in vitro* and *in vivo*.

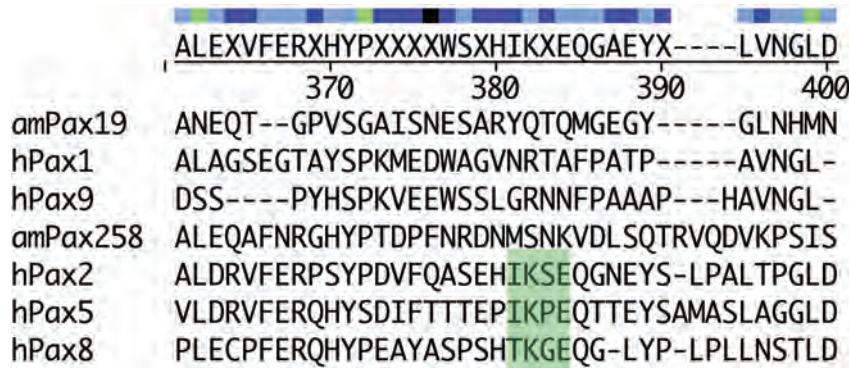


Figure 21. Evolutional conservation of the SUMO motif.

Human (h) Pax 1, 9, 2, 5 and 8 homologues together with amphioxus (am) Pax1/9 and Pax2/5/8 homologues were aligned by Clustal W method (Gonnet algorithm). The conserved SUMO-acceptor motif in human homologues Pax2, Pax5, and Pax8 is highlighted in green. The top lane indicates conservation.

Recombinant Pax5 can be sumoylated *in vitro*

The nuclear pore protein RanBP2 is an E3 ligase, whose structure and catalytic mechanism were thoroughly investigated. The catalytic domain of RanBP2 is natively unfolded and it assumes its structure only upon folding around UBC9 (Pichler *et al.*, 2004). Its catalytically active domain BP2 Δ FG (Figure 22B) consists of two internal repeats, (IR1 and IR2) separated by the M domain of 25 amino acid and of N- and C-terminal flanking regions (Figure 22B, Pichler *et al.*, 2004). The N-terminal flanking region contains a lysine residue that undergoes autosumoylation (Pichler *et al.*, 2004). The N- and C-terminal sequences and IR2 are dispensable for sumoylation, but they provide specificity for the substrates (Pichler *et al.*, 2004). The IR1+M fragment is the catalytic core domain, which recognizes the SUMO1 acceptor site, but cannot discriminate different substrates (Figure 22B, Pichler *et al.*, 2004).

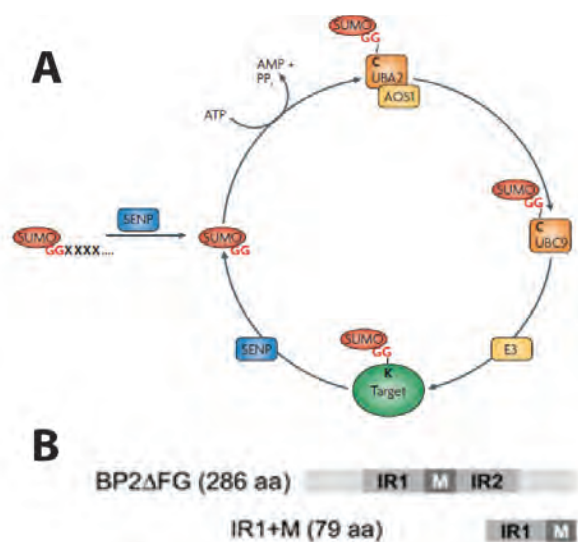


Figure 22. Enzymes of the sumoylation cascade.

(A) Scheme of the sumoylation cycle. (From: Geiss-Friedlander and Melchior, 2007)

(B) Schematic representation of selected RanBP2 fragments used in the *in vitro* experiments. BP2 Δ FG, the catalytic E3 ligase domain, consists of two internal repeats (IR1 and IR2) separated by a 25 amino acid M domain and N- and C-terminal flanking regions (From: Pichler *et al.*, 2004).

We utilised purified recombinant FLAG-Pax5-His and HA-SUMO1 with purified enzymatic components of the sumoylation reaction His-Aos/Uba, Ubc9 and RanBP2 catalytic fragments Δ FG or IR1+M (Table 1). The mixture was incubated at 30°C for 2 hours. This is an established assay (Pichler, 2008) resulting in an approximately 20 kDa size increase of the substrate in a SDS-PAGE gel.

Protein	Mw/kDa	Amount
SUMO1	12	2.2 μ M
His-Aos/Uba2	110	70 nM
Ubc9 (E2)	18	250 nM
IR1+M (E3)	9	20 nM
Δ FG (E3)	32	20 nM
FLAG-hPax5-His	50	570 ng

Table 1. Concentrations of proteins used for the *in vitro* sumoylation assay.

I performed a titration with 4-16 nM IR1+M or 5-80 nM Δ FG E3 ligase components to find the optimal concentration range in which Pax5 is modified. I observed that sumoylation is most efficient up to 20 nM of Δ FG or 16 nM of IR1+M, resulting in the strongest signal intensity of the Pax5-SUMO1 band (Figure 23A). The larger RanBP2 fragment Δ FG, that retained the domain for substrate specificity (Figure 23A), acted as a better E3 ligase on Pax5 than IR1+M. However, Δ FG is prone to autosumoylation, which was especially obvious in concentration above 20 nM (Figure 23A, indicated by the brackets above 95 kDa). Therefore I continued to look for an optimal concentration in the range of 2-5 nM Δ FG (Figure 23A, lower panel). Based on this experiment, I decided to use 4 nM Δ FG in further *in vitro* experiments. Since prolonged incubation of proteins at 30°C can increase their degradation and results in more auto-sumoylated species, I analysed the kinetics of the sumoylation reaction in intervals of 15 minutes to determine the time when sumoylation of Pax5 reaches saturation. Prolonging the incubation for more than 60 minutes did not result in a significant increase in Pax5-SUMO1 band intensity (Figure 23B).

In conclusion, Pax5 could be modified *in vitro* by SUMO1. Notably, the Pax5-specific antibody recognized two bands after the *in vitro* sumoylation reaction. The latter suggests an existence of additional SUMO1 acceptor sites beside the lysine in the consensus IKPE motif. The lower band could correspond to a mono-sumoylated Pax5 (only at the consensus motif), the higher to di-sumoylated Pax5 (both on the consensus and non-consensus motif).

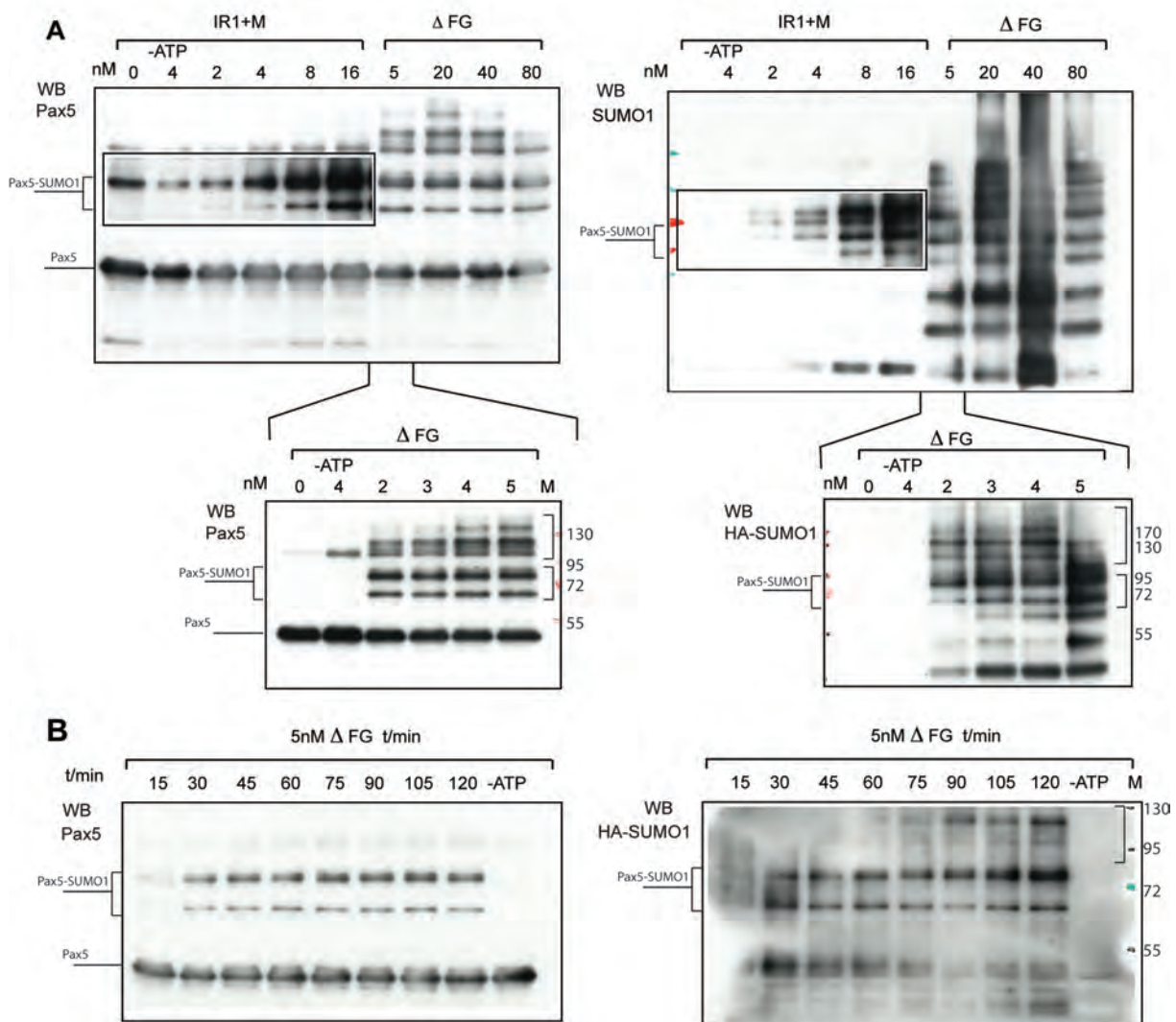


Figure 23. *In vitro* sumoylation assay of recombinant Pax5-His by HA-SUMO1.

The assay was done in the presence of ATP at 30°C for 2 hours. I analysed the products by Western blotting with anti-Pax5 (blots on the left side) and anti-SUMO1 or anti-HA antibodies (blots on the right side). The line indicates two sumoylated Pax5 bands; the bracket above 95 kDa indicates the unspecific bands, products of BP2ΔFG autosumoylation. The assay was controlled with a reaction where ATP (-ATP), or E3 ligase (0 nM) was not included. The lane containing the marker is indicated at the top of the gel (M), with their molecular weight (kDa) on the right.

(A) Titration of sumoylation with 4-16 nM of IR1+M or 5-80 nM of ΔFG (as indicated on top of the picture). For the titration of IR1+M a longer exposure compared to ΔFG is shown. Optimal RanBP2 fragment for *in vitro* sumoylation of Pax5 was ΔFG. The titration with ΔFG was repeated in the concentration range of 1-5 nM. The optimal concentration for the sumoylation of Pax5 is 5 nM of ΔFG.

(B) Kinetics of *in vitro* sumoylation of Pax5. The *in vitro* reaction was allowed to proceed for the amount of time indicated on top of the blots. There is no significant increase in sumoylation if the reaction is prolonged for over 45 minutes.

Pax5 can be sumoylated in HEK293T cells

The unfolded conformation of recombinant proteins could result in unspecific sumoylation, which may not be observed *in vivo* because the acceptor lysine is “hidden” from the enzymatic machinery. Therefore, it was important to confirm that Pax5 could be sumoylated in living cells as well. I tried to answer this question by transiently co-transfecting HEK293T cells with FLAG-Pax5 and His-SUMO1. By Western blot analysis of the total cell extract using a Pax5 specific antibody, I could already detect an additional band at 70 kDa corresponding to the size of Pax5-SUMO1, in addition to the non-sumoylated Pax5 band of 55 kDa (Figure 24A). I purified the pool of sumoylated proteins from cells transfected with Pax5 and SUMO1-His, to investigate by Western blotting, whether it contains Pax5. Indeed the bound fraction of nickel affinity purification contained Pax5-SUMO1, migrating at 70 kDa (Figure 24A). Additionally, by probing with SUMO1 and histidine-tag specific antibodies, I confirmed that the purification precipitated the unconjugated His-SUMO1 itself (Figure 24B) and high molecular SUMO1-conjugated species (Figure 24B)

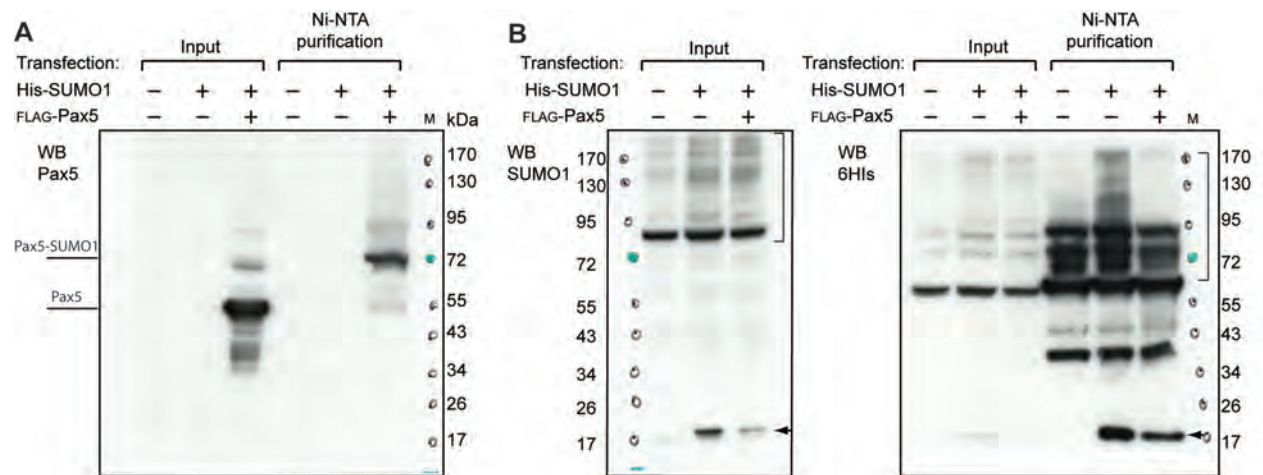


Figure 24. Pax5 is modified by SUMO1 *in vivo*.

HEK293T cells transiently expressing FLAG-Pax5 and His-SUMO1 were lysed under denaturing condition (see Experimental procedures). Immunoblots of crude lysates (Input, 1/100) and nickel affinity purifications (Ni-NTA) are shown. They were analysed with anti-SUMO1 (monoclonal mouse), anti-6xHis (monoclonal mouse) and anti-Pax5 (monoclonal rat) specific antibodies. The molecular weight marker (M) is indicated in kDa.

(A) Western blot analysis of lysates and nickel affinity (Ni-NTA) purification with anti-Pax5 antibody. A band corresponding to unmodified Pax5 and a band of weaker intensity, according to the size corresponding to Pax5-SUMO1, were detected in the lysates (Input). After nickel affinity purification (Ni-NTA) the Pax5-SUMO1-His band was enriched, whereas the unmodified Pax5 was absent. Molecular weight markers are shown (M; kDa).

(B) Western blot analysis of the lysates (Input) and nickel affinity (Ni-NTA) purification with anti-6xHis and anti-SUMO1 antibodies. Transfected SUMO1-His was detected in unconjugated (indicated by the arrowhead at ~17 kD) and conjugated forms (high molecular weight species indicated by the bracket) in lysates of transfected HEK293T cells, and after nickel affinity purification (Ni-NTA).

Since IKPE was the only consensus SUMO1 acceptor motif in Pax5, we expected that K->R mutation would completely abolish sumoylation. For this reason, I mutated the lysine in the consensus SUMO1 motif to an arginine (Pax5-K257R) and performed the nickel affinity purification from lysates of HEK293T cells transfected with the K257R mutant and SUMO1-His. I controlled the efficiency of SUMO1-His conjugation and Ni-NTA purification by Western blot detection of RanGAP1 (Figure 25B), a known sumoylation substrate (Bischoff *et al.*, 1994). By Western blot analysis with Pax5 antibody, I detected two Pax5-SUMO1 bands after Ni-NTA purification, showing that lysines outside of the consensus SUMO-acceptor motif could be sumoylated (Figure 25A).

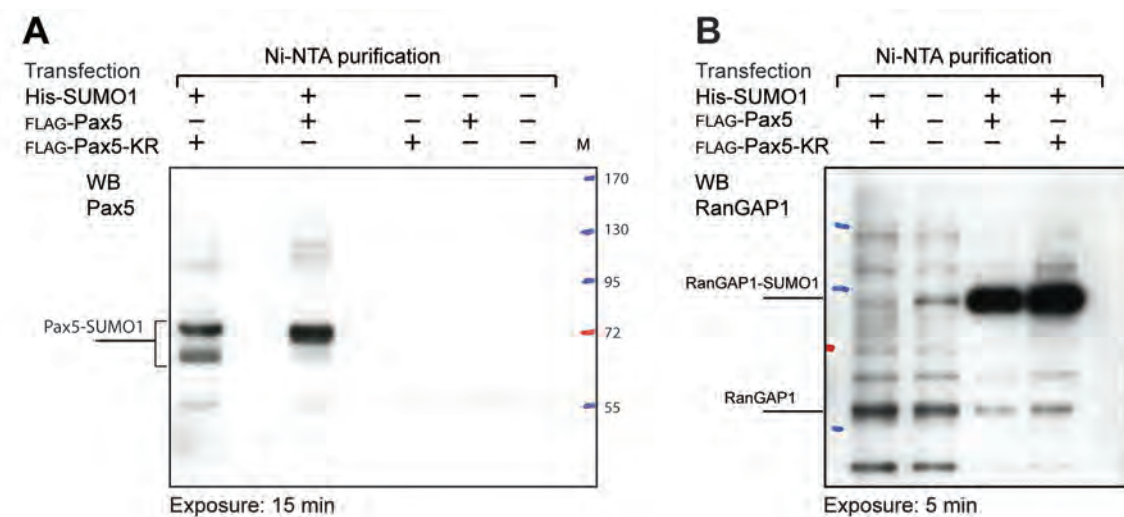


Figure 25. Lysine 257 in the consensus motif is not the only acceptor of SUMO1 in Pax5 protein.

K257R mutant was expressed with His-SUMO1 in HEK293T cells and analysed as described.

(A) Nickel affinity purification (Ni-NTA) in denaturing conditions from lysates of cells expressing wt or K257R mutant Pax5 (Pax5-KR) with (+) or without (-) control co-expression of His-SUMO1. 80% of Ni-NTA purification was analysed with Pax5 antibody. Mutant K257R has two SUMO1 acceptor lysines (indicated by the bracket), whereas wt Pax5 has only one, presumably K257.

(B) I controlled the efficiency of His-SUMO1 conjugation and the amounts of purified SUMO1-protein-conjugates, in different lysate-preparations, by Western blotting (10% of purified proteins) with the RanGAP1 antibody. Both purifications (from lysates expressing wt and K257R mutant Pax5) yielded equal amounts of RanGAP1-SUMO1.

To narrow down the number of putative SUMO1 acceptor lysines, I transfected cells with constructs expressing mutant Pax5 proteins (Dörfler and Busslinger, 1996) with a C-terminal deletion (transactivation and inhibitory domains; Δ Ct), a C-terminal and homeodomain deletions (Δ Ct Δ HD), a deletion of the sequence between the paired domain and the C-terminus (PD-TA) or all sequences deleted except for the paired domain (PD; Figure 26A). Western blot analysis of the cell lysates with anti-Pax5 antibody showed that all mutants were expressed (Figure 26B). After the Ni-NTA purification, the Pax5-antibody detected the full-length Pax5 (FL), Δ Ct and

Δ Ct Δ HD, with a 20 kDa increase in size compared to unmodified proteins. However neither the PD nor the PD-TA mutants were detected after the Ni-NTA purification (Figure 26C, left). The Western blot analysis with SUMO1 antibody also detected bands corresponding to the FL, Δ Ct-SUMO1 and Δ Ct Δ HD-SUMO1 and confirmed successful purification of SUMO1-His-tagged proteins from lysates expressing the PD and PD-TA mutants (Figure 26C, right). In conclusion, all SUMO1-acceptor lysines are in the amino acid sequence between the paired domain and the transactivation domain, excluding the homeodomain.

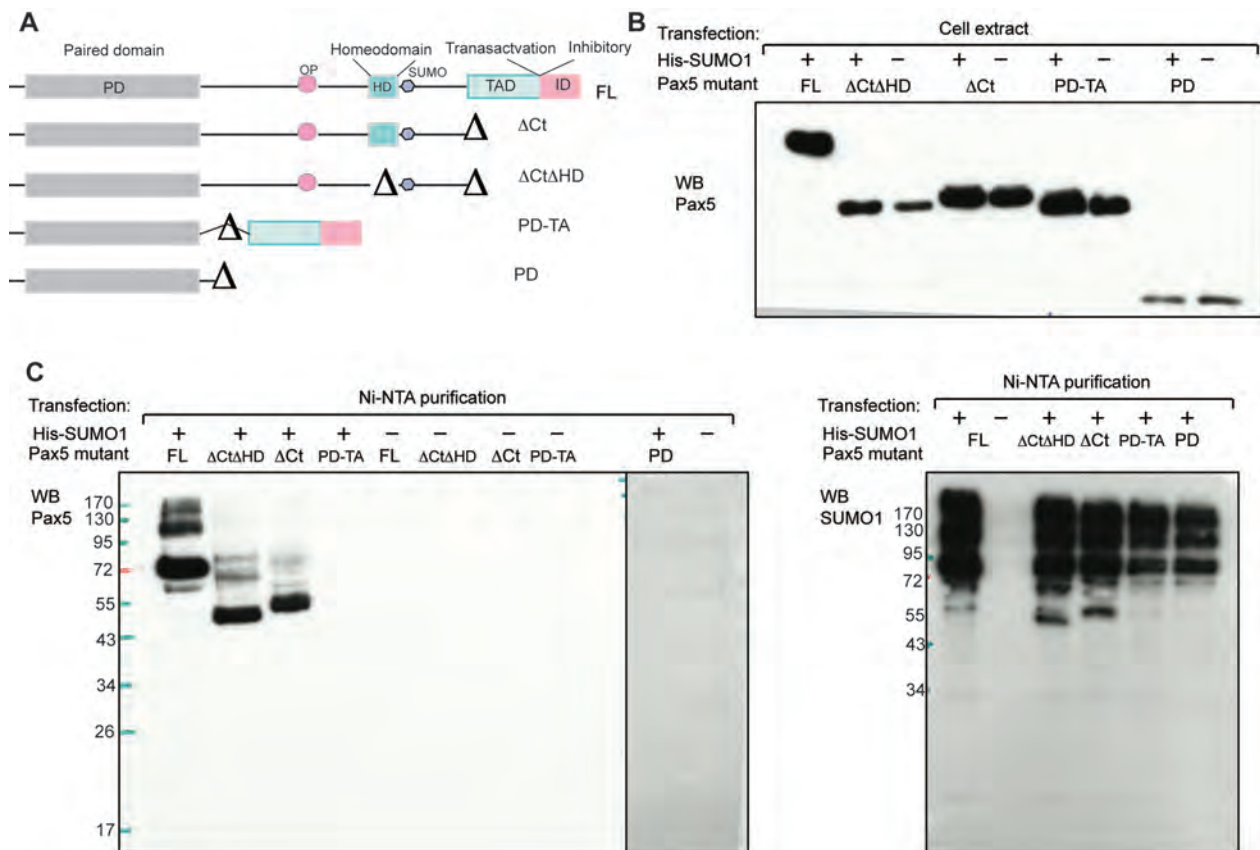


Figure 26. Analysis of Pax5-deletion mutants sumoylation.

SUMO1 acceptor residues are in the sequence between the paired and the transactivation domains, but not within the homeodomain.

Pax5 mutant proteins were transfected into HEK293T cells with or without (negative control) His-SUMO1. Lysates of these cells were prepared under denaturing conditions and used for nickel affinity purification (Ni-NTA). The efficiency of expression and purification of proteins were controlled by Western blotting with anti-SUMO1 (sheep) and anti-Pax5 antibodies (rabbit polyclonal).

(A) Schematic representation of Pax5 deletion mutants used in the experiment: full length protein (FL), deleted: C-terminal domain (Δ Ct), C-terminal and homeo- domains (Δ Ct Δ HD), sequence between the paired and the transactivation domains (PD-TA), sequence after the paired domain (PD). The grey polygon shape indicates the location of the sumoylation motif on K257.

(B) The expression levels of Pax5 deletion mutants in the cell extracts was analysed with polyclonal rabbit anti-Pax5 antibody.

(C) His-SUMO1 can be conjugated to full-length Pax5 and the mutant proteins Δ Ct Δ HD and Δ Ct, as demonstrated by anti-Pax5 (on the left side) and anti-SUMO1 (on the right side) Western blot. Mutants PD-TA and PD could not be detected after Ni-NTA purification suggesting that they lack SUMO-acceptor residues.

In summary, I showed that Pax5 could be sumoylated *in vitro* and *in vivo*. The modification takes place in the sequence between the paired and the transactivation domains. The lysine in the consensus motif (K257) might be the preferential acceptor of SUMO1 *in vivo*, as suggested by a single band detected after Ni-NTA purifications from HEK29T cell lysates. If K257 is mutated, other lysines could serve as SUMO1 acceptors of similar affinity. This finding agrees with my *in vitro* experiments, where I regularly detected two Pax5-SUMO1 bands, one of them with slightly higher intensity. We can predict that K257, in the consensus SUMO1 motif, is recognized in B cells by a specific E3 ligase, and may be modified in response to some cellular signal. The alternative lysines might have similar probabilities to be modified in the unspecific *in vitro* assay.

Previous studies have annotated several domains in Pax5, according to their interactions with the DNA or other proteins (Dörfler and Busslinger 1996). The here-described analysis of posttranslational modifications is in agreement with this annotation as follows. Arginine methylation was reported to contribute to the inhibitory activity of transcription factors, which is consistent with our identification of methylated arginines specifically in the inhibitory domain of Pax5. Modifications reminiscent of chromatin marks, lysine-methylation and lysine-acetylation, are detected in the DNA-binding paired domain. Additionally, the posttranslational modifications, phosphorylation and sumoylation, which are known to change the localization and function of transcription factors in response to cell signalling, are predominantly localized between the paired domain and the transactivation domain within a so far functionally unannotated region. Furthermore, the interplay of these two modifications could have synergistic or agonistic effects on the function of Pax5, providing novel regulatory layers. Therefore I would propose to annotate this region as “the modular domain” (Figure 27).

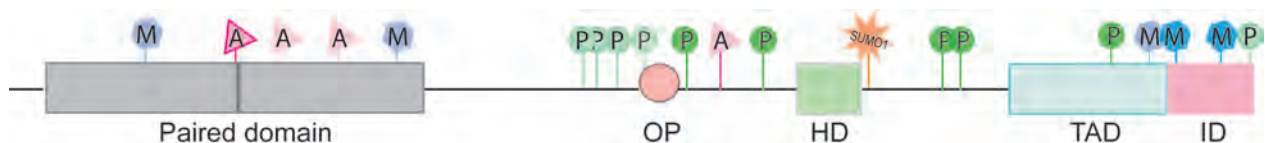


Figure 27. Schematic diagram of Pax5 protein and the identified PTMs.

The domains are depicted approximately to scale; paired domain, octapeptide motif (OP), partial homeodomain (HD), transactivation domain (TAD) and inhibitory domain (ID). Relative positions of identified posttranslational modifications are depicted: methylation (M) in blue, acetylation (A) in pink, phosphorylation (P) in green, sumoylation (SUMO1) in orange. For inspection which residues are modified see Figure 20B and C.

Interaction partners of Pax5

Pax5 recruits chromatin remodelling and modifying complexes

Pax5 fulfils its dual role as a transcriptional activator and repressor by inducing chromatin changes on its target genes. The opposing effects are mediated through recruitment of different chromatin-modifying complexes (McManus *et al.*, 2011). Previous studies, performed in transiently transfected HEK293T cells, showed that Pax5 can interact with TATA-binding protein (TBP) (Eberhard and Busslinger, 1999), Brg1 of the BAF chromatin-remodeling complex (Barlev *et al.*, 2003) and the histone acetyltransferase CBP (CREB-binding protein) (Emelyanov *et al.*, 2002). Additionally, the related transcription factor Pax2 is known to interact with PTIP, a member of the MLL3- and MLL4-containing H3K4 methyltransferase complex (Cho *et al.*, 2007; Patel *et al.*, 2007). Pax5 is also able to recruit members of the Groucho co-repressor family (Eberhard *et al.*, 2000, Linderson *et al.*, 2004), found in histone deacetylase complexes, which explains the repression of transcriptional activities.

I took advantage of our Pax5-Bio protein (Figure 19A) to investigate Pax5-interacting proteins. I performed a streptavidin pulldown followed by identification of precipitated proteins by mass spectrometry. I controlled the specificity of the Pax5-dependant precipitation of proteins by using wild-type pro-B cells expressing the *E.coli* biotin ligase BirA (Figure 28A). Having done five replicas (one example is shown), we filtered out (as described in Materials and Methods section) proteins significantly and reproducibly enriched in Pax5-Bio pulldowns, compared to BirA. The filtering resulted in 300 proteins that I grouped according to their biological function (<http://www.ncbi.nlm.nih.gov/>). Although I used nuclear extracts for the pulldown, many cytoplasmic proteins were copurified. There are several explanations. Enzymatic complexes involved in carbohydrate and fatty acid metabolism are endogenously biotinylated and large proteins. Therefore they represent one of the major contaminating groups ([14%], data not shown). Proteins of the cytoskeleton are also major contaminants ([14%] data not shown) due to their high abundance in the cell. Structural components of the ribosome and the mRNA processing machinery are abundantly copurified ([9% and 14% respectively], data not shown) because of their involvement in gene expression processes and high abundance in the nucleus, compared to transcription factors with much lower expression. From the reasons mentioned above, proteins falling into these four categories were considered to be inspecifically co-purified and were excluded from further analysis.

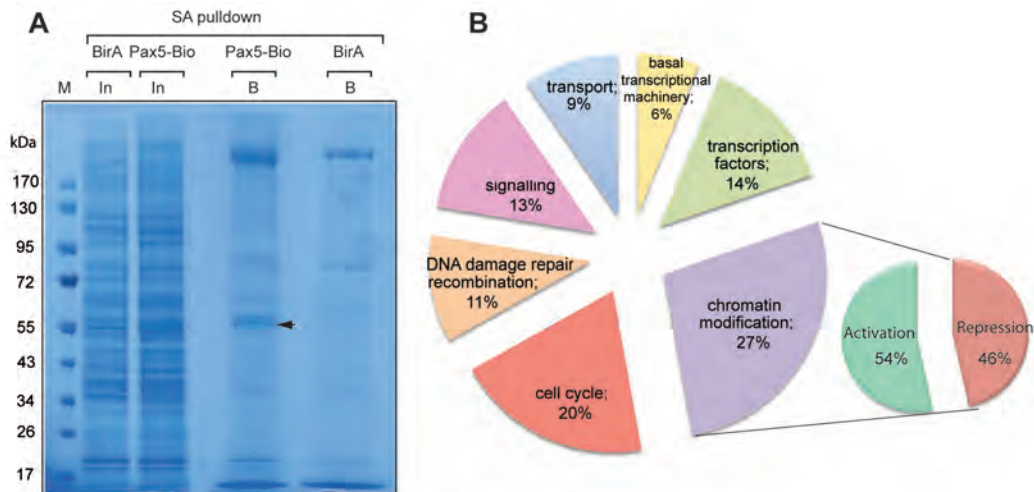


Figure 28. Identification of putative Pax5-interacting proteins by mass spectrometry. Abelson murine leukemia virus (Ab-MuLV) transformed pro-B cells of the *Pax5^{Bio/Bio}* (Pax5-Bio) or control *Rosa26^{BirA/BirA}* (BirA) genotype were used for streptavidin (SA) pulldown from 25 mg of nuclear extracts. The precipitated proteins were subsequently identified by mass spectrometry. (A) Colloidal Coomassie blue-stained gel showing one of the five SA experiments. Shown are lanes containing markers (M), input nuclear extract (In; 25 µg) and proteins eluted after binding to streptavidin beads. The arrowhead indicates Pax5-Bio containing band. (B) Classification of proteins specifically copurifying with Pax5-Bio according to their biological function and cellular localisation, with the criteria used by the Gene Ontology Consortium. Five independent experiments were analysed, and reproducibly identified proteins were selected (criteria are described in Materials and Methods) generating a list of 300 members. Numbers of peptide sequences for all proteins identified within a functional group, across all five experiments, were summed and their relative distribution is shown in the pie chart. The total number of peptide sequence used for this classification was 1244. Proteins annotated as chromatin modifiers were subdivided according to the effects that they have on gene expression (as reported in Pubmed searches).

Transcription factors			sumPAH	
ID	Name	Function	Pax5	BirA
12394	Runx1	Runx-related transcription factor	21	1
15951	Ifi204	Interferon-activable protein 204, coactivator of RUNX2	6	1
83383	HEB	Transcription factor 12, Tcf12	8	0
14025	Bcl11a	B-cell CLL/lymphoma 11A	18	2
17260	Mef2c	Myocyte-specific enhancer factor 2C	13	2
13591	Ebf1	Early B-cell factor 1	16	0
50794	Klf13	Krueppel-like factor 13	4	0
14247	Fli1	Friend leukemia integration 1 transcription factor	4	0
21406	Tcf12	Transcription factor 12	7	0
209200	Dtx3l	Deltex 3-like	3	0
12014	Bach2	BTB and CNC homology 2 isoform 1	3	0
16978	Lrrfip1	Leucine rich repeat (in FLII) interacting protein 1	3	0
14248	Flii	Protein flightless-1 homolog	10	0
20841	Zfp143	Zinc finger protein 143	7	1
330474	Zc3h4	Zinc finger CCCH domain-containing protein 4	5	0
24136	Zeb2	Zinc finger homeobox 1b	6	0
11538	Adnp	Activity-dependent neuroprotective protein	7	4
17190	Mbd1	Methyl-CpG-binding domain protein 1	4	0
22344	Vezf1	Vascular endothelial zinc finger 1	2	2
56463	Snd1	Tudor domain-containing protein 1	25	1
66660	Sltm	SAFB-like transcription modulator	8	2
23856	Dido1	Death-inducer obliterator 1	11	1
21853	Timeless	Protein timeless homolog 1	6	1
52245	Commd2	COMM domain containing 2, supressor of NF-kappaB	3	0

Table 2. Transcription factors identified by mass spectrometry. The sum of unique peptides above homology (sumPAH) identified in 5 Pax5-Bio and BirA (control) experiments is shown.

Chromatin remodelling and modification			sumPAH	
	ID	Name Function	Pax5	BirA
BAF	68094	Smarcc2 BAF170 subunit of BAF complex	22	6
	57376	Smarce1 BAF57 subunit of BAF complex	4	0
	19708	Dpf2 BAF45D subunit of BAF complex	3	1
	207425	Brwd2 Bromodomain and WD repeat-containing protein 2	14	2
HAT	75560	Ep400 E1A binding protein p400, in NuA4 histone acetyltransferase complex	5	0
	100683	Trrap Transformation/transcription domain-associated protein, HAT cofactor	4	0
	98956	Nat10 N-acetyltransferase 10	42	7
	68142	Ino80 DNA helicase INO80 complex homolog 1, coactivator of YY1	3	0
MLL	67772	Chd8 Chromodomain helicase DNA-binding protein 8, in MLL1/MLL complex	8	0
	140858	Wdr5 WD repeat domain 5	17	6
	66867	Hmg20a High mobility group protein 20A, recruitment of MLL complex	1	0
	13018	CTCF CCCTC-binding factor	30	11
HDAC	108155	Ogt O-linked N-acetylglucosamine (GlcNAc) transferase, in MLL5 complex	24	4
	15161	Hcfc1 Host cell factor C1, Set1/Ash2, in MLL5 and Sin3 HDAC	37	4
	213109	Phf3 PHD finger protein 3	10	0
	70998	Phf6 PHD finger protein 6	5	1
NCOR	77683	Ehmt1 Euchromatic histone methyltransferase 1	9	1
	70465	Wdr77 20S PRMT5-arginine methyltransferase, with SUZ12 and H2A/HIST2H2AC	4	1
	107932	Chd4 Chromodomain helicase DNA binding protein 4, in NuRD	15	2
	56086	Set HAT inhibitor of EP300/CREBBP and PCAF-mediated acetylation (H4)	3	0
NCOR	12417	HP1g HP1 gamma (Chromobox homolog 3, Cbx3)	8	2
	20185	NCoR1 Nuclear receptor co-repressor 1	9	0
	81004	Tbl1xr1 Nuclear receptor co-repressor/HDAC3 complex subunit	12	5
	21372	Tbl1x Transducin beta-like protein 1	13	5
NCOR	238317	Gm260 SANT domain-NCOR binding	5	0
	72068	Cnot2 CCR4-NOT transcription, subunit 2; SMRT/NCOR-HDAC3 complexes	5	0
	12034	Phb2 Prohibitin-2, recruits histone deacetylases HDAC1 and HDAC5, Phb	15	2
	18673	Phb Prohibitin, with Phb2	3	0
NCOR	74255	Smu1 Functional spliceosome-associated protein 57	8	3
	227693	Zer1 Zyg-11 homolog B-like protein Cul2-RING ubiquitin ligase complex	8	0
	68926	Ubap2 Ubiquitin-associated protein 2	25	5
	70790	Ubr5 Ubiquitin protein ligase E3 component n-recogin 5	18	0
NCOR	15260	Hira Histone cell cycle regulation defective homolog A1	6	0

Table 3. Proteins involved in chromatin remodelling and modification. Complexes are indicated on the left side.

The sum of peptide numbers in 5 experiments is shown (sumPAH) identified in Pax5-Bio and BirA (control) experiments. Ubiquitin catalytic enzymes are often present in co-repressor complexes. Coloured borders indicated components found in the respective activating (green) or repressive complexes (red).

Basal transcriptional machinery			sumPAH	
ID	Name	Function	Pax5	BirA
245841	Polr2h	RNA polymerase subunit RPABC3	5	1
227606	Tbpl2	TATA box-binding protein-like protein 2	3	0
228980	Taf4a	TFIID subunit 4a	4	1
226182	Taf5	TFIID subunit 5	4	1
21343	Taf6	TFIID subunit 6	10	0
108143	Taf9	TFIID subunit 9	8	0
56771	Med20	Mediator subunit 20	4	1
66580	Esf1	ABT1-associated protein (ABTAP)	3	1
70239	TFIIIC63	General transcription factor 3C polypeptide 5	3	0
12455	Ccnt1	Cyclin T1 subunit of the transcription elongation factor p-TEFb	18	1
17218	Mcm5	DNA replication licensing, necessary for RNAPII transcription	28	5

Table 4. Proteins involved in transcription, the basal transcriptional complex (TFIID and Mediator) identified by mass spectrometry analysis. The sum of peptide numbers in 5 experiments is shown (sumPAH) identified in Pax5-Bio and BirA (control) experiments.

DNA damage/repair			sumPAH	
ID	Name	Function	Pax5	BirA
14375	Ku p70	X-ray repair cross-complementing protein 6, 5'-dRP/AP, Xrcc6	13	6
22596	Ku p80	X-ray repair cross-complementing protein 5, Xrcc5	16	7
50505	Ercc4	DNA excision repair protein ERCC-4	23	2
236930	Ercc6l	DNA excision repair protein ERCC-6-like	3	0
17685	Msh2	DNA mismatch repair protein Msh2	9	2
17688	Msh6	DNA mismatch repair protein Msh6	5	0
17535	Mre11a	Double-strand break repair protein MRE11A	9	2
80905	Polh	DNA polymerase eta	38	9
75764	Slx1b	Structure-specific endonuclease subunit SLX1	4	0
74320	Wdr33	WD repeat domain 33	18	1
68114	Mum1	Mutated melanoma-associated antigen 1	4	1
54380	Smarcal1	Harp helicase, anneals ssDNA strands at stalled replication forks	14	1

Table 5. DNA damage and repair factors identified by mass spectrometry. The sum of peptide numbers in 5 experiments is shown (sumPAH) identified in Pax5-Bio and BirA (control) experiments.

Cell cycle/ DNA replication			sumPAH	
ID	Name	Function	Pax5	BirA
19687	Rfc1	Replication factor C 1 activator of DNA polymerases	60	6
12455	Ccnt1	Cyclin T1	18	1
268697	Ccnb1	G2/mitotic-specific cyclin-B1	3	0
51869	Rif1	Rap1 interacting factor 1 homolog, S-phase checkpoint	33	0
20877	Aurkb	Aurora kinase B	7	0
22367	Vrk1	Serine/threonine-protein kinase	6	4
217718	Nek9	NIMA-related expressed kinase 9	5	0
229584	Pogz	Pogo transposable element with ZNF domain	4	0
72119	Tpx2	Microtubule-associated protein homolog	17	3
17218	Mcm5	Minichromosome maintenance deficient 5	28	5
16571	Kif4	Chromosome-associated kinesin SMC domain	33	8
18536	Pcm1	Pericentriolar material 1 protein isoform 2	7	0
17276	Mela	Melanoma antigen	7	0
100042777	Erh	Enhancer of rudimentary homolog	6	0
12237	Bub3	Budding uninhibited by benzimidazoles 3 homolog	11	0
19290	Pura	Transcriptional activator protein Pur-alpha	8	0
67010	Rbm7	RNA binding motif protein 7	3	0
76478	Hice1	4HAUS augmin-like complex, subunit 8	3	0
15366	Rhamm	Hyaluronan mediated motility receptor, (CD168, Hmnr)	5	1
234069	Pcid2	MAD2 regulator in pre-B	8	0
66413	Psmd6	26S proteasome non-ATPase regulatory subunit 6	6	0
23997	Psmd13	26S proteasome non-ATPase regulatory subunit 13	11	1

Table 6. Proteins involved in cell cycle identified by mass spectrometry. The sum of peptide numbers in 5 experiments is shown (sumPAH) identified in Pax5-Bio and BirA (control) experiments.

Signalling			sumPAH	
ID	Name	Function	Pax5	BirA
320484	Rasal3	RAS protein activator like 3	26	3
26413	Mapk1	Mitogen-activated protein kinase 1	2	0
19229	Ptk2b	Protein tyrosine kinase 2 beta	8	0
16331	Inpp5d	Phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 1	38	0
16332	Inpp1	Phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 2	4	1
66367	Sam	SAM domain in signalling and nuclear proteins	5	1
13494	Drg1	Developmentally-regulated GTP-binding protein 1, Nedd3	11	0
13495	Drg2	Developmentally regulated GTP binding protein 2, GTPase	8	2
23897	Silg111	HCLS1 associated X-1, Hax1	3	0
16859	Lgals9	Lectin, galactose binding, soluble 9	14	0
14694	Gnb2l1	Guanine nucleotide binding protein (G protein), beta polypeptide 2 like 1	20	8
100042856	Gvin1	Interferon-induced very large GTPase 1	22	0
83945	Dnaja3	DnaJ (Hsp40) homolog, subfamily A, member 3	5	1
12388	Ctnnd1	Catenin delta-1 isoform1	5	0
227331	Gigyf2	GRB10 interacting GYF protein 2	6	3
26934	Racgap1	Rac GTPase-activating protein 1	4	0
55935	Fnbp4	Formin binding protein 4	4	0
14026	Evl	Ena/VASP-like protein	3	0

Table 7. Proteins involved in cell signalling identified by mass spectrometry. The sum of peptide numbers in 5 experiments is shown (sumPAH) identified in Pax5-Bio and BirA (control) experiments.

Transport			sumPAH	
ID	Name	Function	Pax5	BirA
18141	Nup50	Nucleoporin 50 kDa	6	0
22218	Sumo1	Small ubiquitin-related modifier 1	6	3
19386	Ranbp2	E3 SUMO-protein ligase RanBP2	18	1
19385	Ranbp1	Ran-specific GTPase-activating protein	3	0
27041	G3bp1	Ras GTPase-activating protein-binding protein 1	11	0
19324	Rab1	Ras-related protein Rab-1A ER and Golgi	5	1
76308	Rab1b	Ras-related protein Rab-1B ER and Golgi	6	1
17274	Rab8a	Ras-related protein Rab-8A	10	3
19325	Rab10	Ras-related protein Rab-10 secretory vesicles	4	1
77407	Rab35	Ras-related protein Rab-35	7	1
17931	Ppp1r12a	Protein phosphatase 1 regulatory subunit 12A	11	0
75608	Chmp4b	Charged multivesicular body protein 4b	6	0
59042	Cope	Coatamer subunit epsilon	2	1
12847	Copa	Coatamer subunit alpha	18	3
50797	Copb2	Coatamer subunit beta'	7	0
11771	Ap2a1	AP-2 protein transport complex subunit alpha-1	4	0
11772	Ap2a2	AP-2 complex subunit alpha-2	2	0
71770	Ap2b1	AP-2 complex subunit beta	2	0
11764	Ap1b1	Golgi adaptor HA1/AP1 adaptin beta subunit	3	0
545030	Wdfy4	WD repeat and FYVE domain containing 4	9	1

Table 8. Proteins mediating transport between cellular compartments

In agreement with the biological function of Pax5, most of the remaining proteins are involved in the regulation of transcription (Figure 28B; Tables 2-4). Moreover, even 27% of them are reported to be members of chromatin-modifying complexes. According to the literature, 54% of these proteins contribute to gene activation and the rest to repression (Figure 28B, Table 3). Hence I proceeded with confirming some of these findings by streptavidin pulldowns and co-immunoprecipitations followed by Western blot analysis. In agreement with previous studies and our result from mass spectrometry (Table 4), I confirmed that Pax5-Bio was copurified with members of the basal transcriptional complex (TFIID), TBP and TAF6 (Figure 29A). In the reciprocal precipitation with antibodies specific for both TBP and TAF4, I could detect copurification of Pax5 with TAF4 and TBP respectively (Figure 29B), confirming that Pax5 is a component of the TFIID complex. The components of SWI/SNF-like BAF complex (BAF170, BaF57, BAF45) were also detected in mass spectrometry (Table 3). In order to confirm that Pax5 interacts with the BAF complex, I performed an immunoprecipitation using an antibody specific for Brg1, one of the core components of the BAF complex. Brg1 co-precipitated Pax5 as well as other components of the BAF complex, BAF57 and BAF170, indicating that Pax5 is indeed in the BAF-chromatin remodelling complex (Figure 29B). Furthermore, our previous work demonstrates that a number of activated Pax5 target genes acquire active histone marks such as H3K9ac and H3K4me in a Pax5-dependent manner (McManus *et al.*, 2011). This suggests that Pax5 is involved in recruiting histone modifiers to its target regions. Moreover, we identified some components of the MLL-complex (Chd8, Wdr5) and histone acetyltransferase (HAT) complexes by mass spectrometry (Table 3). I therefore tested the existence of the interactions between Pax5 and components of these complexes. Using the Pax5-Bio pulldown assay, I could show that PTIP co-purifies with Pax5-Bio (Figure 29A) and reversely antibodies specific for PTIP co-precipitate Pax5 (Figure 29C). As described above, PTIP is involved in recruiting the MLL-complex. I, therefore, used one of the core components of the MLL-complex RbBP5 to show that Pax5 is involved in MLL-complex recruitment through the interaction with PTIP. Antibody specific for RbBP5 could copurify Pax5 although the amount of precipitated Pax5 was significantly less compared to the precipitation with the anti-PTIP antibody (Figure 29C). This might suggest that RbBP5 is not the direct interaction partner of Pax5 but contained within the complex that is recruited to Pax5 through the interaction with PTIP. I further confirmed the interaction between Pax5 (wt or Bio-tagged) and FLAG-tagged-PTIP in transiently transfected HEK293T cells (Figure 29E). This experimental system will allow us in future to investigate which domain(s) are involved in this interaction. CBP is a coactivator protein with histone acetyl transferase (HAT) activity. I confirmed the previous findings that Pax5 can interact with CBP (Emelyanov *et al.*,

2002) by showing that immunoprecipitation with a CBP-antibody copurified Pax5 (Figure 29C). Although transcriptional activation by Pax5 can be largely explained by the above findings, it is less well understood how Pax5 could be repressing its target genes. Groucho family protein, Grg4, is so far the only interaction partner that was shown to be involved in Pax5-dependent gene repression (Eberhard *et al.*, 2000, Linderson *et al.*, 2004). Interestingly, we detected the interaction of three core components of the NCoR1 corepressor complex (NCoR1, Tbl1xr1 and Tbl1x; Table 3) in the mass spectrometry analysis. The NCoR complex is associated with HDAC3 histone deacetylase activity and therefore can explain the Pax5-dependent removal of H3K9ac mark at the repressed target genes. I confirmed this finding by showing that NCoR specific antibodies co-precipitated Pax5 (Figure 29D).

Since some of the interaction partners described above can also bind the DNA, it is important to distinguish whether the interaction with Pax5 is mediated solely by protein-protein interaction or involving DNA bindings. To this end, I tested all the interactions in the absence of DNA by preparing the nuclear extracts in the presence of benzonase that digests away all nucleic acids. A comparison in Figure 29D shows that the co-precipitation was not affected by the presence of benzonase.

This study demonstrated that Pax5 interacts in pro-B cells with components of the basal transcription factor complex TFIID, the chromatin remodelling BAF complex, the histone acetyltransferase CBP and PTIP, which recruits the MLL-containing H3K4 methyltransferase complex to the chromatin (Cho *et al.*, 2007; Patel *et al.*, 2007). Moreover, these complexes are rapidly recruited to enhancers and promoters of Pax5-activated target genes, in response to Pax5 activation in the Pax5-ER induction system, (McManus *et al.*, 2011). Furthermore, we showed a novel mechanism of Pax5-mediated gene repression that involves the NCoR complex. Similarly to the case of activated target genes, our Pax5-ER induction system demonstrated that NCoR is rapidly recruited to promoters and enhancers of Pax5-repressed target genes (McManus *et al.*, 2011). Taken together, our results show that Pax5 coordinates the epigenetic and transcriptional control of its target genes by recruiting chromatin-remodelling and histone-modifying complexes to their regulatory elements (McManus *et al.*, 2011).

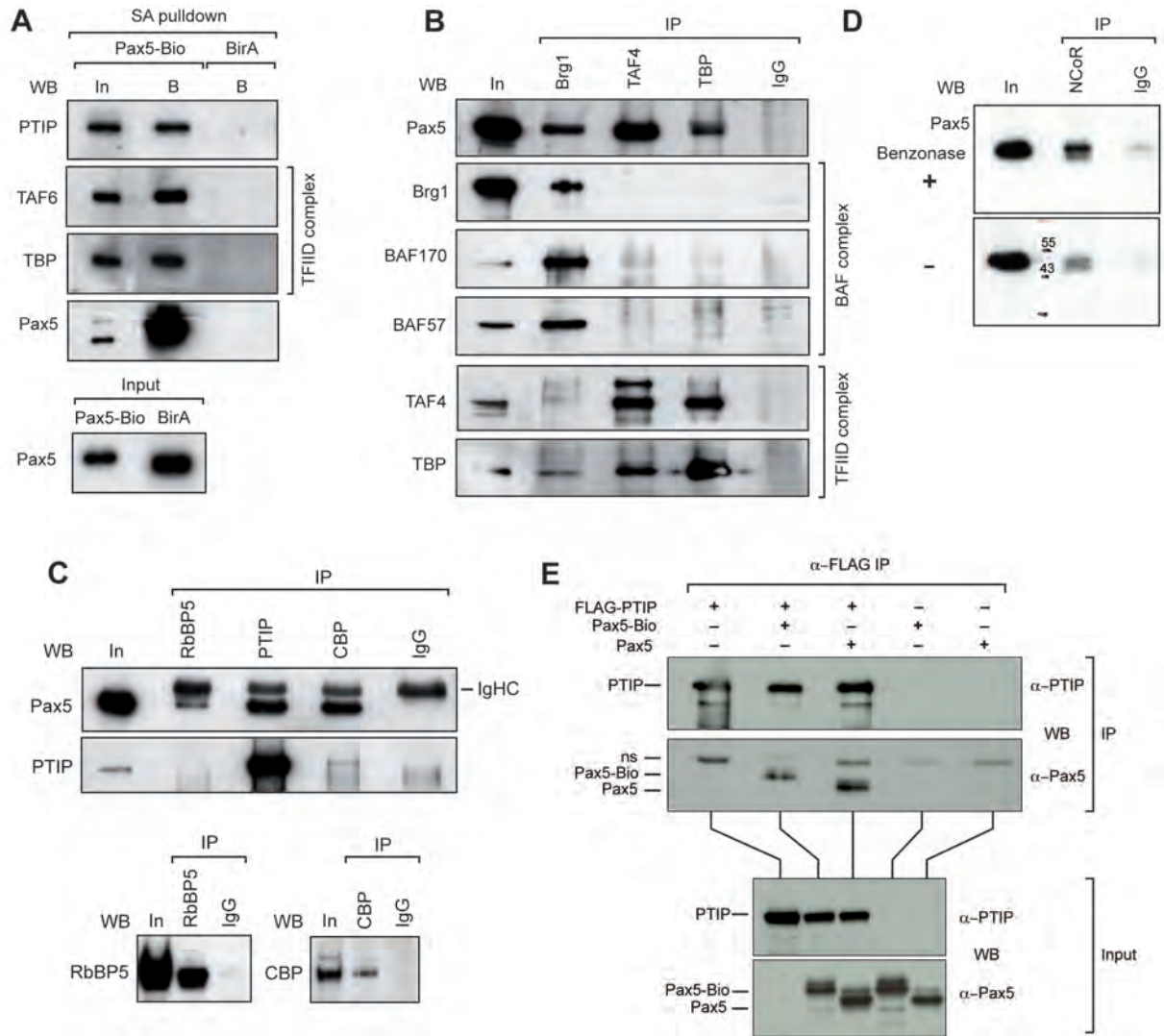


Figure 29. Pax5 recruits histone-modifying, chromatin-remodelling and basal transcription factor complexes.

(A) Copurification of PTIP, TAF6 and TBP with Pax5-Bio. Abelson murine leukemia virus (Ab-MuLV)-transformed pro-B cells of the *Pax5^{Bio/Bio}* (Pax5-Bio) or control *Rosa26^{BirA/BirA}* (BirA) genotype were used for streptavidin (SA) pulldown of nuclear extracts. The input (In) fraction (1/100) and streptavidin-bound (B) precipitate were analyzed by Western blotting (WB) with antibodies (Abs) detecting the indicated proteins. Pax5 was present in similar amounts in both input fractions.

(B,C) Co-immunoprecipitation of Pax5 from nuclear extract of Ab-MuLV-transformed *Pax5^{Bio/Bio}* (B) and *Rag2^{-/-}* (C) pro-B cells with Brg1, TAF4, TBP, RbBP5, PTIP or CBP antibodies. In (B), Pax5 was visualized in the immunoprecipitate (IP) by Western blotting with a biotinylated rat anti-Pax5 mAb (detected with streptavidin-coupled horse radish peroxidase). Input (In; 1/100) and immunoprecipitation with rabbit IgG were used as controls. Only one tenth of the immunoprecipitated fractions were used for Western blotting with the Brg1 antibody. In (C), Pax5 was detected with unlabeled rat anti-Pax5 mAb, which was visualized with an anti-rat IgG Ab that cross-reacted with the heavy-chain (IgHC) of the rabbit IgG Abs (left). Only one tenth of the immunoprecipitated fractions were used for Western blotting with RbBP5 and CBP antibodies (right).

(D) Co-immunoprecipitation of Pax5 and PTIP in transfected 293T cells. CMV-based expression vectors coding for FLAG-tagged PTIP, biotin-tagged Pax5 or wild-type Pax5 were transiently transfected into 293T cells, and whole cell extracts were prepared 48 hr later followed by immunoprecipitation (IP) with the anti-FLAG M2 Affinity Gel (Sigma). Bound proteins were eluted in two bead volumes of 0.3 mg/ml FLAG peptide (Sigma). PTIP and Pax5 were detected in the input and precipitated fractions by Western blot (WB) analysis with the respective antibodies.

(E) Co-immunoprecipitation of Pax5 with NCoR antibodies from nuclear extracts of *Rag2^{-/-}* pro-B cells. The nuclear extracts were prepared with or without benzonase as indicated on the left side (benzonase +/-). Pax5 was visualised in the immunoprecipitate (IP) by Western blotting with a biotinylated rat anti-Pax5 mAb (detected with streptavidin-coupled horse radish peroxidase). Input (In; 1/100) and immunoprecipitation with rabbit IgG were used as controls.

Interaction of Pax5 with non-POU domain containing, octamer-binding protein

In addition to proteins previously known to interact with Pax5, or expected to interact according to their involvement in chromatin modifying-complexes, we identified, by mass spectrometry, proteins involved in processes other than transcription.

Nono (non-POU domain-containing octamer-binding protein, p54nrb) was present in all mass spectrometry replicate experiments. The sum of Nono-belonging peptide sequences across all replicates was among the highest (data not shown). Additionally, Pspc1 (Paraspeckle component 1) a protein reported to interact with Nono was also identified, although with lower reproducibility (Table 9). By streptavidin pulldown (as described above) and Western blot, I could confirm an interaction between Pax5 and Nono (Figure 30A). It resisted lysis of nuclei with 2% Chaps detergent, suggesting that this interaction could be of higher affinity because high detergent content easily disrupts weaker protein-protein engagements. However, in an experiment where nuclear extracts were prepared in the presence of benzonase, the interaction of Nono was significantly decreased (Figure 30B). On the contrary, all other Pax5-interactions described in this study, resisted benzonase treatment. This suggests two possibilities: (1) Nono coprecipitates unspecifically with Pax5 via nucleic acids, (2) Nono and Pax5 together perform a regulatory function which includes a crucial RNA component.

Nono is a member of a DBHS (*Drosophila melanogaster* behaviour, human) family. Together with PSF/Sfpq and PSPC1, Nono is a core component of the paraspeckle complex formed on a scaffold of a long nuclear non-coding RNA (ncRNA), *NEAT1* (Chen and Carmichael, 2009). Paraspeckles retain RNA molecules that have adenosine (A) to inosine (I) hyperedited 3'UTR in the nucleus (Zhang and Carmichael, 2001). These nuclear bodies have been associated with the loss of pluripotency in embryonic stem cells (Chen and Carmichael, 2009). Other processes within the nucleus involving Nono are transcription initiation, RNA processing, transcription elongation and termination (Shav-Tal and Zipori, 2002). A recent finding shows that Nono can recruit the mSin3A-HDAC complex and inhibit progesterone receptor (PR) mediated transactivation (Dong *et al.*, 2009). Which of the numerous functions of Nono include Pax5 remains an open question. Several hypotheses can be proposed. Based on the finding of Pspc1 (described above), we can postulate that Pax5 could localize to paraspeckles where it strongly interacts with Nono, whereas Pspc1 is co-precipitated indirectly via heterodimerization with Nono. The putative interaction probably depends on a RNA molecule, the scaffolding ncRNA or a transcript of a regulated gene. The function of this putative localization remains elusive. Alternatively, Nono could be a core component necessary for recruitment of mSin3A-HDAC co-repressor complex on genes negatively regulated by Pax5, in analogy to the progesterone

receptor. Lastly we can postulate that the interaction has multiple functions whose interplay results in one or more biological functions.

RNA processing			sumPAH	
ID	Name	Function	Pax5	BirA
53610	Nono	Non-POU domain-containing octamer-binding protein	33	8
66645	Pspc1	Paraspeckle component 1	5	0

Table 9. Identification of Nono and Pspc1 by mass spectrometry. The sum of peptide numbers in 5 experiments is shown (sumPAH).

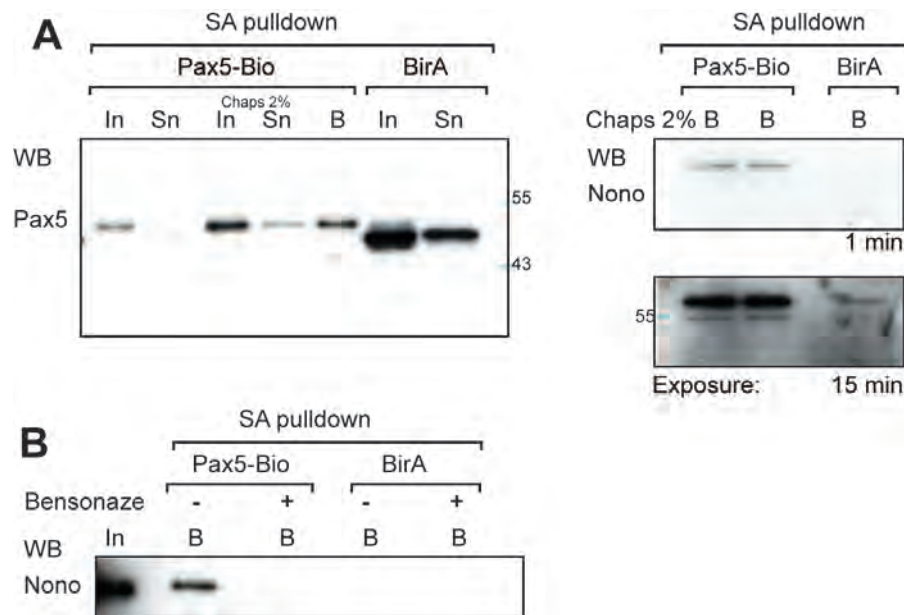


Figure 30. Pax5 interacts with non-POU domain-containing octamer-binding protein (Nono).

Copurification of Nono with Pax5-Bio. Ab-MuLV-transformed pro-B cells, Pax5-Bio or control BirA, were used for streptavidin (SA) pulldown of nuclear extracts prepared with 0.2% or 2% Chaps in the nuclear extraction buffer.

(A) The input (In; 1/100), supernatant (Sn; 1/100) and streptavidin-bound (B; 1/10) fractions were analysed by Western blotting (WB) with anti-Pax5 antibodies (left). Detection of Nono in the streptavidin-bound precipitate by Western blot analysis with an anti-Nono antibody is shown in the right panel. The majority of streptavidin bound precipitate was analysed with anti-Nono antibody. Nono co-purified with Pax5-Bio equally well when the nuclei were extracted with 0.2 or 2 % Chaps, and it was not detected in the control SA pulldown from BirA nuclear extract. Two different exposure lengths are indicated below the blots.

(B) Streptavidin pulldown of nuclear extracts prepared with or without (+/-) benzonase was analysed with anti-Nono antibody. Nono co-purified with Pax5 only if benzonase was not included in the nuclear extract preparation.

Interaction of Pax5 with CTCF

Mass spectrometry analysis also identified CTCF as a so far unknown, potential partner protein of Pax5. The CCCTC-binding factor is a transcriptional regulator with 11 highly conserved zinc finger motifs, which are used in different combinations to bind various DNA target sequences. CTCF binds insulator DNA elements. At “barriers” insulators it prevents spreading of heterochromatin to nearby active regions, whereas at “enhancer blocking” insulators CTCF-binding prevents contact with a distal enhancer (Wallace and Felsenfeld, 2007). Together with cohesin, CTCF is responsible for chromatin loop formation, which confines all promoters and enhancers to a small space and enables a high level of transcriptional activity (Hadjur *et al.*, 2009; Majumder and Boss, 2009). I reciprocally validated the direct interaction of Pax5 and CTCF by streptavidin pulldown of Pax5-Bio protein (Figure 31A) and co-immunoprecipitation with a CTCF-specific antibody (Figure 31B), followed by Western blot analysis.

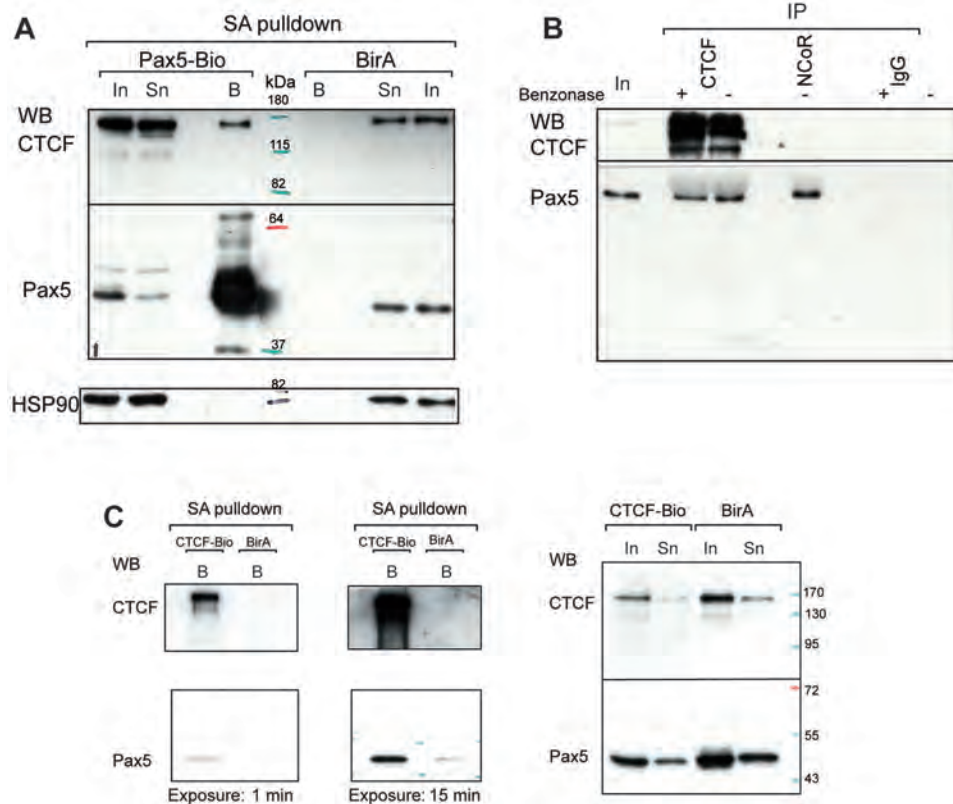


Figure 31. Pax5 interacts with CTCF.

(A) Copurification of CTCF with Pax5-Bio. Ab-MuLV-transformed pro-B cells of *Pax5^{Bio/Bio}* (Pax5-Bio) or control *Rosa26^{BirA/BirA}* (BirA) genotype were used for streptavidin (SA) pulldown from 5 mg of nuclear extracts. The proteins in the input (In) fraction (1/100), supernatant fraction (Sn; 1/100) not bound to streptavidin beads and streptavidin-bound (B) precipitate were analyzed by Western blotting (WB) with antibodies detecting CTCF (rabbit polyclonal) and Pax5 (rat monoclonal).
 (B) Co-immunoprecipitation of Pax5 with NCoR or CTCF antibodies. Nuclear extracts of *Rag2^{-/-}* pro-B cells were prepared with or without (+/-) benzonase. Pax5 was visualized in the immunoprecipitate (IP) by Western blotting with a biotinylated rat anti-Pax5 mAb (detected with streptavidin-coupled HRP). Input (In; 1/100) and rabbit IgG immunoprecipitate were used as controls.
 (C) Copurification of Pax5 with CTCF-Bio. Ab-MuLV-transformed pro-B cells of the *Ctcf^{Bio/+}* (CTCF-Bio) or control *Rosa26^{BirA/BirA}* (BirA) genotype were used for streptavidin (SA) pulldown from 5 mg of nuclear extracts. The pulldown was analysed by Western blot with anti-Pax5 and anti-CTCF antibodies. Different film exposure lengths are indicated below the blots.

Through collaboration with Dr. Niels Galjart (Erasmus MC) we had access to *Ctcf*^{Bio/+} knock-in mice (van de Nobelen *et al.*, 2010). This allowed me to performed a streptavidin pulldown with extracts of CTCF-Bio pro-B cells and showed by Western blot analysis that Pax5 was co-precipitated (Figure 31C). With this experiment, I confirmed that co-immunoprecipitation of Pax5 with an anti-CTCF antibody was not due to unspecific recognition by the polyclonal anti-CTCF antibody that I used. Additionally, I demonstrated that the interaction is direct and DNA-independent, as the antibody could co-immunoprecipitate similar amounts of Pax5 with CTCF after benzonase treatment of the nuclear extract (Figure 30B).

I next wanted to identify the domain of CTCF that interacts with Pax5. For this purpose I created six deletion mutants of CTCF (Figure 32A) where the following domains were retained: (1) N-terminus and zinc-fingers domain (NT-ZF), (2) zinc fingers and the C-terminus (ZF-CT), (3) N-terminus fused directly to the C-terminus with deleted zinc fingers (NT-CT), (4) zinc – fingers (ZF), (5) N-terminus (NT), (6) C-terminus (CT). I cloned these deletion mutants together with a C-terminal FLAG-V5-Biotin tag into CMV-based expression vectors. The mutants were co-expressed with the full-length Pax5 protein into transiently transfected HEK293T cells (Figure 32B). I analysed their interaction with Pax5 in immunoprecipitation with FLAG-M2-agarose, followed by FLAG-peptide elution and Western blot analysis with a Pax5-specific antibody. The assay was done in the presence of benzonase. The C-terminal domain (CT) was a very weak interaction partner (Figure 32C). In contrast, the N-terminal domain (NT) of CTCF interacted with Pax5 with the highest affinity, even superior then the full length CTCF protein (Figure 32C). Although the zinc finger domain (ZF) *per se* did not interact with Pax5, it stabilised the interaction of other domains (Figure 32C; compare Pax5 band signals on the WB in ZF-CT vs. CT and NT-ZF vs. NT lanes), probably by localizing CTCF to the DNA where it had a higher probability to encounter Pax5. Western blot of the whole cell lysates that I used for the immunoprecipitation showed that all mutants were expressed in HEK293T cells (Figure 32B), and purified with the FLAG agarose (Figure 32D) therefore the decrease in the interaction-affinity is not the consequence of lower expression levels of the mutants.

In summary, this experiment showed that the N-terminus of CTCF interacts with Pax5 more strongly compared to the C-terminus, while the zinc-finger domain of CTCF is dispensable for the interaction with Pax5.

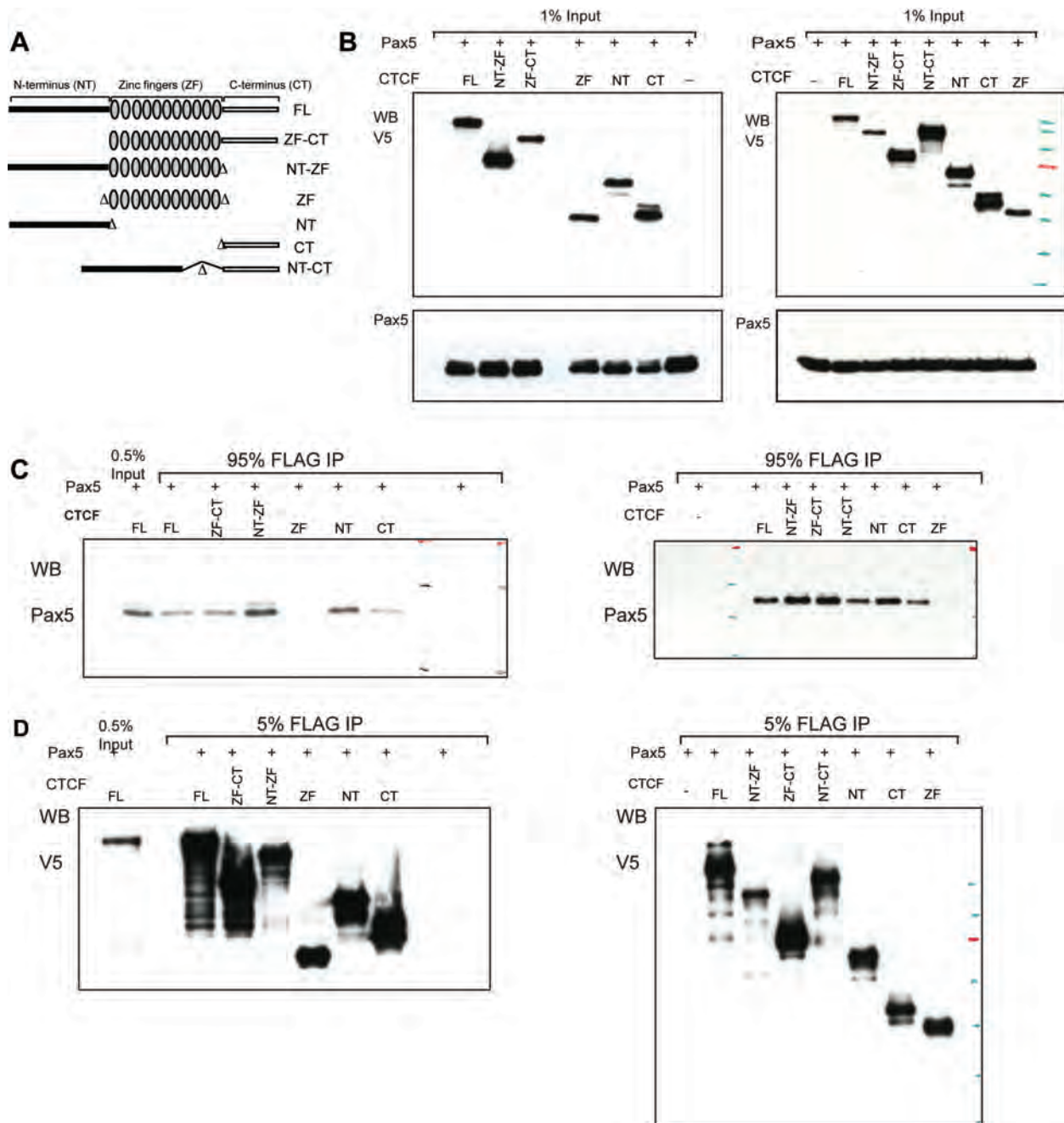


Figure 32. Pax5 interacts with the N-terminal domain of CTCF in HEK293T cells.

CMV-based expression vectors coding for FLAG-V5-bio tagged CTCF or its deletion mutants and wild-type Pax5 were transiently transfected into HEK293T cells. Whole cell extracts were immunoprecipitated (IP) with the anti-FLAG M2 Affinity Gel (Sigma), in the presence of benzonase, and eluted with FLAG peptide (Sigma). Western blots (WB) were analysed with anti-V5 and anti-Pax5 antibodies. Two experimental and biological replicas are shown.

(A) Schemes of full-length CTCF (FL) and FLAG-V5-bio tagged mutants with the following domains deleted: C-terminus (NT-ZF), N-terminus (ZF-CT), zinc-finger (NT-CT), zinc fingers and C-terminus (NT), N-terminus and zinc fingers (CT), N- and C-terminus (ZF).

(B) Western blot of the input extract used for the FLAG-immunoprecipitation (In; 1/100 of the extract used for FLAG-IP) showing that all CTCF deletion mutants (upper blot, anti-V5) and Pax5 (lower blot) are expressed. The mutants have a functional tag detected with anti-V5 antibody.

(C) 95% of FLAG-IP was analysed with anti-Pax5 antibody. In the control IP (only Pax5 was transfected, CTCF [-]), there is no unspecific precipitation of Pax5. Zinc finger domain (ZF lane) of CTCF does not interact with Pax5, while the N-terminal (NT and NT-ZF lanes) domain shows higher affinity for Pax5.

(D) 5% of FLAG-IP was analysed with anti-V5 antibody showing that CTCF deletion mutants were successfully purified.

In order to investigate potential protein complexes where Pax5 and CTCF could co-localise, I took advantage of the *Ctcf^{Bio/+}* mouse. I performed streptavidin pulldowns with CTCF-Bio and Pax5-Bio nuclear extracts and identified the precipitated proteins by mass spectrometry (Figure 33A). Proteins identified by mass spectrometry analysis of both Pax5-Bio and CTCF-Bio pro-B cells, but not identified in the control BirA pulldown, were distributed into functional categories (Figure 33B) according to Gene Ontology. Most of the CTCF/Pax5-shared proteins were involved in processes of DNA replication and cell cycle as exemplified by cyclins (*Ccnt1*), replication-activating helicases (*Rfc2,3,4,5*), proteases of the anaphase promoting complex and finally the cohesin complex (Figure 33B, Table 13).

Genome-wide ChIP-chip studies discovered that cohesin occupancy was enriched across the genome at CTCF sites (Parelho *et al.*, 2008; Stedman *et al.*, 2008; Wendt *et al.*, 2008). Moreover cohesin is recruited by CTCF and can thereafter stabilise the looping of complex loci (Hadjur *et al.*, 2009). In agreement with these studies our mass spectrometry analysis identified all 4 subunits of cohesin, *Smc1*, *Smc3*, *Rad21/Scc1* and *Scc3/SA2* as the top scoring putative CTCF-Bio interaction partners (Table 13). Although CTCF directly interacts only with the SA2 subunit of cohesin (Xiao *et al.*, 2011), we found higher enrichment for *Smc1a* and *Smc3*, which are components of the cohesin ring. The reason might be that other proteins, like Pax5, interact with both CTCF and *Smc1a/Smc3* subunits, or that their overrepresentation is simply a technical bias of the mass spectrometry analysis. Surprisingly, mass spectrometry analysis of Pax5-Bio streptavidin pulldown experiment that was done in parallel, also identified all cohesin subunits, except *Scc3/SA2*, although their enrichment was lower than with CTCF-Bio. By comparing the number of peptides belonging to cohesin subunits, identified in Pax5-Bio and CTCF-Bio mass spectrometry analysis (Table 13), we can assume that the interaction between Pax5 and cohesin is weaker than that of CTCF and cohesin. Other proteins potentially interacting with both Pax5 and CTCF are involved in regulation of gene expression (Table 11 and 12), they include transcription factors and components of chromatin regulator-complexes, for example, transcription factors *Runx/Cbfb* or *Chd8* helicase of the MLL-containing H3K4 methyltransferase complex, which has been previously reported to be a CTCF-interaction partner (Ishihara *et al.*, 2006). This might be a consequence to the fact that both factors are involved in gene-regulation processes.

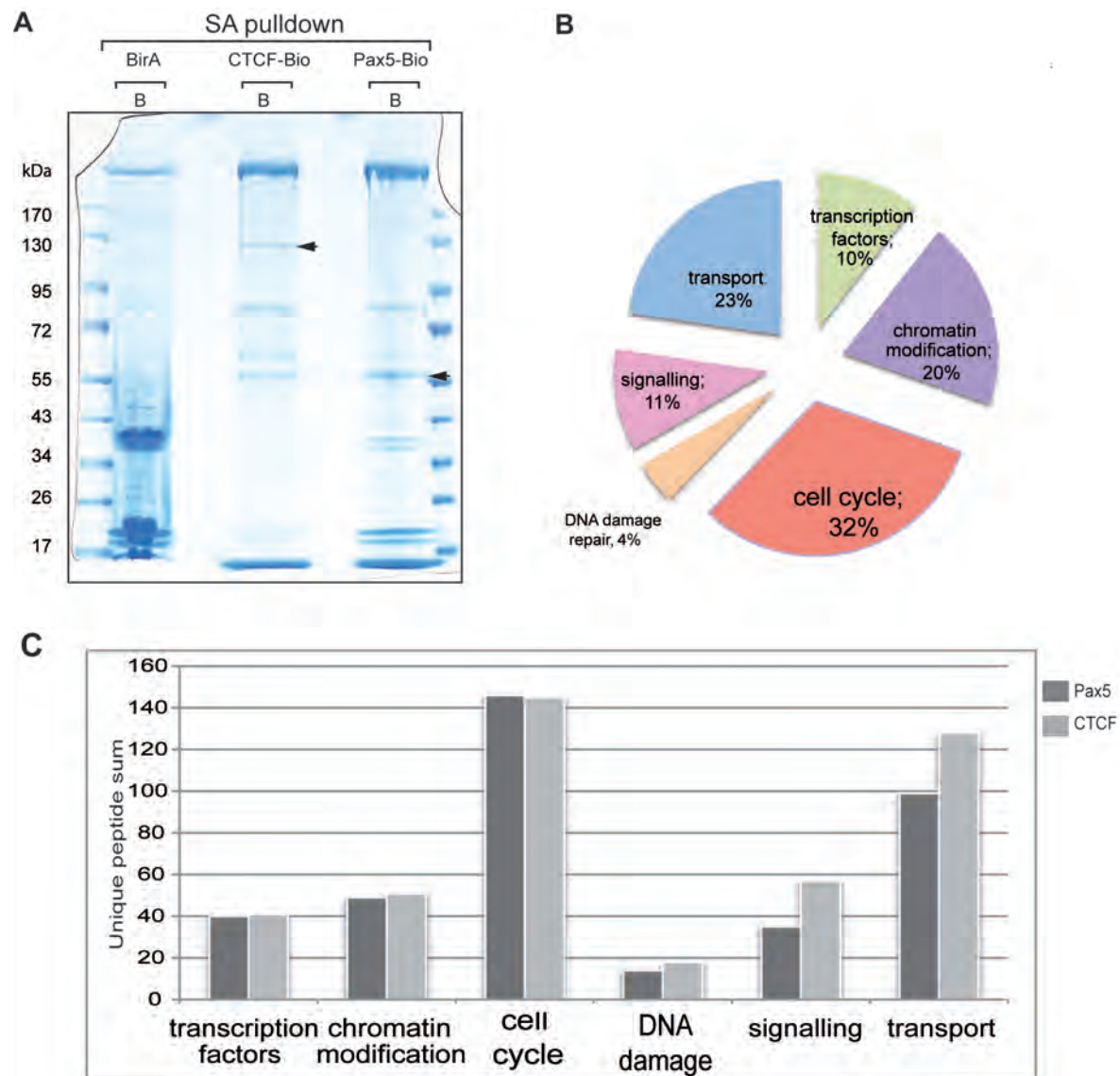


Figure 33. Identification of proteins that interact with both Pax5 and CTCF, by mass spectrometry.

Abelson murine leukemia virus transformed pro-B cells of the *Pax5^{Bio/Bio}* (Pax5-Bio), *Ctcf^{Bio/+}* (CTCF-Bio) or control *Rosa26^{BirA/BirA}* (BirA) genotype were used for streptavidin pulldown from 25 mg of nuclear extracts. The purified proteins were subsequently identified by mass spectrometry.

(A) Colloidal Coomassie blue-stained gel of streptavidin (SA) pulldown experiments. Shown are lanes containing markers (M) and proteins eluted from the streptavidin beads (B). The arrowheads indicate bands where Pax5 or CTCF were identified by mass spectrometry.

(B) Gene Ontology classification of proteins specifically identified by mass spectrometry in Pax5-Bio and CTCF-Bio SA pulldowns, but not in the control (BirA) SA pulldown. Numbers of identified peptides of all proteins classified to the same GO group were summed and the distribution visualised by the pie chart shown.

(C) Bar diagram comparing total numbers of peptide sequences of all proteins classified in the same functional group in Pax5-Bio and CTCF-Bio pulldowns.

Chromatin remodelling and modification - overlap of Pax5-Bio and CTCF-Bio				sumPAH	
	ID	Name	Function	Pax5	CTCF
BAF	68094	Smarcc2	BAF170 subunit of BAF complex	0	2
	20587	Smarchb1	BAF47 subunit of BAF complex	2	2
	83796	Smardc2	BAF60B subunit of BAF complex	2	3
MLL	67772	Chd8	Chromodomain helicase DNA binding protein 8	5	8
	19046	Ppp1cb	MLL5-L member protein phosphatase 1, catalytic subunit, beta isoform	2	3
	19045	Ppp1ca	MLL5-L member protein phosphatase 1, catalytic subunit, alpha isoform	6	2
NCoR	15161	Hcfc1	Host cell factor C1	7	6
	433759	Hdac1	Histone deacetylase 1	6	6
	20185	Ncor1	Nuclear receptor co-repressor 1	0	3
	21372	Tbl1x	Transducin (beta)-like 1 X-linked	2	5
	68926	Ubap2	Ubiquitin-associated protein 2	3	3
	70790	Ubr5	Ubiquitin protein ligase E3 component n-recogin 5	2	2

Table 11. Chromatin regulators copurified with Pax5-Bio and CTCF-Bio. These proteins were not identified in the control SA pulldown with BirA cells. The numbers refer to peptides identified for the indicated proteins. If the number for Pax5-Bio is 0 the protein was identified in a previous mass spectrometry analysis. Coloured borders indicated components found in the respective activating (green) or repressive complexes (red).

Transcription factors - overlap of Pax5-Bio and CTCF-Bio				sumPAH	
ID	Name	Function	Pax5	CTCF	
12399	Runx3	Runt related transcription factor 3	4	3	
12400	Cbfb	Core binding factor beta	4	3	
20848	Stat3	Signal transducer and activator of transcription 3	4	8	
229700	Rbm15	RNA binding motif protein 15	2	6	
22661	Zfp 148	Zinc finger protein 148, transcription factor ZBP-89	5	5	
70579	Zc3h11a	Zinc finger CCCH type containing 11A	8	5	
56805	ZBTB33	Kaiso, zinc finger and BTB domain containing 33	13	12	

Table 12. Transcription factors copurified with Pax5-Bio and CTCF-Bio. These proteins were not identified in the control SA pulldown with BirA cells. The numbers of peptide sequences are shown.

Cell cycle/ replication - overlap of Pax5-Bio and CTCF-Bio				sumPAH	
ID	Name	Function	Pax5	CTCF	
24061	Smc1a	SMC protein 1A, SmcB	12	32	
13006	Smc3	Structural maintenance of chromosomes 3	7	28	
	SCC3	Cohesin subunit SA-2, SCC3 homolog 2	0	9	
19357	SCC1	Double-strand-break repair protein Rad21 homolog	3	6	
70099	Smc4		15	9	
215387	Ncaph	Non-SMC condensin I complex, subunit H	4	4	
19718	Rfc2	Replication factor C (activator 1) 2	9	4	
69263	Rfc3	Replication factor C (activator 1) 3	10	7	
106344	Rfc4	Replication factor C (activator 1) 4	13	6	
72151	Rfc5	Replication factor C (activator 1) 5	6	4	
17215	Mcm3	Minichromosome maintenance deficient 3	5	2	
17220	Mcm7	DNA replication licensing factor MCM7	6	4	
51869	Rif1	Rap1 interacting factor 1 homolog, S-phase checkpoint	14	7	
12455	Ccnt1	Cyclin T1	6	3	
107951	Cdk9	Cyclin-dependent kinase 9	7	7	
18854	Pml	Promyelocytic leukemia	3	2	
19181	Psmc2	Proteasome (prosome, macropain) 26S subunit, ATPase 2	2	2	
19184	Psmc5	Protease (prosome, macropain) 26S subunit, ATPase 5	3	4	
67089	Psmc6		8	3	
26442	Psma5	Proteasome (prosome, macropain) subunit, alpha type 5	2	4	
77891	Ube2s	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 13	9	5	

Table 13. Proteins involved in cell cycle and DNA replication. These proteins were not identified in the control SA pulldown with BirA cells. SA2, cohesin subunit that directly interacts with CTCF did not copurify with Pax5-Bio, but is listed here as a cohesin component. The vertical border line indicates cohesin subunits.

RNA processing - overlap of Pax5-Bio and CTCF-Bio			sumPAH	
ID	Name	Function	Pax5	CTCF
53610	Nono	Non-POU domain-containing octamer-binding protein	6	8
66645	Sfpq	Polypirimidine tract-bindin protein-associated-splicing factor, PSF	10	6

Table 14. Paraspeckle components copurified with Pax5-Bio and CTCF-Bio. These proteins were not identified in the control SA pulldown with BirA cells.

CTCF-Bio - unique transcription factors			
ID	Name	Function	CTCF
14886	Gtf2i	TFII-I, general transcription factor Iii	11
70239	Gtf3c5	TFIIICepsilon, General transcription factor IIIC, polypeptide 5	2
269252	Gtf3c4	Cohesin subunit SA-2, SCC3 homolog 2	2
779429	Taf9b	Transcription initiation factor TFIID subunit 9B	2
13709	E74	E47-like factor 1 (Ets-domain)	8
22780	Ikzf3	Aiolos	7
17261	Mef2D	Myocyte enhancer factor 2D	2
20850	Stat5a	Signal transducer and activator of transcription 5A	2
20371	Foxp3	Forkhead box P3	2
303905	Parp9	Replication factor C (activator 1) 5	5
213464	Rbbp5	Retinoblastoma-binding protein 5	2
14886	Gtf2i	TFII-I, general transcription factor Iii	11

Table 15. Proteins copurified only with CTCF-Bio. These proteins were not identified in the control SA pulldown with BirA cells.

In summary, we report for the first time, the interaction between Pax5 and CTCF. Pax5 co-purified with the CTCF-antibody despite the treatment of the nuclear extract with benzonase, showing that DNA does not unspecifically mediate the interaction. Interestingly, the N-terminus of CTCF recruits Pax5 more efficiently than the full length CTCF protein does, which may suggest that the C-terminal domain of CTCF has a suppressive effect on the interaction with Pax5. CTCF is known to co-localise with cohesin in a genome-wide manner (Parelho *et al.*, 2008; Stedman *et al.*, 2008; Wendt *et al.*, 2008). CTCF connects its bound elements through dimerisation via the zinc fingers (Yusufzai *et al.*, 2004), while cohesin stabilises the formed loops and is responsible for long-range interactions (Hadjur *et al.*, 2009). Expression of a cell type specific factor can elicit formation of additional loops, through preferential recruitment of CTCF to the binding sites of this factor (Hadjur *et al.*, 2009; Sekimata *et al.*, 2009; Choi *et al.*, 2011). We hypothesize that Pax5, by directly interacting with CTCF bound to PAIR elements, can employ the dimerization and the consequent DNA-looping function of CTCF to stimulate *Igh* locus contraction. Our preliminary result that both Pax5 and CTCF are able to interact with cohesin supports this hypothesis.

RESULTS II

CTCF, Pax5 and E2A bind to conserved PAIR sequences

Although regulatory elements in the 3' region of the *Igh* locus that play a role in V(D)J recombination have been characterised, no such elements located in the V_H region have been known until recently. Our group discovered conserved repeat elements upstream of the V_H3609 genes that have Pax5-dependent active chromatin marks in pro-B cells and were therefore named Pax5-activated intergenic repeat (PAIR) elements (Ebert *et al.*, 2011). In addition to Pax5, the PAIR elements were bound by CTCF and E2A in ChIP-Chip and *in vivo* footprint analyses. I contributed to this study by analysing the binding of E2A and Pax5 to the PAIR elements by electrophoretic mobility shift assay (EMSA). Sequence conservation revealed that 10 of the 14 PAIR elements contain CTCF-binding sites homologous to the CTCF consensus recognition sequence (Figure 34A). The E2A-binding sites of ten PAIR elements match the consensus E2A binding motif. To verify the binding of E2A, I used the E2A recognition sequence of PAIR7 as a probe for EMSA with a pre-B cell nuclear extract (Figure 34B). The PAIR7 probe interacted with a protein identified as E2A, as it resulted in a supershift with an E2A antibody. Next, I performed titration experiments (Figure 34B) with unlabelled competitor oligonucleotides containing high affinity E2A-binding sites (μ E5 and κ E2), in the *Igh* E μ and the *Igk* enhancers (Murre *et al.*, 1989), and a mutant E2A binding sequence from PAIR5. This demonstrated that E2A binds the sequences in PAIR7 with similarly high affinity as the μ E5 and κ E2 sites. However, the PAIR5 sequence did not compete for E2A binding indicating that E2A does not interact with the mutant sequences of PAIR5 and PAIR1. Nine PAIR elements contained Pax5-binding sequences homologous to the Pax5 consensus recognition sequence (Cobaleda *et al.*, 2007). The protein binding to the corresponding PAIR7 probe was identified as Pax5 because an antibody raised against the Pax5 paired domain abolished its binding in a pro-B cell nuclear extract (Figure 33C). Titration with unlabelled competitor DNA (Figure 34C) demonstrated that Pax5 binds to the site in PAIR7 with similar affinity as the reference Pax5-binding site in the human *CD19* promoter (Kozmik *et al.*, 1992). PAIR10 has a single-nucleotide substitution, which impaired Pax5 binding. In contrast, an insertion of an A residue in the Pax5-binding motif of PAIR8 prevented Pax5 binding in agreement with a previous mutagenesis analysis of Pax5-binding sites (Czerny *et al.*, 1993). For comparison, a 3 bp substitution in the mutant PAIR7M sequence was as inefficient in competing for Pax5 binding as the sequence from PAIR8 (Figure 34C). In summary, most PAIR sequences contain functional CTCF, E2A and Pax5-binding sites, whereas nucleotide substitutions or insertions weaken or abolish protein binding in only a few of these elements.

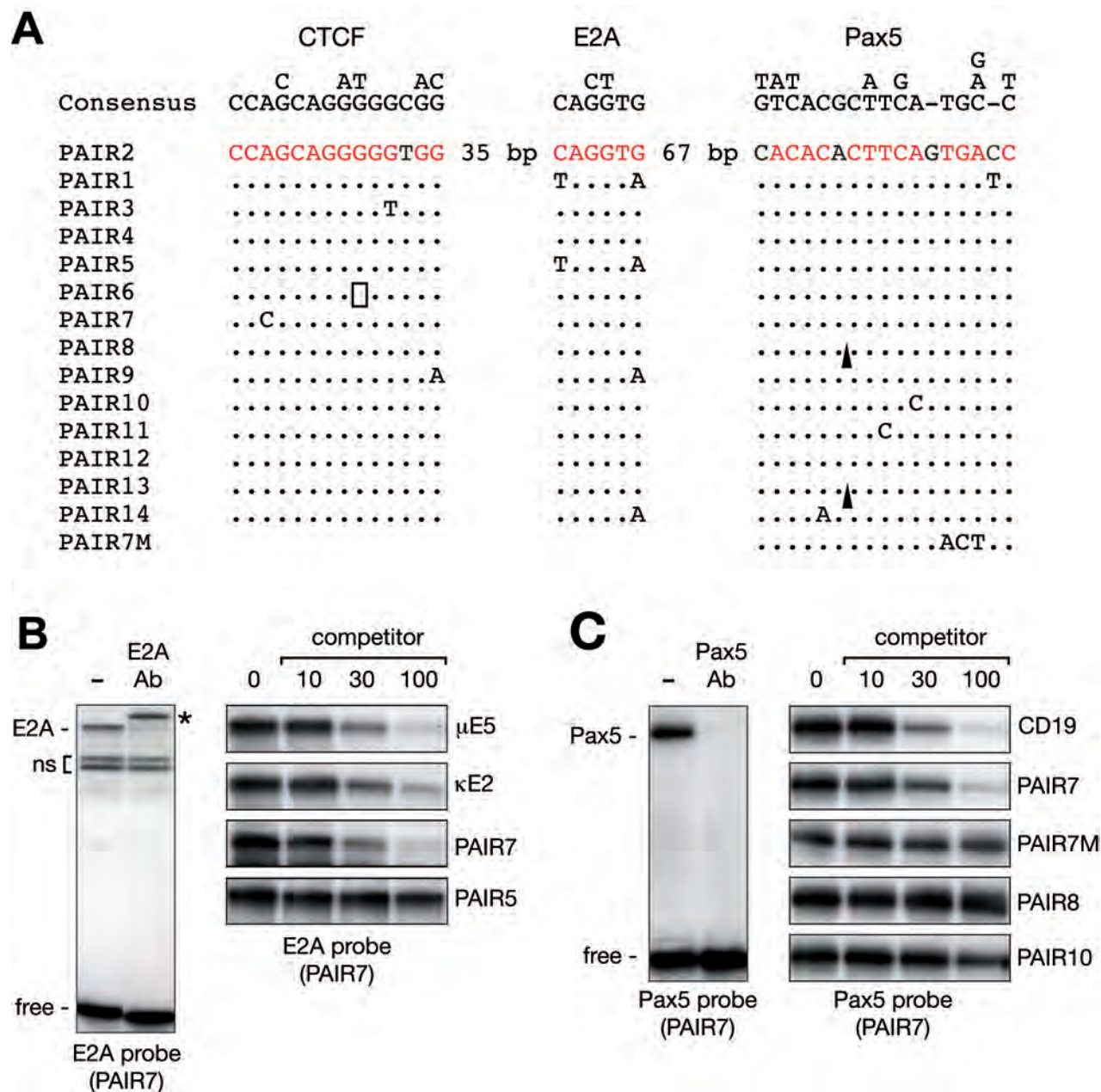


Figure 34. CTCF, Pax5, and E2A Bind to Conserved PAIR Sequences

(A) Sequence conservation. Nucleotides of PAIR2 with homologies to the consensus sequence of CTCF (Cuddapah *et al.*, 2009), Pax5 (Cobaleda *et al.*, 2007), or E2A (TRANSFAC Database) are indicated in red, and dots denote identical nucleotides in other PAIRs. An empty box indicates the absence of a G-residue in PAIR6, and arrowheads denote the insertion of an A-residue in PAIR8 and PAIR13.

(B and C) High-affinity binding of E2A (B) and Pax5 (C) to PAIR elements. A double-stranded oligonucleotide containing the E2A or Pax5 recognition sequence of PAIR7 was used as a probe for electrophoretic mobility shift assay (EMSA) with a nuclear extract of in vitro-cultured 70Z/3 pre-B cells (B) or wild-type pro-B cells (C), respectively. The presence of an E2A antibody (Ab) resulted in a supershift of the E2A-DNA complex (mark by an asterisk), whereas the addition of a Pax5 antibody prevented Pax5 binding. Double-stranded oligonucleotides containing the E2A- and Pax5-binding sites of other PAIRs were analyzed at 10-, 30-, and 100-fold molar excess for competition of protein binding. The PAIR7M oligonucleotide contained a 3-bp substitution in the Pax5-binding site (A). Recognition sequences of the human CD19 promoter (Kozmik *et al.*, 1992) or the mouse *Igh* E μ (μ E5) and *Igk* iE κ (κ E2) enhancers (Murre *et al.*, 1989) served as reference high-affinity binding sites for Pax5 and E2A, respectively. The E2A probe detected two non-specific (ns) proteins. All oligonucleotides used are listed in Materials and methods. (From: Ebert *et al.*, 2011)

Quality controls of the circular chromosome conformation capture (4C) method

Spatial organisation of the *Igh* locus plays an important role in the control of ordered and lineage-specific assembly of the discontinuous immunoglobulin gene segments. Considering the length of the V_H region, if the recombination occurred stochastically, it would be strongly biased towards the proximal V_H gene segments, since the distal ones are up to 2.5 Mb away from the D_H-J_H region. To overcome these constraints the *Igh* locus assumes a contracted topology at the pro-B cell stage. Fluorescence *in situ* hybridization (FISH) analysis demonstrated that contraction indeed serves to bring the distal V_H gene segments in proximity to the D_H region at the pro-B cell stage, while in *Pax5*^{-/-} pro-B cells they localise at a distance corresponding to an extended topology of the *Igh* locus (Fuxa *et al.*, 2004). However, the gene segments that colocalise in the FISH analysis do not necessarily directly interact, since the resolution of this method is less than 0.2 μm (Göndör *et al.*, 2008). The chromosome conformation capture (3C) technology (Dekker *et al.*, 2002) has revolutionised studies on chromatin looping and spatial organisation of the genome, which allows the examination of interactions between proximal molecules at the resolution of less than 5 Å (Göndör *et al.*, 2008). Therefore we utilised the circular chromosome conformation capture method (4C) (Simonis *et al.*, 2006) to compare the spatial organisation of the *Igh* locus in *Rag2*^{-/-} pro-B cells and *Rag2*^{-/-} *Pax5*^{-/-} pro-B as well as mouse embryonic fibroblasts (MEFs).

In brief, for 4C library preparation (Figure 35A), the cells are crosslinked with formaldehyde to fix the DNA-protein interactions in a cell population. The crosslinked chromatin is cleaved with a six base-pair recognising restriction enzyme and these fragments are then ligated at low DNA concentration. Under these conditions ligations between crosslinked fragments are strongly favoured over ligations between random fragments. After ligation the crosslinks are reversed and the DNA is purified. To improve the resolution of the interactions, the ligation products are cleaved with a frequently cutting secondary restriction enzyme and religated in dilution to form small circular DNA molecules. This represents the 4C-DNA library then amplified with primers that are specific for the restriction fragment of interest but that amplify the adjacent unknown sequences. Previously, the PCR products were analysed by quantitative PCR or Chip microarrays (Simonis *et al.*, 2006). Nowadays, next generation sequencing provides a more comprehensive analysis of the entire DNA interactome of our locus of interests without any *a priori* knowledge of interacting regions.

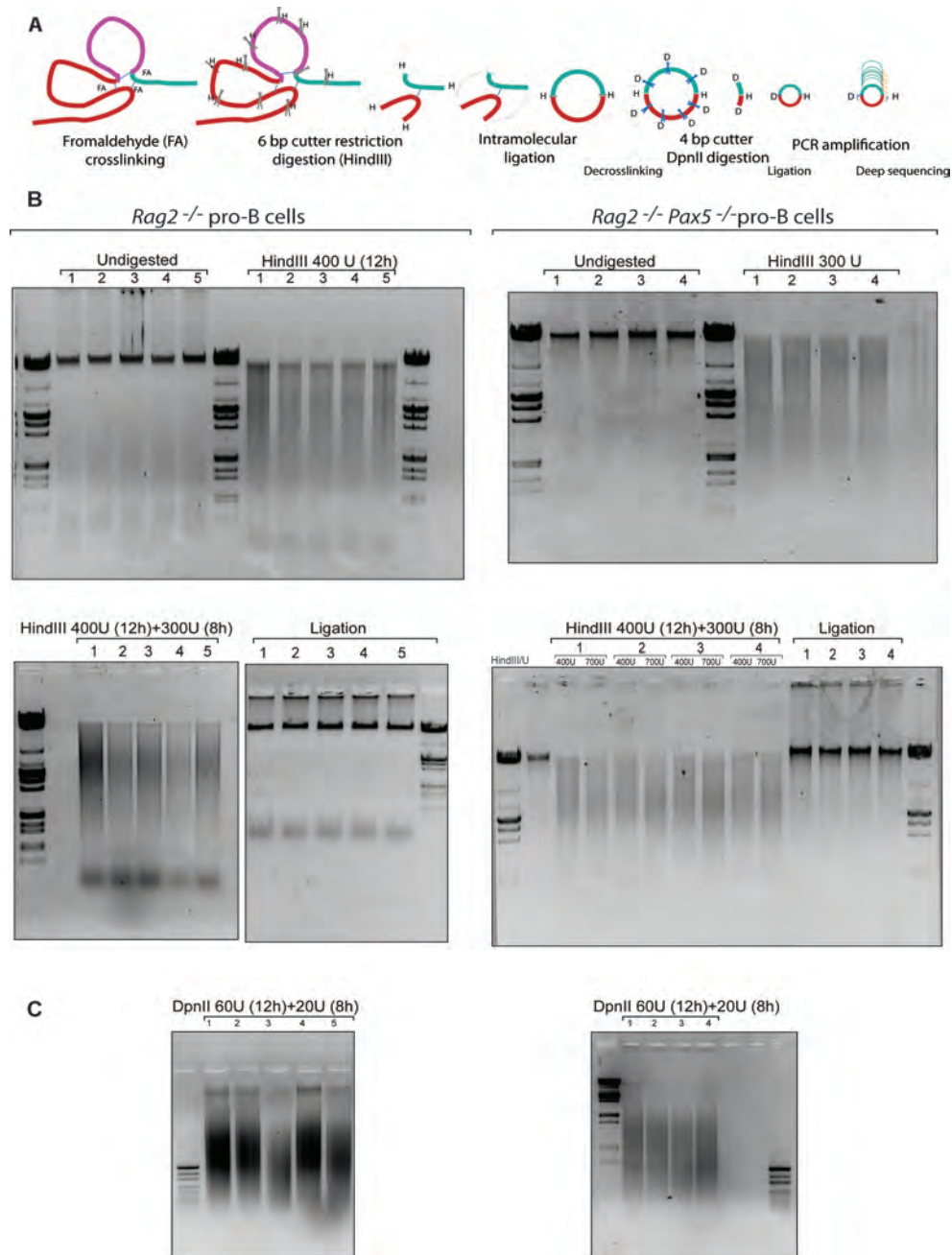


Figure 35. Workflow of the circular chromosome conformation capture (4C) library preparation.

(A) Schematic diagram of the workflow of 4C-library preparation. The products of steps indicated in bigger font, crosslinking, HindIII digestion, ligation and DpnII digestion, are shown in (B) and (C). The products of 4C-PCR amplification are shown in figure 36C.

Ten million cells were crosslinked with 2% formaldehyde (PBS, 10% FCS) and lysed. The nuclei (Undigested) were digested with HindIII and ligated in high dilution. After decrosslinking and RNase treatment, the purified DNA was digested with DpnII and religated creating the template for the investigation of DNA interactions. In each step of the procedure one percent of the products was purified and visualised by agarose gel electrophoresis. The cell type used for 4C-library preparations is indicated: left side *Rag2*^{-/-} pro-B cells, right side *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells. The numbers below the brackets indicate replicates of 4C-library preparations.

(B) HindIII restriction digestion and subsequent ligation efficiency analysed by comparing purified input (Undigested) and restriction digested DNA (HindIII) on a 0.6% agarose gel. Input chromatin (left) was digested sequentially with 700 U of HindIII in total. Incubation with 400 U of HindIII over night (right) still leaves a lot of undigested DNA (upper gel, right). Addition of fresh 300 U of HindIII and prolonging the incubation for another 8 hours improves the digestion efficiency (lower gel, left). Ligated HindIII restriction fragments shift up as one band (lower gel, on the right). The bands of low molecular weight and weaker intensity represent the RNA.

(C) Decrosslinked ligation products were treated with RNase A and incubated with 80 U of DpnII in total. Digestion efficiencies after the incubation with 60 U of DpnII over night (left) and 20 U over day (right) are shown, separated on 1.5 % agarose gels.

One of the caveats of the 4C technology is that lower efficiency in each of the enzymatic reactions can create a bias that could potentially lead to wrong conclusions. Therefore it is important to carefully monitor each step of the procedure. We chose the 6 bp cutter HindIII as the first restriction enzyme (Figure 35B). This restriction digestion is the most sensitive step, as it needs to be as complete as possible. Regulatory DNA elements bind transcription factors and thus contain fewer histones. Hence their hypersensitivity to nuclease digestion could create a bias for these sites in 4C experiments (Simonis *et al.*, 2007). The bias is avoided by digesting overnight in a large excess of the restriction enzyme (Simonis *et al.*, 2007). The efficiency strongly depends on the compaction of the chromatin in the cell type chosen for the study. Cell lines that have been cultivated for longer periods tend to yield low digestion frequencies. In contrast, the chromatin of primary or short-term cultured cells can be up to 90% digested. Because of this reason, I used *ex vivo* isolated *Rag2*^{-/-} pro-B cells, which were expanded in culture for no longer than 4 days. I improved the efficiency of digestion by sequentially adding the enzyme (Figure 35B, compare upper and lower gel pictures). After ligation the smear of HindIII-digested DNA fragments shifted back to the position of uncleaved chromatin as only one high molecular weight band was detected (Figure 35B, lower gel). As the secondary restriction enzyme I sequentially added the 4-bp cutter DpnII. The majority of the DNA was digested (Figure 35C). Finally, the second ligation created a 4C library consisting of circular molecules used for the PCR amplifications (not shown).

In summary, with minor modifications, as sequential addition of restriction enzymes, I optimised the 4C protocol for the utilisation of short-term cultured *Rag2*^{-/-} pro-B-cells. Moreover, the same protocol also proved to be successful for *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells and short-term cultured mouse embryonic fibroblasts (MEFs).

PCR amplification of the 4C library

Ligation frequencies of the restriction fragments are analysed by PCR using primers specific for the region of interest (anchor HindIII fragment). The primer anchored exactly in the HindIII site consists of a sequence specific for the element of interest and an extension of the sequencing adapter that allows direct deep sequencing of the PCR products (Figure 36A). The second primer is designed close to, but not necessarily at the DpnII restriction site and contains the second extension important for the deep sequencing (Figure 36A).

In order to investigate the topological organisation of the *Igh* locus formed prior to V(D)J recombination, we designed primers anchored at six different positions throughout the locus (Figure 36B); two were localised in the 3' end, one in the proximal V_H region and three in PAIR elements of the distal V_H region. Interactions of the regulatory elements at the 3' end of the locus have already been investigated (Ju *et al.*, 2007). Thus the anchor points in the HS3B and the *Eu* enhancer could serve as controls. However, previous studies required *a priori* assumptions about the interacting regions, based on other data (DNAase hypersensitivity, transcription factor binding). Hence they focused on more local interactions. Our application of deep sequencing should extend the detection of the interactions into the distal 5' end of the locus up to the "putative 5' regulatory elements" the PAIRs. The primer in the proximal V_H region is anchored in a HindIII fragment containing a pseudogene (PG.4.28) and a binding site for CTCF (Figure 36B). Since this site is neither a regulatory element nor an active gene, we expect lower interaction frequencies with the regulatory elements of the locus. In the distal V_H region we focused on the PAIR elements (Figure 36B, right). Although it would be ideal to analyse all PAIRs individually, and their contribution to the overall contraction of the *Igh* locus, we were limited in this intention by several constraints originating from the structure of the V_H region. LINE elements constitute 40.4 % of the mouse V_H region. Moreover, this region has a greater proportion of interspersed repeats (52.4%) than most of the mouse genome (39%) (Johnston *et al.*, 2006). The PAIRs themselves are flanked by LINE repeats (Maria Novatchkova, personal communication). For this reason we succeeded to design unique primers anchored in only three PAIR elements: PAIR4, PAIR5 and PAIR8 (Figure 36B, indicated by arrows). We analysed the interactions of these six anchor fragments in *Rag2*^{-/-} pro-B cells, *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells and MEFs.

The primers amplify the unknown fragments that are juxtaposed with the anchor fragment in the three dimensional nuclear space. Therefore, the abundance of the PCR products directly correlates with the frequency of such encounters (depicted on Figure 36A, on the right side of the scheme), which emphasizes the importance of an accurate quantification and normalisation.

Independently of the anchor restriction fragment to be analysed and the cell type used for the preparation of the 4C library, two types of junctions are always over-represented (Simonis *et al.*, 2007) and give rise to the most prominent PCR products. The first most abundant junction, with the neighbouring HindIII fragment, results from the incomplete restriction digestion (Figure 36C, green boxes) and can constitute up to 20-30 % of all the junctions (Simonis *et al.*, 2007). The second product is a consequence of a circularisation of the anchor restriction fragment, self-ligation (Figure 36C, red boxes) and accounts for 5-10 % of all junctions (Simonis *et al.*, 2007). The percentage of other junctions quickly drops with increasing genomic distance, to below 0.1% unless the two sites are engaged in a specific interaction (Simonis *et al.*, 2007). To accurately quantify rare events that often occur in less than 1000 cells, it is necessary to include many genome equivalents in the PCR amplification (Simonis *et al.*, 2007). For this reason, I pooled products of 16 individual PCR amplifications. I used around 200 ng of the 4C library for an individual PCR. Considering that I routinely use 10^7 of cells for the preparation of each library replica, which yields 20 to 50 μg of 4C DNA, a pool of 16 PCR amplifications represents a view of approximately $6.4 \times 10^5 - 1.6 \times 10^6$ cells. Since the genome is diploid, this is equal to approximately $1.28 \times 10^6 - 3.2 \times 10^6$ ligation events of the anchor HindIII fragment. Before proceeding with deep sequencing I controlled, by agarose gel electrophoresis, the efficiency of the individual PCRs (not shown) and an aliquot from the pool of 16 PCRs (Figure 36C).

From the pattern of the PCR products on the agarose gel (Figure 36C), we concluded that the primers we designed in the *Igh* locus efficiently amplified the 4C libraries prepared from *Rag2*^{-/-} pro-B cells, *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells and MEFs. We could detect the two prominent bands representing the most frequent junctions arising from incomplete digestion (Figure 36C, green) and self-ligation (Figure 36C, red) in all three 4C libraries after amplifying with the same primer. Other products formed a smear with a characteristic pattern for each primer pair (Figure 36C). The abundance of these products differs depending on the cell type and was, together with the identity of the HindIII fragments, analyzed by deep sequencing.

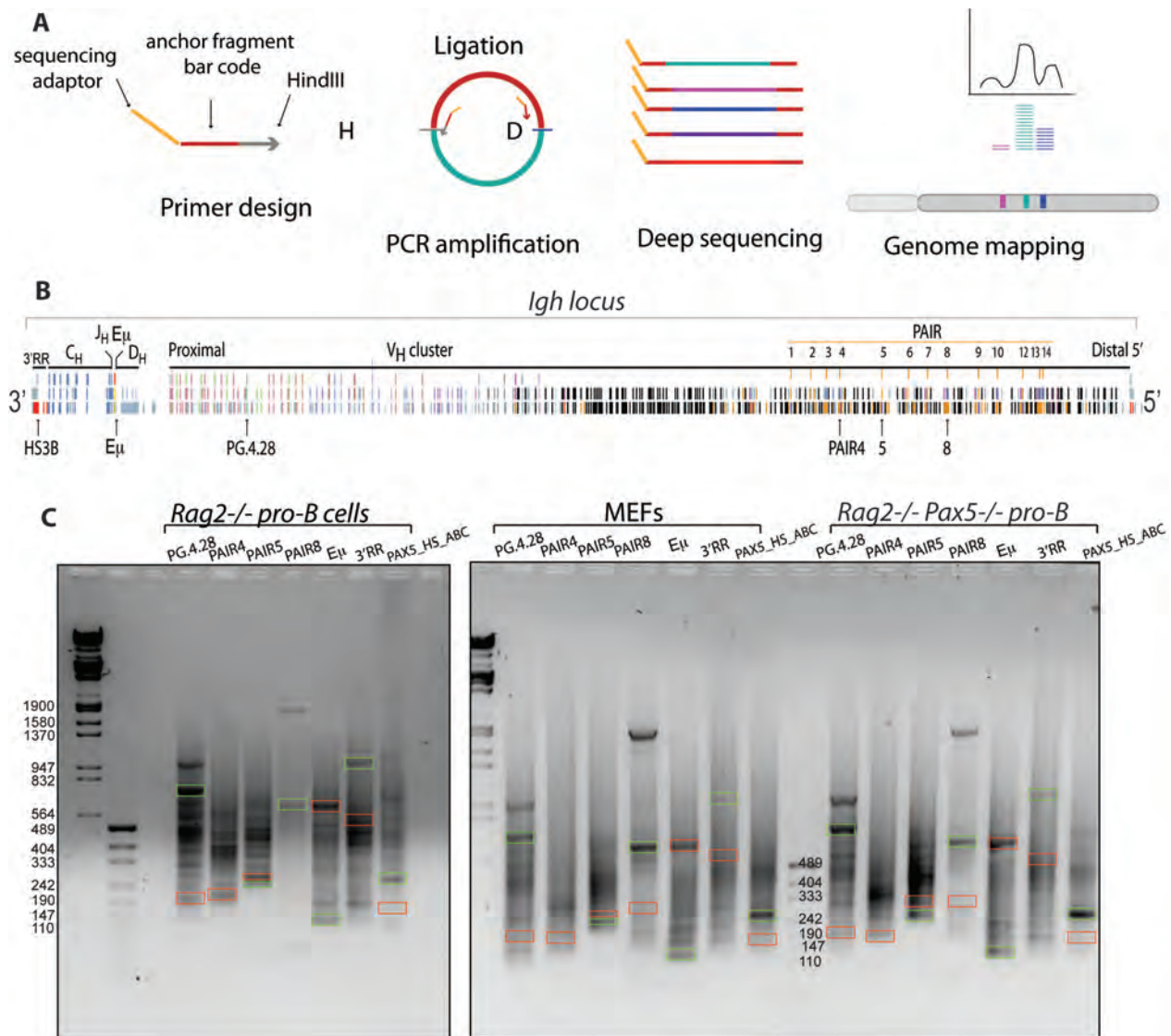


Figure 36. Primer design and agarose gel electrophoresis of 4C-PCR products, before deep sequencing.

(A) The scheme of a HindIII site-specific primer for deep sequencing, which consists of three types of sequences: deep sequencing adaptor (orange), allows direct sequencing after PCR amplification of the library; bar code sequence (red), specific for each anchor HindIII fragment, usually located in the element of interest; recognition sequence for HindIII (grey). Products of the second ligation are PCR-amplified with the HindIII (H) site-specific sequencing primer, and the primer designed close to, but not necessarily at the DpnII (D) site. The amplification products consist of the sequencing adaptor (orange) "bar code" sequence" (red) and the unknown "interacting" sequence (various colours). Deep sequenced reads are mapped to the mouse genome. The number of sequence reads mapped to a gene correlates with frequency it interacts with the anchor fragment.

(B) Schematic diagram of the *Igh* locus indicating the position of the primers used for 4C (indicated by arrows below the scheme) together with the location of the 3' DNase hypersensitive region, the E_H enhancer, C_H, J_H, D_H genes and the V_H gene segments. Red colour indicates DNase hypersensitive sites, orange bars indicate the location of PAIR elements within the V_H gene cluster. Other colours refer to different V_H gene families, of which the V_HJ558 (black) and V_H3609 (pink) are the most relevant for this study.

(C) 4C libraries prepared from *Rag2*^{-/-} or *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells and MEFs were PCR-amplified with primers designed in different regions of the *Igh* locus as depicted in (B). Sixteen PCRs of the same primer pair (indicated on the top of the gel) were pooled and analysed by agarose gel electrophoresis. The PCR amplification with the same primers results in a reproducible pattern of products on a 1.5 % agarose gel, independently of the cell type used for preparing the 4C library. The boxes indicate the two prominent overrepresented bands resulting from the self-ligation (red) and incomplete digestion (green) of the HindIII fragment used for the primer design. The size of the DNA marker is indicated on the left side (*Rag2*^{-/-}) or within the gel (MEF, *Rag2*^{-/-} *Pax5*^{-/-}).

Genome distribution and normalisation of deep-sequenced 4C reads in *Rag2*^{-/-} pro-B cells, *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells and MEFs

As already mentioned above, the primer anchored exactly in the HindIII site contains a “bar code” sequence specific for the respective restriction fragment and an extension that allows direct deep sequencing of the products (Figure 36A). Only a small aliquot of the pooled PCRs was utilised for single-end deep sequencing with a read length of 76 nucleotide. Those aliquots that were amplified with different primers, but from a library of the same cell type, were mixed in equal ratios and sequenced in the same lane of the flow cell in the Solexa Genome Analyzer. In such multiplexed samples, the primer’s “bar code” ensures specific identification of all sequences ligated to one unique anchor fragment. The sequences were mapped to the mouse genome resulting in approximately 12×10^6 , 9×10^6 and 13.7×10^6 reads for *Rag2*^{-/-} pro-B cells, *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells and MEFs, respectively (Figure 37A). Although we tried to mix products of all PCRs that corresponded to the same cell type in equal ratios, some of them were overrepresented in the multiplexed sample, based on the number of reads mapped to their anchor fragment (Figure 37B). In the multiplex sample of *Rag2*^{-/-} pro-B cells’ 4C library almost ½ of all reads mapped to the PAIR8 fragment and more than 1/3 to the PG.4.28 fragment. In the MEF sample, the same two primer-products were overrepresented, PAIR8 with slightly less than ½ and PG.4.28 with more than 1/3 of total reads. In the *Rag2*^{-/-} *Pax5*^{-/-}-pro-B cell 4C library, more than 2/3 of the total reads mapped to the PG.4.28.

The reads that were mapped to the HindIII fragments flanking the anchor fragment were subtracted from the total counts as they originated from the experimental artifacts incompletely digested and self-ligated fragments. Next, we compared where the majority of reads mapped in the genome of different cells. For this we mapped the reads to the *Igh* locus, to the chromosome 12 (excluding the *Igh* locus) or to the rest of the genome (excluding the whole chromosome 12). We plotted the percentage of reads mapped to these three genomic compartments into a stacked chart (Figure 38A). For PAIR4 and 3’RR anchor fragments, we visually inspected the distribution of reads throughout chromosome 12 in the genome browser view (Figure 38B). In the *Rag2*^{-/-} pro-B cells, the majority of reads mapped to the *Igh* locus (~70%), the same was true for the *Rag2*^{-/-} *Pax5*^{-/-} cells, although the percentage in the *Igh* locus dropped to ~58% at the expense of an increase in the rest of chromosome 12 and the genome. Strikingly, in the MEFs, more reads mapped outside (~40%), than within (~38%), the *Igh* locus (Figure 38A and B). These observations lead to several conclusions. Firstly, the primers that we designed specifically amplified fragments within the *Igh* locus, as seen for the *Rag2*^{-/-} pro-B cell library.

Secondly, the more a cell type differed from the *Rag2*^{-/-} pro-B cell phenotype, the more random were the interactions of the *Igh* locus reference fragments, which spread outside of the *Igh* locus throughout chromosome 12 (Figure 38B). Since the MEFs show a significantly different pattern in the distributions of reads (Figure 38), we supposed that the observed higher cross-linking frequencies were the consequence of inactive chromatin at the *Igh* locus, which resulted in chromatin compaction rather than specific interactions in non-lymphoid cells. Therefore, we excluded MEFs from further analysis.

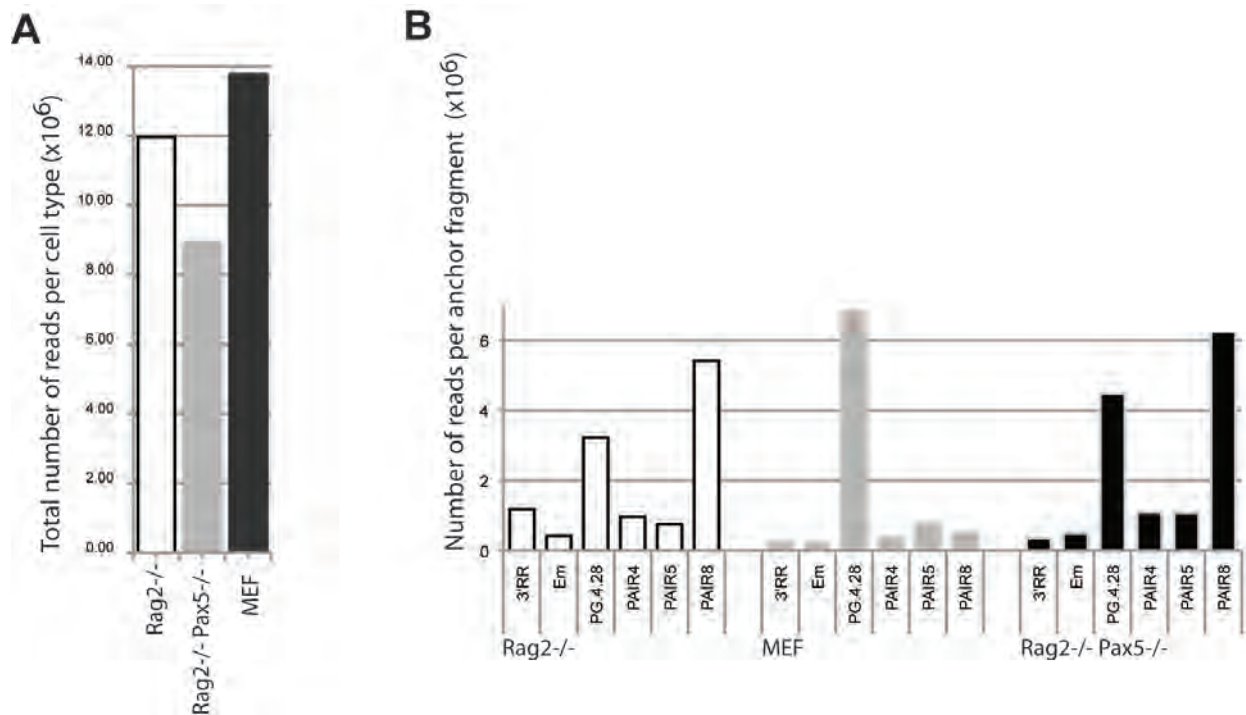


Figure 37. Numbers and primer distribution of deep sequencing reads mapped to HindIII fragments.

All primers contain a deep sequencing adaptor and a HindIII site. The sequence 5' to the HindIII site is specific for each anchor fragment, which allows multiplexing of all PCRs that were done separately on a library from the same cell type, in equal ratios for deep sequencing in one lane.

(A) Bar chart showing total numbers of reads (in millions) that could be uniquely aligned, obtained from multiplexed 4C-PCR samples, corresponding to *Rag2*^{-/-} pro-B cells (white), *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells (grey) and MEFs (black).

(B) Contributions of PCR products, mapped from different view points, to the total number of reads in *Rag2*^{-/-} pro-B cells (white), *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells (grey) and MEF (black) multiplexed 4C-PCR samples.

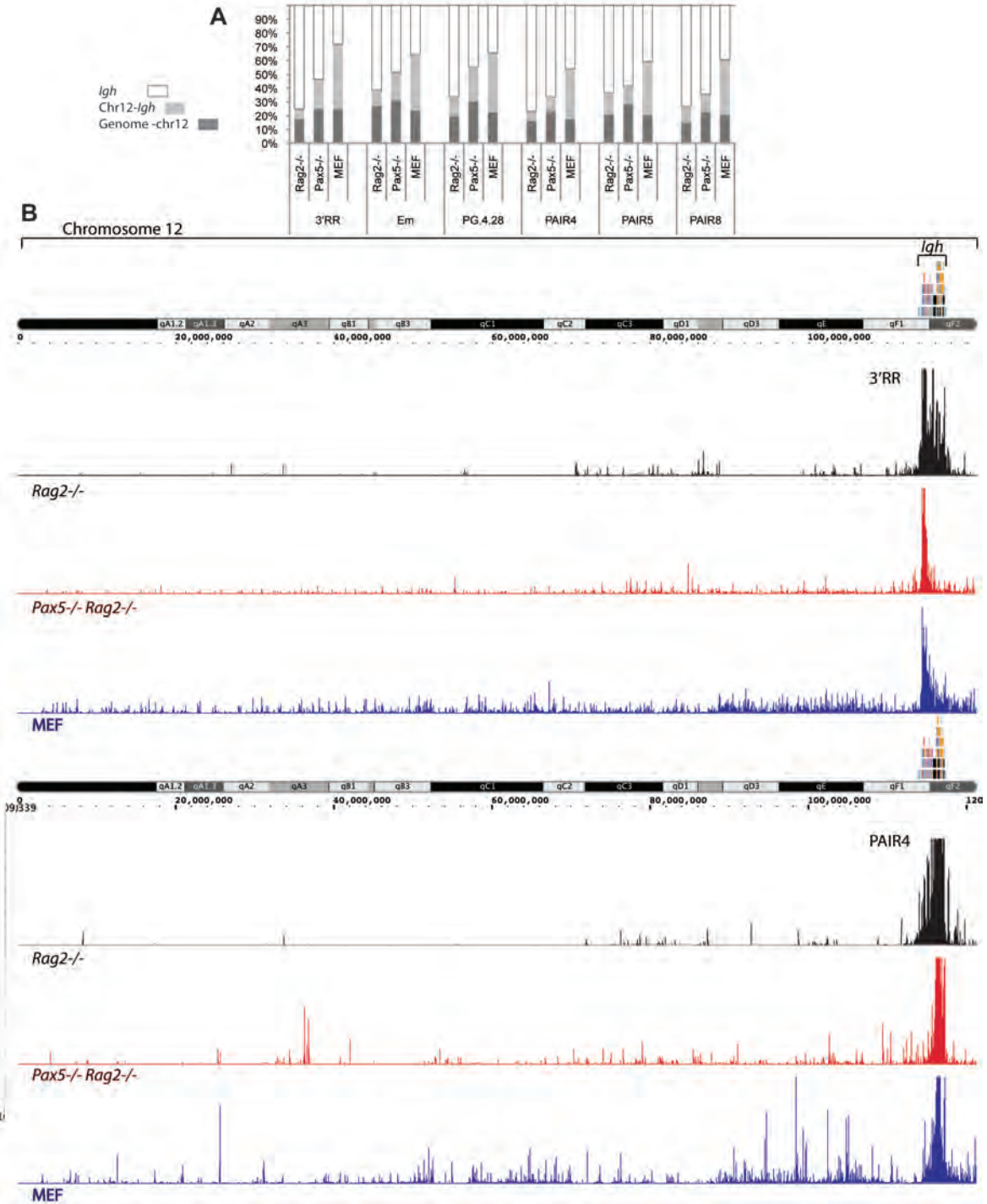


Figure 38. Genome distribution of deep sequencing reads mapped to HindIII fragments.

(A) Reads mapping to the self-ligation and incompletely digested products were excluded. Shown is the percent of sequence reads, mapped to the genome sequences excluding chromosome 12 (dark grey), chromosome 12 sequences excluding the *Igh* locus (light grey) or the *Igh* locus (white). The number of reads mapped to the *Igh* locus decreases in *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells, and even more so in MEFs, in comparison to *Rag2*^{-/-} pro-B cells.

(B) Distribution of reads mapped from 3'RR and PAIR4 anchor fragments across chromosome 12 in *Rag2*^{-/-} (black) or *Rag2*^{-/-} *Pax5*^{-/-} (red) pro-B cells and MEFs (blue). The *Igh* is situated on the chromosome 12 close to the telomere. The picture recapitulates the result shown in the chart (A). The majority of reads is mapped to the *Igh* locus. In *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells more reads are mapped outside of the *Igh* than in *Rag2*^{-/-} pro-B cells. In MEFs substantially more reads are mapped outside of the *Igh* locus.

The differences in the numbers of mapped reads using different primers for 4C sequencing (Figure 37B) demonstrated that direct comparison of experiments by raw read counts could lead to wrong conclusions. Therefore we divided reads mapped to the *Igh* locus (excluding the self-ligation (SL) and the incompletely digested fragments (ID) by the number of reads mapped to: (1) the entire genome, (2) the genome excluding chromosome 12, (3) the chromosome 12 excluding the *Igh* locus. We compared the values calculated by these normalizations, now referred to as relative cross-linking frequencies, for a HindIII fragment A that is equally present in both *Rag2*^{-/-} pro-B cells and *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells, and for a fragment B that is overrepresented only in *Rag2*^{-/-} pro-B cells (Figure 39; Table 16).

Fragment A				
Normalisation	Raw counts	Genome	Genome-Chr12	Chr12-Igh
<i>Rag2</i> ^{-/-}	11 100	2 200	15 000	19 000
<i>Rag2</i> ^{-/-} <i>Pax5</i> ^{-/-}	1 000	2 800	12 000	22 000

Fragment B				
Normalisation	Raw counts	Genome	Genome-Chr12	Chr12-Igh
<i>Rag2</i> ^{-/-}	25 000	5 000	34 000	40 000
<i>Rag2</i> ^{-/-} <i>Pax5</i> ^{-/-}	0	0	0	0

Table 16. Normalised interaction frequencies of the PAIR8 anchor fragment and HindIII fragments A and B. Fragment A is equally present in both *Rag2*^{-/-} pro-B cells and *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells, while fragment B is strongly overrepresented only in the *Rag2*^{-/-} pro-B cells.

In the case of fragment A, raw counts show that the crosslinking frequency is more than 10 times higher in the *Rag2*^{-/-} compared to the *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells. However, after normalisations, the cross-linking frequencies became more similar (Figure 39B, C and D; Table 16). Normalisations against the genome excluding the chromosome 12 (Figure 39C) and against the chromosome 12 excluding the *Igh* (Figure 39D) brought the crosslinking frequencies of fragment A to PAIR8 to the closest value, with only 1.25 and 1.15 fold difference, respectively, between the *Rag2*^{-/-} and *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells. On the contrary, the interaction frequency with fragment B, overrepresented in *Rag2*^{-/-} pro-B cells remained consistently higher in these cells than in the *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells even after normalisations (Figure 39C and D; Table 16).

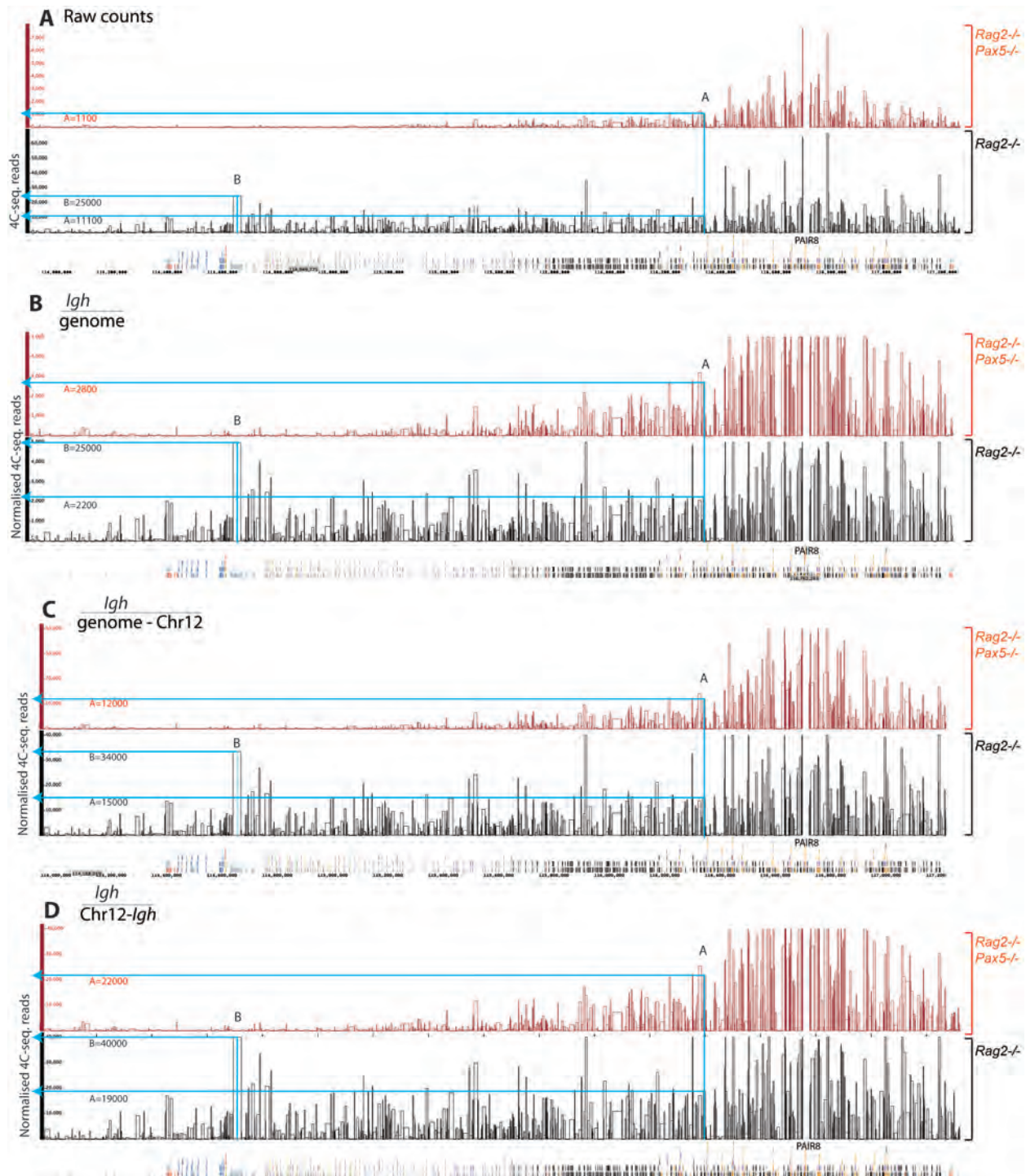


Figure 39. Comparison of different normalisation types between experiments.

Reads corresponding to the self-ligation and the incompletely digested products were excluded from the calculations. Formulas, indicated above the diagrams, were multiplied by 10^6 . Fluctuation in numbers of mapped reads (scale on the left side) in *Rag2*^{-/-} (black) or *Rag2*^{-/-} *Pax5*^{-/-} (red) pro-B cells after different normalisations. Total number of reads mapped within the *IgH* locus from the anchor fragment containing PAIR8 (A, raw counts) was normalised by dividing it by: (B) total number of reads in the genome, (C) total number of reads in the genome excluding chromosome 12, (D) total number of reads in the chromosome 12 excluding the *IgH* locus. Values obtained for fragment A, (present in both cell types) and fragment B (overrepresented in *Rag2*^{-/-} pro-B cells) are indicated. Blue lines connect the respective fragments to the y-axis of the diagram. Note that in (A) the scale is different for *Rag2*^{-/-} (max=70000) and *Rag2*^{-/-} *Pax5*^{-/-} (8000). The annotation of the *IgH* locus is described in Figure 36B.

This analysis showed that the differences between the numbers of mapped reads for one anchor fragment in different cell types can be normalised by dividing the reads mapped in the *Igh* locus by the reads mapped to the rest of the genome (excluding the chromosome 12), or by the reads of chromosome 12 (excluding the *Igh*). Normalizing against the chromosome 12 brought the crosslinking frequencies of a cell type-independent interacting fragment to the closest value in samples from different cells. Therefore, we will use this normalisation to measure the cross-linking frequencies of our anchor HindIII fragments to other fragments in the *Igh* locus. Since the *Igh* locus is subdivided in functional domains (Figure 36B), we assumed that some interactions might involve entire regions, rather than single HindIII fragments. Moreover, some fragments might be inaccessible to the restrictions enzyme because of occupancy by a transcription factor resulting in mapping of an interaction to the neighboring fragment. Because of these reasons, we calculated the running mean of the cross-linking frequencies of HindIII fragments within a 20-kb window across the *Igh* locus.

4C interactions of the 3' regulatory elements in the *Igh* locus

The best-characterised *cis*-regulatory elements in the *Igh* locus are the E μ enhancer and the 3' regulatory region. The E μ enhancer is important for D-J $_H$ rearrangement and essential for V $_H$ to D $_H$ -J $_H$ recombination (Perlot *et al.*, 2005). It seems to form the origin for the spreading of active chromatin marks that are important for locus accessibility (Chakraborty *et al.*, 2007). The 3'regulatory region consists of a series of hypersensitive sites HS3A, HS1,2, HS3B and HS4. These enhancers are important at later stages of B cell development for the maintenance of the *Igh* expression in mature B cells and plasma cells. Furthermore, combined deletion of HS3B and HS4 affected transcription and class switching to all isotypes except IgG1 (Ju *et al.*, 2007). Four additional DNase hypersensitive sites (HS4-HS7), where CTCF can bind, are located farther away at the 3' end (Garett *et al.*, 2004). HS5 and HS6 form an insulator separating the *Igh* locus as a chromatin domain (Garett *et al.*, 2004). In resting splenic B cells and a plasmacytoma cell line, all these hypersensitivity sites have been shown to physically interact with each other and with the recombined VDJ $_H$ intron (Ju *et al.*, 2007), although the E μ was dispensable for this interaction (Ju *et al.*, 2007).

The HindIII fragment that we used as an anchor to study the interactions of the 3'RR contained HS3B and HS4, which contains a binding site for Pax5 (Figure 41A, the part of the diagram where reads are missing indicated by the blue line). The anchor fragment with E μ additionally contains the J $_H$ 4 gene segment and can be bound by Pax5 (Figure 41B, the part of the diagram where reads are missing). In the diagrams of the running means of crosslinking

frequencies across the *Igh* locus, similar crosslinking frequencies were observed for *Rag2*^{-/-} and *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells within the V_H-D_H region and the proximal V_H genes (Figure 40A and B). Therefore, the interactions of the HS3B and Eμ within this region do not depend on the expression of Pax5. However, the crosslinking frequencies in the V_H region gradually decreased in *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells starting from the proximal 7183.7.10V_H gene to almost no interaction in the distal V_H region (Figure 40A and B).

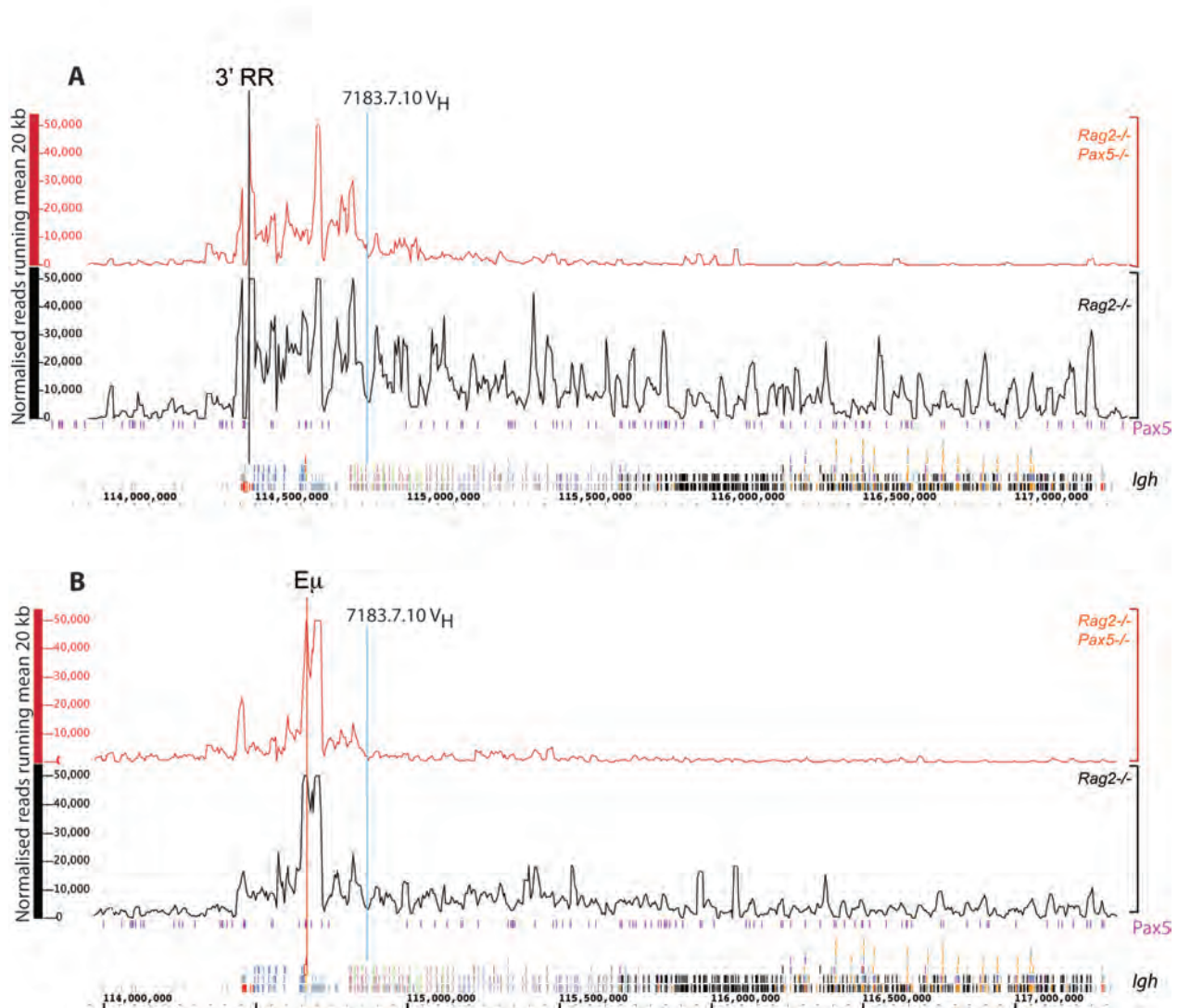


Figure 40. 4C-seq analysis in the *Igh* locus with the anchor HindIII fragments containing the 3'RR or Eμ.

The vertical line indicates the location of the anchor fragments. Reads mapped to HindIII fragments were normalised by dividing the number of reads mapped to the *Igh* locus with the total number of reads mapped to chromosome 12, excluding the *Igh* locus and the self-ligation products, and multiplied by 10⁶. The values of the running mean within a 20 kb window are plotted on the y axis of the diagram. They correlate with the relative cross-linking frequencies of the anchor fragments containing the 3' RR in (A) or Eμ in (B) and other HindIII fragments in the *Igh* locus of *Rag2*^{-/-} (black) or *Rag2*^{-/-} *Pax5*^{-/-} (red) pro-B cells. The violet lines indicate positions of Pax5-binding sites in the *Igh* locus (Ebert, unpublished). The vertical blue line indicates the position of the 7183.7.10 V_H gene. For annotation of the *Igh* locus see Figure 36B.

The E μ enhancer interacted with all hypersensitive sites in the 3'RR independently of Pax5 (Figure 41B). However in the *Rag2*^{-/-} pro-B cells, E μ is more frequently crosslinked to the fragment containing HS4, HS3B and a Pax5-binding site than to the other fragments. On the contrary, in *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells, E μ interacted more frequently with the CTCF-binding HS6/HS7, reported to have an insulator function (Figure 41B). The HS3B anchor fragment interacted with the fragment containing the E μ and with the E μ -neighbouring fragments, slightly more frequently in the *Rag2*^{-/-} pro-B cells. Both the E μ and 3' RR interacted with the C $_H$ gene segments (Figure 41A and B).

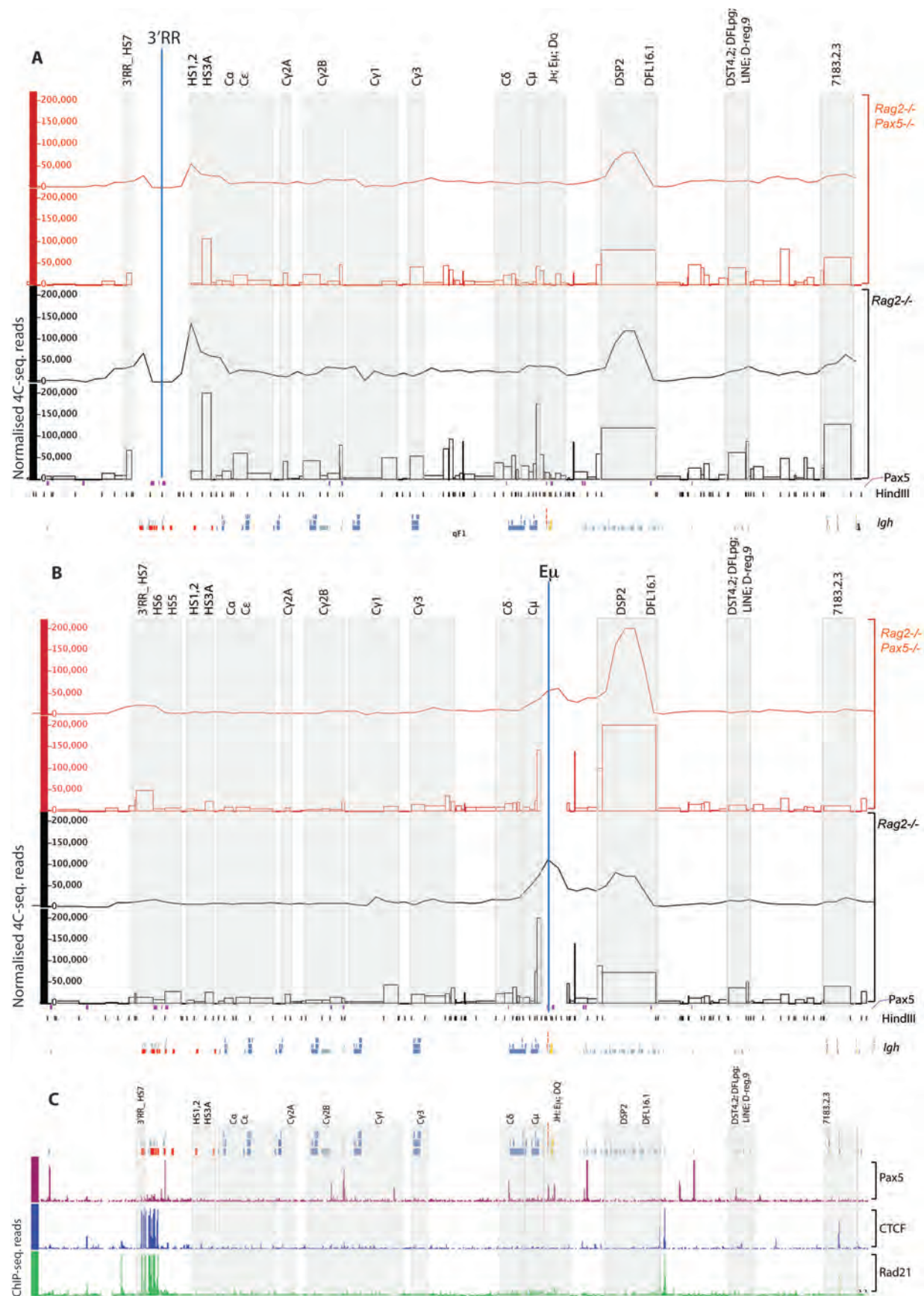
The D $_H$ gene segments displayed a very high crosslinking frequency to E μ and the 3'RR. Unfortunately, all the DSP genes, together with the DFL16.1, were located within one large HindIII fragment (Figure 41A and B). For this reason, we were not able to investigate their individual interactions. In the V $_H$ -D $_H$ region two fragments were most frequently crosslinked to HS3B and the E μ . One of them contains the DF16.1pg (Featherstone *et al.*, 2010), while both have binding sites for Pax5 (Figure 41C), although their cross-linking frequencies are similar in *Rag2*^{-/-} and *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells.

In the proximal V $_H$ region the crosslinking frequencies were 6 - 40 fold higher in *Rag2*^{-/-} pro-B cells than *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells. With increasing the genomic distance from the 3' anchor fragment, this trend became even more obvious especially after the 7183.7.10 gene (Figure 40, indicated by the blue line).

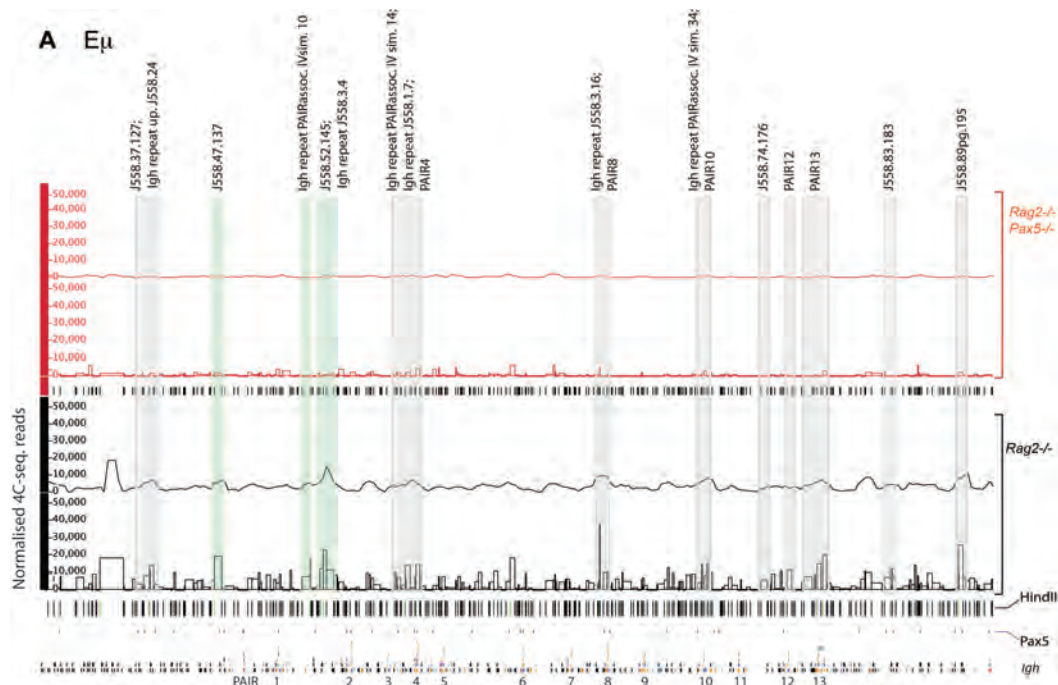
Figure 41. Transcription factor binding sites and relative crosslinking frequencies of the 3'RR and E μ to individual HindIII fragments in the 3' end of the *Igh* locus.

The 3'RR, C $_H$, J $_H$, D $_H$ gene segments and the D $_H$ -V $_H$ intergenic regions are magnified. Areas of the diagrams where reads are missing represent the anchor element (indicated by the blue line) and the neighbouring HindIII fragments that were removed for the analysis. The annotation lanes (labelled on the right side) from bottom to top indicate: *Igh* - the genes in the *Igh* locus; HindIII - HindIII fragments; Pax5 - violet bars indicate Pax5-binding sites (Ebert, unpublished). Bar diagrams show relative crosslinking frequencies of the 3'RR (A) or E μ (B) to individual HindIII fragments. Line diagrams indicate the running mean values of cross-linking frequencies to HindIII fragments within 20-kb windows. The cell type is colour coded, *Rag2*^{-/-} in black, *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells in red. Shaded rectangles highlight the regions of high crosslinking frequencies with annotation on the top.

(C) Binding sites of transcription factors in the regions with high crosslinking frequency to the anchor fragments highlighted by rectangles in (A) and (B): Pax5 (violet) [Anja Ebert, unpublished data], CTCF (blue) and Rad21 (green) [Ebert *et al.*, 2011]. Chromatin was precipitated from *Rag2*^{-/-} pro-B cells with streptavidin magnetic beads (Pax5-Bio) or anti-CTCF and anti-Rad21 antibodies followed by deep sequencing (performed by Anja Ebert).



Out of 432 fragments in the V_H region only 31 had a ratio of crosslinking frequencies between $Rag2^{-/-}$ and $Rag2^{-/-} Pax5^{-/-}$ pro-B cells below 7 and 55 fragments had a ratio below 10 (data not shown). For bioinformatics analysis of sequence similarity, we selected 151 fragments with the crosslinking frequency above 10000 and with a ratio of crosslinking frequencies between $Rag2^{-/-}$ and $Rag2^{-/-} Pax5^{-/-}$ pro-B cells above 6. Notably, for the 55 HindIII fragments shorter than 2000 bp, there were hardly any known annotations, and if known, the annotations were not strikingly similar. On the other hand, grouping of longer fragments to the same gene family was more likely to occur by chance. However, among the top 50 interacting fragments, 10 contained CTCF-binding sites and 2 fragments contained Pax5-binding sites. Furthermore, among the selected 151 HindIII fragments, 24 contained CTCF binding sites. Out of the 151 selected fragments, 13 consisted of repeats only and 55 consisted of more than 80% repeats. For sequence analysis we used the set of 115 HindIII fragments which had more than 100 bp of non-repetitive sequences. Among those, a group of sequences with strong similarity to each other consisted of: (1) PAIR elements 1, 4, 6, 8, 10, 13, 14, (2) PAIR-related elements, for example PAIRassocIVsim_32, *Igh* repeat_J558.3.24, and others. A second, bigger but less coherent group of similar sequences clustered most likely due to the contained V_H gene segments mainly of the J558 family (Figure 42B). In addition, some individual sequences appeared to form sequence-similar groups, which contained for example S107.2, SM7, V_H11 , J558. In summary, PAIR elements and the V_H genes defined the most obvious sequence similar clusters in the HindIII fragments of high crosslinking frequency with the 3' RR.



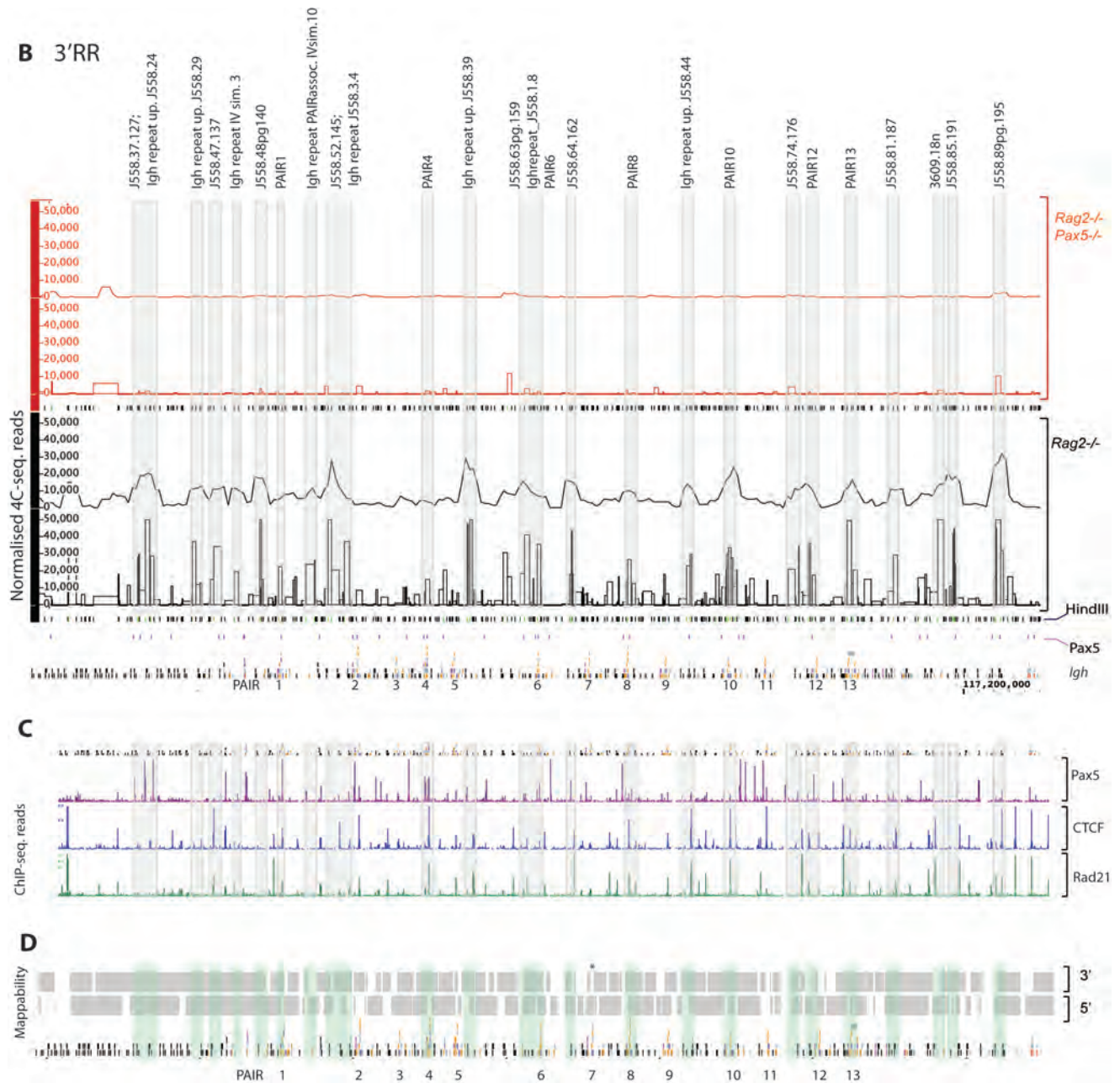


Figure 42. Relative cross-linking frequencies of the 3'RR and Eμ to HindIII fragments in distal V_H region.

The annotation lanes (labelled on the right side) from bottom to top indicate: *lgh* - the genes in the *lgh* locus; Pax5 - violet bars indicate Pax5-binding sites (Ebert, unpublished); HindIII - HindIII fragments. Bar diagrams show relative crosslinking frequencies of the Eμ (A) and 3'RR (B), of individual HindIII fragments in distal V_H region. Line diagrams indicate the running mean values of crosslinking frequencies to HindIII fragments within 20-kb windows. The cell type is colour coded, *Rag2*^{-/-} in black, *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells in red. Coloured rectangles highlight the regions of high crosslinking frequencies, annotated on the top.

(C) ChIP-sequencing result of binding sites of transcription factors Pax5 (violet) [Anja Ebert, unpublished], CTCF (blue) and Rad21 [Ebert *et al.*, 2011] in the distal V_H region. Chromatin was precipitated from *Rag2*^{-/-} pro-B cells with anti-CTCF and anti-Rad21 (green) antibodies and analysed by deep sequencing. The rectangles surround regions of high crosslinking frequency to the 3'RR.

(D) Mappability of the HindIII fragments at their 3' (upper row) or 5' end (lower row) in the distal V_H region. Gaps indicate sequences that could not be mapped by single-end Solexa sequencing with a read length of 76 nucleotides.

The recent analysis of Ebert *et al.* (2011) confirmed that PAIR elements 4, 6, 7 and 12 give rise to antisense transcripts in pro-B cells. Moreover, the core element of PAIR4 can function as a potent Pax5-dependent enhancer, while PAIR7 has promoter activity (Ebert *et al.*, 2011). Now, I have shown that 3'regulatory elements can interact with distantly located PAIR elements. This finding supports our assumption that PAIR elements could have a role in the regulation of *Igh* locus contraction. Surprisingly, I found that PAIR elements 8, 10 and 13 (the same mutation as in 8) can be equally engaged in these interactions although they could not efficiently compete for Pax5 binding in EMSA experiments. I propose that these elements can nonetheless be recruited to the interaction foci via CTCF, since Pax5 and CTCF can interact in pro-B cells. Another possibility is that the DNA conformation *in vivo* might be altered in association with the chromatin to allow binding of Pax5 despite the mutations in the recognition motif. I found no interaction with PAIR7 (Figure 42B) although it has no mutations in the binding motifs of the transcription factors CTCF, E2A or Pax5 (Figure 34A, sequence alignment), and was characterised as a functional PAIR element. We suppose that the reason could be the low mappability of this region (Figure 42D). Other frequently crosslinked fragments, the repeats upstream of J558 genes or the J558 genes themselves, often overlap with or are flanked by Pax5- or CTCF-binding sites (Figure 42C). Their crosslinking frequency is sometimes higher than for the mentioned PAIR elements. This could again be a consequence of better mappability of the respective regions than the PAIR elements (for example compare the interactions of J558.37.127 and PAIR4 (Figure 42B) to the mappability plot (Figure 42D)).

In summary, we showed that the regulatory elements at the 3'end of the *Igh* locus, the 3' DNase hypersensitive sites (3'RR) and the E μ enhancer, can physically interact with the C $_H$ and the D $_H$ regions of the *Igh* locus. The 3'RR may primarily form the focus of these interactions, thereby also recruiting the E μ enhancer. These findings are consistent with previous studies that showed the importance of long-range interactions between the 3'enhancers for the expression of the rearranged *Igh* gene and for class switch recombination. We assume that they are in a poised state already at an early stage of B cell development, and expect that they would interact more frequently in mature B cells and plasma cells upon stimulation. Most importantly, for the first time we confirmed that the long-range physical interactions of the 3' RR extend to the distal V $_H$ regulatory elements, possibly through the interaction of CTCF with Pax5. Cohesin would be important to stabilise the formed loops. As support for this assumption, we found that the 3' enhancers interact with the PAIR elements in the distal V $_H$ gene cluster, which contain Pax5-, CTCF- and Rad21-binding sites (Figure 42C).

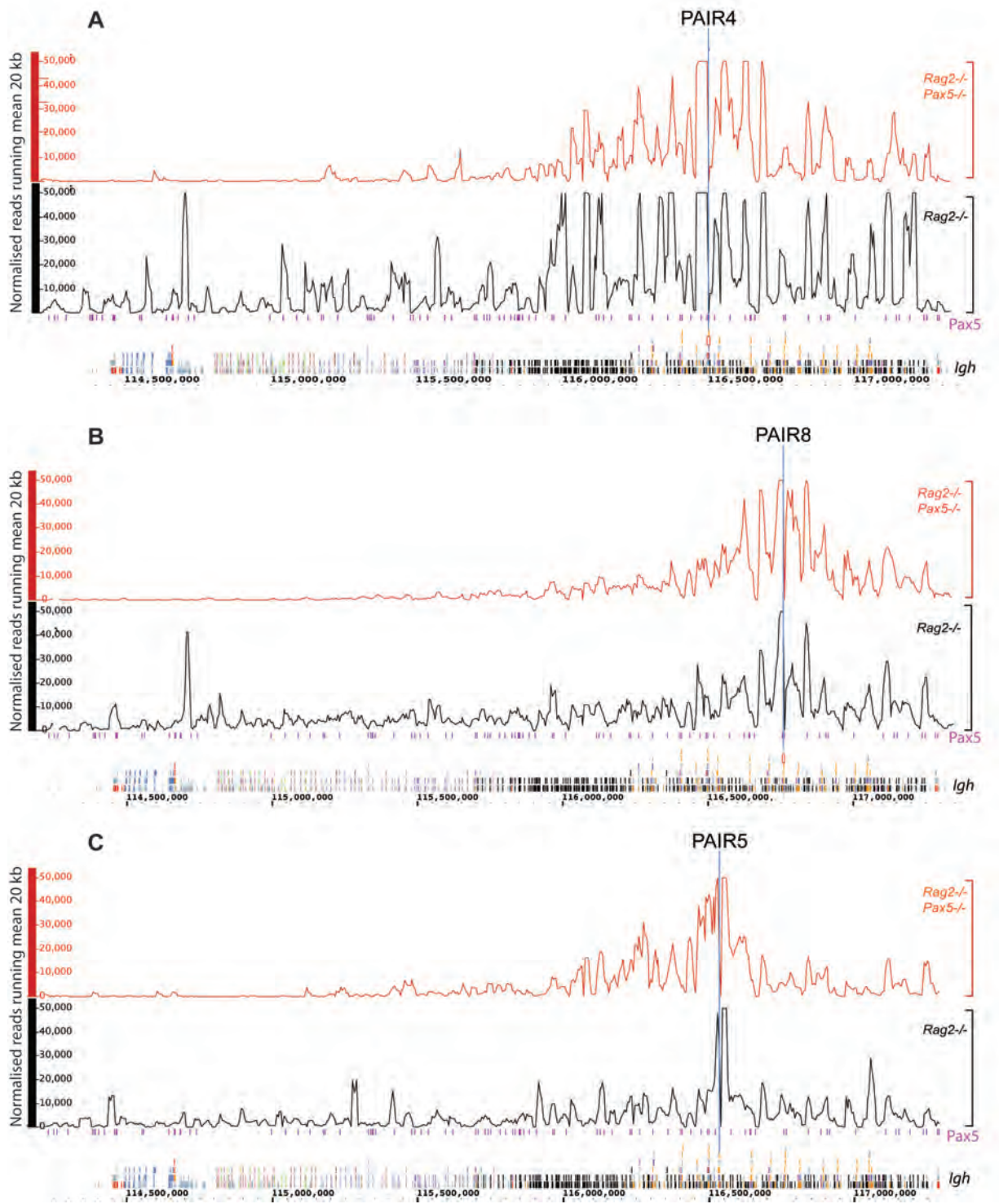
4C interactions of PAIR elements 4, 5 and 8

PAIR elements are present 14 times in the *Igh* locus mostly upstream of V_H3609 genes (Figure 36B). The core sequence contains binding motifs for transcription factors Pax5, E2A and CTCF. Since PAIR elements have binding sites for the transcription factors that are involved in V(D)J recombination, and CTCF and Rad21, we postulated that they are involved in long-range interactions with the regulatory elements located at the 3' end of the locus and in the V_H - D_H region. In order to investigate, their interactions we designed primers in HindIII fragments containing PAIR elements 4, 5 and 8. Unfortunately the repetitive nature of the V_H region prevented us from designing unique primers for other PAIR elements.

As shown by plotting the running mean values of the crosslinking frequencies through the entire locus, the PAIR elements 4, 5 and 8 frequently interacted with elements in the 5' end of the *Igh* locus, independently of the expression of Pax5 (Figure 43A, B and C). In the proximal V_H region the interactions were more frequent in *Rag2*^{-/-} pro-B cells compared to *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells. Finally, the running mean interaction frequencies across the entire locus indicated that PAIR elements interacted with the D_H gene segments, the E_μ enhancer and the 3'RR (Figure 43A, B and C). Notably the PAIR element 4 has overall higher crosslinking frequencies than PAIR8 and even more so than PAIR5. Furthermore, the interactions of PAIR4 are more frequent in *Rag2*^{-/-} pro-B cells than in *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells even in the distal V_H region, whereas there is hardly any difference in crosslinking frequencies between the two cell types for the other PAIR elements in this region.

In the distal V_H region all PAIR elements displayed a similar pattern of interactions, which occurred most frequently with other PAIR and PAIR-related elements.

At the 3' end of the *Igh* locus, the crosslinking frequencies were much higher in the *Rag2*^{-/-} pro-B cells than in *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells (Figure 43, Figure 44A). All three PAIR elements interacted with the 3'RR, the C_H region and with the sequences containing the E_μ enhancer and the J_H gene segments. The interactions with the fragments containing HS3B/HS4 and HS7 were more frequent than with the other hypersensitive sites in the 3'RR (Figure 44A, B and C). In the C_H region, PAIR4 weakly interacted with C_α , C_ϵ , $C_\gamma1$, C_δ and C_μ , while the interaction with $C_\gamma1$ occurred with the highest frequency (Figure 44C). PAIR elements 5 and 8 differed from PAIR4 in that they interacted more with $C_\gamma2c$ and $C_\gamma2b$ instead of $C_\gamma1$ (Figure 44A and B). The interaction with E_μ and the J_H segments occurred at a rather low frequency. PAIR elements 4 and 8 interacted with the DQ52 gene or its promoter (Figure 44A and C).



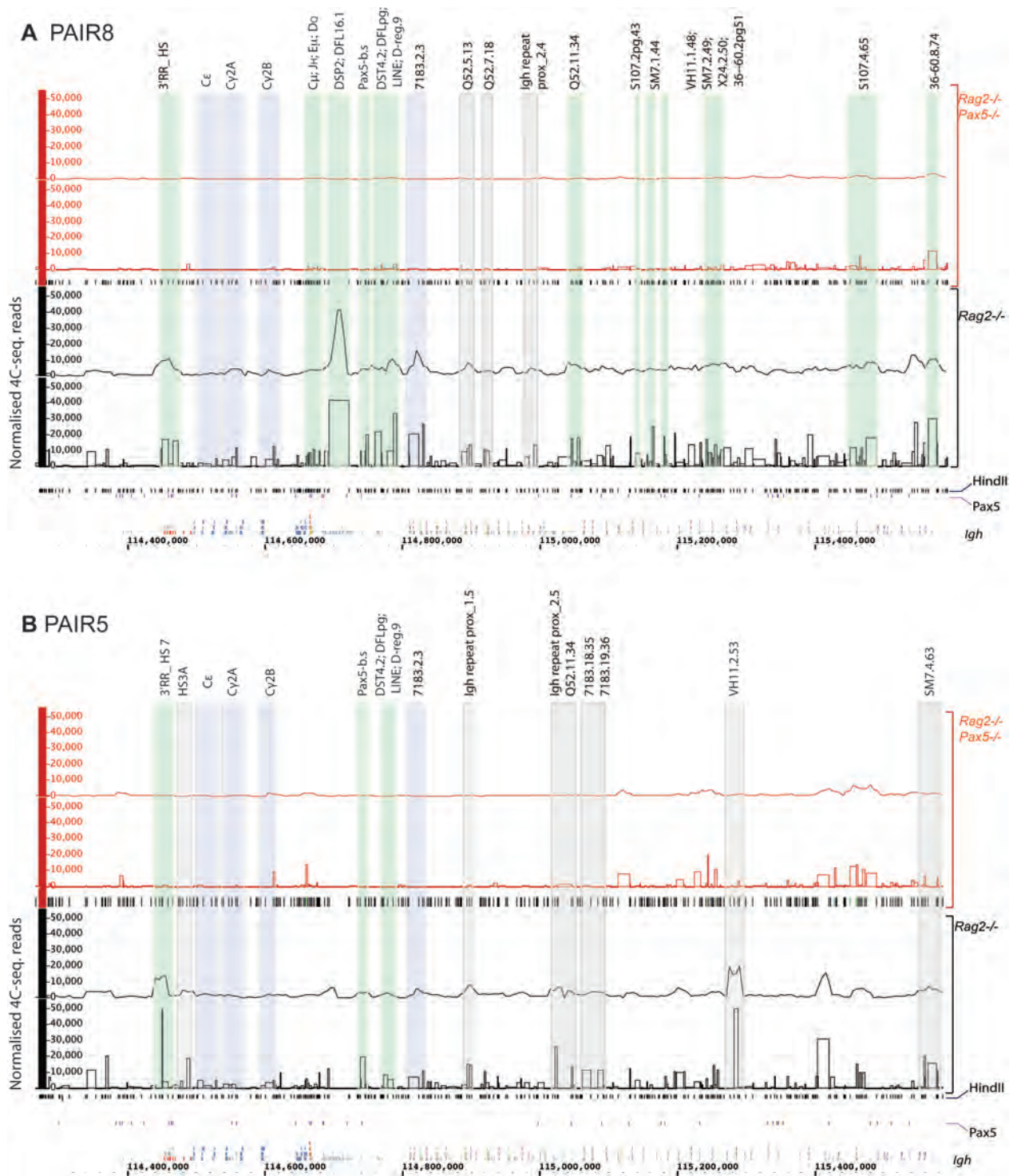
The most frequent interaction of PAIR elements 4 and 8 was with the DSP gene segments (Figure 44A and C). They consist of repeat sequences and are less frequently used for recombination. However, the most 5' DFL16.1 gene and the most 3' DQ52 gene are very frequently recombined (Chakraborty *et al.*, 2007). DFL16.1 is flanked by hypersensitive regions that contain CTCF-binding sites (Figure 44D). They are supposed to regulate germline transcription (Giallourakis *et al.*, 2010) and, since the two CTCF-binding sites are in the reverse orientation, they might have differential roles in insulating the DJ_H regions from the V_H region at the pre-pro-B cell stage and in juxtaposition of these two regions prior to V_H-DJ_H recombination in pro-B cells (Featherstone *et al.*, 2010). We would like to propose that PAIR elements contribute to this regulation. In the remaining V_H-D_H region, PAIR elements frequently crosslinked with a fragment containing a Pax5-binding site. Most frequently, PAIR8, but others also, interacted with the DFLpg and the surrounding repeats (Figure 44A). Lastly they all interacted with a 5' located region that contains a CTCF-binding site (LINE Dreg, Figure 44). In the proximal V_H region, some gene segments frequently interacted with all three PAIR elements, whereas others uniquely interacted with one of the PAIRs. Most of them are within HindIII fragments that contain binding sites for CTCF, Pax5 or both factors (Figure 44D).

To summarise, we showed that the PAIR elements could interact with each other, other repeats and gene segments in the distal V_H region of the *Igh* locus. Since the interactions in the V_H region occurred at similar frequency in *Rag2*^{-/-} and in the *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells, we suppose they are Pax5-independent and rely on CTCF and its dimerisation function. In contrast, the frequency of interactions decreased towards the 3' end. However, in *Rag2*^{-/-} pro-B cells, the long-range interactions were maintained in high frequency at the known regulatory elements and gene segments involved in V(D)J recombination and poised for later class switch recombination. Finally, we note that PAIR4, the only of the three analysed elements with conserved consensus motifs for all three transcription factors, interacted more frequently with elements in the proximal *Igh* domain than PAIR5 and PAIR8.

Figure 43. 4C-sequencing analysis in the *Igh* locus with the anchor HindIII fragments containing PAIR elements.

The vertical line indicates the location of the respective PAIR elements present in the anchor fragments.

Reads mapped to HindIII fragments were normalised by dividing the number of reads mapped to *Igh* locus with the total number of reads mapped to chromosome 12, excluding the *Igh* locus and the self-ligation products, and multiplied by 10⁶. The values of the running mean within a 20-kb window are plotted on the y-axis. They correlate with the relative cross-linking frequencies of the anchor fragments containing PAIR elements and other HindIII fragments in the *Igh* locus of *Rag2*^{-/-} (black) or *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells (red). The annotation lane with violet bars indicates positions of Pax5-binding sites in the *Igh* locus (Anja Ebert, unpublished data). (A) PAIR4 (B) PAIR8 (C) PAIR5.



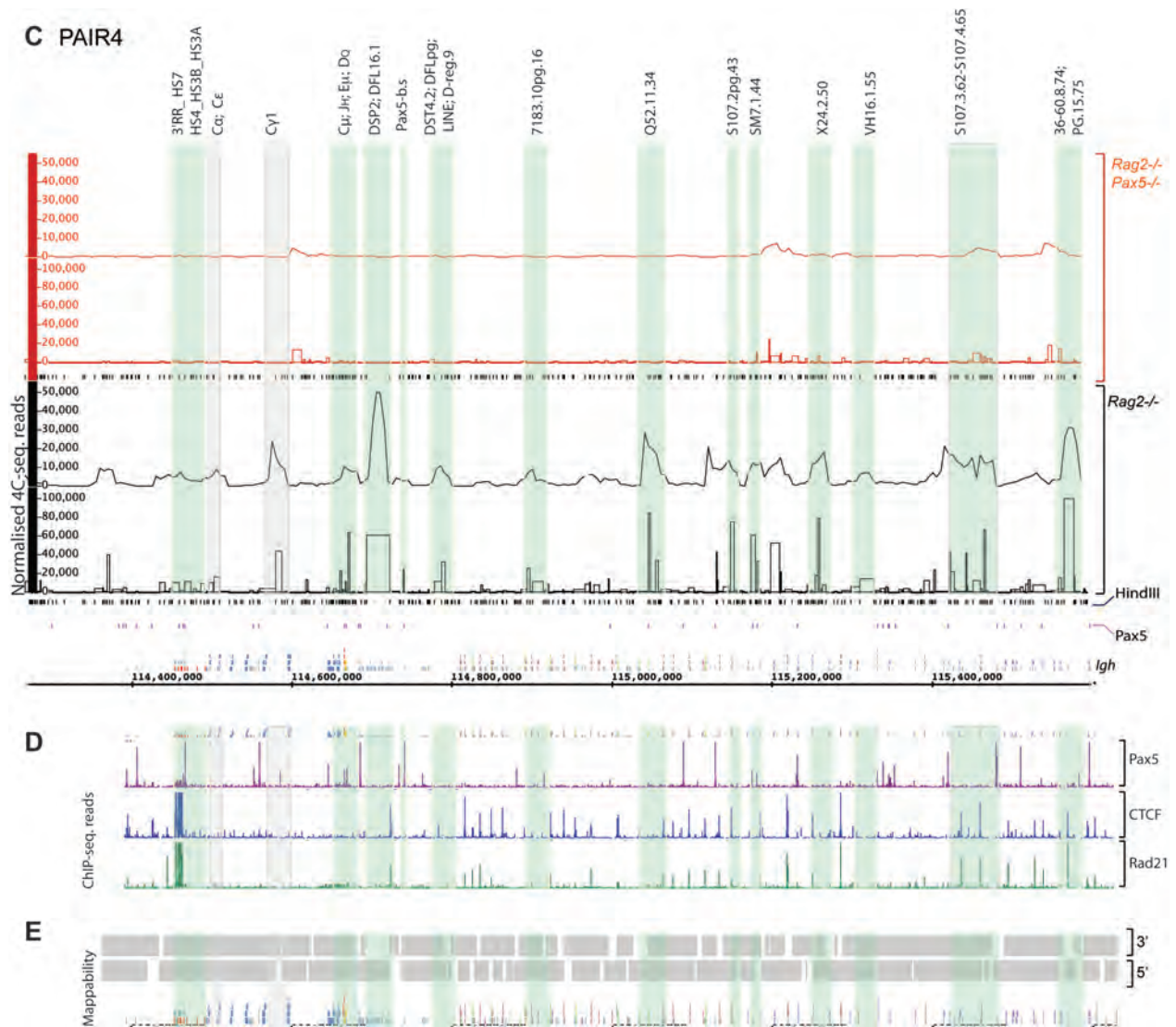


Figure 44. Relative crosslinking frequencies of PAIR elements in the 3' region.

The magnified region contains the C_H , J_H , D_H gene segments, D_H - V_H intergenic region and proximal V_H genes. Annotation lanes (labelled on the right side) from bottom to top indicate: *Igh* - the genes in the *Igh* locus; Pax5 - violet bars indicate Pax5-binding sites (Ebert, unpublished), HindIII - HindIII fragments. Bar diagrams show relative crosslinking frequencies of PAIR8 (A), PAIR5 (B) and PAIR4 (C) to individual HindIII fragments. Note that the scale (y-axis) differs for PAIR4 (max 100 000) compared to PAIR8 and PAIR5 (max 50 000). Line diagrams indicate running mean PAIR-crosslinking frequencies to HindIII fragments within 20-kb windows. The cell type is colour coded, *Rag2*^{-/-} pro-B cells in black, *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells in red. Coloured rectangles highlight the regions of high crosslinking frequencies, annotated on the top: green - regions frequently crosslinked to both PAIR4 and PAIR8; grey - regions frequently crosslinked to only one of the PAIR elements; blue - regions frequently crosslinked to both PAIR8 and PAIR5.

(D) Binding sites of transcription factors in the region containing the C_H , J_H , D_H gene segments, D_H - V_H intergenic region and proximal V_H genes: violet - Pax5 (Anja Ebert, unpublished data); blue - CTCF; green - Rad21 (Ebert *et al.*, 2011). Chromatin was precipitated from *Rag2*^{-/-} pro-B cells with anti-CTCF and anti-Rad21 antibodies and analysed by deep sequencing. The rectangles surround regions of high crosslinking frequency to the PAIR4 (as shown in (A)).

(E) Mappability of the HindIII fragments from their 3' (upper row) or 5' end (lower row). The rectangles surround regions of high crosslinking frequency to the PAIR4. Gaps indicate sequences that could not be mapped by single-end deep sequencing with a read length of 76 nucleotides.

Interactions of the pseudogene PG.4.28 from the proximal V_H region

In addition to known regulatory elements, we analysed the frequencies and the pattern of interactions of a gene that, according to our assumption, had no regulatory role in the process of V(D)J recombination. The pseudo V_H gene segment PG.4.28 is located in the proximal V_H region, near a CTCF-binding site, flanked by V_HQ52.9.29 gene segment at the 3' side and V_H7183.16.27 gene segment at the 3'side.

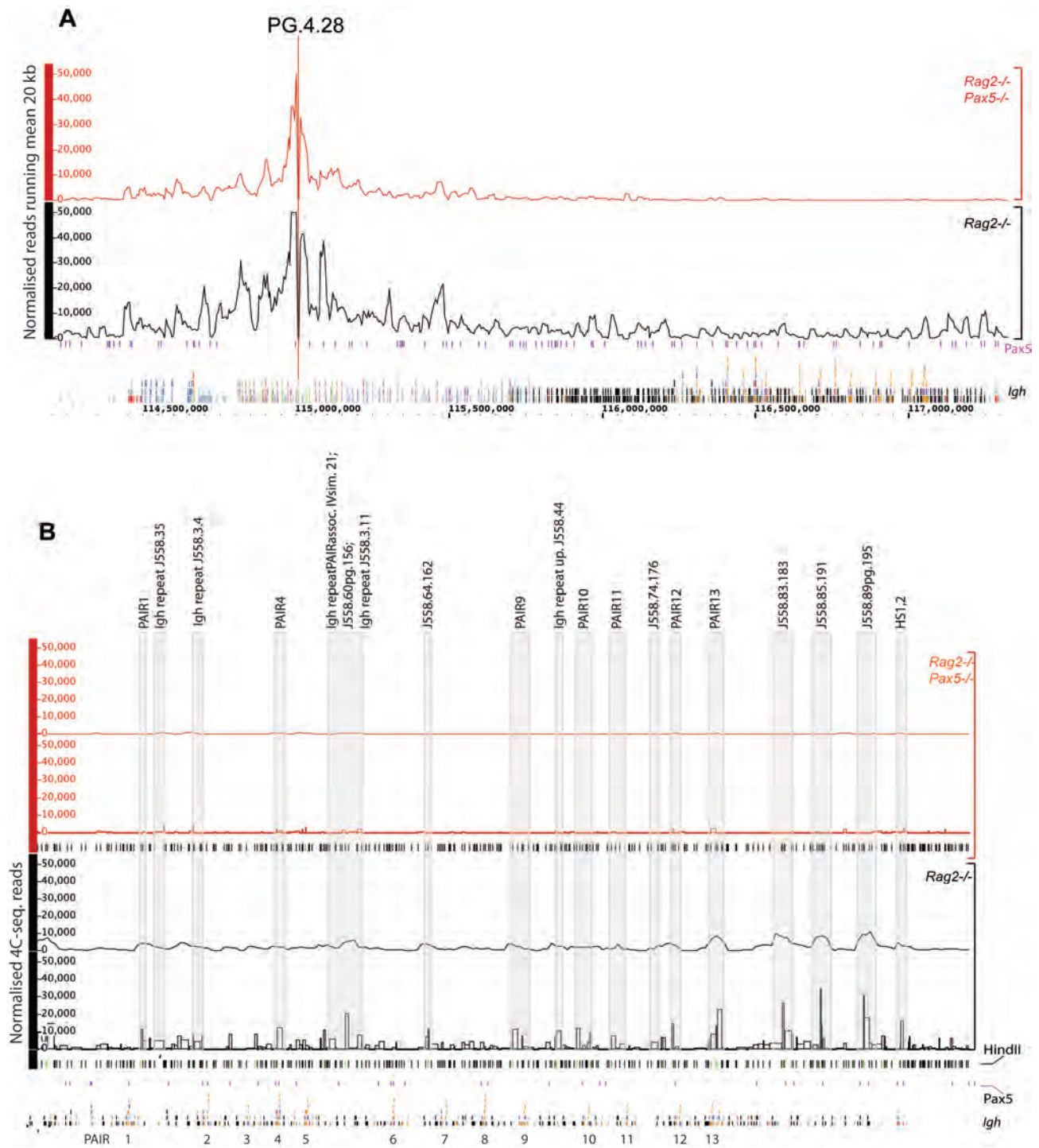
The running mean plot of crosslinking frequencies showed that the interactions decreased with increasing genomic distance and that they are almost absent in the distal V_H region of *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells (Figure 45A). At the resolution of individual HindIII fragments, the pattern of interactions in the distal V_H region was very similar to that observed for the E_μ enhancer and the 3'RR. However, their frequencies were substantially lower (Figure 45B, 46B). In the 3' end of the locus, interactions with the 3'RR, C_H genes, E_μ, J_H and D_H region were the most frequent events. Compared to PAIR4, the interaction of PG.4.28 with the 3'RR was of similar frequency, however PAIR4 showed a stronger interaction with the C_γ1 gene, the region containing the E_μ enhancer, and J_H and D_Q gene segments as well as the DSP gene region (Figure 46A). Furthermore, the interactions of PG.4.28 occurred as frequently in this proximal region in *Rag2*^{-/-} pro-B cells and *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells in contrast to PAIR4, which exhibited significantly enriched interactions only in *Rag2*^{-/-} pro-B cells. Moreover, the location of PAIR4 is more distant from this proximal region than of PG.4.28, which reduces the probability that the interactions observed could be a consequence of random collisions.

Figure 45. 4C-sequencing analysis of the *Igh* locus with the anchor HindIII fragments containing the PG.4.28.

The line transects the location of the PG.4.28.

(A) Reads mapped to HindIII fragments were normalised by dividing the number of reads mapped to *Igh* locus with the total number of reads mapped to chromosome 12, excluding the *Igh* locus and the self-ligation products, and multiplied by 10⁶. The values of the running mean within a 20-kb window are plotted on the y-axis. They correlate with the relative cross-linking frequencies of the anchor fragments containing PG.4.28 with other HindIII fragments in the *Igh* locus of *Rag2*^{-/-} (black) or *Rag2*^{-/-} *Pax5*^{-/-} (red) pro-B cells. The annotation lane with violet bars indicates the positions of Pax5-binding sites in the *Igh* locus (Anja Ebert, unpublished data).

(B) Relative cross-linking frequencies of the PG.4.28 gene segment to individual HindIII fragments in the distal V_H region. The annotation lanes (labelled on the right side) from bottom to top indicate: *Igh* - the genes in the *Igh* locus; Pax5 - violet bars indicate Pax5-binding sites (Anja Ebert, unpublished data); HindIII - HindIII fragments. Column diagrams show relative crosslinking frequencies of PG.4.28 to individual HindIII fragments in distal V_H region. Line diagrams indicate the running mean values of crosslinking frequencies to HindIII fragments within 20-kb windows. The cell type is colour coded, *Rag2*^{-/-} pro-B cells in black, *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells in red. Coloured rectangles highlight the regions of high crosslinking frequencies, annotated on the top. They overlap with regions that also interacted with the 3'RR.



In conclusion, other elements of the *Igh* locus like PG.4.28 may form long-range interactions in the presence of Pax5 transcription factor. These interactions of regulatory elements located in proximal (3'RR) and distal domains of the *Igh* locus (PAIR) occur more frequently and independently of genomic distances.

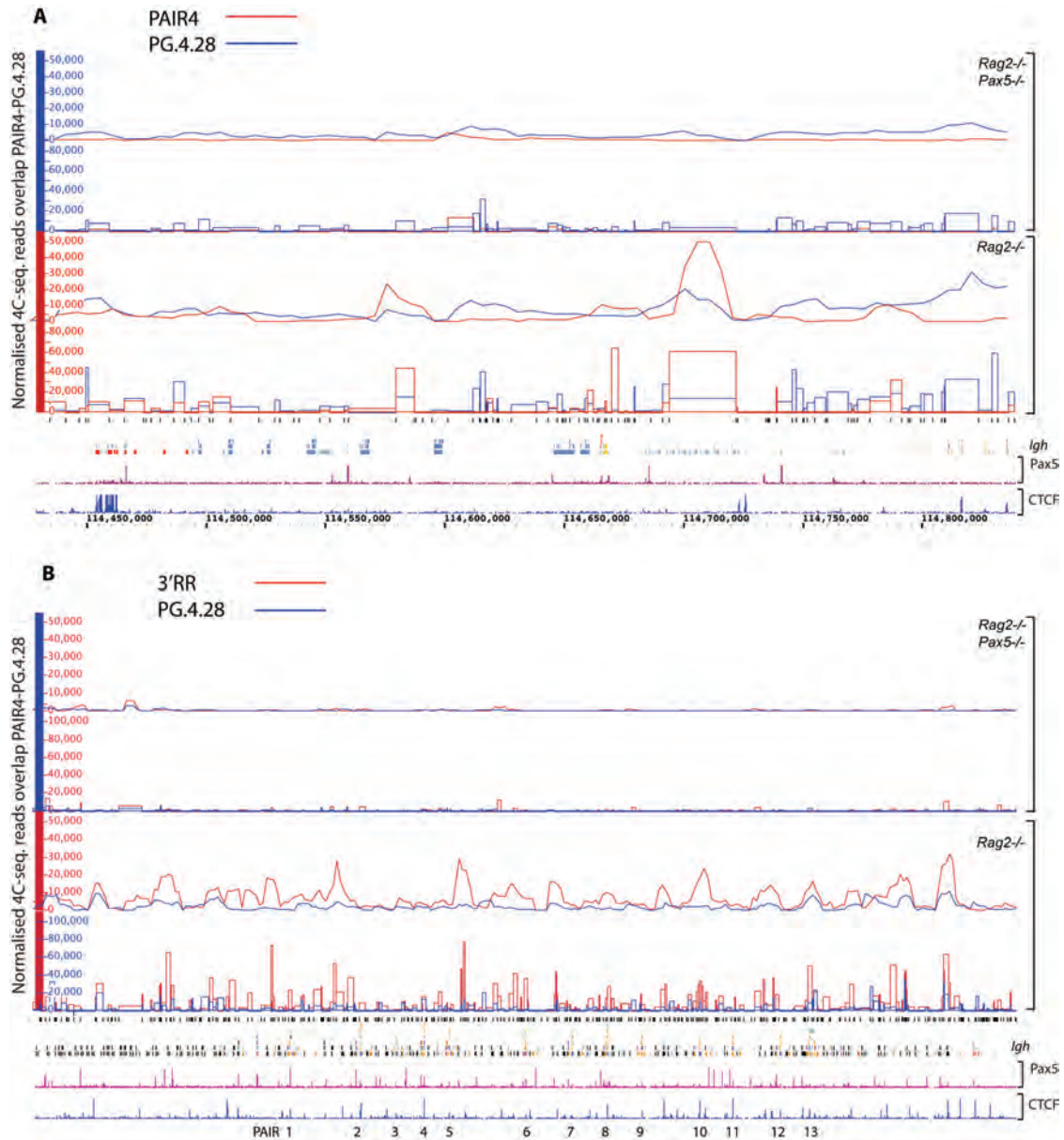


Figure 46. Relative crosslinking frequencies of PG.4.28 compared to PAIR4 or 3' RR

The 20-kb running mean, or normalised numbers of reads for individual HindIII fragments mapped in blue from PG.4.28 and in red from PAIR4 (A) or 3' RR (B), were compared by overlapping the diagrams in the genome browser tool. Indicated below the diagrams are: HindIII fragments; scheme of the *Igh* locus; transcription factor-binding sites for CTCF in blue (Ebert *et al.*, 2011) and Pax5 in violet (Anja Ebert, unpublished).

(A) Relative crosslinking frequencies of PG.4.28 compared to PAIR4 in the 3' domain of the *Igh* locus. The region in view contains the C_H, J_H, D_H gene segments, D_H-V_H intergenic region and proximal V_H gene segments.

(B) Relative crosslinking frequencies of PG.4.28 compared to 3'RR in the distal 5' V_H region of the *Igh* locus. The locations of PAIR elements are indicated below the ChIP-sequencing lane.

DISCUSSION I

For posttranslational modification analysis of Pax5, we needed large quantities of the purified protein. Therefore we transformed pro-B cells with Abelson murine leukemia virus (Ab-MuLV) (Rosenberg *et al.*, 1975). For the interpretation of the results obtained with these cells it is important to briefly characterise their features. Ab-MuLV is a replication-defective retrovirus encoding a fusion between Gag and a N-terminally truncated form of the nonreceptor protein tyrosine kinase c-Abl. The resulting v-Abl has constitutive tyrosine kinase activity. The transformed B cells become independent of IL-7 receptor signalling, proliferate in culture and fail to mature to the pre-B cell stage (Muljo and Schlissel, 2003). They have downregulated Rag1 and Rag2, therefore do not undergo *Igk* rearrangement (Muljo and Schlissel, 2003). Furthermore, v-Abl interferes with pre-BCR signalling by downregulating the crucial components of this pathway *Syk*, *Igα*, *Igβ* and *BLNK* (Muljo and Schlissel, 2003). The RAS-MEK-ERK signalling pathway is fully active in Ab-MuLV transformed pro-B cells (Kuo *et al.*, 2011). The Ras proteins are a family of GTPases which relay signals from the tyrosine receptor kinases to elsewhere in the cell. They are often required for transmitting signals inducing cell proliferation or differentiation. Activated Ras initiates a cascade leading to mitogen-activated protein (MAP) kinase phosphorylation that can alter the pattern of gene expression. Most importantly it activates genes required for cell proliferation.

Our analysis of posttranslational modifications in these cells showed that Pax5 is constitutively phosphorylated on four serine residues (S189, S206, S283, S344) that are located within recognition motifs for members of the MAPK family. Moreover, these motifs are supposed to engage in interactions with proteins containing class IV WW domains (for example in Nedd4 ubiquitin ligase) that could provide a platform for the assembly of multiprotein networks (Ingham *et al.*, 2005). WW domains are small protein modules composed of approximately 40 amino acids (Macias *et al.*, 2001), which contain elements of SH3 and SH2 domains by recognising proline-rich ligands and being regulated by phosphorylation (Macias *et al.*, 2001). The name WW refers to two signature tryptophan (W) residues (Macias *et al.*, 2001). For the interaction to occur, the ligands of Class IV WW domain require phosphorylation of a serine and threonine in the binding motif (Ilsley *et al.*, 2001). In agreement with identifying phosphorylation in motifs for MAPK, the mass spectrometry analysis of putative Pax5-interacting proteins identified mitogen-activated protein kinase 1 (Mapk1) and several other components of the Ras-MAPK signalling cascade (Table 7) including RAS protein activator 3 like (Rasa13), Drg1/2 (Developmentally-regulated GTP-binding protein 1), Gnb2211 (Guanine nucleotide binding protein [G protein]). At

present, we do not know whether the mapped phosphorylation pattern is also observed in *ex vivo* pro-B cells that survive in response to IL-7R signalling in contrast to transformed Ab-MuLV pro-B cell line. We assume this to be the case, since MAPK signalling is also active in proliferating pro-B cells, which resemble Ab-MuLV pro B cells. We suppose that a change in the PTM pattern might occur at the transition to the pre-B cell stage. Pre-BCR signalling induces cell proliferation, but at the same time activates a negative feedback loop that acts through BLNK and stimulates differentiation and cell growth (Herzog *et al.*, 2009). It is possible that the function of Pax5 can be modulated through PTMs at different developmental stages to control different processes in B cells such as V(D)J recombination, pre-BCR signalling or migration to the periphery. Treatment of Ab-MuLV pro-B cells with STI571 (Gleevec) results in G₁ cell cycle arrest and developmental progression to a pre-B cell like stage (Muljo and Schlissel, 2003). Genes encoding the components of the pre-BCR are reactivated as well as Rag1 and Rag2 enabling the cells to proceed with *Igk* rearrangement (Muljo and Schlissel, 2003). Some of these genes are targets of Pax5. Therefore, it is possible that some of the modifications that we detected in Ab-MuLV pro-B cells have an inhibitory effect on Pax5, which is relieved after treatment with STI571. This would be analogous to Foxo phosphorylation and subsequent sequestration to the cytoplasm during pre-BCR signalling in order to prevent the activation of *Rag* genes (Herzog *et al.*, 2009). Although we suppose that the effect on Pax5 might be less drastic, since it significantly contributes to the pro-B cell transcriptional program, it would be interesting to investigate the posttranslational modifications of Pax5 after treating the MuLV pro-B cells with STI571. The only disadvantage is that the cells undergo apoptosis soon after STI571 treatment since v-Abl signalling activates anti-apoptotic genes (Muljo and Schlissel, 2003). However, it should be possible to expand the cells prior to short-term treatment with STI571 and mass spectrometry analysis.

Finally, the most drastic change in B cell development occurs upon stimulation with an antigen, which induces terminal differentiation of mature B cells into immunoglobulin-secreting plasma cells. During this process, the Pax5-dependent gene expression program is lost at the expense of a plasma cell-specific program controlled by the transcription factor Blimp1 (Lin *et al.*, 2002). At the transient pre-plasmablast stage, the transcriptional function of Pax5 is inactivated prior to active repression of the *Pax5* gene (Kallies *et al.*, 2007). Two mechanisms to explain this posttranslational inactivation of Pax5 were proposed, by: (1) an inactivating posttranslational modification or, (2) interaction with an unknown protein (Kallies *et al.*, 2007). Since PTMs often act by creating novel interaction surfaces, both of the proposed changes may act in concert. To investigate the PTMs of Pax5 at this stage, we stimulated the mature B cells

with LPS *in vitro*. Unfortunately, we were not able to obtain enough cells for mass spectrometry analysis because the pre-plasmablast is a transient stage whereas the LPS-stimulation is accompanied by high frequency of cell death. In the future, we will raise antibodies against the MAP-kinase-target residues (S189, S206, S283, S344) to investigate their role at different stages of B cell development and how they modulate the activity and protein-interaction potential of Pax5. We also detected phospho-residues that seemed to occur at a lower frequency downstream a different signalling pathway (Y179, Y388, S171, T172, S174). They might have a faster turnover rate and act as switches that can be activated in specific but more rare challenges.

The mass spectrometry analysis also identified two methylated arginines (R359 and R377) in the inhibitory domain. These modifications might explain the molecular mechanism by which the inhibitory domain modulates the potential of the neighbouring transactivation domain to activate transcription. Previous studies have shown that the C-terminal regulatory region of Pax5, fused to the paired domain, is on its own sufficient to activate transcription (Dörfler and Busslinger, 1996). An increase in transcriptional activation was observed if the inhibitory domain was deleted (Dörfler and Busslinger, 1996). The C/EBP β harbours a similar regulatory module consisting of juxtaposed transactivation and inhibitory domains. The arginines in the inhibitory domain of C/EBP β were substituted with alanines (R3A) or leucines (R3L), which with their additional methyl group mimic the increased hydrophobicity after methylation (Kowenz-Leutz *et al.*, 2010). The R3A mutant successfully interacted with the BAF (Brm1) and mediator (Med23) complex subunits, whereas the interaction of the R3L mutant with these complexes was strongly decreased (Kowenz-Leutz *et al.*, 2010). The R3A substitution significantly increased the activation potential of C/EBP β (Kowenz-Leutz *et al.*, 2010). Additionally, MAP-kinase phosphorylation of the inhibitory domain impaired the interaction with the methyltransferase PRMT4/CARM1 (Kowenz-Leutz *et al.*, 2010). Based on this study a model proposes that C/EBP β is primarily inactive or displays repressor functions. MAP-kinase signalling transiently abrogates C/EBP β methylation and permits co-activator recruitment and gene activating functions (Kowenz-Leutz *et al.*, 2010). Thus the functional crosstalk between Ras/MAPkinase signalling and arginine methylation allows fast switches in gene activation important for differentiation.

In the line with this study, the results of our mass spectrometry analysis imply that a similar mechanism might be involved in regulating the dual role of Pax5 in gene expression. The methylation of R359 and R377 in the inhibitory domain of Pax5 might interfere with recruitment of BAF-, MLL- or TFIID complexes to Pax5 target genes. Moreover, the mass spectrometry

analysis identified a putative interaction with a component of the PRMT5-arginine methyltransferase complex, Wdr77 (Table 3). To test the role of arginine methylation, we will generate R359A and R377A single and double mutants of the Pax5-ER fusion protein and compare their ability to recruit BAF, MLL or TFIID complexes to Pax5 target genes. We will also perform co-immunoprecipitation assays.

By *in vitro* and co-transfection assays, I showed that Pax5 could be sumoylated as predicted by the finding of SUMO1-consensus motif (IKPE; Figure 20C) in the sequence of Pax5. However, I realised that K257 is not the only SUMO1 acceptor lysine in the Pax5 protein. I predict that two other lysine residues can be modified by SUMO1, since I detected two modified bands, probably corresponding to mono- and di-SUMO1-Pax5 protein. Since SUMO1 cannot form chains, two different lysines must act as acceptors. By using Pax5 mutant proteins, I determined that the SUMO1 modification occurs in the region between the paired domain and the transactivation domain. Apart from K257, there are four other lysine residues in this region. I analysed their potential to act as SUMO1 acceptors by using SUMOsp 2.0 software (Ren *et al.*, 2009; <http://sumosp.biocuckoo.org/>). The motifs that resulted from this search are highlighted in dashed boxes in Figure 20C. The DDMKANL motif scored even higher than the SUMO1 consensus motif (data not shown). The reason might be that it is followed by a sequence containing proline-directed MAPkinase recognition motif LTSPTPADI. Therefore it might act as a phosphorylation dependent SUMO motif (Gregoire *et al.*, 2006). Another attractive candidate is the DTNKRKRDE sequence. Strikingly, it overlaps with the nuclear localisation signal sequence. It is recognised as two SUMO1 motifs, but the second of the lysines has a higher potential for being sumoylated based on the score and the following DE residues, which remind of the negatively charged amino acid-dependent SUMO motif (NDSM) ΨK(E/D)XXEIEEE. Moreover, this lysine was seen acetylated in our PTM analysis suggesting there might be a crosstalk between acetylation and sumoylation. The presence of SUMO1-acceptors in the NLS implies that sumoylation could affect the localisation of Pax5. Moreover two potential SUMO1 acceptors might have differential usage, one might mediate import and the other export out of the nucleus. Different developmental or environmental cues might activate the expression of E3 ligases that have affinity for one or the other lysine in the NLS. Interestingly, we detected RanBP2 in our mass spectrometry analysis of Pax5-interacting proteins. Ranbp2 is a GTP binding protein associated with the nuclear pore, which can act as an E3 SUMO ligase implying that sumoylation and import are linked events for some substrates (Pichler *et al.*, 2002). The role of other SUMO1-acceptor sites of Pax5 might be more directly linked to transcriptional control. In general, sumoylation of transcription factors is correlated with increased recruitment of HDACs

and repression functions. Previous studies have shown that Pax5 interacts with the co-repressor Daxx. Recently the mechanism of Daxx function has been elucidated. Daxx associates with SUMO-modified CBP (Kuo *et al.*, 2005) via its SUMO-interacting motif (SIM) (Lin *et al.*, 2006) and suppresses transcription by recruiting HDAC2. The previous finding that Daxx had different effects on the transactivation function of Pax5 in different B cell lines could result from interaction with different SUMO1-modified lysines of Pax5. I hypothesise the existence of different “Pax5-SUMO1” isoforms, whose prevalence depends on the developmental stage or specific stimuli. In conclusion, I propose that the different PTMs phosphorylation, methylation, sumoylation could have antagonistic and synergistic combinatorial effects on the function of Pax5. The presence of additional putative SUMO1 acceptors in the vicinity of a phosphorylation motif and the alternative modification of a lysine residue in the NLS by acetylation or SUMO1 strongly imply there might be a complicated combinatorial posttranslational modulation of the activity of Pax5.

Our analysis of posttranslational modifications can provide preliminary insights into mechanisms that determine the dual role of Pax5 in the regulation of transcription. However, further experiments will be needed to decipher the biological significance of the identified modifications: (1) biochemical analysis of serine and putative SUMO-acceptor lysine mutants in transactivation and co-recruitment assays, (2) analysis of the PTM state of mature B cells upon antigen stimulation, and (3) generation of mouse models with mutations of the residues undergoing PTM.

The paired domain, partial homeodomain, the octapeptide motif and the C-terminal regulatory domain of Pax5 (transactivation and inhibitory domains) show a high degree of evolutionary conservation. Moreover the paired domain and regulatory domain alone are sufficient for Pax5-dependent transcription activation in transiently transfected B cells. They constitute the “transcription factor” signature of Pax5. The paired domain binds the DNA, whereas the regulatory domain is responsible for transactivation. The latter domain, on its own, already has a dual entity of transactivation and inhibition at the same time, probably through intrinsic affinity for different co-regulatory complexes. The intervening sequences (143-304) between the conserved domains seem to act as a playground for evolutionary forces that shaped the different functions of Pax proteins. Residues accepting phospho- or SUMO-groups are being created, removed or retained within these sequences during evolution. As a result every Pax transcription factor in mammals has a unique signature that modulates its transcriptional activity.

Interacting proteins of Pax5

Developmental fate decisions are determined through the interplay of transcription factors and epigenetic modifiers, which together determine cellular identity (Schuettengruber *et al.*, 2007). Pax5 is a transcription factor that controls B-cell identity by simultaneously activating B-cell specific genes and repressing B-lineage inappropriate gene at lineage commitment (Nutt *et al.*, 1999; Delogu *et al.*, 2006; Schebesta *et al.*, 2007). Pax5 regulates gene expression by controlling the chromatin state at its target genes. It orchestrates these chromatin and transcription changes by recruiting chromatin-remodeling, histone-modifying and basal transcription factor complexes to its target genes, which defines Pax5 as a critical regulator of gene transcription in early B-cell development (McManus *et al.*, 2011).

By combining *in vivo* biotinylation of Pax5 with streptavidin-mediated protein precipitation I identified putative Pax5-interacting proteins in Ab-MuLV transformed pro-B cells by mass spectrometry analysis. *In vivo* biotinylation was previously used for identifying protein-complexes interacting with transcription factors, although these experiments were performed with stably transfected cell lines ectopically expressing the biotin-tagged proteins (de Boer *et al.*, 2003; Kim *et al.*, 2008; Soler *et al.*, 2010). In contrast, I isolated pro-B cells from a *Pax5*^{Bio/Bio} knock-in mouse expressing a biotin-tagged transcription factor, which was generated in our laboratory (McManus *et al.*, 2011). We identified more than 300 putative Pax5-interacting proteins, which co-purified in at least three out of six replicas that we analysed, and were enriched compared to the control precipitation. After grouping these proteins according to gene ontology categories, we could exclude ~30% of them since they were metabolic enzymes or structural components of the cytoskeleton, functions unlikely involving Pax5. From this data we conclude that the streptavidin tag, although ensuring more efficient purification of transcription factors, compared to antibody precipitation, results in high, unspecific background binding of endogenously biotinylated cytoplasmic proteins. Although I prepared nuclear extracts to avoid purification of the biotinylated cytoplasmic carboxylase enzymes, predominantly residing in the cytosol, they were overrepresented in my samples. I suppose this was partially because of (1) partial cross-contamination of the nuclear extracts with the cytosol, (2) multi-complex structure of metabolic enzymes and (3) the large size of individual components of these enzymes, which resulted in many tryptic peptides. Although in ChIP experiments high background binding inherent to the streptavidin tag can be avoided by using stringent washing buffers containing SDS, this was not an option for detecting interacting proteins, because I did not employ a crosslinking step. High background can be avoided using a V5-FLAG-Biotin-tag followed by tandem affinity purification (Wang *et al.*, 2009). The first step would be immunoprecipitation with M2 FLAG-agarose followed

by affinity elution with the FLAG-peptide and subsequent streptavidin pulldown (Wang *et al.*, 2009). Additionally, better distinction of specific interactions could be achieved by heavy isotope labeling of experimental compared to control samples (for example SILAC method) prior to quantitative mass spectrometry (Pan *et al.*, 2009).

We considered interacting proteins to be functionally relevant for Pax5 activities if they belonged to the following groups: (1) transcription factors [14%], (2) chromatin regulators [27%], (3) basal transcriptional machinery [6%], (4) DNA damage and repair [11%], (5) cell cycle [20%], (6) signalling [13%] and (7) transport [9%]. Overall these different protein classes may reflect the different functions of Pax5. In agreement with the identification of phosphorylation sites in Pax5, we found proteins involved in intracellular signaling. Depending on the activation of different signaling pathways and the consequent posttranslational modification (sumoylation, phosphorylation), Pax5 may be shuttled from its place of action, the nucleus, to the cytoplasm, or reverse through interaction with transporter proteins, although direct evidence of nuclear-cytoplasmic shuttling is yet missing for Pax5. In the nucleus, Pax5 can collaborate with the basal transcriptional complexes and other transcription factors to activate target genes. We identified Runx1, Ebf1 and Bcl11a as Pax5-interacting factors with important B-cell development function. These interactions were indirectly confirmed by ChIP-sequencing studies in our laboratory, which found the frequent presence of Runx1 motif in close proximity to Pax5-binding sites and identified gene-regulatory elements simultaneously bound by Pax5 and Ebf1 in pro-B cells (Revilla, Vilagoš and Bilić, unpublished data). The majority of putative interacting proteins are epigenetic regulators present in chromatin-remodeling and histone-modifying complexes, consistent with the role of Pax5 in orchestrating transcriptional and epigenetic changes of the B-cell gene expression program. Interestingly, we also identified proteins involved in DNA damage response and repair. For instance Ku70 and Ku80 proteins are involved in the ligation of coding joints of *Igh* gene segments after Rag1/2 cleavage and may thus link Pax5 with the V(D)J recombination process as previously suggested (Zhang *et al.*, 2006). Finally, the finding of many proteins involved in DNA replication and cell cycle control could reflect a novel role of Pax5, which we intend to further investigate. By co-precipitation and Western blotting I confirmed that Pax5 could interact with the chromatin remodeling BAF complex, the H3K4me3 methyltransferase complex MLL3/4, the histone acetyl transferase complex (CBP) and the basal transcriptional complex TFIID. Moreover these complexes are rapidly recruited to Pax5-target genes upon induction of Pax5 activity (McManus *et al.*, 2011). This recruitment is accompanied by H3K9 acetylation and H3K4 methylation (McManus *et al.*, 2011). Importantly, our mass spectrometry analysis identified more than one component of each of the Pax5-interacting

complexes (Table 3). Although previous studies showed that Pax5 represses its target genes through interaction with the Groucho corepressor, we could not investigate this in our system due to the lack of good-quality Groucho antibodies. Instead, I could identify the NCoR1 corepressor complex associated with HDAC3 activity as a Pax5-interacting complex that is recruited to repressed Pax5 target genes (McManus *et al.*, 2011). Besides NCoR1, we found other components of the associated complex in our mass spectrometry analysis Tbl1xr1, Tbl1x, Gm260 and Cnot2 (Table 3). Moreover we also identified components of other repressive complexes with HDAC activity Phb2 (HDAC), Chd4 (NuRD) and Set (Table 3).

In conclusion, the mass spectrometry analysis identified proteins reflecting the entire spectrum of Pax5 activities in the cell from its transport to the nucleus, modulation by signaling pathways and involvement in V(D)J recombination to its ultimate role in orchestrating transcriptional and epigenetic gene expression changes controlling the B cell fate. Pax5 activates regulatory elements of its target genes through direct recruitment of chromatin remodelling, histone-modifying and basal transcription factor complexes. Moreover, Pax5 directly recruits the NCoR1 histone deacetylase complex to repressed target genes. Now that we have unraveled the molecular mechanism of Pax5-dependent transcriptional regulation, further experiments can focus on decoding other processes involving Pax5, for example, in *Igh* locus contraction or in DNA replication and cell cycle control.

CTCF is a ubiquitous transcriptional regulator with numerous functions and many interaction partners that binds multiple sites genome-wide probably due to combinatorial use of its 11 zinc-finger domains. It plays multiple roles in the regulation of gene expression. Most of its activities derive from the ability of CTCF to dimerise (Yusufzai *et al.*, 2004) and bring distant sites into proximity within the nucleus resulting in DNA loop formation (reviewed by Choi *et al.*, 2011). Binding of CTCF often isolates a promoter from an enhancer as in the imprinted *Igf2/H19* locus (Hadjur *et al.*, 2009). In contrast, evidence is emerging that CTCF-mediated looping can bring many promoters and enhancers of co-regulated loci into close proximity (Majumder *et al.*, 2008; Hadjur *et al.*, 2009; Sekimata *et al.*, 2009). This type of regulation is primed, under certain differentiation stimuli, by expression of a specific transcription factor, usually an interaction partner of CTCF (Majumder *et al.*, 2008; Hadjur *et al.*, 2009; Sekimata *et al.*, 2009). In addition to transcription factors, CTCF interacts with various epigenetic regulators, among them the DNA helicase component of the MLL-complex Chd8 (Ishihara *et al.*, 2006), Suz12, a member of polycomb group repressive complex 2 (Li *et al.*, 2008) or Sin3A associated with HDAC recruitment (Lutz *et al.*, 2000). While these interactions seem to be mostly transient and cell context-dependent, a *bona fide* interaction partner of CTCF is the cohesin complex, which co-

localises at most CTCF-binding sites (Parelho *et al.*, 2008; Wendt *et al.*, 2008; Stedman *et al.*, 2008). The cohesin complex is critical for the long-range interactions mediated by CTCF (Hadjur *et al.*, 2009). It consists of a tripartite ring formed by Smc1, Smc3 and Scc1/Rad21 that embrace the sister chromatids in *trans* (reviewed by Wendt and Peters, 2009). The structure is further stabilized, opened or closed by the SA2 subunit (Wendt and Peters, 2009). The C-terminal domain of CTCF directly interacts with the SA2 subunit of cohesin, whereas the other subunits co-purify with CTCF through this SA2 interaction (Xiao *et al.*, 2011). Similarly to cohesion during mitosis, the cohesin ring, composed of Smc1, Smc3 and Rad21, can encircle DNA strands in *cis* at distant CTCF sites to form chromatin loops (Wendt and Peters, 2009). The SA2 subunit would act as bridge between CTCF and the cohesin circle also in this situation (Xiao *et al.*, 2011). Secondary proteins could bind to both CTCF and cohesin and stabilize the interaction (Xiao *et al.*, 2011).

I have confirmed by reciprocal Pax5-Bio and CTCF-Bio streptavidin-pulldown and immunoprecipitation experiments, followed by mass spectrometry or Western blotting, that Pax5 and CTCF directly interact in pro-B cells. The interaction is stronger if the C-terminal domain of CTCF was deleted. Therefore, it is conceivable that the C-terminal sequence may suppress the interaction with Pax5. The C-terminal domain has casein kinase II phosphorylation sites. It would be interesting to analyse whether mutation of these sites would be sufficient for increasing the affinity of CTCF for Pax5. Another possibility is that a certain interaction partner binding to this domain partially impairs the interaction with Pax5. In contrast, the C-terminal domain is essential and sufficient for the interaction of CTCF with cohesin (Xiao *et al.*, 2011). Indeed, we identified all of the cohesin subunits by mass spectrometry analysis in the CTCF-Bio streptavidin-pulldown experiment. Interestingly, some of the cohesin subunits Smc1, Smc3 and Rad21 were co-purified with Pax5-Bio as well, although the confirmation of this potential interaction by Western blotting has not been done. A site where the Pax5-CTCF-cohesin complex might join forces to execute a developmentally controlled long-range interaction in pro-B cells is the *Igh* locus. In this regard, the PAIR elements in the distal V_H region are particularly interesting with their pro-B cell specific binding of Pax5, Pax5-dependent active chromatin marks and antisense transcription (Ebert *et al.*, 2011). CTCF, cohesin and E2A bind the PAIR elements already in uncommitted progenitors (Ebert *et al.*, 2011). At the pro-B cell stage, Pax5 recognizes its binding motif in the PAIR sequence (Ebert *et al.*, 2011), but could also be coordinately recruited through its interaction with CTCF and cohesin on the PAIR elements. Pax5 binding could induce an activating cascade. Chromatin regulatory MLL- and HAT-complexes could activate the chromatin and create additional binding sites for the BAF-complex. Finally, RNA pol II could

actively transcribe the PAIR elements. Moreover, components of the MLL complex and RNA pol II could also directly interact with CTCF, and E2A, enhancing their recruitment to more distant elements within the V_H region and thus leading to the contraction of the entire locus. This amplification cascade generated by chromatin-transcription-crosstalk is reminiscent of the signaling cascade, where acetylation and methylation are creating binding sites for additional scaffolding proteins, in analogy to phosphorylation, while the antisense transcripts could play a role of secondary messengers by further promoting chromatin accessibility of adjacent *Igh* sequences.

In conclusion, the mass spectrometry analysis has paved the way for future investigations of Pax5 function. Although we identified some transcription factors that could be acting in concert with Pax5 (Mef2c, Klf13), we lacked antibodies enabling confirmation of these interactions and analysis of genes that they regulate. ChIP-sequencing experiments in our laboratory (Revilla, Vilagoš and Bilić; unpublished data) revealed that Runx1 and Ebf1 often bind in proximity to Pax5 sites, which can be considered as indirect confirmation of the mass spectrometry result. We have preliminary co-precipitation evidence that Bcl11a, a transcription factor important for B-cell development, can interact with Pax5 (unpublished data). However we still do not know the genes that Bcl11a regulates. Surprisingly we identified, by mass spectrometry, a high enrichment of Pax5-interacting proteins involved in DNA replication and cell cycle regulation. Moreover, they were also enriched with CTCF, which is known to associate with centrosomes during mitosis (Zhang *et al.*, 2004), supporting the possibility that the finding is not a mere consequence of DNA-mediated unspecific binding. Future work will investigate whether this finding has a biological relevance by performing co-immunoprecipitation/co-localisation experiments with B cells synchronized at different stages of the cell cycle.

In summary, with this study, I have deepened the understanding of the molecular mechanism employed by Pax5 to regulate the B-cell transcriptional program. I have confirmed previously known, and added new players in this network, and finally opened many new questions to be answered by future experimenters.

DISCUSSION II

We successfully employed the circular chromosome conformation capture (4C) technology to investigate the interactions of DNA elements that regulate the accessibility and subsequent recombination of the discontinuous gene segments in the *Igh* immunoglobulin heavy-chain (*Igh*) locus. To our knowledge, this is the first study of long-range interactions in the entire *Igh* locus, which were identified by the 4C technology coupled with deep sequencing. We studied the interaction frequencies of elements strategically located in different domains of the *Igh* locus: (1) enhancer elements of the 3' end, (2) a proximal V_H pseudogene segment and (3) PAIR elements in the distal V_H region. In this study, a common feature of all the anchor fragments that we analysed was that their interactions with nearby sequences were frequent and independent of the expression of the Pax5 transcription factor, suggesting that they result from random collisions. On the contrary, the long-range interactions, up to 3 Mb in genomic distance, were detected with high frequency only in *Rag2*^{-/-} pro-B cells as a consequence of locus contraction mediated by Pax5.

We noticed a discrepancy between our EMSA and 4C experiments as PAIR elements 8 and 5, which have a mutation in Pax5- and E2A-binding motifs, respectively, still engaged in long-range interactions. Furthermore, Pax5 was binding to PAIR8 in a ChIP-sequencing experiment (Ebert, unpublished data), although it showed impaired binding in the EMSA experiment. Based on these results, we predict that Pax5 could be recruited to PAIR elements through its interaction with CTCF, which might be sufficient even if the Pax5-binding motif is mutated. Another explanation might be that other factors and the local chromatin structure alter the DNA conformation of the motif, allowing Pax5 to bind despite the mutation. I have cloned a DNA fragment that contains the binding motifs for all three factors. In order to investigate the existence of mutual cooperation, I have mutated the binding motifs of Pax5, CTCF and E2A individually and in different combinations. The same issue could also be resolved by *in vivo* footprint analysis of the PAIR8 element.

Our 4C analysis confirmed the previously reported long-range interactions of the enhancer elements at the 3' end of the *Igh* locus, the 3'RR and E μ (Ju *et al.*, 2007). The HS3B and HS4 of the 3'RR stimulate expression of the rearranged *Igh* gene (Pinaud *et al.*, 2001) and class switch recombination (CSR) through physical interactions (Ju *et al.*, 2007). These interactions involving the 3'RR, E μ and the D_H region might organise a poised chromatin domain that could become fully active upon antigen stimulation and CSR. We also identified an interaction of the 3' enhancers with the D_H genes and elements in the V_H-D_H region that contain Pax5- or CTCF-

binding sites. Most importantly, we show for the first time that the 3'RR directly interacts beyond the V_H-D_H region with the distal V_H gene cluster. Only in *Rag2*^{-/-} pro-B cells do these interactions occur at high frequencies, indicating that they are a consequence of Pax5-dependant locus contraction. The most prominent interactions are with the PAIR elements and V_HJ558 genes. In the reciprocal experiment, the PAIR elements 4 and 8 also engage in interactions with the same groups of fragments in the 3' region of the *Igh* locus, for example with the 3'RR, E μ , D_H and V_H-D_H region. In contrast to the other PAIR elements, PAIR5 does not interact with the D_H gene segments probably due to the absence of a functional E2A-binding site. As expected, the PAIR elements interact with each other even in the absence of Pax5, most likely through dimerisation of CTCF bound to its sites in these elements. In line with this evidence, we would propose that the PAIR elements, through the factors bound to them, regulate and are responsible for the distal interactions with the 3'RR, E μ and the D_H gene regions. We noticed higher interaction frequencies when PAIR4 was used as an anchor fragment compared to PAIR5 or PAIR8. The reason might be that PAIR4 has no mutation in any motifs of the three factors, which could provide this element with better ability to mediate interactions. Therefore this result may suggest the existence of mutual synergism between factors bound to each single PAIR element.

The pseudogene PG.4.28, located in the proximal V_H region, seemed to interact with the same set of elements at the 3' end of the locus and in the 5' distal V_H region as did the PAIR elements and the 3' enhancers. The CTCF binding site within this fragment probably contributes to these interactions. Nevertheless, the interaction frequencies of PG.4.28 were much lower, despite the smaller genomic distance from the 3'RR compared to PAIR elements and from the PAIR elements compared to the 3'RR.

The V(D)J recombination proceeds in a developmentally ordered and cell type-specific manner. The intergenic region between the most 3' V_H gene and the most 5' D_H gene harbours elements that prevent proximal V_H-to-DJ_H recombination in T cells by suppressing antisense transcription (Featherstone *et al.*, 2010; Giallourakis *et al.*, 2010). The proposed mechanism includes separation of the DJ_H regions from the V_H region by the ~100 kb V_H-D_H sequences. CTCF-binding sites 5' of the DFL16.1 gene are proposed to function as insulators (Giallourakis *et al.*, 2010) in non-B cells, while activation of a lymphocyte-specific hypersensitive site 3' of DFL16.1 might be involved in looping out of V_H-D_H region to juxtapose the DJ_H regions next to the V_H region (Featherstone *et al.*, 2010). Although it would have been interesting to test the interactions of the hypersensitive sites surrounding the DFL16.1, the choice of HindIII restriction enzyme did not allow this, as all D_H gene segments except for D_HQ52 are located within one large HindIII fragment. We suspect that the size of this fragment and the repetitive nature of the

DSP genes interfere with the mappability of the two neighbouring smaller HindIII fragments containing the two CTCF-binding sites located 5' of the DFL16.1 gene. These CTCF-binding sites are in opposite orientation, which could provide them with a distinct role at different stages of B cell development. At the progenitor stage and in T cells, CTCF could, upon interaction with the 3'RR, insulate the region to prevent premature V_H-D_H rearrangement. Subsequently to D_H-J_H joining, a cell type-specific factor could induce binding of CTCF to an activating site (Featherstone *et al.*, 2010). In line with this hypothesis, the following model based on 3D-FISH (Jhunjhunwala *et al.*, 2008) and preliminary 3C studies has been proposed (Degner-Leisso and Feeney, 2010). The two CTCF binding sites 5' of DFL16.1 and the sites in the 3'RR are proposed to form a loop retaining the D_H-J_H-C_H region isolated from the V_H region (Degner-Leisso and Feeney, 2010). During V_H-to-DJ_H rearrangement, a new set of loops would form to bring the V_H region close to the rearranged DJ_H segment (Degner-Leisso and Feeney, 2010). In our 4C analysis of *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells, we detected interactions of the 3'RR and the E_μ enhancer, outside of the V_H-D_H region, with proximal V_H genes. Only 5' to the proximal V_H7183.7.10 gene, the interactions of the 3'RR drastically decrease in *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells. The discrepancy might originate in the type of cells used for the studies. The 3D-FISH analysis was done on *E2A*^{-/-} cells that have an earlier block in B cell development than *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells. *E2A* is required for D_H-J_H rearrangement; therefore the topology might reflect the stage of suppressing V_H-D_H joining prior to D_H-J_H recombination. Otherwise, the results of our study agree with the 3D-FISH model of the *Igh* locus topology in pro-B cells (Jhunjhunwala *et al.*, 2008). According to this, the *Igh* locus is organised in three separate domains in *E2A*^{-/-} pre-pro B cells. In pro-B cells, the proximal and distal V_H gene regions merged and are juxtaposed in close proximity to the DJ_H region, allowing all V_H genes to encounter D_H elements with similar probability (Jhunjhunwala *et al.*, 2008). Likewise, in our 4C study of the 3'RR in *Rag2*^{-/-} pro-B cells, the HS3B/HS4 sites of the 3'RR interacted at similar frequencies with the proximal, middle and distal V_H genes. Additionally, we noticed a few more frequently interacting regions, which might be involved in the formation of larger loops or domains.

Our 4C analysis of the *Igh* locus in pro-B cells strongly supports the idea that the intergenic PAIR elements have a crucial regulatory role in the process of V(D)J recombination. Similar to the analysis of the complex MHCII locus (Majumder *et al.*, 2010), CTCF molecules bound to PAIR elements may randomly dimerise at their nearby binding sites in the V_H region in the *Rag2*^{-/-} *Pax5*^{-/-} pro-B cells. In pro-B cells, Pax5 binds to the PAIR elements and engages in protein-protein interactions with CTCF bound to PAIRs, but also other more distal elements

throughout the locus. These interactions are further strengthened by CTCF-CTCF dimerisation (Yusufzai *et al.*, 2004) and finally binding of the cohesin complex, which embraces the formed DNA loops (reviewed by Wendt and Peters, 2009). Thus, a stable multiprotein-DNA complex is formed, “the V(D)J recombination hub”, where all the components are collaborating in order to confine the large gene domains of the entire *Igh* locus into a small 3D space, which provides an equal opportunity to all V_H genes to undergo V_H-DJ_H recombination.

In the future, we would like to validate the interactions of the 3' enhancer and the PAIR elements. We will use a 4-bp cutting restriction enzyme for 4C analysis. Moreover, this approach will allow us to dissect the individual interactions of the gene segments in the D_H region, which will provide a more detailed map of the reported putative interactions. Hopefully, this will also allow us to measure to what degree the different PAIR elements contribute to the regulation of *Igh* locus contraction and whether the observed differences correlate with conservation of the binding motifs for Pax5, E2A and CTCF in the PAIR elements.

We are also interested to use the 4C-sequencing method to examine the *Igh* locus contraction in mutant mice that display a V(D)J recombination defect such as for example *Ikaros*^{-/-}, *Eed*^{-/-} and *Yy1*^{-/-} mice. Furthermore, all evidence indicates that CTCF is crucial for *Igh* locus contraction. However, due to its important cell cycle functions, we were unable to obtain adequate numbers of viable *Rag2*^{-/-} *Ctcf*^{-/-} pro-B cells for 4C and V(D)J recombination analyses. To circumvent this obstacle, we are in the process of generating a conditional dominant-negative CTCFΔCt allele. As the C-terminus of CTCF is important for recruiting the SA2 subunit of cohesin (Xiao *et al.*, 2011), the CTCFΔCt protein may interfere with the formation of DNA loops and consequently with *Igh* locus contraction. Finally, we hope that these experiments will help us to unravel the missing links and dissect in more detail the complex protein-DNA interaction network that regulates *Igh* locus contraction and V(D)J recombination at the pro-B cell stage.

EXPERIMENTAL PROCEDURES

Isolation and immortalisation of bone marrow pro-B cells

Cells were isolated from bones of 4 weeks old mice of *Pax5^{Bio}* (Pax5-Bio), *CTCF^{Bio/+}*, *Rosa26^{BirA/BirA}* (CTCF-Bio) or control *Rosa26^{BirA/BirA}* (BirA) genotype. One half of total cells isolated from a single mouse was resuspended in 1 ml of the Abelson murine leukemia virus (Ab-MuLV) supernatant containing 1% polybrene. The cells were incubated at 37°C for 3.5 hours when 1 ml of pro-B cell medium was added (IMDM, 20%FCS, 1% PSG, 50mM β -mercaptoethanol, 0.5% cyproxin, 1% IL-7). After 3 days 1 ml of fresh medium was added. At day 7 after the infection, the infected cells started to expand. They detached from the surface of the plate and built pro-B cell like clusters. At this point the cells were split to bigger cell culture plates in the pro-B cell medium. After two weeks the cells became independent of IL-7 receptor signaling and were expanded in IMDM without IL-7 and with 10 % FCS. In this medium the cells could be maintained for an indefinite time, with regular splitting of the culture in an interval of 2 days. We confirmed by flow cytometry analysis that they expressed CD19 and B220 receptors at their cell surface.

Nuclear extract preparation and streptavidin pulldown for mass spectrometry analysis

Nuclear extracts were prepared as described previously (Dignam *et al.*, 1983). Approximately 12×10^9 of Abelson murine leukemia virus (Ab-MuLV)-transformed pro-B cells of the *Pax5^{Bio/Bio}* (Pax5-Bio) or control *Rosa26^{BirA/BirA}* (BirA) genotype were harvested and washed with PBS. The cell pellet was carefully resuspended in 5 volumes of ice cold buffer A (10 mM Hepes, 10 mM KCl, 0.1 mM EDTA, 0.1mM EGTA, 0.75 mM spermidine, 0.15 mM spermine, protease inhibitor cocktail (Roche), 1 mM DTT, 10 mM NaF, 20 mM β -glycerophosphate, 10 mM sodium pyrophosphate, 1 mM orto-vanadate, 0.1 μ M PMSF) and incubated on ice for 10 min. Swollen cells were spun at 250 x g and lysed in 2 volumes of buffer A by 15 strokes of the Dounce homogeniser. The nuclei were harvested by centrifugation at 500 x g for 10 min at 4°C. The pellet was lysed in 4 volumes of buffer C (420 mM NaCl, 20 mM Hepes, 25 % glycerol, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 10 mM NaF, 20 mM β -glycerophosphate, 10 mM sodium pyrophosphate, 1 mM orto-vanadate, 0.1 μ M PMSF). DNA associated proteins were extracted by tumbling for 45 min at 4°C. Aggregated DNA was removed by centrifugation at 16100 x g for 30 min at 4°C. The supernatant was collected, the salt concentration lowered to 150 mM by drop-wise addition of the dilution buffer (420 mM NaCl, 20 mM Hepes, 25 % glycerol, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 10 mM NaF, 20 mM β -glycerophosphate, 10 mM sodium

pyrophosphate, 1 mM orto-vanadate, 0.1 μ M PMSF). The concentration of proteins in the extract was determined by Bradford assay using BSA as a standard. An amount of 25 mg (for the detection of interacting proteins) or 50 mg (for the detection of posttranslational modifications) of proteins was used for the streptavidin affinity pulldown. Streptavidin coated magnetic beads (Dynabeads M-280 SA, Invitrogen) were washed 3 times in PBS and blocked in 0.2 mg/ml BSA. After equilibration in the binding buffer, the beads were incubated over night in the experimental (Pax5-Bio) or control nuclear extract (BirA). Streptavidin bound protein complexes were separated with a magnetic separation rack (Invitrogen), followed by 5-6 washes of beads with Heng 250 buffer (250 mM KCl, 0.3% NP-40, 10 mM HEPES-KOH, 1.5 mM MgCl, 0.25 mM EDTA, 20% glycerol, 1 mM DTT, 10 mM NaF, 20 mM β -glycerophosphate, 10 mM sodium pyrophosphate, 1 mM orto-vanadate, 0.1 μ M PMSF). The precipitated proteins were eluted from the beads by boiling in 2x SDS sample buffer (5% (wt/vol) SDS, 150 mM Tris-HCl pH 6.7, 30% (vol/vol) glycerol, 720 mM β -mercaptoethanol and 0.0025% (wt/vol) bromophenol blue), and separated in the 4-12 % Bis-Tris NuPage gel (NuPAGE® Novex, Invitrogen) by electrophoresis with the MOPS SDS running buffer (NuPAGE® MOPS, Invitrogen). The fractioned proteins were fixed (40 % ethanol, 10 % acetic acid) in the gel and visualised by Colloidal Coomassie staining (10% ammonium sulphate, 2% orto-phosphoric acid, 20% methanol, 0.2% Coomassie brilliant blue G-250 (Serva) over night at 4°C. The discrete protein bands were cut out of the gel and prepared for mass spectrometry analysis.

Isolation of SDS-PAGE fractioned proteins and tryptic digestion for mass spectrometry analysis

The gel slices containing discrete protein bands were cut in 2-3 mm pieces, placed into 0.6 PCR tubes (AXYGEN) and washed at room temperature (RT) for 10 min: (1) in 200 μ l 50 mM tetra ammonium bicarbonate (TEAB; NH_4HCO_3 , Sigma) two times, (2) in 200 μ l 50 mM TEAB / 30% acetonitrile (HPLC grade, Supra-Gradient, Biosolve B.V.). The gel pieces were then shrank in 60% acetonitrile, reduced in 100 μ l of 1 mg/ml DTT (Sigma)/50 mM TEAB at 57°C for 30 min and alkylated in 200 μ l 5 mg/ml ioadacetamide/50 mM TEAB for 35 min at RT in the dark. The supernatant was removed and the washing steps repeated as described above. The gel-pieces were dried in the speed vacuum centrifuge for 5-7 min and digested with 12 ng/ μ l trypsin (Gold, Promega) in 50 mM TEAB at 37°C over night. The tryptic supernatant was collected and stored at 4°C. The remaining tryptic peptides were extracted from the gel pieces by sonicating them twice in 20 μ l of 5% formic acid for 10 min in a cooled ultrasonic bath. The procedure resulted in ~60 μ l of the tryptic gel digest. For detection of posttranslational modifications the Pax5-

containing gel slice was incubated with three enzymes: trypsin, chymotrypsin and subtilisin.

The nano-HPLC (high-performance liquid chromatography) system used for the separation and analysis of the tryptic peptides was an UltiMate 3000 Dual Gradient HPLC system (Dionex), equipped with a nanospray source (Proxeon), coupled to a LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific). The LTQ FT was operated in data-dependent mode using a high-resolution full scan in the Orbitrap followed by tandem mass spectrometry scans of the five most abundant ions in the linear ion trap. Acquired data (XCalibur RAW-file) was converted into Sequest DTA-files using the program extract_msn and further merged into Mascot generic files using Mascot Daemon (Matrix Science). Database searches were performed with the Mascot 2.2.0 software using a non-redundant mouse database.

To confirm the existence of posttranslational modifications in pro-B cells, the Pax5-band was analysed in three replicas. The analysis of co-purified proteins was done in 6 replicas. One mass spectrometry analysis of Pax5-Bio and BirA purifications was done in parallel with a streptavidin pulldown from CTCF-Bio nuclear extract.

Mass spectrometry data analysis

Proteins identified in all purifications were merged into one list. This list was filtered so that only proteins that meet the following criteria are selected: (1) not a known contaminant (ex. keratin, trypsin), (2) the average number of peptides observed in the Pax5-Bio samples is three times above the average number of peptides in the BirA samples and (3) the protein must be identified in at least two Pax5-Bio samples with at least two peptides identified in at least one of the two Pax5 samples. The filtering procedure generated a list of 314 proteins. Their abundance in the sample can be displayed with the number of uniquely identified tryptic peptides matching the sequence of a specific protein. The proteins selected by this procedure were grouped according to their Gene Ontology annotation (GO), which is reported in the NCBI database (<http://www.ncbi.nlm.nih.gov/gene>). The relative contribution of each functional group to the Pax5-interactome was displayed in the following way: (1) numbers of unique peptides (PAH) of all proteins in the same GO group were summed, (2) the distribution was visualised by an Excel-pie chart.

Analysis of proteins identified both in Pax5-Bio and CTCF-Bio streptavidin pulldowns

Proteins identified in the control BirA purification were excluded. The proteins uniquely identified in CTCF-Bio and Pax5-Bio purifications were analysed by the Venny tool for comparing lists (Oliveros, 2007; <http://bioinfogp.cnb.csic.es/tools/venny/index.html>). The proteins that overlapped in both lists were grouped according to GO as described above.

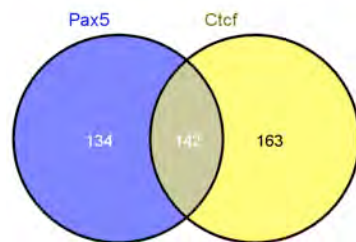


Figure 47. Output of the Venny tool.

The lists of proteins found uniquely in mass spectrometry analyses of Pax5-Bio or CTCF-Bio streptavidin pulldown were compared by the Venny tool. The list of the overlapping 142 proteins was downloaded, and the GO annotation of the proteins was investigated.

Nuclear extract preparation with benzonase and validation of Pax5-interacting proteins

The streptavidin pulldown results in high background binding of endogenously biotinylated proteins. Most of them reside in the cytosol, but can sometimes be extracted during the Dounce homogenisation. Therefore the above described protocol for nuclear-extract preparation was replaced with the following one, resulting in less contamination by cytosolic proteins (Dyer and Herzog, 1995; with minor modifications). The protocol involves a gentler isolation of nuclei from pro-B cells. Furthermore, the possibility of a DNA-mediated association of proteins with Pax5 was excluded by including benzonase in the extract preparation. Most of Pax5 in a cell is tightly bound to DNA, therefore the inclusion of benzonase additionally increased the enrichment of Pax5 protein in the extract (Figure 47).

Pro-B cells ($1-2 \times 10^8$) were lysed in a buffer containing 10 mM Tris pH 8.0, 0.32 M Sucrose, 50 mM KCl, 20 mM NaCl, 3 mM CaCl_2 , 2 mM MgCl_2 , 0.1% NP-40, 1 mM DTT, 2 mM 6-aminocaproic acid (6AA), 0.15 mM spermine, 0.5 mM spermidine, 0.5 μM PMSF, 10 mM NaF, 10 mM β -glycerophosphate and protease inhibitor cocktail. The nuclei were harvested by centrifugation for 5 minutes at 500 x g and lysed in a buffer containing 10 mM HEPES pH 7.9, 20% glycerol, 2.0 mM MgAc, 0.36 M KCl, 0.2% CHAPS, 10 mM NaF, 0.5 mM DTT, 2 mM 6-aminocaproic acid (6AA), 0.5 μM PMSF, 10 mM NaF, 10 mM β -glycerophosphate and protease inhibitor cocktail. The endonuclease Benzonase (1016540001, Merck) was added (450 units per 2×10^8 cells) and the suspension was incubated at 4°C for 45 min followed by centrifugation at

16100 x g for 10 min. The extract was diluted with 3 volumes of 0.36 M KCl lysis buffer (see above). The protein concentration was determined by the Bradford assay using BSA as a standard.

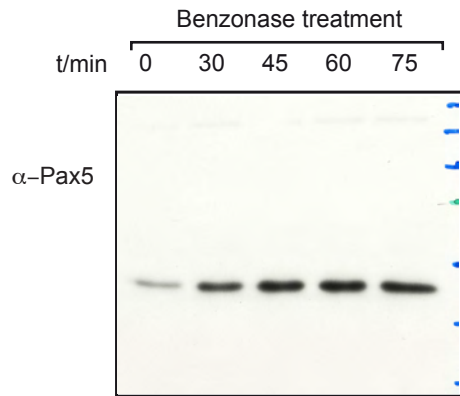


Figure 48. Nuclei were resuspended in 350 mM KCl buffer. Per 2×10^8 pro-B cells, 450 units of benzonase were added. Nuclear proteins were extracted for the indicated amount of time (t/min). Time 0 indicates that benzonase was not included; this extract was prepared only with high salt extraction for 75 minutes.

Immunoprecipitation and streptavidin pulldown for Western blot analysis

Streptavidin magnetic beads (Dynabeads M-280 SA, 11206D, Invitrogen) and protein A beads (10002D, Invitrogen) were blocked with 2 mg/ml BSA and washed three times with PBS. Nuclear extracts were incubated with streptavidin beads or purified antibodies at 4°C over night. The antibody bound complexes were collected by addition of protein A beads for another 3 hours. The beads were washed 5 times in 5 ml of buffer (20 mM Tris pH 8, 250 mM KCl, 2 mM MgCl₂, 10% glycerol, 1 mM DTT, 2 mM 6-aminocaproic acid, 10 mM NaF, 10 mM β-glycerophosphate, 1 mM sodium pyrophosphate and 0.1% protease inhibitor cocktail).

In CTCF and NCoR immunoprecipitation experiments the interaction with Pax5 was investigated in preparations with or without benzonase. The antibody complexes were collected by incubation with 40 μl of protein-A sepharose beads (17-0780-01, GE Healthcare) for 2 hours. The washing buffer in these two experiments contained 150 mM NaCl (other components were as described).

The precipitated proteins were eluted by boiling in 2x SDS sample buffer and separated by SDS-PAGE followed by Western blotting.

Co-immunoprecipitation of Pax5 with PTIP and CTCF in HEK293T cells

The biotin-tagged CTCF protein was previously generated in our laboratory (Barbara Werner) under the control of a Murine Stem Cell Virus (MSCV) promoter. To use a system that gives high protein expression levels, the cDNA was re-cloned into a CMV (Cytomegalovirus) based expression vector. Mutant CTCF proteins were generated by PCR mutagenesis using the proofreading DNA polymerase (Pfu, Stratagene). Sequences coding for different domains of CTCF were amplified with Sall and EcoRI restriction enzyme overhangs and cloned into the CMV expression vector mentioned above. The zinc fingers deletion mutant (NT-CT) was generated by three separate PCRs: (1) separate amplification of the N- and the C-terminal domains with primers containing an overlapping complementary sequence (blue font in Table 17), (2) gel-purified products (Qiagen Gel Extraction Kit) were mixed in equal molar ratios and the annealed DNA molecule was amplified.

Domain	Overhang	Primer name	5' Primer sequence	F/R
N-terminus (NT)	Sall	NT-1	taggtcgaccaccatgGAAGGTGAGGCG	F
	EcoRI	NT-2	gcggaattcACACTGGAATGTTTTCTTTACA	R
	C-terminus	NT-3	GCCATCTGGACCAGCCTGGAATGTTTTCTTA	R
Zn finger (ZF)	Sall	ZF-1	gaggtcgaccaccatgACATTCCAGTGTGA	F
	EcoRI	ZF-2	acagaattcCTCTACGCCATCTGGACCAGC	R
C-terminus (CT)	Sall	CT-1	gcggtcgaccaccatgGCAGATAACTGTG	F
	EcoRI	CT-2	gaggaattcCCGGTCCATCATGCTGAGGAT	R
	N-terminus	CT-3	TAAGAAAACATTCCAGGCTGGTCCAGATGGC	F

Table 17. Sequences of primers designed for the cloning of CTCF mutants.

In the primer sequences, the restriction-site overhangs are indicated in small caps. The overlapping sequence in the primers used for generating the NT-CT mutant is depicted in blue. The direction of the primer is indicated in the last column (F; forward), (R; reverse).

Mutant name	Deleted domain	Primer combination
NT-ZF	C-terminus (Δ CT)	NT-1; ZF-2
CT-ZF	N-terminus (Δ NT)	ZF-1; CT-2
NT	Zn-finger; C-terminus (Δ ZF Δ CT)	NT-1; NT-2
ZF	N-terminus; C-terminus (Δ NT Δ CT)	ZF-1; ZF-2
CT	N-terminus; Zn-finger (Δ NT Δ ZF)	CT-1; CT-2
NT-CT	Zn finger (Δ ZF)	PCR1: NT-1; NT-3 PCR2: CT-3; CT-2 PCR3: NT-1; CT-2

Table 18. Primer combinations used for the cloning of CTCF mutants.

The names of the mutants refer to the domains that were retained.

In transient transfection assays using Lipofectamine 2000 (Invitrogen), HEK293T cells were grown in 10 cm tissue culture plates until reaching 90% confluency. At this point they were co-transfected with 10 to 15 μ g (depending on the expression efficiency of CTCF-mutant proteins) of CTCF-plasmid DNA and 4 μ g of Pax5-plasmid DNA. When Fugene 6 (Roche) was used for transfection, cells were grown until being 60% confluent and transfected with 3 μ g of FLAG-PTIP and 1 μ g of Pax5 CMV-based expression vectors.

The cells were harvested after 48 hours, washed twice with PBS and lysed in a buffer containing 150 mM NaCl, 50 mM Tris-HCl pH 8, 5 mM MgCl₂, 1 mM EDTA, 0.5 % NP-40, 2 mM 6-aminocaproic acid, 10 mM NaF, 10 mM β -glycerophosphate, 1 mM sodium pyrophosphate and protease inhibitor cocktail (Roche). To enhance the extraction of nuclear proteins, one volume of this buffer containing 600 mM KCl was added drop-wise to the lysate, followed by a rotation at 4°C for 45 minutes. Cellular debris was spun down at 16100 x g for 30 minutes. The extract was diluted to 150 mM salt concentration and incubated, with washed and blocked (in 1 mg/ml BSA for 2 hours) ANTI-FLAG® M2 Affinity Gel (Sigma) at 4°C for 3 hours (CTCF and its mutants) or over night (PTIP), in the presence of benzonase. The beads were washed five times with a buffer containing 250 mM KCl, 20 mM Tris-HCl pH 8.0, 5 mM MgCl₂, 1 mM EDTA, 10 % glycerol, 2 mM 6-aminocaproic acid, 10 mM NaF, 10 mM β -glycerophosphate, 1 mM sodium pyrophosphate and protease inhibitor cocktail (Roche) and twice with TBS. Bound proteins were eluted by incubation in two bead volumes of 0.3 mg/ml FLAG peptide (diluted in TBS) for 30 minutes at 4°C, the eluate was resuspended in 6x SDS-sample buffer and analysed by Western blotting.

***In vitro* sumoylation assay**

In vitro SUMO1 modification assay was done as previously described (Pichler *et al.*, 2008). All recombinant proteins used in the assay (except Pax5) were purified and provided by Dr. Andrea Pichler's laboratory.

Recombinant Pax5 protein was purified from BL21 bacterial strain in denaturing conditions. Five hundred nanograms of purified Pax5 was combined with 2.2 μ M SUMO1, 70 nM His-Aos1/Uba2 (E1), 250 nM Ubc9 (E2) and 4-16 nM of RanBP2 IR1+M or 5-80 RanBP2 Δ FG (catalytically active domains of RanBP2 E3 ligase; Pichler *et al.*, 2004) in 20 μ l of transport buffer (20 mM HEPES, 110 mM potassium acetate, 2 mM magnesium acetate, 1 mM EGTA, 1 mM DTT, 0.05 % (v/v) Tween, 0.2 mg/ml ovalbumine (Sigma, grade VI). The reaction was initiated with addition of 0.5 mM ATP and allowed to proceed at 30°C, initially for 2 hours. ATP-deficient reaction mixtures were used as negative

controls. The reaction was terminated by adding one volume of 2x SDS sample buffer and boiling at 95°C for 5 minutes. Further optimizations resulted in decreasing the concentration of RanBP2ΔFG to 5 nM. To analyse the sumoylation-kinetics of Pax5, the reaction mixture was divided to 9 aliquots and incubated at 30°C. In intervals of 15 minutes, the reaction of one aliquot was stopped by was resuspending it in 2x SDS sample buffer and boiling at 95°C and stored on ice. The products were fractioned by SDS-PAGE and analysed by Western blotting with Pax5 and SUMO1 specific antibodies.

***In vivo* sumoylation assay in transiently transfected HEK293T cells**

The investigation of *in vivo* sumoylation was performed as previously described (Knipscheer *et al.*, 2009) with minor modifications. The lysine (K257) in the consensus SUMO1 motif of Pax5 protein was mutated to an arginine by PCR-mutagenesis with overlapping primers that introduced a point mutation in the codon for K257.

N-terminal and C-terminal sequences of Pax5-coding cDNA were amplified with Pfu polymerase (Pfu, Stratagene) from a CMV expression vector (Eberhard *et al.*, 1999). The primers had restriction-site overhangs to allow cloning into a CMV expression vector: (1) the 5' oligo of the N-terminus coding sequence Sall, (2) the 3' oligo of the C-terminus coding sequence FLAG-EcoRI. The 3' oligo of the N-terminus coding sequence and the 5' oligo of the C-terminus coding sequence overlapped and contained the point. The sizes of the PCR products were: (1) the N-terminus coding 833 bp, (2) the C-terminus coding 460 bp. They were mixed in equal molar ratios (N-terminus 100 ng, C-terminus 55 ng) giving rise to an annealed molecule, which was extended by Taq polymerase in 5 PCR cycles (95°C 2 min, 95°C 30s, 72°C 2 min 5x, 72°C 2 minutes). The enriched mutant Pax5-K257R cDNA was subsequently amplified with 5'Sal-NT and 3'CT-FLAG-EcoRI primers in 35 PCR cycles (annealing at 50°C and extension at 72°C for 2 minutes). Finally, the Pax5-K257R PCR product was gel purified, digested with EcoRI and Sall restriction enzymes and cloned into a CMV expression vector.

Domain	Overhang	Start domain	Primer sequence	F/R	T _a /°C	t _{ex} /s
NT-PD-SUMO-motif	Sall	NT	GCGTCGACGAACTTTTCCCTGTCC	F	49	90
		KR_AAG-AGG	GTCTGCTCGGGCCTGATGGGCTCTGT	R		
SUMO-motif-TD-CT	EcoRI	KR_AAG-AGG	ACAGAGCCCATCAGGCCCGAGCAGAC	F	49	150
		CT	GCAGAATTCTCAGTGACGCTTGTC	R		

Table 19. Primer combinations used for the generation of Pax5-K257R mutant.
The names of the mutants refer to the domains that were retained.

HEK 293T cells were grown to 60% confluence in 10 mm plates and transfected with 1 µg of Pax5-FLAG and 6 µg SUMO1ΔC-6xHis or control empty vectors, by using 21 µl Fugene (Roche). After 48 hours, the cells were harvested, washed twice in PBS and 1/10 was taken to analyse the expression levels of the transiently expressed proteins. The remaining cells were lysed in 5 ml of a buffer consisting of 6 M Guanidinium-HCl, 10 mM Tris pH 8.0, 100 mM sodium phosphate pH 8.0, 5 mM β-mercaptoethanol, 5mM imidazole and 20 mM N-ethylmaleimide (Pierce [23030]; SUMO-protease inhibitor), gently mixed for 15 minutes at room temperature, sonicated and centrifuged at 4000 x g for 1 hour. The lysates were utilised for Ni²⁺ affinity purification (Ni-NTA) with 100 µl of Ni²⁺NTA sepharose beads (Qiagen [30210]) at 4°C over night. The beads were washed with 5 ml of a buffer containing 8 M Urea, 10 mM Tris pH 8, 100 mM sodium phosphate pH 8.0, 0.1% (vol/vol) Triton X-100, 5 mM β-mercaptoethanol with increasing concentration of imidazole (5, 10 and 20 mM) in each washing step. The proteins were affinity eluted by incubating in SDS-sample buffer supplemented with 200 mM imidazole for 20 minutes at room temperature. The input and purified proteins were fractioned on SDS-PAGE and analysed by Western blotting with anti-Pax5, anti-SUMO1, anti-6xHis and anti-RanGAP1 antibodies.

Antibodies

The rabbit polyclonal anti-Pax5 antibody used was directed against the paired domain of Pax5 (amino acid residues 17-145) (Adams *et al.*, 1992). The following antibodies were used for immunoblot analyses: anti-Brg1 (mouse mAb clone G-7, Santa Cruz [sc-17796X]), anti-BAF170 (affinity-purified rabbit Ab, G. Crabtree), anti-BAF57 (unfractionated rabbit serum, G. Crabtree), anti-TBP (mouse mAb clone TBP-3G3 from L. Tora), anti-TAF4 (mouse mAb clone TAF4-2B9 from L. Tora), anti-TAF6 (mouse mAb clone TAF6-2G7 from L. Tora), anti-Pax5 (rat mAb clone 1H9 from S. Nutt), anti-CBP (affinity-purified rabbit Ab, Santa Cruz [A-22, sc-369X]), anti-PTIP (affinity-purified rabbit Ab from K. Ge), anti-RbBP5 (affinity-purified rabbit Ab, Bethyl Laboratories [A300-109A]), anti NCoR (polyclonal rabbit serum, Abcam [ab24552]), anti-CTCF (affinity purified rabbit Ab, Biomedica [07-729]), anti-V5 (monoclonal mouse Ab from D. Meijer), anti-SUMO1 (monoclonal mouse Ab, Invitrogen [33-2400]), anti-SUMO1 (sheep affinity purified Ab from R. Hay), anti 6xHis (mouse mAb, Santa Cruz Biotechnology [sc-8036]), anti-RanGAP1 (19C7, Zymed), anti-NONO (affinity purified rabbit Ab, Sigma [N8789]).

Electrophoretic mobility shift assays (EMSA)

Nuclear extracts were prepared from wild-type pro-B or 70Z/3 pre-B cells as described (Decker *et al.*, 2009). Double-stranded oligonucleotides corresponding to the Pax5- or E2A-binding sites of PAIR7 were end-labeled by Klenow polymerase. For analysis of Pax5 binding, the labeled probe (2 fmoles) was incubated with 1.5 µg of nuclear extract in 20 µl of binding buffer (10 mM HEPES pH 7.9, 100 mM KCl, 1 mM DTT, 4% Ficoll, 250 µg/ml BSA, 0.1 µg/µl poly[d(I-C)]) for 30 min on ice. E2A binding was analyzed with 2.5 µg of nuclear extract in a modified binding buffer (10 mM HEPES pH 7.9, 50 mM KCl, 1.25 mM MgCl₂, 0.1 mM EDTA, 0.5 mM DTT, 4% Ficoll, 100 µg/ml BSA, 0.025 µg/µl poly[d(I-C)]). Double-stranded oligonucleotides containing the E2A and Pax5-binding sites of other PAIR elements were analysed at 10-, 30, 100-fold molar excess for competition of protein binding (Table). Recognition sequences of the human *CD19* promoter (Kozmik *et al.*, 1992) or the mouse *Igh* Eµ (µE5) and *Igk* iEκ (κE2) enhancers (Murre *et al.*, 1989) served as reference high affinity binding sites for Pax5 and E2A, respectively (Table).

Protein-DNA complexes were separated on a 6% polyacrylamide gel in 0.25x TBE buffer (22 mM Tris-borate pH 8.3, 0.5 mM EDTA) at 250 V for 2 h at 25°C. Gels were dried and subjected to autoradiography.

Locus	Oligo sequence 5'-3'
PAIR7	tcgacTATATTGTATGGTCACTGAAGTGTGTGATATTTGGAAG tcgacTTCCAAATATCACACACTTCAGTGACCATACAATATAg
CD19	tcgacCCCCGCAGACACCCATGGTTGAGTGCCCTCCAGGCCCGg tcgacGGGCCTGGAGGGCACTCAACCATGGGTGTCTGCGGGGg
PAIR8	tcgacTGTATTGTATGGTCACTGAAGTTGTGTGTTATATGGAAG tcgacTTCCATATACCACACAACCTTCAGTGACCATACAATACAg
PAIR7M	tcgacTATATTGTATGGAGTCTGAAGTGTGTGATATTTGGAAG tcgacTTCCAAATATCACACACTTCAGACTCCATACAATATAg
PAIR10	tcgacTGTATTGTATGGTCACTGAAGTGTGTGATATTTGGAAG tcgacTTCCAAATATCACACACTTCAGTGACCATACAATACAg
PAIR5	tcgacAGCTGTAGTAGGTACAGAGAATg tcgacATTCTCTGTACCTACTACAGCTg

Table 20. Oligonucleotides with Pax5 recognition sequence.

The recognition motif is in bold while the mutated residues are in red.

Locus	Oligo sequence 5'-3'
PAIR7	tcgacAACAATAACAGGTGCAGAGAAAg tcgacTTTCTCTGCACTGTTATTGTTg
PAIR5	tcgacAGCTGTAGTAGGTACAGAGAATg tcgacATTCTCTGTACCTACTACAGCTg
mE5	tcgacAGCAGGAGCAGGTGTTCTCTAGg tcgacCTAGAGAACACCTGCTCCTGCTg
κE2	tcgacCTCCCAGGCAGGTGCCCCAGATg tcgacATCTGGGCCACCTGCCTGGGAGg

Table 21. Oligonucleotides with E2A recognition sequence.

The recognition motif is in bold while the mutated residues are in red.

Isolation of Pro-B cells for 4C analysis

Bone marrow was isolated from *Rag2^{-/-}* or *Rag2^{-/-} Pax5^{-/-}* mice, of the C57BL6/J background, and stained with B220 antibody magnetic beads followed by magnetic cell sorting. The cells were cultured on OP9 (JCZP) cells for 2 days in standard pro-B cell conditions (Nutt *et al.*, 1997) IMDM, 2% FCS, primaton, IL-7, peniclin, streptomycin, ciproxin, β -mercaptoethanol.

For establishing a culture of mouse embryonic fibroblasts (MEFs) embryos were isolated at day 14 from mice of the C57BL6/J background and washed in cold PBS. Inner organs and the head were removed. The remaining tissue was cut in pieces and digested with trypsin (Sigma) while incubating at 37°C for a few minutes. The cells were washed, centrifuged and resuspended in the feeder medium (DMEM, 10% FCS, peniclin, streptomycin, ciproxin). After 3 days in culture the cells were harvested.

Circular Chromosome Conformations Capture (4C) method

Digestions with restriction enzymes HindIII and DpnII were done at 37°C over night followed by further 8 hours over the next day. The ligation reactions were done at 16°C over night followed by room temperature for 1 hour. The total amounts of restriction enzymes (700 U HindIII, 80 U DpnII) were added sequentially: 300 U HindIII (Roche [10798983001]) or 40 U DpnII (10709751001, Roche) over night, twice 200 U of HindIII or 20 U DpnII on the following day. The efficiency of every enzyme reaction in the procedure was monitored by analysing 1% of the products on by agarose gel electrophoresis, before proceeding with protocol. For this purpose, the DNA was extracted with phenol/chloroform/isoamyl alcohol (Sigma) and precipitated with ethanol.

The DNA template for chromosome conformations capture analysis was prepared as previously described (Simonis *et al.*, 2007) with minor changes. Ten million cells (*Rag2^{-/-}*, *Rag2^{-/-} Pax5^{-/-}* pro-B or MEF) were harvested, centrifuged at 1500 rpm for 5 minutes and fixed in PBS supplemented with 2% formaldehyde and 10% FCS, for 10 minutes, while tumbling. Crosslinking was quenched on ice by adding 0.125 M Glycine. The cells were washed twice with PBS containing 0.125 M Glycine and lysed in 5 ml of buffer consisting of 10 mM Tris-HCl pH 7.5, 10 mM NaCl, 5 mM MgCl₂, 0.1 mM EGTA, complete protease inhibitor cocktail). After incubation on ice for 1.5 hour, the cells were homogenised with 15 strokes of Dounce homogenizer and centrifuged at 400 x g for 5 minutes. Nuclei were washed once in 0.5 ml of 1.2x HindIII restriction buffer B, then incubated in 0.5 ml of buffer B supplemented with 0.3 % SDS, at 37°C. After 2 hours up to 2% Triton X-100 was added to sequester SDS while shaking for 2 hours at 37 °C. The nuclei were then digested with HindIII. The digestion efficiency was analysed on a

0.6% agarose gel. The enzyme was inactivated at 65°C for 10 min. The digested DNA was ligated with 100 U of T4 DNA ligase (Roche [10716359001]) in 6.5 ml. The ligated chromatin was de-crosslinked over night at 65°C in the presence of 300 mg of proteinase K. The decrosslinked DNA, was digested with 80 U of DpnII and ligated with 200 U of T4 DNA ligase in 15 ml. This final reaction resulted in the 4C library. To remove contaminating buffer components that interfere with optical density (OD260) measurements, the 4C library samples were purified using the QIAquick nucleotide removal column (Qiagen). The quality of the preparation was analysed by PCR with primers that give a reproducible pattern of products when run on a 1.5% agarose gel. The PCR products should appear as a smear with two prominent bands: one representing the 10-20% of the investigated restriction site that was not digested (incompletely digested) and the other representing the events where the investigated fragment was ligated to its own end (self-ligated). Lighter bands (less abundant products) should give a reproducible pattern in replicate PCR reactions. A linear increase in the amount of 4C-DNA library used for the PCR, should result in a linear increase of products. The highest amount of 4C DNA, for which the reaction is still in the linear range, allowing a maximum of 200 ng, is used for 16 PCR amplifications with primers that have deep-sequencing adapters. These are pooled, purified (Qiagen column) and sequenced.

PCR parameters are as follows:

1 cycle 94 °C for 2 minutes, followed by
30 cycles of: 94 °C for 15 seconds,
55 °C for 1 minute,
68 °C for 3 minutes, followed by
1 cycle of 68 °C for 7 minutes.

Primer design for 4C in the mouse *Igh* locus

Primers were designed within a HindIII-DpnII fragment that contains the site of interest. The recommended length of the fragment is 300 to 5000 base pairs. Fragments shorter than 300 base pairs circulate and ligate inefficiently, therefore they are detected less successfully. The HindIII-site primer (an example is given below) overlaps with the restriction site (highlighted in bold) then extends 18-21 base pairs into the DpnII-HindIII fragment. This primer has the deep sequencing adaptor (in blue), which allows direct deep sequencing of the 4C-PCR amplification products. The base pairs upstream the HindIII-recognition site contain the site of interest unique for each “anchor fragment”. Therefore they serve as a “bar code” for identification of the

respective fragment. This allows multiplexing PCR amplifications by different primers done on a 4C library from the same cell type. The multiplexed sample can then be sequenced in one lane of the Solexa Genome Analyzer. The primer close to the DpnII end of the anchor fragment can be located anywhere up to 300 base pairs from the restriction site. The extension on its 5' end (in blue) is important for the deep sequencing reaction. Primers were inspected for uniqueness in the mouse chromosome 12.

HindIII site-primer:

5'**AATGATACGGCGACCACCGAACACTCTTTCCTACACGACGCTCTTCCGATCT**TTGTAGTGGAACAA
TACTCT**AAGCTT** 3'

DpnII primer:

5' **CAAGCAGAAGACGGCATACG**ACTTAGAGAGAATAAGATTTGTCTAATGT 3'

In order to investigate the topological organisation of the *Igh* locus by 4C method we designed primers anchored at six different positions throughout the locus; two were located in the 3' end, one in the proximal V_H region and three in the distal V_H region, within the PAIR elements.

Solexa sequencing

4C libraries prepared from *Rag2*^{-/-}, *Rag2*^{-/-} *Pax5*^{-/-} or MEF cells were amplified separately with different “anchor” primers. Products of 16 replicate PCR amplifications with the same primer were pooled together. Only a small aliquot of the pooled sample was utilised for single-end deep sequencing with 76-nucleotide read length. Those aliquots that were amplified with different primers, but from a library of the same cell type, were mixed in equal ratios and sequenced in the same lane of the Solexa Genome Analyzer. In such multiplexed samples, the primer’s “bar code” ensures specific identification of all sequences ligated to that anchor fragment. The sequences were mapped to the mouse genome.

Bioinformatic analysis

The reads that were mapped to the HindIII fragments flanking the anchor fragment were subtracted from the total counts because they originated from the experimental artifacts, the incompletely digested and the self-ligated fragments. Then we divided the number of mapped reads, for each anchor fragment in different 4C libraries to three genomic compartments: (1) the genome, (2) the genome excluding the chromosome 12 and (3) the chromosome 12 excluding the *Igh* locus.

The PCR amplification products from the same 4C library could not be always multiplexed in the same ratio. This resulted in differences in the number of mapped reads between samples. Therefore we divided reads mapped to the *Igh* locus by the number of reads mapped to the compartment we wanted to normalize to and multiplied the values by million. This kind of normalization set all the values, in the compartment we normalised to, to one as depicted by grey shading in the table below.

Reads/10 ⁶	Cell type	genome - chr12	chr12(- <i>Igh</i>)	<i>Igh</i>	Sum
raw	<i>Rag2</i> ^{-/-}	5	5	10	20
raw	<i>Pax5</i> ^{-/-}	5	5	10	20
raw	MEF	5	10	5	20
genome	<i>Rag2</i> ^{-/-}	0.25	0.25	0.5	1
genome	<i>Pax5</i> ^{-/-}	0.25	0.25	0.5	1
genome	MEF	0.25	0.5	0.25	1
genome - chr12	<i>Rag2</i> ^{-/-}	1	1	2	4
genome - chr12	<i>Pax5</i> ^{-/-}	1	1	2	4
genome - chr12	MEF	1	2	1	4
chr12 - <i>Igh</i>	<i>Rag2</i> ^{-/-}	1	1	2	4
chr12 - <i>Igh</i>	<i>Pax5</i> ^{-/-}	1	1	2	4
chr12 - <i>Igh</i>	MEF	0.5	1	0.5	2

Table 22. Scheme of normalisations.

Since the *Igh* locus is subdivided in functional domains, we assumed that some interactions might involve entire regions, rather than single HindIII fragments. Moreover, some fragments might be inaccessible to the restrictions enzyme because of occupancy by a transcription factor, which could result in mapping of an interaction to the neighboring fragment. Because of these reasons, we calculated the running mean of raw and normalised read counts for all HindIII fragments within a 20 kb-window across the *Igh* locus.

REFERENCES

- Adams, B., Dorfler, P., Aguzzi, A., Kozmik, Z., Urbanek, P., Maurer-Fogy, I., and Busslinger, M. (1992). Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev* 6, 1589-1607.
- Afshar, R., Pierce, S., Bolland, D. J., Corcoran, A., and Oltz, E. M. (2006). Regulation of IgH gene assembly: role of the intronic enhancer and 5'DQ52 region in targeting DHJH recombination. *J Immunol* 176, 2439-2447.
- Akashi, K. (2009). Lymphoid lineage fate decision of hematopoietic stem cells. *Ann N Y Acad Sci* 1176, 18-25.
- Alkuraya, F. S., Saadi, I., Lund, J. J., Turbe-Doan, A., Morton, C. C., and Maas, R. L. (2006). SUMO1 haploinsufficiency leads to cleft lip and palate. *Science* 313, 1751.
- Atchison, L., Ghias, A., Wilkinson, F., Bonini, N., and Atchison, M. L. (2003). Transcription factor YY1 functions as a PcG protein in vivo. *EMBO J* 17, 1347-1358.
- Bain, G., Maandag, E. C. R., Izon, D. J., Amsen, D., Kruisbeek, A. M., Weintraub, B. C., Krop I., Schlissel, M. S., Feeney, A. J., and van Roon, M. (1994). E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. *Cell* 79, 885–892.
- Barberis, A., Widenhorn, K., Vitelli, L., and Busslinger, M. (1990). A novel B-cell lineage-specific transcription factor present at early but not late stages of differentiation. *Genes Dev* 4, 849-859.
- Barlev, N. A., Emelyanov, A.V., Castagnino, P., Zegerman, P., Bannister, A. J., Sepulveda, M. A., Robert, F., Tora, L., Kouzarides, T., and Birshstein, B. K. (2003). A novel human Ada2 homologue functions with Gcn5 or Brg1 to coactivate transcription. *Mol Cell Biol* 23, 6944-6957.
- 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* 28, 2271-2284.
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., and Plath, K. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315-326.
- Bischoff, F. R., Klebe, C., Kretschmer, J., Wittinghofer, A., and Ponstingl, H. (1994). RanGAP1 induces GTPase activity of nuclear Ras-related Ran. *Proc Natl Acad Sci U S A* 91, 2587-2591.
- Bolland, D. J., Wood, A. L., Johnston, C. M., Bunting, S. F., Morgan, G., Chakalova, L.,

- Fraser, P. J., and Corcoran, A. E. (2004). Antisense intergenic transcription in V(D)J recombination. *Nat Immunol* 5, 630-637.
- Bouchard, M., Schleiffer, A., Eisenhaber, F., and Busslinger, M. (2003). Evolution and Function of Pax Genes. *Encyclopedia of the human genome*.
 - Bryder, D. and Sigvardsson, M. (2010). Shaping up a lineage--lessons from B lymphopoiesis. *Curr Opin Immunol* 22, 48-53.
 - Caretti, G., Di Padova, M., Micales, B., Lyons, G. E., and Sartorelli V. (2004). The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. *Genes Dev* 18, 2627-2638.
 - Chakraborty, T., Chowdhury, D., Keyes, A., Jani, A., Subrahmanyam, R., Ivanova, I., and Sen, R. (2007). Repeat organization and epigenetic regulation of the DH-Cmu domain of the immunoglobulin heavy-chain gene locus. *Mol Cell* 7, 842-850.
 - Chakraborty, T., Perlot, T., Subrahmanyam, R., Jani, A., Goff, P. H., Zhang, Y., Ivanova, I., Alt, F. W., and Sen, R. (2009). A 220-nucleotide deletion of the intronic enhancer reveals an epigenetic hierarchy in immunoglobulin heavy chain locus activation. *J Exp Med* 206, 1019-1027.
 - Chen, L. L., and Carmichael, G. G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol Cell* 35, 467-478.
 - Chi T. (2004). A BAF-centred view of the immune system. *Nat Rev Immunol*, 12, 965-977.
 - Cho, Y.-W., Hong, T., Hong, S., Guo, H., Yu, H., Kim, D., Guszczynski, T., Dressler, G.R., Copeland, T.D., and Kalkum, M. (2007). PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem* 282, 20395-20406.
 - Choi, N. M., Majumder, P., and Boss, J. M. (2011). Regulation of major histocompatibility complex class II genes. *Curr Opin Immunol* 23, 81-87
 - Chowdhury D, and Sen R. (2003). Transient IL-7/IL-7R signaling provides a mechanism for feedback inhibition of immunoglobulin heavy chain gene rearrangements. *Immunity* 18, 229-241.
 - Cobaleda, C.^b, Schebesta, A., Delogu, A., and Busslinger, M. (2007b). Pax5: the guardian of B cell identity and function. *Nat Immunol* 8, 463-470.
 - Cobaleda, C.^a, Jochum, W., and Busslinger, M. (2007a). Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* 449, 473-477.

- Cuddapah, S., Jothi, R., Schones, D. E., Roh, T. Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* 19, 24-32.
- Czerny, T., Schaffner, G., and Busslinger, M.. (1993). DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev* 10, 2048-2061.
- Degner-Leisso, S. C, and Feeney, A. J. (2010). Epigenetic and 3-dimensional regulation of V(D)J rearrangement of immunoglobulin genes. *Semin Immunol* 6, 346-352.
- 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.
- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311.
- DeKoter, R. P., Singh, H., (2000). Regulation of B lymphocyte and macrophage development by graded expression of PU.1. *Science* 288, 1439-1441.
- Delogu, A., Schebesta, A., Sun, Q., Aschenbrenner, K., Perlot, T., and Busslinger, M. (2006). Gene repression by Pax5 in B cells is essential for blood cell homeostasis and is reversed in plasma cells. *Immunity* 24, 269-281.
- Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* 11, 1475-1489.
- Dong, X., Yu, C., Shynlova, O., Challis, J. R., Rennie, P. S, and Lye, S. J. (2009). p54nrb is a transcriptional corepressor of the progesterone receptor that modulates transcription of the labor-associated gene, connexin 43 (Gja1). *Mol Endocrinol* 8, 1147-60.
- Donohoe, M. E., Silva, S. S., Pinter, S. F., Xu, N., and Lee, J. T. (2009). The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* 7251, 128-312.
- Dörfler P, and Busslinger, M. (1996). C-terminal activating and inhibitory domains determine the transactivation potential of BSAP (Pax-5), Pax-2 and Pax-8. *EMBO J* 15, 1971-1982.
- Dyer, R. B., and Herzog, N. K. (1995). Isolation of Intact Nuclei for Nuclear Extract Preparation from Fragile B-Lymphocyte Cell Lines. *Biotechniques* 19, 192-195.
- Dykstra, B., Kent, D., Bowie, M., McCaffrey, L., Hamilton, M., Lyons, K., Lee, S. J., Brinkman, R., and Eaves, C. (2007). Long-term propagation of distinct hematopoietic

differentiation programs in vivo. *Cell Stem Cell* 1(2), 218-229.

- Eberhard, D., and Busslinger, M. (1999). The partial homeodomain of the transcription factor Pax-5 (BSAP) is an interaction motif for the retinoblastoma and TATA-binding proteins. *Cancer Res* 59, 1716s-1724s; discussion 1724s-1725s.
- Eberhard, D., Jimenez, G., Heavey, B., and Busslinger, M. (2000). Transcriptional repression by Pax5 (BSAP) through interaction with corepressors of the Groucho family. *Embo J* 19, 2292-2303.
- 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.
- Emelyanov, A.V., Kovac, C. R., Sepulveda, M. A., and Birshtein, B. K. (2002). The interaction of Pax5 (BSAP) with Daxx can result in transcriptional activation in B cells. *J Biol Chem* 277, 11156-11164.
- 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* 13, 9327-9338.
- Fisher, A. G., and Grosschedl, R. (2007). Distinct promoters mediate the regulation of Ebf1 gene expression by interleukin-7 and Pax5. *Mol Cell Biol* 27, 579-94.
- 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.
- Gareau, J. R., and Lima, C. D. (2010). The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol* 12, 861-871.
- Garrett, F. E., Emelyanov, A. V., Sepulveda, M. A., Flanagan, P., Volpi, S., Li, F., Loukinov, D., Eckhardt, L.A., Lobanenko, V. V., and Birshtein, B. K. (2005). Chromatin architecture near a potential 3' end of the Igh locus involves modular regulation of histone modifications during B-cell development and in vivo occupancy at CTCF sites. *Mol. Cell Biol* 25, 1511-1525.
- Garvie, C. W., Hagman, J., and Wolberger, C. (2001). Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol Cell*, 6, 1267-1276.
- Geiss-Friedlander, R., and Melchior, F. (2007). Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol* 12, 947-956.

- Georgopoulos, K., Moore, D. D., and Derfler, B. (1992). Ikaros: an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. *Science* 258, 808–812.
- Giallourakis, C. C., Franklin, A., Guo, C., Cheng, H. L., Yoon, H. S., Gallagher, M., Perlot, T., Andzelm, M., Murphy, A. J., Macdonald, L. E., Yancopoulos, G.D., and Alt, F. W. (2010). Elements between the IgH variable (V) and diversity (D) clusters influence antisense transcription and lineage-specific V(D)J recombination. *Proc Natl Acad Sci U S A* 107, 22207-22212.
- Gill, G. (2004). SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? *Genes Dev* 17, 2046-2059.
- Girdwood, D. W., Tatham, M. H. and Hay, R. T. (2004). SUMO and transcriptional regulation. *Semin. Cell Dev. Biol.* 15, 201–210.
- Göndör, A., Rougier, C., and Ohlsson, R. (2008). High-resolution circular chromosome conformation capture assay. *Nat Protoc* 2, 303-213.
- Gordon, S. J., Saleque, S., and Birshtein, B. K. (2003). Yin Yang 1 is a lipopolysaccharide-inducible activator of the murine 3' Igh enhancer, hs3. *J Immunol* 171, 5549-5557.
- Grégoire, S., Tremblay, A. M., Xiao, L., Yang, Q., Ma, K., Nie, J., Mao, Z., Wu, Z., Giguère, V., and Yang, X. J. (2006). Control of MEF2 transcriptional activity by coordinated phosphorylation and sumoylation. *J Biol Chem* 281, 4423-4433.
- Hadjur, S., Williams, L. M., Ryan, N. K., Cobb, B. S., Sexton, T., Fraser, P., Fisher, A. G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 460, 410-3.
- Hagman, J., Belanger, C., Travis, A., Turck, C. W., and Grosschedl, R. (1993). Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. *Genes Dev* 5, 760-773.
- 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.
- Hay, R. T. (2005). SUMO: a history of modification. *Mol Cell* 1, 1-12.
- He, T., Hong, S.Y., Huang, L., Xue, W., Yu, Z., Kwon, H., Kirk, M., Ding, S. J., Su, K., and Zhang, Z. (2011). Histone Acetyltransferase p300 Acetylates Pax5 and Strongly Enhances Pax5-mediated Transcriptional Activity. *J Biol Chem* 286, 14137-14145.
- Heavey, B., Charalambous, C., Cobaleda, C., and Busslinger, M. (2003). Myeloid lineage switch of Pax5 mutant but not wild-type B cell progenitors by C/EBP α and GATA factors. *EMBO J* 22, 3887-3897.

- Herzog, S., Hug, E., Meixlsperger, S., Paik, J. H., DePinho, R. A., Reth, M., and Jumaa, H. (2008). SLP-65 regulates immunoglobulin light chain gene recombination through the PI(3)K-PKB-Foxo pathway. *Nat Immunol* 6, 623-631.
- Herzog, S., Reth, M., and Jumaa, H. (2009). Regulation of B-cell proliferation and differentiation by pre-B-cell receptor signalling. *Nat Rev Immunol* 3, 195-205.
- Hesslein, D. G., Pflugh, D. L., Chowdhury, D., Bothwell, A. L., Sen, R., and Schatz, D. G. (2003). Pax5 is required for recombination of transcribed, acetylated, 5' IgH V gene segments. *Genes Dev* 1, 37-42.
- Hietakangas, V., Anckar, J., Blomster, H. A., Fujimoto M., Palvimo, J. J., Nakai, A., and Sistonen, L. (2006). PDSM, a motif for phosphorylation-dependent SUMO modification. *Proc Natl Acad Sci U S A.* 103, 45-50.
- Ho, L., and Crabtree, G. R. (2010). Chromatin remodelling during development. *Nature* 7280, 474-484.
- Höflinger, S., Kesavan, K., Fuxa, M., Hutter, C., Heavey, B., Radtke, F., and Busslinger M. (2004). Analysis of Notch1 function by in vitro T cell differentiation of Pax5 mutant lymphoid progenitors. *J Immunol*, 6, 3935-3944.
- Holmes, M. L., Pridans, C., and Nutt, S. L. (2008). The regulation of the B-cell gene expression programme by Pax5. *Immunol Cell Biol* 1, 47-53
- Horcher, M., Souabni, A., and Busslinger, M. (2001). Pax5/BSAP maintains the identity of B cells in late B lymphopoiesis. *Immunity* 6, 779-790.
- Hsu, L. Y., Lanning, J., Liang, H. E, Greenbaum, S., Cado, D., Zhuang, Y., Schlissel, M. S. (2003). A conserved transcriptional enhancer regulates RAG gene expression in developing B cells. *Immunity* 1, 105-117.
- Ilsley, J. L., Sudol, M., and Winder, S. J. (2002). The WW domain: linking cell signalling to the membrane cytoskeleton. *Cell Signal* 14, 183-9.
- Ingham, R. J., Colwill, K., Howard, C., Dettwiler, S., Lim, C. S., Yu, J., Hersi, K., Raaijmakers, J., Gish, G., Mbamalu, G., Taylor, L., Yeung, B., Vassilovski, G., Amin, M., Chen, F., Matskova, L., Winberg, G., Ernberg, I., Linding, R., O'donnell, P., Starostine, A., Keller, W., Metalnikov, P., Stark, C., and Pawson, T. (2005). WW domains provide a platform for the assembly of multiprotein networks. *Mol Cell Biol* 16, 7092-7106.
- Inlay, M. A., Bhattacharya, D., Sahoo, D., Serwold, T., Seita, J., Karsunky, H., Plevritis, S.K., Dill, D.L., and Weissman, I.L. (2009). Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. *Genes Dev* 23, 2376-2381.

- Ishihara, K., Oshimura, M., and Nakao M. (2006). CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell* 23, 733-742.
- Ivanov, G. S., Ivanova, T., Kurash, J., Ivanov, A., Chuikov, S., Gizatullin, F., Herrera-Medina, E. M., Rauscher, F. 3rd, Reinberg, D., and Barlev, N. A. (2007). Methylation-acetylation interplay activates p53 in response to DNA damage. *Mol Cell Biol* 27, 6756-6769.
- Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 33, Suppl:245-54.
- Jenuwein, T, and Allis, C. D. (2001). Translating the histone code. *Science* 5532, 1074-80.
- Jhunjunwala, 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.
- Jhunjunwala, 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.
- 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* 3, 419-431.
- Johnson, K., Pflugh, D. L., Yu, D., Hesslein, D. G., Lin, K. I., Bothwell, A. L., Thomas-Tikhonenko, A., Schatz, D. G., and Calame, K. (2004). B cell-specific loss of histone 3 lysine 9 methylation in the V(H) locus depends on Pax5. *Nat Immunol* 8, 853-861.
- Johnston, C. M., Wood, A. L., Bolland, D. J., and Corcoran, A.E. (2006). Complete sequence assembly and characterization of the C57BL/6 mouse Ig heavy chain V region. *J. Immunol* 176, 4221-4234.
- Ju, Z., Volpi, S. A., Hassan, R., Martinez, N., Giannini, S. L., Gold, T., and Birshtein, B. K. (2007). Evidence for physical interaction between the immunoglobulin heavy chain variable region and the 3' regulatory region. *J Biol Chem* 48, 35169-3578.
- Kallies, A., Hasbold, J., Fairfax, K., Pridans, C., Emslie, D., McKenzie, B. S., Lew, A. M, Corcoran, L. M., Hodgkin, P. D, Tarlinton, D. M, and Nutt, S. L. (2007). Initiation of plasma-cell differentiation is independent of the transcription factor Blimp-1. *Immunity* 5, 555-566.
- Kee, B. L. (2009). E and ID proteins branch out. *Nat Rev Immunol* 3, 175-184.
- Kim J, Sif S, Jones B, Jackson A, Koipally J, Heller E, Winandy S, Viel A, Sawyer A, Ikeda, T., Kingston, R., and Georgopoulos, K. (1999). Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes. *Immunity* 3, 345-355.
- Kirstetter, P., Thomas, M., Dierich, A., Kastner, P., and Chan, S. (2002). Ikaros is critical

for B cell differentiation and function. *Eur J Immunol* 3, 720-730.

- Kondo, M., Weissman, I. L., and Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* 5, 661-672.
- Knipscheer P, Klug H, Sixma TK, Pichler A. (2009). Preparation of sumoylated substrates for biochemical analysis. *Methods Mol Biol* 497, 201-210.
- 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.
- Kowenz-Leutz, E., Pless, O., Dittmar, G., Knoblich, M., and Leutz, A. (2010). Crosstalk between C/EBPbeta phosphorylation, arginine methylation, and SWI/SNF/Mediator implies an indexing transcription factor code. *EMBO J* 6, 1105-1115.
- Kowenz-Leutz, E., Twamley, G., Ansieau, S., and Leutz, A. (1994). Novel mechanism of C/EBP beta (NF-M) transcriptional control: activation through derepression. *Genes Dev* 22, 2781-2791.
- Kozmik, Z., Wang, S., Dörfler, P., Adams, B., and Busslinger, M. (1992). The promoter of the CD19 gene is a target for the B-cell-specific transcription factor BSAP. *Mol. Cell. Biol.* 12, 2662-2672.
- Kuo, H. Y., Chang, C. C, Jeng, J. C., Hu, H. M., Lin, D. Y., Maul, G. G., Kwok, R. P., and Shih, H. M. (2005). SUMO modification negatively modulates the transcriptional activity of CREB-binding protein via the recruitment of Daxx. *Proc Natl Acad Sci U S A* 47, 16973-16978.
- Kuo, T. C., Chavarria-Smith, J. E., Huang, D., and Schlissel, M. S. (2011). Forced expression of CDK6 confers resistance of pro-B ALL to Gleevec treatment. *Mol Cell Biol*
- Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenko, V., Reik, W., and Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A*. 103, 10684-10689.
- Kwon, K., Hutter, C., Sun, Q., Bilic, I., Cobaleda, C., Malin, S., and Busslinger M. (2008). Instructive role of the transcription factor E2A in early B lymphopoiesis and germinal center B cell development. *Immunity* 6, 751-762.
- Lee, H. Y., Johnson, K. D., Fujiwara, T., Boyer, M. E., Kim, S. I., and Bresnick, E. H. (2009). Controlling hematopoiesis through sumoylation-dependent regulation of a GATA factor. *Mol Cell* 6, 984-95.
- Leutz, A., Pless, O., Lappe, M., Dittmar, G., and Kowenz-Leutz, E. (2010). Crosstalk

between phosphorylation and multi-site arginine/lysine methylation in C/EBPs.

Transcription 1, 3-8.

- Lin, D. Y., Huang, Y. S., Jeng, J. C., Kuo, H. Y., Chang, C. C., Chao, T. T., Ho, C. C., Chen, Y. C., Lin, T. P., Fang, H. I., Hung, C. C., Suen, C. S., Hwang, M. J., Chang, K. S., Maul, G. G., and Shih, H. M. (2006). Role of SUMO-interacting motif in Daxx SUMO modification, subnuclear localization, and repression of sumoylated transcription factors. *Mol Cell* 3, 341-354.
- Lin, K. I., Angelin-Duclos, C., Kuo, T. C., and Calame, K. (2002). Blimp-1-dependent repression of Pax-5 is required for differentiation of B cells to immunoglobulin M-secreting plasma cells. *Mol Cell Biol* 22, 4771-80.
- Lin, Y.C., Jhunjhunwala, S., Benner, C., Heinz, S., Welinder, E., Mansson, R., Sigvardsson, M., Hagman, J., Espinoza, C.A., and Dutkowski, J. (2010). A global network of transcription 4/11/10-24-factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat Immunol* 11, 635-643.
- Linderson, Y., Eberhard, D., Malin, S., Johansson, A., Busslinger, M., and Pettersson, S. (2004). Corecruitment of the Grg4 repressor by PU.1 is critical for Pax5-mediated repression of B-cell-specific genes. *EMBO Rep* 5, 291-296.
- Liu, H., Schmidt-Suppran, M., Shi, Y., Hobeika, E., Barteneva, N., Jumaa, H., Pelanda, R., Reth, M., Skok, J., Rajewsky, K., and Shi, Y. (2007). Yin Yang 1 is a critical regulator of B-cell development. *Genes Dev* 21, 1179-1189.
- Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R., and Desiderio, S. (2007). A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary *Genes Dev* 21, 1179-1189.
- Macias, M. J., Wiesner, S., and Sudol, M. (2002). WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett* 513, 30-7.
- Majumder, P., Boss, J. M. (2010). CTCF controls expression and chromatin architecture of the human major histocompatibility complex class II locus. *Mol Cell Biol* 17, 4211-1223.
- Majumder, P., Gomez, J. A., Chadwick, B. P., and Boss, J. M. (2008). The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J. Exp. Med.* 205, 785-798.
- Malin, S., McManus, S., Cobaleda, C., Novatchkova, M., Delogu, A., Bouillet, P., Strasser, A., and Busslinger, M. (2009). Role of STAT5 in controlling cell survival and immunoglobulin gene recombination during pro-B cell development. *Nat Immunol* 2, 171.

- Manis, J. P., van der Stoep, N., Tian, M., Ferrini, R., Davidson, L., Bottaro, A., and Alt, F. W. (1998). Class switching in B cells lacking 3' immunoglobulin heavy chain enhancers. *J Exp Med* 8, 1421-1431.
- Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* 469, 343-349. Review.
- Massari, M. E., and Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol* 2, 429-440.
- Matthews, A. G., Kuo, A. J., Ramón-Maiques, S., Han, S., Champagne, K. S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D. N., Walter, K. L., Utz, P. J., Shi, Y., Kutateladze, T. G., Yang, W., Gozani, O., and Oettinger, M. A. (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* 13, 1106-1110.
- 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*
- Michaelson, J. S., Giannini, S. L., and Birshtein, B. K. (1996). Identification of 3' alpha-hs4, a novel Ig heavy chain enhancer element regulated at multiple stages of B cell differentiation. *Nucleic Acids Res* 6, 975-981.
- Mikkola, H. K., Klintman, J., Yang, H., Hock, H., Schlaeger, T. M., Fujiwara, Y., and Orkin, S. H. (2003). Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature* 421, 547-551.
- Mikkola, I., Heavey, B., Horcher, M., and Busslinger, M. (2002). Reversion of B cell commitment upon loss of Pax5 expression. *Science* 297, 110-113.
- Monroe, J. G., and Dorshkind, K. (2007). Fate decisions regulating bone marrow and peripheral B lymphocyte development. *Adv Immunol* 95, 1-50.
- 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* 20, 11577-11582.
- 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* 1, 31-37.
- Muñoz, I. M., and Rouse, J. (2009). Control of histone methylation and genome stability by PTIP. *EMBO* 3, 239-245.

- Murre, C., McCaw, P. S., Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* 56, 777-783.
- 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.
- Nutt, S. L., Morrison, A. M., Dorfler, P., Rolink, A., and Busslinger, M. (1998). Identification of BSAP (Pax-5) target genes in early B-cell development by loss- and gain-of-function experiments. *Embo J* 17, 2319-2333.
- Nutt, S. L., Urbanek, P., Rolink, A., and Busslinger, M. (1997). Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus. *Genes Dev* 11, 476-491.
- Osipovich, O., and Oltz, E. M. (2010). Regulation of antigen receptor gene assembly by genetic-epigenetic crosstalk. *Semin Immunol* 6, 313-322.
- Palaparti, A., Baratz, A., and Stifani, S. (1997). The Groucho/transducin-like enhancer of split transcriptional repressors interact with the genetically defined amino-terminal silencing domain of histone H3. *J Biol Chem* 272, 26604-26610.
- 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* 2, 190-194.
- Pan, C., Kumar, C., Bohl, S., Klingmueller, U., and Mann M. (2009). Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. *Mol Cell Proteomics* 3, 443-50.
- Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer S., Gregson, H. C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B. S., Yokomori, K., Dillon, N., Aragon, L., Fisher, A. G., and Merkenschlager, M. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 3, 422-433.
- Patel, S. R., Kim, D., Levitan, I., and Dressler, G. R. (2007). The BRCT-domain containing protein PTIP links PAX2 to a histone H3, lysine 4 methyltransferase complex. *Dev Cell* 13,580-592.
- Pawlitzky, I., Angeles, C. V., Siegel, A. M., Stanton, M. L., Riblet, R., Brodeur, P. H. (2006). Identification of a candidate regulatory element within the 5' flanking region of the mouse IgH locus defined by pro-B cell-specific hypersensitivity associated with binding of PU.1, Pax5, and E2A. *J Immunol* 176, 6839-51.

- Perlot, T., Alt, F. W., Bassing, C. H., Suh, H., and Pinaud, E. (2005). Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc Natl Acad Sci U S A.* 102,14362-14367.
- Perlot, T. and Alt, F.W. (2008). Cis-regulatory elements and epigenetic changes control genomic rearrangements of the *IgH* locus. *Adv. Immunol.* 99, 1-32.
- Perlot, T., Pawlitzky, I., Manis, J. P., Zarrin, A. A., Brodeur, P. H., and Alt, F. W. (2010). Analysis of mice lacking DNaseI hypersensitive sites at the 5' end of the IgH locus. *PLoS One* 11, e13992.
- Peserico, A, and Simone, C. (2011). Physical and functional HAT/HDAC interplay regulates protein acetylation balance. *J Biomed Biotechnol*, 371832.
- Pichler, A. (2007). Analysis of sumoylation. *Methods Mol Biol* 446, 131-138.
- Pichler, A., Gast, A., Seeler, J. S., Dejean, A., and Melchior, F. (2002). The nucleoporin RanBP2 has SUMO1 E3 ligase activity. *Cell* 108, 109-20.
- Pichler, A., Knipscheer, P., Saitoh, H., Sixma, T. K., and Melchior, F. (2004). The RanBP2 SUMO E3 ligase is neither HECT- nor RING-type. *Nat Struct Mol Biol* 10, 984-991.
- Ramírez, J., Lukin, K., and Hagman, J. (2010). From hematopoietic progenitors to B cells: mechanisms of lineage restriction and commitment. *Curr Opin Immunol* 2, 177-184.
- Ren, J., Gao, X., Jin, C., Zhu, M., Wang, X., Shaw, A., Wen, L., Yao, X., and Yu Xue. (2009). Systematic study of protein sumoylation: Development of a site-specific predictor of SUMOsp 2.0. *Proteomics* 9, 3409-3412.
- Reynaud, D., Demarco, I. A., Reddy, K. L., Schjerven, H., Bertolino, E., Chen, Z., Smale, S. T., Winandy, S., and Singh, H. (2008). Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. *Nat Immunol* 8, 927-936.
- Riising, E. M., Boggio, R., Chiocca, S., Helin, K., and Pasini, D. (2008). The polycomb repressive complex 2 is a potential target of SUMO modifications. *PLoS One* 7, e2704.
- Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet.* 38, 413-443. Review.
- Rodriguez, M. S., Dargemont, C., and Hay, R. T. (2000). SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. *J Biol Chem* 16, 12654-12659.
- Rodriguez, P., Braun, H., Kolodziej, K. E., de Boer, E., Campbell, J., Bonte, E., Grosveld, F., Philipsen, S., and Strouboulis, J. (2006). Isolation of transcription factor complexes by in vivo biotinylation tagging and direct binding to streptavidin beads. *Methods Mol Biol* 338, 305-323.

- Roessler, S., Györy, I., Imhof, S., Spivakov, M., Williams, R. R., and Busslinger, M., (2007). Distinct promoters mediate the regulation of Ebf1 gene expression by interleukin-7 and Pax5. *Mol Cell Biol* 27, 579-94.
- Rolink, A. G., Nutt, S. L., Melchers, F., and Busslinger, M. (1999). Long-term in vivo reconstitution of T-cell development by Pax5-deficient B-cell progenitors. *Nature* 401, 603-606.
- Rolink, A., Kudo, A., Karasuyama, H., Kikuchi, Y., and Melchers, F. (1991). Long-term proliferating early pre B cell lines and clones with the potential to develop to surface Ig-positive, mitogen reactive B cells in vitro and in vivo. *EMBO J* 2, 327-36.
- Rosenberg, N., Baltimore, D., and Scher, C. D. (1975). In vitro transformation of lymphoid cells by Abelson murine leukemia virus. *Proc. Natl. Acad. Sci. USA* 72, 1932–1936.
- Saleque, S., Singh, M., Little, R.D., Giannini, S. L., Michaelson, J. S., and Birshtein, B. K. (2001). Dyad symmetry within the mouse 3' IgH regulatory region includes two virtually identical enhancers (C alpha3'E and hs3). *J Immunol* 10, 4780-7.
- Satijn, D. P., Hamer, K. M., den Blaauwen, J., and Otte, A. P. The polycomb group protein EED interacts with YY1, and both proteins induce neural tissue in *Xenopus* embryos. *Mol Cell Biol* 4, 1360-9.
- Schaniel, C., Bruno, L., Melchers, F., and Rolink, A. G. (2002). Multiple hematopoietic cell lineages develop in vivo from transplanted Pax5-deficient pre-B I-cell clones. *Blood* 99, 472-478.
- Schebesta, A., McManus, S., Salvagiotto, G., Delogu, A., Busslinger, G.A., and Busslinger, M. (2007). Transcription factor Pax5 activates the chromatin of key genes involved in B cell signaling, adhesion, migration and immune function. *Immunity* 27, 49-63.
- Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N. F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J. A., Umlauf, D., Dimitrova, D. S., Eskiw, C. H., Luo, Y., Wei, C. L., Ruan, Y., Bieker, J. J., and Fraser, P. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet.* 1, 53-61.
- Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. *Cell* 128, 735-745.
- Seet, C. S., Brumbaugh, R. L., and Kee, B. L. (2004). Early B cell factor promotes B lymphopoiesis with reduced interleukin 7 responsiveness in the absence of E2A. *J Exp Med* 199,1689-1700.
- Sekimata, M., Pérez-Melgosa, M., Miller, S. A., Weinmann, A. S., Sabo, P. J., Sandstrom, R., Dorschner, M. O., Stamatoyannopoulos, J. A., and Wilson, C. B. (2009). CCCTC-

binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon- γ locus. *Immunity* 31, 551-564.

- Semerad, C., Mercer, E., Inlay, M., Weissman, I., Murre, C. (2009). E2A proteins maintain the hematopoietic stem cell pool and promote the maturation of myelolymphoid and myeloerythroid progenitors. *Proc Natl Acad Sci USA*. 106, 1930–1935.
- Shalizi, A., Gaudillière, B., Yuan, Z., Stegmüller, J., Shirogane, T., Ge, Q., Tan, Y., Schulman, B., Harper, J. W., and Bonni, A. (2006). A calcium-regulated MEF2 sumoylation switch controls postsynaptic differentiation. *Science* 311, 1012-7.
- Shav-Tal, Y., and Zipori, D. (2002). PSF and p54(nrb)/NonO--multi-functional nuclear proteins. *FEBS Lett*. 531, 109-114.
- Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* 38, 1348-1354.
- Simonis, M., Kooren, J., and de Laat, W. (2007). An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* 4, 895-901.
- Sims, R. J. 3rd, and Reinberg, D. (2008). Is there a code embedded in proteins that is based on post-translational modifications? *Nat Rev Mol Cell Biol* 10, 815-20. Review.
- Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H. D., and Cordes, S. P. (2009). A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell* 138, 885-97.
- Smith, E., and Shilatifard, A. (2010). The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. *Mol Cell* 40, 689-701. Review.
- Soler, E., Andrieu-Soler, C., de Boer, E., Bryne, J. C., Thongjuea, S., Stadhouders, R., Palstra, R. J., Stevens, M., Kockx, C., van Ijcken, W., Hou, J., Steinhoff, C., Rijkers, E., Lenhard, B., and Grosveld, F. (2010). The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes Dev* 24, 277-289.
- Souabni, A., Cobaleda, C., Schebesta, M., and Busslinger, M. (2002). Pax5 promotes B lymphopoiesis and blocks T cell development by repressing Notch1. *Immunity* 6, 781-93.
- Spooner, C. J., Cheng, J. X., Pujadas, E., Laslo, P., and Singh, H. (2009). A recurrent network involving the transcription factors PU.1 and Gfi1 orchestrates innate and adaptive immune cell fates. *Immunity* 31, 576-586.

- Sprung, R., Chen, Y., Zhang, K., Cheng, D., Zhang, T., Peng, J., and Zhao, Y. (2008). Identification and validation of eukaryotic aspartate and glutamate methylation in proteins. *J Proteome Res* 7, 1001-1006.
- Sridharan, R., and Smale, S. T. (2007). Predominant interaction of both Ikaros and Helios with the NuRD complex in immature thymocytes. *J Biol Chem* 282, 30227-30238.
- Stedman, W., Kang, H., Lin, S., Kissil, J. L., Bartolomei, M. S., and Lieberman, P. M. (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J* 27, 654-666.
- Su, IH., Basavaraj, A., Krutchinsky, A. N., Hobert, O., Ullrich, A., Chait, B. T., and Tarakhovsky, A. (2003). Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nat. Immunol.* 4, 124-131.
- Subrahmanyam R and Sen R. (2010). RAGs' eye view of the immunoglobulin heavy chain gene locus. *Semin Immunol.* 22, 337-45. Review.
- Tokoyoda, K., Egawa, T., Sugiyama, T., Choi, B. I., and Nagasawa, T. (2004). Cellular niches controlling B lymphocyte behavior within bone marrow during development. *Immunity* 20, 707-718.
- van de Nobelen, S., Rosa-Garrido, M., Leers, J., Heath, H., Soochit, W., Joosen, L., Jonkers, I., Demmers, J., van der Reijden, M., Torrano, V., Grosveld, F., Delgado, M. D., Renkawitz, R., Galjart, N., Sleutels, F. (2010). CTCF regulates the local epigenetic state of ribosomal DNA repeats. *Epigenetics Chromatin* 3, 19.
- Vincent-Fabert, C., Fiancette, R., Pinaud, E., Truffinet, V., Cogné, N., Cogné, M., and Denizot, Y. (2010). Genomic deletion of the whole IgH 3' regulatory region (hs3a, hs1,2, hs3b, and hs4) dramatically affects class switch recombination and Ig secretion to all isotypes. *Blood* 116, 1895-8.
- Wallace, J. A., and Felsenfeld, G. (2007). We gather together: insulators and genome organization. *Curr Opin Genet Dev* 17400-7.
- Wang, J., Cantor, A. B., and Orkin, S. H. (2009). Tandem affinity purification of protein complexes in mouse embryonic stem cells using in vivo biotinylation. *Curr Protoc Stem Cell Biol*, Chapter 1:Unit1B.5.
- Wang, Z., Zang, C., Cui, K., Schones, D. E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes, *Cell* 138, 1019–1031.
- Wendt, K. S., and Peters, J. M. (2009). How cohesin and CTCF cooperate in regulating gene expression. *Chromosome Res* 17, 201-14. Review.

- Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K., and Peters, J. M. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 796-801.
- Whitehurst, C. E., Schlissel, M. S., and Chen, J. (2000). Deletion of germline promoter PD beta 1 from the TCR beta locus causes hypermethylation that impairs D beta 1 recombination by multiple mechanisms. *Immunity* 13, 703-14.
- Woo, C. J., Kharchenko, P. V., Daheron, L., Park, P. J., and Kingston R. E. (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell* 140, 99-110.
- Xiao, T., Wallace, J., and Felsenfeld, G. (2011). Specific Sites in the C Terminus of CTCF Interact with the SA2 Subunit of the Cohesin Complex and Are Required for Cohesin-Dependent Insulation Activity. *Mol Cell Biol* 31, 2174-83.
- Xie, X., Mikkelsen, T. S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E. S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc Natl Acad Sci U S A*. 104, 7145-50.
- Xu, C. R., Schaffer, L., Head, S. R., and Feeney, A. J. (2008). Reciprocal patterns of methylation of H3K36 and H3K27 on proximal vs. distal IgVH genes are modulated by IL-7 and Pax5. *Proc Natl Acad Sci U S A* 105, 8685-8690.
- Xu, H. E., Rould, M. A., Xu, W., Epstein, J. A., Maas, R. L., and Pabo, C. O. (1999). Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev* 13, 1263-1275.
- Yang, S. H., and Sharrocks, A. D. (2004). SUMO promotes HDAC-mediated transcriptional repression. *Mol Cell* 13, 611-7.
- Yang, S. H., Galanis, A., Witty, J., and Sharrocks, A. D. (2006). An extended consensus motif enhances the specificity of substrate modification by SUMO. *EMBO J*. 25, 5083–5093.
- Yang, X. J. (2005) Multisite protein modification and intramolecular signaling. *Oncogene* 10, 1653-1162. Review.
- Yang, X. J., and Grégoire, S. (2006). A recurrent phospho-sumoyl switch in transcriptional repression and beyond. *Mol Cell*, 23, 779-86. Review.

- Ye, W., Bouchard, M., Stone, D., Liu, X., Vella, F., Lee, J., Nakamura H, Ang. S. L., Busslinger, M., and Rosenthal, A. (2001). Distinct regulators control the expression of the mid-hindbrain organizer signal FGF8. *Nat Neurosci* 4, 1175-81.
- Yusufzai, T. M., Tagami, H., Nakatani, Y., and Felsenfeld, G. (2004). CTCF Tethers an Insulator to Subnuclear Sites, Suggesting Shared Insulator Mechanisms across Species. *Molecular Cell* 13, 291-298
- Zhang, F. P., Mikkonen, L., Toppari, J., Palvimo, J. J., Thesleff, I., and Jänne, O.A. (2008). Sumo-1 function is dispensable in normal mouse development. *Mol Cell Biol* 28, 5381-90.
- Zhang, R., Burke, L.J., Rasko, J. E., Lobanenko, V., and Renkawitz, R. (2004). Dynamic association of the mammalian insulator protein CTCF with centrosomes and the midbody. *Exp Cell Res* 294, 86-93.
- Zhang, Z., and Carmichael, G. G. (2001). The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell* 106, 465-75.
- Zhang, Z., Espinoza, C. R., Yu, Z., Stephan, R., He, T., Williams, G. S., Burrows, P. D., Hagman, J., Feeney, A. J., and Cooper, M. D. (2006). Transcription factor Pax5 (BSAP) transactivates the RAG-mediated V(H)-to-DJ(H) rearrangement of immunoglobulin genes. *Nat Immunol*, 7, 616-24.
- Zhou, V. W., Goren, A., and Bernstein, B. E. (2010). Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* 12, 7-18.
- Zhuang, Y., Soriano, P., and Weintraub, H. (1994). The helix-loop-helix gene E2A is required for B cell formation. *Cell* 79, 875-884.

CREDITS FOR CONTRIBUTION TO THIS WORK

The research described here would not have been possible without previous or simultaneous efforts from several people.

- Giorgia Salvagiotto generated the *Pax5*^{Bio/Bio} mouse used for the study of Pax5-interacting proteins.
- Maria Novatchkova designed the primers for the 4C PCR and assisted in the bioinformatic analysis.
- Ido Tamir performed all the normalisations of the 4C-sequencing reads. Ido designed the genome browser that was used for displaying the 4C sequencing results in this work.
- Anja Ebert generated all the unpublished ChIP-sequencing data presented in this work. Furthermore, without her discovery of the PAIR elements, 4C analysis of the *Igh* locus would have been just a random walk.
- Andreas Sommer and his team are responsible for the deep sequencing analysis.
- The mass spectrometry analysis was performed by the Protein Chemistry Facility under the supervision of Karl Mechtler.
Ines Steinmacher assisted with the tryptic digestion of gel-fractionated proteins for mass spectrometry.
Michael Mazanek operated the mass spectrometer.
Richard Imre performed the Mascot database searches.

ACKNOWLEDGMENTS

Scientific

Thanks to:

- Meinrad Busslinger for the opportunity to work in his research group as a PhD student and a postdoc in the future.
- Wouter de Laat who provided the training in the 4C method, and continued to advise me during the optimisation of the 4C protocol for utilisation on pro-B cells.
- Frank Grosveld and Eric Soler for sharing their experience and improvements in the preparation and analysis of 4C samples.
- Niels Galjart who provided the *Ctcf*^{Bio/+} mouse on a collaborative basis.
- Andrea Pichler, for all the advice, enthusiasm and reagents in the sumoylation subproject.
- All members of the Busslinger research group for the great working atmosphere at present and in the past.

Personal

Thanks to:

- My one and only sister, Nataša, for all the love and care she has been giving me throughout my life. *Nataša, ako ima nešto dobro u meni, postoji zbog tebe, hvala.*
- My friends back in Zagreb, for staying by my side during the last 4 years. Especially to Sanja and Marina for all the adventures in forests and cities ☺...to be continued, I hope.....
- To Sabrina, for being my friend from the first day that I stepped into the IMP.
- To my Croatian colleagues, who are here in Vienna, for shared feelings and thoughts about the things we miss “*somewhere over the rainbow*”.

Curriculum vitae

Jasna Medvedović

- Date of birth: 22.08.1983
- Nationality: Croatian
- Place of birth: Zagreb, Croatia

Education

- June 2002 – May 2007 University of Zagreb, Faculty of Science, Department of Biology; 327 ECTS, GPA 4,95 (of maximum 5.0);
Degrees: *Biologiae ingeniarius diplomate probatus – sectione biologiae molecularis*,
Biologiae ingeniarius diplomate probatus – sectione oecologiae
- October 2006 – February 2007 Diploma research at The Ruđer Bošković Institute (RBI) in Zagreb, Laboratory for Molecular Ecotoxicology, Prof. Dr Tvrtko Smital
Diploma thesis: Identification and functional characterisation of p-glycoprotein (ABCB1) in mussel species *Unio pictorum* (LINNAEUS 1758)
- February – April 2007 Internship at The Institute of Biochemistry II in Frankfurt, Molecular Signaling Group, Prof. Dr. Ivan Đikić
- June 2007 – June 2011 PhD research at The Institute of Molecular Pathology in Vienna, Stem cell commitment in hematopoiesis Group, Prof. Dr. Meinrad Busslinger

Scholarships

- 2003 – 2005 Scholarship of the Ministry of Science, Education and Sports
- 2005 – 2006 Scholarship of the City of Zagreb received
- 2005 Scholarship of the University of Zagreb granted

Meetings and symposia

- November 2007 Molecules to Mind, PhD Symposium, Vienna Biocenter; participant
- March 2008 1st annual EuTRACC (European Transcriptome, Regulome and Cellular Commitment Consortium) meeting, Antwerp; participant
- October 2008 4th International Conference on Gene Regulation in Lymphocyte Development, Rhodos; participant
- November 2008 Life under (re)construction, PhD Symposium, Vienna Biocenter; co-organizer
- March 2009 2nd annual EuTRACC (European Transcriptome, Regulome and Cellular Commitment Consortium) meeting, Prague; participant
- March 2010 Epigenetic Regulation in Cell Fate & Disease, IMP, Vienna; participant
- April 2010 Gene Expression and Signaling in the Immune System; Cold Spring Harbor; participant
- December 2010 EuTRACC 2nd Young Scientists Meeting, Dubrovnik; co-organizer

Associated with bat biology:

- August 2005 10th European Bat Research Symposium, Galway;
- August 2008 11th European Bat Research Symposium, Cluj Napoca;
- January 2009 Symposium on Bat Migration, Berlin, Germany

Courses

- September 2009 FEBS Immunology Summer School, Hvar
- February 2010 EuTRACC Proteomics Course, Rotterdam

Publications

- 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, 175-87.
- 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* (Epub ahead of print.)