THE UNIVERSITY OF CHICAGO

THE ROLE OF GAGA MOTIF DOMAINS IN CHROMATIN REMODELING OF IGκ DURING B CELL DEVELOPMENT

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM:
DEVELOPMENT, REGENERATION, AND STEM CELL BIOLOGY

BY
KAITLIN MCLEAN

CHICAGO, ILLINOIS
DECEMBER 2020

To the strong women in my life who have inspired, encouraged and believed in me

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First, I'd like to express my sincerest gratitude to my advisor, Dr. Marcus Clark. His mentorship guided my development as a scientist, for which I will be forever grateful. He helped me learn how to see the minute details of a project as a part of a larger picture. The course of the PhD was neither simple nor smooth, but Marcus never gave up on me. He encouraged me and never lost faith in me, even when at times I felt I had lost faith in myself. It was comforting throughout the PhD knowing that Marcus's open door was always waiting just down the hall (or zoom call) when I had questions or wanted to go over data. I am grateful for the opportunities he gave me, from teaching me how to write grants, introducing me to other scientists at conferences, and sharing his perspectives on talks, that have greatly helped my personal growth. Marcus simultaneously allowed me freedom to explore my ideas, and provided the structure I needed to actually accomplish those goals. He was incredibly generous in providing all the resources I needed, and never put a ceiling on my personal development as a scientist.

I also want to thank my wonderful committee members for their support and encouragement. First, to my committee chair, Megan McNerney, I want to thank you for being not only a valuable member of my committee, but also for being a role model. It was an incredible privilege for me to have on my committee who was herself successfully following the path of physician-scientist that I hope to tread. Megan was always a realist, and helped me to moderate the scope of my project into something I could accomplish and be proud of, especially when my ambitions overran the realities of the time constraints of a PhD. Next, to Barbara Kee, thank you for your excellent immunological insights on my project, and for your measured and sound advice. Finally, I'd like to thank Alex Ruthenberg. Alex taught the first graduate course at the University of Chicago that truly sparked my passion. He opened up to my eyes the world of epigenetics and chromatin biology, a field for which my love and wonder only grows by the day. He is one of the best-read and most careful scientists I have ever met, and it was an honor to have him on my committee. Alex, it was an absolute

privilege to TA for you, and I've learned so much through your example. Thank you for your unsuppressed enthusiasm for work. I cannot express how much it meant.

Next, I need to thank my fellow lab members. The lab was always immensely collaborative and collegial, and it is not an exaggeration to say that without them, my project would not have been possible. Thanks especially to Malay Mandal, who was a critical mentor to me throughout my entire PhD. My work built off of the elegant work he had previously published. He helped me through each and every step of the project, from the initial CRISPR design to the data analysis. Malay knows everything there is to know about B cell development, and has an intuitive feel for what experiments will work that study alone cannot achieve. It was a privilege to write a review with him, as I learned an incredible amount in the process. He is a profoundly generous scientist and mentor, and I am so grateful to have been able to lean on his expertise throughout my graduate work. Additionally, the other members of the B cell development group, Domenick Kennedy and Michael Okoreeh were so wonderful to work with. I could not ask for a better or more collaborative group. Thanks to our lab mom, Margaret Veselits. The lab would not operate without her work as lab manager, and it is a testament to her skills that the lab always ran so smoothly. Thank you too to Andrew Kinloch and Yuta Asano for their help with my protein work in the lab, and just generally for being such pleasant labmates. Finally, but certainly not least, I want to thank my lab sister, Reba Abraham. I was so happy that she joined the lab, as she became someone that I felt I could always confide in and lean on for emotional support when experiments weren't working. She was also someone that I could laugh with, and her presence made the lab not just a productive, but a joyful place to be. Some of my best memories in the lab are of Reba and I laughing at ourselves, and I will cherish those memories.

I was counseled from the start of graduate school to place importance on self-care, and this would not have been possible without the many and excellent friends who supported me outside of grad school. First, to one of my very best friends, Aziza Suleymanzade, with

whom I would get lunch or coffee just about every day. I firmly believe that using coffee time with friends as a time to vent frustrations or explore blue-sky ideas is a critical part of ones professional as well as personal development, and I can't think of a better person for that than Aziza. She is one of the most brilliant, as well as one of the most kind, individuals I have ever met, and I will treasure her friendship forever. Another grad school friend, Lari de Wet, has also been a wonderful friend, and I loved our movie nights as a chance to just relax and enjoy the company of someone who makes you really happy. To the community of friends I made through dancing, thank you so much. Though too numerous to name everyone, dance was one of the best outlets I had to stay active and achieve a good work-life balance. The dance community at the University of Chicago is wonderful, and I owe such a debt of gratitude to it. Finally, thank you to my best friends from college, and I hope best friends for life, Rebecca Miller and Sabrina Jean-Baptiste. You two are completely wonderful, and have supported and encouraged me for over 10 years now. I am so inspired by how amazing you both are, and am encouraged to know that even when we're not physically close, we can remain close friends.

Finally, I want to thank my family, without whom I would not be where I am today. In particular, I want to thank my mom, Hattie McLean, for her unwavering support, and her confidence in me that knows no bounds. I have always thought that she overestimated my capabilities, and yet still she persists. My mom taught me that no dream was too big, and the only thing at which one was certain to fail was the thing that one did not attempt. Thank you also to my brother, Ryan McLean, and my dad, David McLean, for the encouragement and levity you provide. Though I find inspiration in all my family, I find it especially in my grandmother, Julia Loufek Kimport. A code breaker and pilot during WWII, when such things were simply not done by women, she has been a living proof of the ability of "ordinary people" to perform extraordinary things.

ABSTRACT

V(D)J recombination is a spatially and temporally restricted process, and depends heavily on epigenetic regulation to provide this control. The failure of V(D)J recombination would result in an immunodeficiency, while the lack of regulation of this process can result in genomic instability and oncogenic transformation. Brwd1 is an epigenetic reader required for normal B cell development, and specifically for Igκ recombination. Brwd1 is recruited to activating histone marks, and when binding is coincident with GAGA DNA repeats, it appears to participate in epigenetic remodeling by altering nucleosome positioning and enhancing DNA accessibility. During normal Igκ recombination, Brwd1 binds at the Jκ gene cluster and repositions nucleosomes 5' to each Jκ segment, exposing the RSS, allowing for RAG recruitment.

Here, we examine whether GAGA repeat domains are required for the chromatin remodeling activity of Brwd1. Using CRISPR-Cas9, we removed the 5' GAGA motif from Jκ1 and Jκ2 in mice. This resulted in not only a dramatic loss of usage of Jκ1 in the Jκ1-GAGA-deletion mice, but also in a partial developmental block during Igκ recombination and lower overall kappa usage. Furthermore, the architecture of nucleosome positioning is altered in the Jκ locus in small pre-B cells from mice with the Jκ1 GAGA deletion. The Jκ2 GAGA deletion also shows decreased usage of the gene segment proximal to the deletion, but with a less severe overall defect.

Our work adds new insights to the accessibility hypothesis of ordered recombination, and demonstrates the importance of absolute primary nucleosome positioning at Jκ in cells preparing for recombination. Furthermore, it provides evidence that GAGA motif domains are required for the role of Brwd1 in chromatin remodeling. Finally, our results inspired a new perspective on how cryptic RSSs throughout the genome are shielded from off-target recombination by the RAG recombinase, thus aiding in the maintenance of genomic integrity.

CHAPTER 1

INTRODUCTION

1.1 B Cell Development

1.1.1 Differentiation of B Cells from Hematopoetic Stem Cells

The vertebrate immune system consists of both an innate and an adaptive immune system, both of which are essential to fending off attacks from foreign pathogens [1]. The adaptive immune system is unique in that it continues to develop and learn throughout the lifetime of the organism as it encounters different pathogens, forming an immunological memory. Two major types of lymphocytes, B cells and T cells, comprise the adaptive immune system. Each B cell expresses a B cell receptor (BCR) that is unique and antigen-specific. After the binding of antigen to the BCR, the B cell will proliferate and differentiate into plasma cells, which undergo an alternative splicing event to delete the BCR transmembrane domain. This will allow for the secretion of the BCR as antigen-specific antibodies [1].

B cells are derived from the multipotent hematopoietic stem cells (HSCs) that give rise to all blood cells (Figure 1.1). HSCs are true somatic stem cells and have tremendous capacity both for self-renewal and for differentiation into all blood cell lineages. To produce B cells, HSCs first differentiate into the more lineage-restricted multipotent progenitor cells (MPPs), which give rise to common myeloid progenitors (CMPs) and common lymphoid progenitors (CLPs). A subset of CLPs is then committed to the B-cell lineage, and gives rise to pre-pro and then pro-B stages. The commitment of an MPP first to the lymphoid fate then to the B cell fate requires the support of a number of transcription factors and cytokines produced by the stroma surrounding the developing cell, known as the bone marrow (BM) microenvironment. In particular, PU.1 and E2A are critical for suppressing myeloid lineages

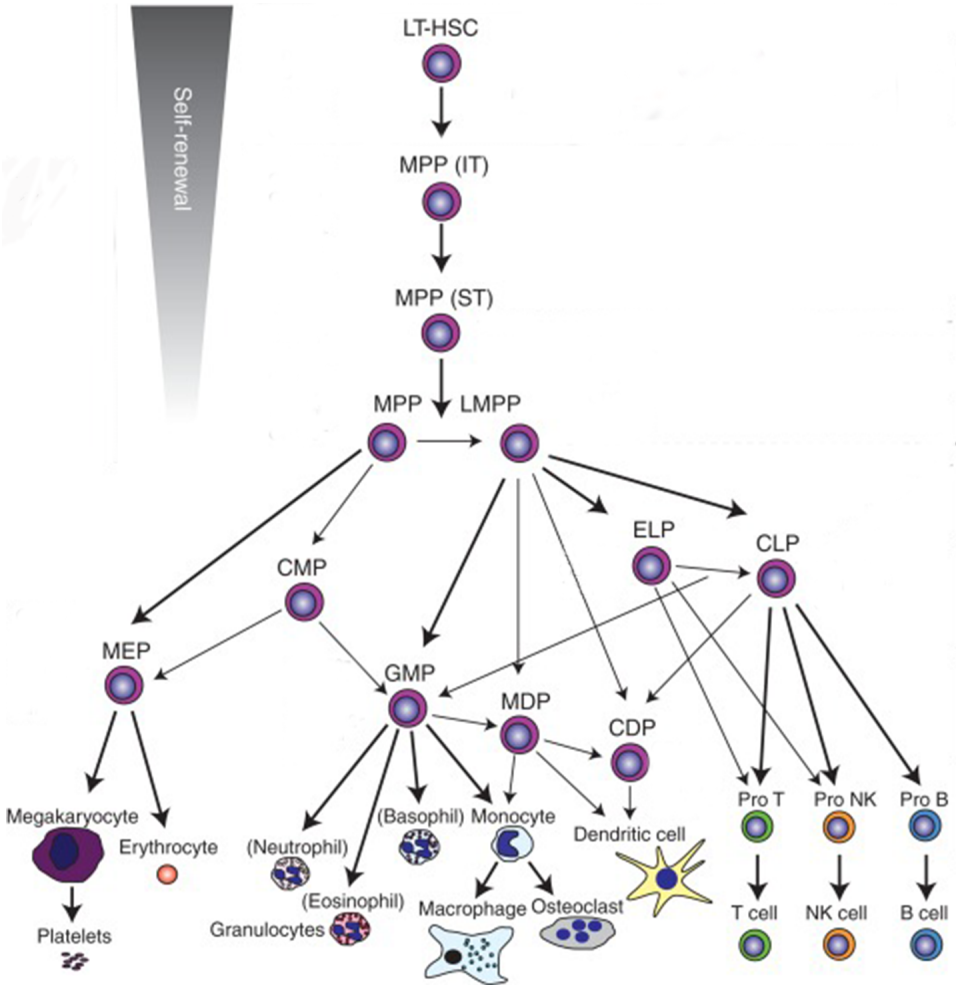

Figure 1.1: B Cells are derived from Hematopoetic Stem Cells
All differentiated blood cell lineages are derived from long-term, self-renewing hematopoietic stem cells. The well-established pathways of differentiation are shown in bold arrows, whereas more recently described differentiation pathways are displayed in thin arrows. HSC, hematopoietic stem cell; MPP, multipotent progenitor; LT-, long-term repopulating; IT-, intermediate-term repopulating; ST-, short-term repopulating; LMPP, lymphoid-primed MPP; ELP, early lymphoid progenitor; CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte–macrophage progenitor; MEP, megakaryocyte–erythrocyte progenitor; CDP, common dendritic progenitor; MDP, monocyte–dendritic cell progenitor; NK, natural killer cell. Adapted from Rieger *et al*, 2012.

and priming for B-lineage commitment. They induce factors like early B-cell factor 1 (EBF1) and Pax5. The progenitor cell becomes stably committed to the B cell fate when a feedback loop of Pax5 and IKAROS is established to maintain EBF1 expression [2].

The hallmark of B lymphopoiesis is the sequential rearrangement of the immunoglobulin heavy chain (IgH: Igμ) locus followed by the immunoglobulin light chain loci (IgL: Igκ or Igλ). Rearrangement of Igμ starts with diversity (D) and joining (J) gene segments in pre-pro B cells (Figure 1.2). Subsequent recombination of variable (V) to rearranged D(J) is then completed in pro-B cells, at which point, cells become committed to the B lineage. Developing B cells are then driven to proliferate by interleukin-7 receptor (IL-7R) signaling in response to IL-7 secreted by BM stromal cells. Successfully rearranged heavy chain then associates with surrogate light chain (SLC: λ5 and VpreB) and the signaling molecules Igα and Igβ to form the pre-B cell receptor complex (pre-BCR) expressed on the cell surface of pre-B cells [3, 4]. Pre-B cells then undergo a proliferative burst associated with both pre-BCR and IL-7 receptor signaling [5]. Subsequently, large pre-B cells exit cell cycle, and small pre-B cells initiate IgL recombination, attempting recombination first at the immunoglobulin κ-chain locus (Igκ) [5, 6]. This occurs when there is concurrent repression of pre-BCR expression [3, 7]. The product of the rearranged Igκ light chain then associates with Igμ heavy chain to form B cell receptor (BCR) on the surface of immature B cells. Autoreactive early immature B cells bearing Igκ light chain can diminish autoreactivity by consecutive rearrangements of available Vκ and Jκ gene segments at the Igκ locus and subsequently Vκ-Jκ joining. This process of receptor editing helps ensure a diverse peripheral repertoire that is tolerant of self [8, 9]. The selected immature B cells then migrate from the BM to the spleen and lymph nodes as mature B cells to become organized in B cell follicles [9, 10].

During the developmental progression of B cells, the pre-B stage is a critical developmental checkpoint. The SLC probes Igμ fitness, and the expression of the pre-BCR enables the pre-B cell pool to enter into a proliferative burst in a IL-7 dependent manner. These

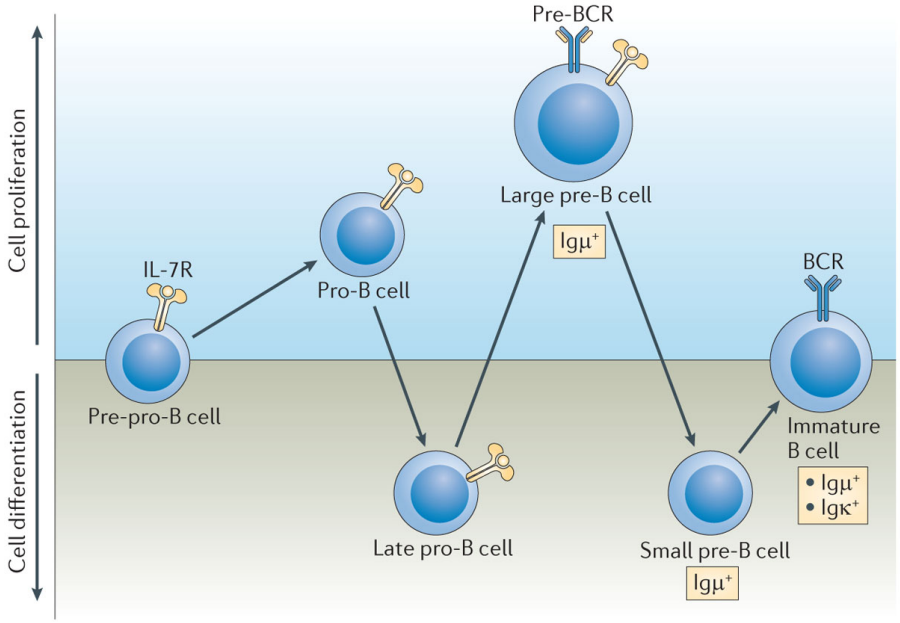

Figure 1.2: B Cell Lymphopoiesis
To avoid genomic instability and oncogenic transformation, it is essential that the processes of differentiation and proliferation be strictly segregated during B cell development. The process of V(D)J recombination that will give rise to a mature and unique B cell receptor begins in pre-pro-B cells, which undergo recombination of the D and J heavy chain (Igμ) segments. Pro B cells are then driven to proliferate by IL-7R signaling. Igμ recombination is completed in late pro B cells, which cease proliferating to recombine the V gene segment to the D(J) segment. This successful round of V(D)J recombination leads to expression of Igμ, which then assembles with the surrogate light chain to form the pre-BCR in large-pre B cells. Signaling from the pre-BCR along with IL-7R drives large pre-B cells to undergo a short burst of proliferation. Following cell cycle exit, the progenitors transition to the small-pre B cell stage, at which light chain recombination occurs. Cells that successful recombine Igκ then join the pool of immature B cells, which will undergo further selection mechanisms like receptor editing before becoming fully mature B cells. (Adapted from Clark *et al*, 2014).

cells first proliferate as large pre-B cells and then recombine the IgL loci as resting small pre-B cells [3, 11]. Proliferation and DNA rearrangement are strictly segregated processes, as concurrent replication and introduction of double-strand breaks during recombination would compromise genomic integrity [12]. Mutations or alterations that affect this checkpoint can result in development of pre-B cell leukemias, primary immunodeficiency, and systemic autoimmunity.

1.1.2 IL7-R Signaling is critical for survival and proliferation during early B cell development

IL-7 is a crucial cytokine secreted by BM stromal cells that plays an important role in B cell lineage commitment and development [13]. The IL-7R is expressed on early B cell progenitors and composed of the IL-7Rα chain, which confers specificity to IL-7, and common-γ chain receptor [14]. IL-7R plays diverse roles at different developmental stages and is essential for the growth, proliferation and survival of all the progenitor stages from CLPs to large pre-B cells [3, 5, 15, 16]. Mice lacking IL-7R have severe impairment in B lymphopoiesis [15, 16]. While B cell development is thought to be less dependent on IL-7R signaling in humans, most of the regulatory responses downstream of IL-7R are similar in mice and humans [17, 18]. Indeed, many of the patients with IL-7Rα mutations have low levels of serum immunoglobulin suggesting defective peripheral B cell function.

IL-7 binds to the IL-7Rα chain and induces the dimerization of the α and γ chains bringing associated Janus kinases (JAK1 and JAK3) together and stimulating their transphosphorylation and activation [16, 19]. Activated JAK kinases recruit the transcription factors (TFs) STAT5A and STAT5B [20, 21]. In addition to JAK-STAT pathway, IL-7R signaling also activates the phosphoinositide-3 kinase (PI3K)-PKB (protein kinase B, aka AKT) pathway [15, 16, 22] (Figure 1.3). Deletion of the PI3K regulatory subunit p85α, or the catalytic

subunits p110α and p110δ, impairs B cell development [23, 24].

One of the major functions of STAT5 activation in the JAK-STAT pathway is to ensure the survival of pro-B cells (Figure 1.3). STAT5 activates the pro-survival factors myeloid cell leukemia sequence 1 (MCL1) and B cell lymphoma-2 (BCL2) [21, 25, 26]. In addition to stimulating survival, STAT5 enhances the proliferation of B cell progenitors by inducing cyclin D3 (encoded by Ccnd3) required for proliferation of both pro-B and pre-B cells [27, 28] (Figure 1.3). Additionally, the activation of the PI3K-AKT pathway phosphorylates and promotes the nuclear export of the forkhead box protein O (FOXO) family of transcription factors that induce pro-apoptotic protein BCL-2 interacting mediator of cell death (BIM) (encoded by bcl211b) [29, 30]. AKT also directly phosphorylates and inactivates the pro-apoptotic factor BCL-2 antagonist of cell death (BAD) [31]. Therefore, IL-7R signaling promotes the survival of pro-B cells by both upregulating survival signaling and repressing apoptotic signaling.

IL-7R signaling is crucial for preventing premature *Ig*κ recombination. IL7R signaling inhibits RAG expression via PI3K-AKT-mediated phosphorylation and inactivation of TFs FOXO1 and FOXO3a, which directly activate Rag expression [22, 32, 33]. Expression of RAG1 and RAG2 is absolutely necessary for immunoglobulin gene recombination [34]. Two other downstream effectors induced by IL-7R signaling, STAT5 and Cyclin D3, are essential for inhibiting premature Igκ recombination [22, 35, 36, 37, 27] (Figure 1.2). They do this by intricately coordinating the epigenetic landscape of Igκ, making it inaccessible to the RAGs while the cell is proliferating.

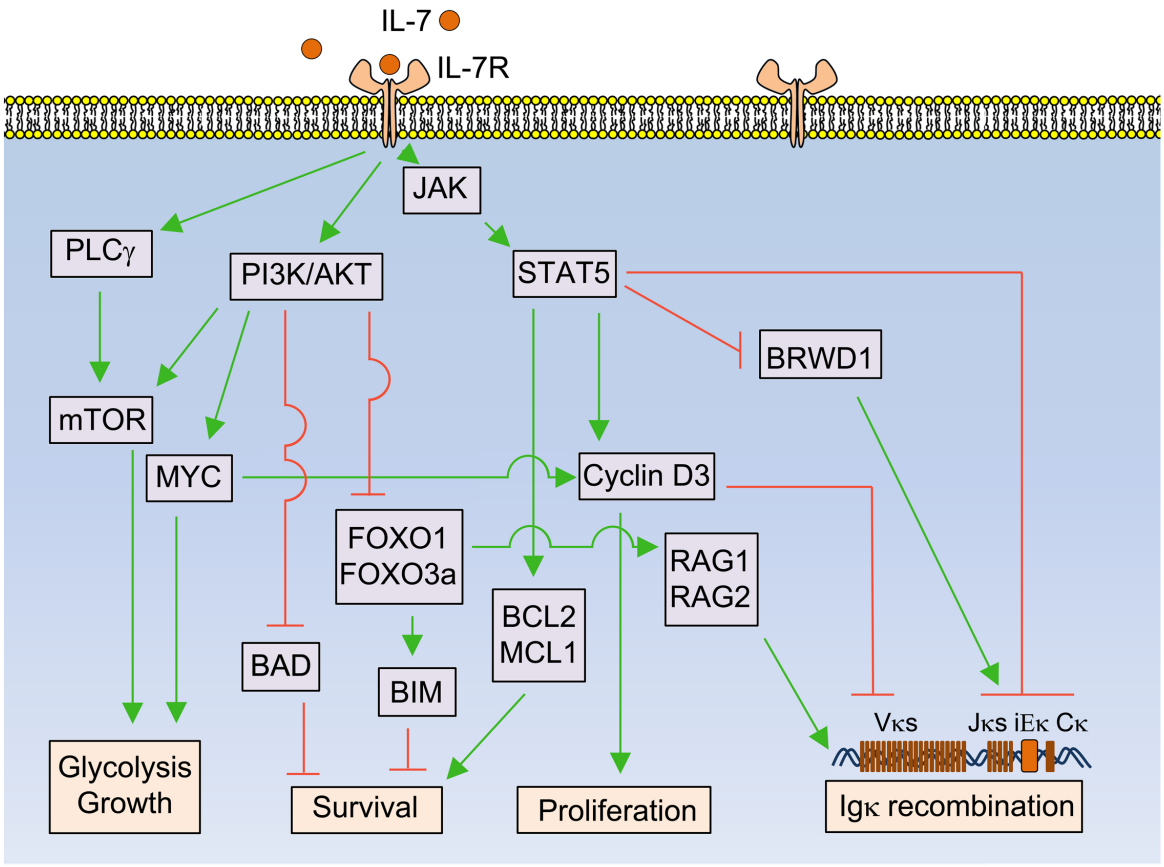

Figure 1.3: IL-7R Signaling Drives Proliferation and Survival While Preventing Premature Igκ Recombination Programs During B Cell Development
The stages of B cell development represent sequential phases of proliferation and differentiation. The survival and proliferation of pro-B and large pre-B cells is driven by signaling from the IL-7R receptor. IL-7R has activates two major signaling pathways, JAK-STAT5 and PI3K/AKT. JAK kinase(s) phosphorylates STAT5, which then stimulates transcription of cyclin D3, which promotes proliferation, and of BCL2/MCL1, which promotes survival. Additionally, STAT5 also serves to repress BRWD1 and Igκ accessibility and recombination. PI3K)/AKT pathway inhibits FOXO1 and FOXO3a, which is an inducer of recombination-activating gene products RAG1 and RAG2 essential for recombination. The PI3K/AKT pathway, in addition to repressing recombination, plays a critical role in cell survival by repressing BAD and BIM, the pro-apopotic proteins. Finally, IL-7R signaling serves to promote cell growth and glycolysis via PI3K/AKT signaling which upregulates mTOR and Myc, both of which drive glycolysis and cell growth. mTOR is also upregulated by PLCg, which is also a target of IL-7R signaling. Adapted from McLean *et al*, 2020.

1.1.3 The Role of the Bone Marrow Microenvironment in B cell Development

The pre-BCR expressing large pre-B cells are characterized by an initial proliferative phase in vivo with four to five rounds of cell division [38]. However, pre-B cells require both pre-BCR and IL-7 for expansion [5]. The proliferative burst is largely attributed to PI3K-AKT pathway and MYC-dependent induction of cyclin D3, which are activated downstream of IL-7R signaling [3, 10, 13, 22, 28, 39]. The proliferative pre-B cell pool does not expand and survive well without IL-7 in culture even though the cells express pre-BCR [22, 28, 37]. Thus, while pre-BCR expression is associated with a proliferative burst in large pre-B cells, it is unclear if this burst is directly mediated by the pre-BCR.

Following the proliferation of pre-B cells, Igκ recombination requires exit from the cell cycle to maintain genomic integrity, and consequently, numerous studies have been conducted regarding cell-autonomous pre-BCR signaling and gene regulatory mechanisms limiting proliferation [3, 10, 12, 13]. Based on the observation that pre-BCR signaling silences the expression of the SLC, and subsequently the expression of pre-BCR [40], it was postulated that loss of SLC levels via multiple cell divisions is responsible for attenuated pre-BCR mediated proliferative signaling and subsequent exit from cell cycle [40, 41]. However, constitutive expression of SLC in pre-B cells in vivo demonstrated that downregulation of the pre-BCR is not required for the cell cycle exit, rather expression of pre-BCR and signaling downstream is necessary for cessation of proliferation [7].

The dominance of IL-7R signaling over pre-BCR signaling prompts the question of how pre-B cells are able to overcome IL-7R signaling. The answer lies in the localization of progenitor B cells within BM microenvironment where B cells develop. The BM microenvironment provides the extracellular cues necessary to determine cell fate [42, 43, 44]. Additionally, extracellular cues may coordinate with cell intrinsic factors to define a specific signaling pro-

gram or by changing the signaling threshold for a differentiation event like cell cycle exit and immunoglobulin light chain recombination. Therefore, it is postulated that positioning away from IL-7 expressing stromal cells within the BM reinforces the switch from IL-7R signaling to pre-BCR signaling to progress through development [42, 22, 45]. In support of this model, it was found that downstream of pre-BCR, IRF4 induces the expression of CXC chemokine receptor 4 (CXCR4), which confers responsiveness to CXC-chemokine ligand 12 (CXCL12) [22, 45].

An early study of the BM microenvironment suggested that IL-7-producing stromal cells are distinct and spatially distributed away from CXCL12-producing stroma [42]. However, later studies showed that most of the stromal cells that highly express *Il-7* also express *Cxcl12* [44, 46]. A recent study of BM stromal cells at single cell resolution further showed that the fraction of mesenchymal stromal cells that highly express both *Il-7* and *Cxcl12* is quite small (approximately 14%), but most of those that highly express *Cxcl12* express very low levels of *Il7* [47, 48]. Tying together these observations, work from our lab identified three distinct major populations of stromal cells, IL-7^{neg}/lowCXCL12^{high}, IL-7^{int}CXCL12^{high} and IL-7^{high}CXCL12^{low} by high power field confocal microscopy [49]. Examination of section of whole BM single planes revealed widespread distribution of each cell type. Overall, the BM is a mosaic of IL-7 and CXCL12 producing cells with varying degrees of *Il-7* and *Cxcl12* expression, creating some niches that are relatively high in IL-7 and some that are high in CXCL12 [48, 49] (Figure 1.4). The small pre-B cells specifically reside in those niches enriched for IL-$7^{\text{neg/low}}$CXCL12^{high} stromal cells [49].

There is a significant difference in the localization of pro-B and pre-B cells in BM niches [44, 46, 49] (Figure 1.4). Pro-B cells, which express higher amounts of focal adhesion kinase (FAK) and very late antigen 4 (VLA4) compared to pre-B cells are more adherent to vascular cell adhesion molecule 1 (VCAM-1) in IL-7^{high} expressing stroma [43]. Proliferating B cell progenitors are also mostly localized in IL-7^{high} stroma in BM [49]. This is consistent with

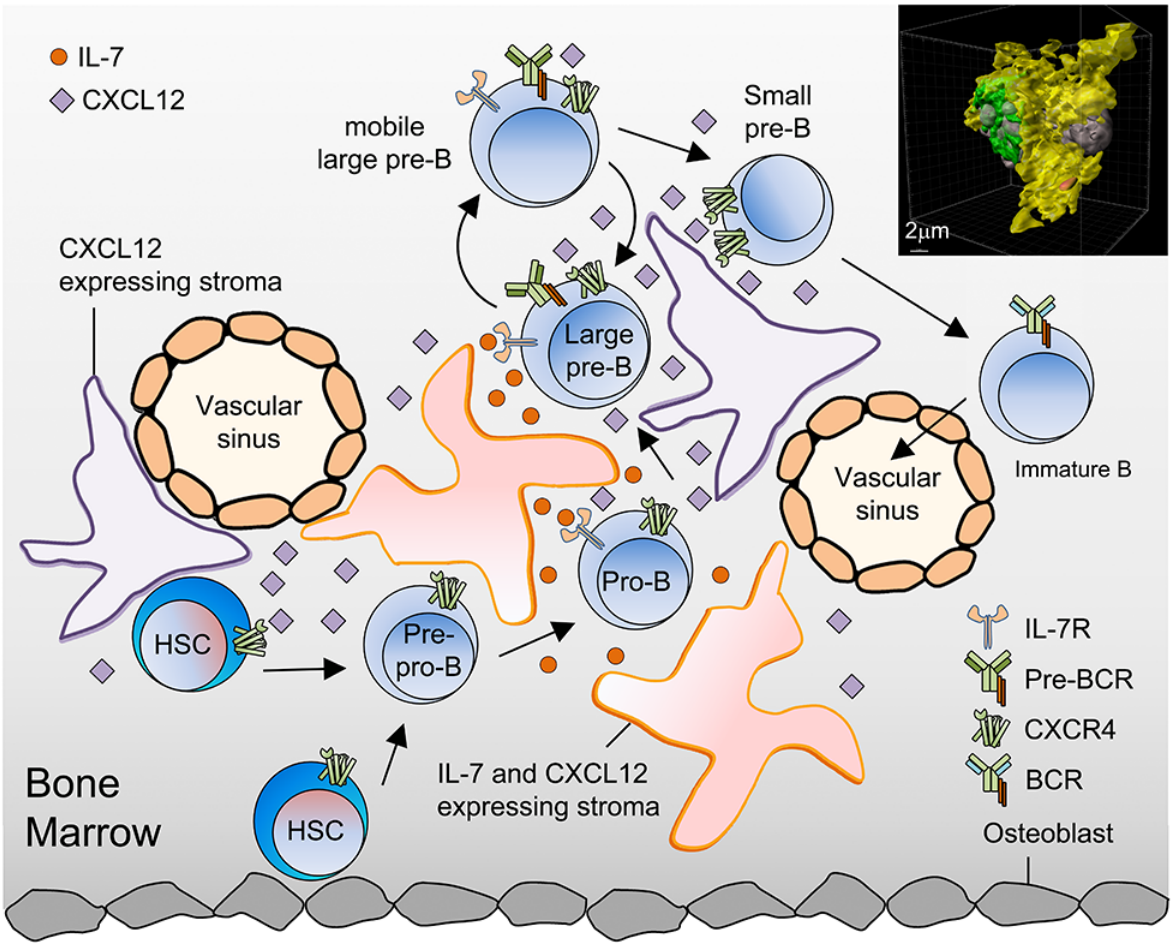

Figure 1.4: The ability of pre-BCR signaling to overcome IL-7R signaling is highly dependent on the CXCL12 and IL-7-rich microniches within the bone marrow
Pre-pro-B cells migrate to IL-7^{high} microniches from $CXCL12^{high}$ niches and differentiate into pro-B cells. Pro-B cells reside in IL-7^{high} niches where they upregulate IL-7R signaling and proliferate. Large pre-B cells then upregulate CXCR4 and become increasingly sensitive to CXCR12. These highly motile large pre-B cells undergo chemotaxis away from IL-7^{high} niches toward IL-$7^{low}CXCL12^{high}$ niches, where they activate pre-BCR and CXCR4 signaling and are able to escape from the cell cycle and recombine the immunoglobulin light-chain genes. Newly generated immature B cells downregulate CXCR4 and exit the BM for peripheral development. Ligand–receptor contact in the figure depicts active signaling of that receptor. High-power field confocal microscopy with 3D reconstruction showed that small pre-B cells (green) are in tight contact with CXCL12 (yellow), and with high local accumulations of extracellular CXCL12 (inset). Adapted from McLean *et al*, 2020.

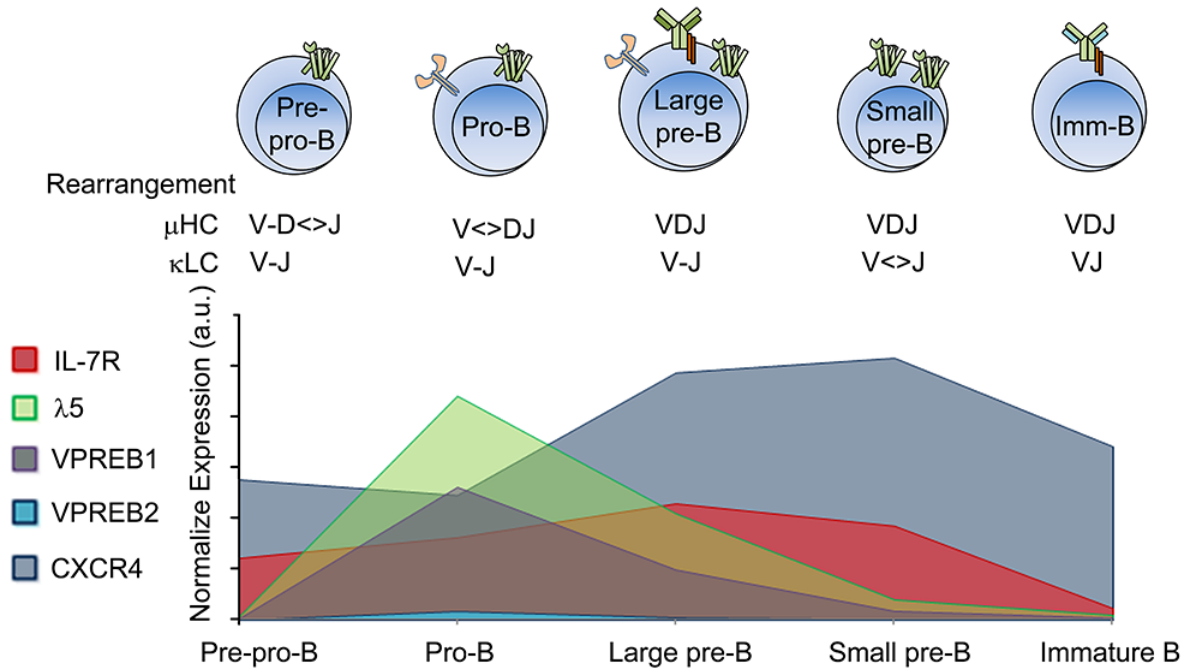

Figure 1.5: Graphical representation of the expression of the three essential early B cell development receptors: IL-7R, pre-BCR , and CXCR4.
Expression of IL-7R peaks at the large pre-B cell stage, and is gradually downregulated during later developmental stages; IL-7R signaling is dependent on the presence of the IL-7 ligand in the BM microniche. Expression of the pre-BCR components ($\lambda 5$ and VPREB1/VPREB2) spike at the pro-B cell stage; however, VH to DHJH recombination is ongoing at that stage. As a result, the functional pre-BCR complex is formed at the large pre-B cell stage when the rearranged heavy chain is available to make the pre-BCR. The expression of the third receptor, CXCR4, increases in the large pre-B cell stage where it helps the motile large pre-B cells to escape IL-7R signaling. Subsequently, CXCR4 downstream signaling arrests cell proliferation and completes Igκ recombination as well as repressing pre-BCR components to generate immature B cells with a functional BCR. Adapted from McLean *et al*, 2020.

the observation that cell surface expression of IL-7R and CXCR4 are reciprocally regulated during developmental progression from pro-B to large pre-B to small pre-B cells (Figure 1.5). The latter localize in CXCL12^{high} stroma [49]. Both IL-7R and CXCR4 are downregulated in immature B cells. Although pro-B and large pre-B cells can migrate along an IL-7 gradient, direct comparison of in vitro chemotaxis revealed that both large and small pre-B cells respond strongly to CXCR4 [49]. Indeed, intravital two-photon microscopy in the calvarial BM showed that the pro-B cells are non-motile and pre-B cells are mostly motile [46]. Large pre-B cells show the strongest chemotaxis to CXCL12, whereas small pre-B cells with even higher CXCR4 cell surface densities are in intimate contact with CXCL12 producing stroma [49]. This repositioning of pre-B cells in CXCL12 rich stroma is severely impaired in mice where CXCR4 is conditionally deleted in large pre-B cells [49]. Immature B cells reside in the CXCL12 expressing area but not in intimate contact with CXCL12. Therefore, CXCR4 is required to shift the signaling from IL-7R to pre-BCR, for normal development of small pre-B cells and their positioning away from high IL-7 niches.

1.1.4 Pre-BCR signaling guides the differentiation of pre-B cells and Igκ recombination

Successful assembly of the V-D-J gene segments in pro-B cells results in the expression of Igμ heavy chain (μHC) and surface expression of the pre-BCR. Pre-BCR-mediated signaling (Figure 1.6) is activated in a cell autonomous manner, and pre-BCR surface levels seem to regulate both cell proliferation and survival [50, 51, 52]. The pre-BCR is composed of two identical membrane anchored μHC subunits and two SLC subunits (SLC, a complex of λ5 and VpreB molecules) bound to each of the μHCs and the signaling subunits Igα and Igβ [53, 54]. Mice lacking SLC inefficiently produce heavy and light chains that confer autoreactivity, suggesting that one function of the pre-BCR is to select against auto-reactivity, making the pre-B stage a tolerance checkpoint [55].

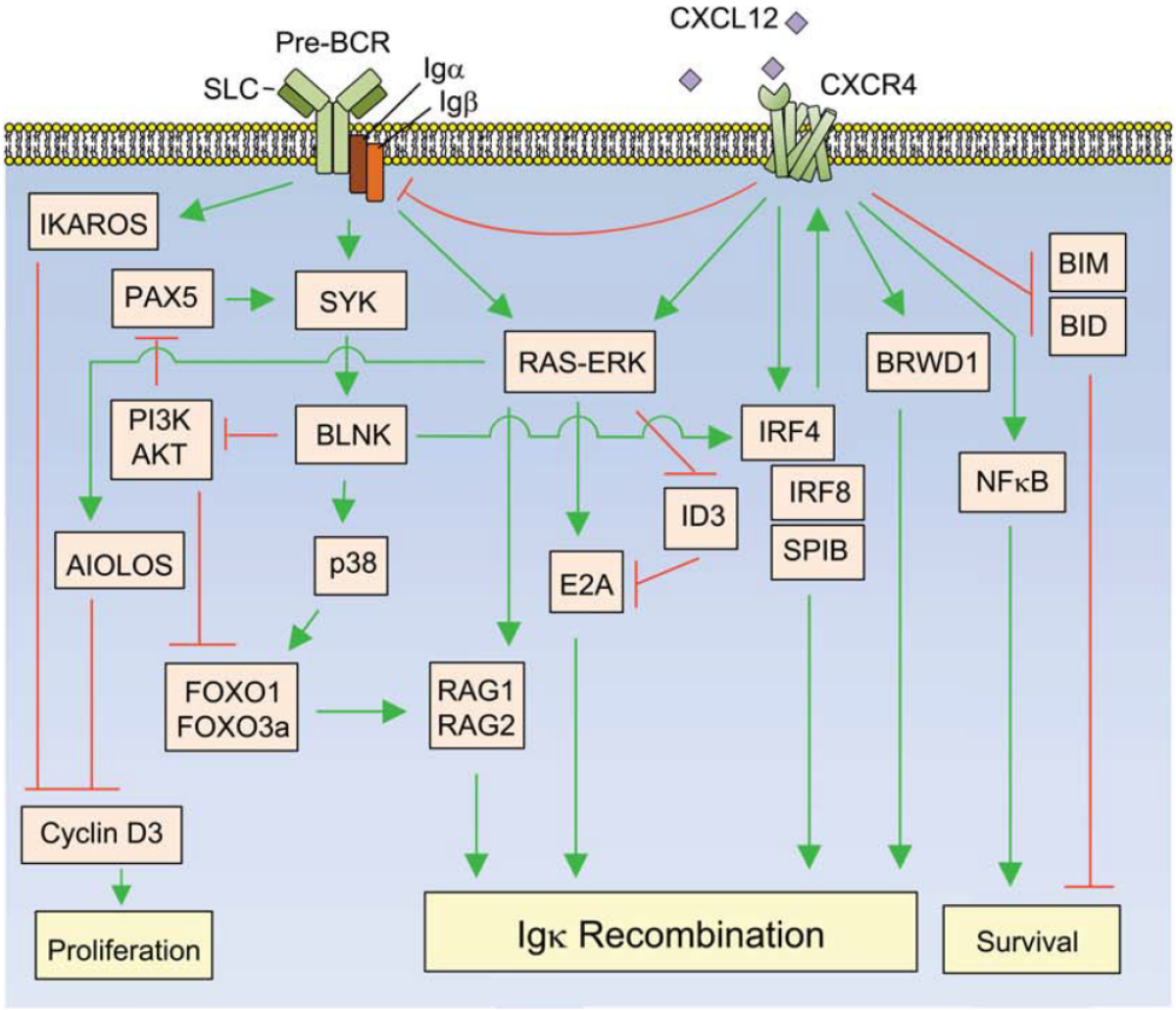

Figure 1.6: The Pre-BCR and CXCR4 Work Together to Help Developing B Cells Escape Cell Cycle and Recombine the Immunoglobulin Light Chain
The pre-BCR serves to counteract the proliferative effects of IL-7R signaling by recruiting spleen tyrosine kinase (Syk), which phosphorylates B cell linker protein (BLNK), which then inhibits PI3K/AKT. The pre-BCR further depresses proliferative signaling by upregulating IKAROS and AIOLOS (via RAS-ERK), both of which repress Cyclin D3. The pre-BCR also induces Igκ recombination by both derepressing and activating FOXO1, which then induces RAG1/2 and transcription factor IRF4. The RAS-ERK signaling activated by the pre-BCR also serves to upregulate E2A and RAG1/2, both of which are required for Igκ recombination. Signaling from CXCR4, which is activated in response to the binding of the ligand CXCL12, supports and perpetuates signaling from the pre-BCR. CXCR4 helps to drive Igκ recombination by upregulating RAS/ERK signaling module, BRWD1 and IRF4/IRF8/SpiB. Additionally, CXCR4 signaling supports cell survival by repressing BIM/BID, and upregulating NFκB. Adapted from McLean *et al*, 2020.

Regulation of cell cycle by the pre-BCR

Initially, pre-BCR signaling (Figure 1.6) activates Src kinases such as Lyn, Fyn and Blk that phosphorylate ITAMs of the cytoplasmic tails of Igα and Igβ facilitating recruitment and activation of the spleen tyrosine kinase (Syk) [56, 57]. Syk signaling leads to the induction of B cell linker protein (BLNK also known as SLP-65), Bruton tyrosine kinase (BTK) and PLCγ2. BLNK activation is coupled to Syk by recruitment to the phosphorylated ITAM of Igα [58]. Loss of Syk, BLNK, BTK or PLCγ2 results in a block in early B cell development [59, 60, 61]. Thus, the central event of pre-BCR signaling is activation of the Syk-BLNK module, which eventually functions to cease proliferation downstream of pre-BCR.

The Syk- BLNK axis also represses the PI3K-AKT pathway, which was upregulated by IL-7R signaling and actively represses PAX5. The inhibition of PI3K-AKT depresses PAX5 and FOXO1, both of which are necessary for optimal expression of SYK and BLNK. BLNK also induces activation of the mitogen-activated protein kinase p38, which phosphorylates and augments FOXO1 activity [22]. FOXO1 and PAX5 are also necessary for the induction of interferon-regulatory factor 4 (IRF4) expression through SYK–BLNK signaling [22]. IRF4 along with its binding partner IRF8 negatively regulate pre-B-cell proliferation through induction of expression of the TFs Ikaros (encoded by Ikzf1) and Aiolos (encoded by Ikzf3) [62, 28]. Ikaros and Aiolos directly suppress c-Myc expression and induce expression of the cell cycle inhibitor p27 to downregulate cyclin D3 in pre-B cells, facilitating exit from cell cycle [39]. Additionally, the low energy state generated by extensive proliferation activates AMPK to antagonize mTORC1 and restricts protein synthesis. This regulatory function likely contributes to Igκ recombination as AMPK can directly phosphorylate and activate RAG1 [63].

Pre-BCR mediated regulation of Igκ recombination

Following proliferation, cell cycle exit of the pre-B cells is necessary but not sufficient to induce Igκ recombination [45]. Constitutive expression of the pre-BCR results in induction of Igκ recombination [7] (Figure 1.6). In Rag2-/- pro-B cells, transgenic expression of μHC induces Igκ locus accessibility, which is a requirement of efficient Igκ recombination [64, 38]. Furthermore, deficiency of one or more pre-BCR signaling components enhances proliferation and reduces Igκ recombination [61, 65]. These observations suggest the active role of the pre-BCR in light chain recombination.

Downstream of the pre-BCR, the RAS-ERK pathway plays an important role in Igκ recombination by inducing RAGs [28, 66, 67]. Additionally, this signaling increases expression of the transcription factor E2A (encoded by Tcf3) while repressing its inhibitor ID3. This enhances free E2A, which then activates the Igκ enhancers, and regulates accessibility and Igκ germline transcription [28, 68, 69]. Binding of E2A at the intronic enhancer of Igκ locus (iEκ) activates this regulatory element and allows the recruitment of co-transcriptional activators protein (CBP) and p300 to decorate the H3 histones present in the flanking Cκ and Jκ regions with acetyl groups (H3Ac), thereby making the region accessible to the recombination machinery [28, 68, 70, 71]. Furthermore, the RAS-ERK pathway directly induces phosphorylation of serine 10 in histone H3 (H3S10p), which in combination with E2A mediated acetylation of H3K9 and H3K14 (H3K4AcK14Ac) sets up a specific epigenetic landscape (H3K9AcS10pK14Ac) for recruitment of the epigenetic reader BRWD1 (Bromodomain and WD repeat-containing protein 1) at the putative recombination center at Jκ [72, 73]. BRWD1 then re-positions the nucleosomes relative to GAGA motifs (where ERK can also directly recruited to recruit RAG complex binding and making Jκ segments accessible for recombination [74, 72].

In addition to E2A, pre-BCR signaling induces the expression of IRF4 and IRF8 required

for Igκ recombination, silencing pre-B cell proliferation and suppressing the SLC [62, 45, 75, 76]. IRF4 and IRF8 also complement the function of E2A by binding and activating both of the Igκ enhancers, iEκ and the 3' enhancer located 9kb downstream of iEκ [45].

1.1.5 CXCR4 signaling works with pre-BCR signaling to regulate early B cell differentiation and Igκ recombination

The movement of pre-B cells away from IL-7^{high} microenvironments by CXCR4 can explain the initiation and subsequent dominance of pre-BCR signaling. However, there are several questions that remain unanswered. First, given that activation of pre-BCR signaling is associated with concurrent repression of pre-BCR expression, the model fails to explain how the pre-BCR provides continuous signals for recombination of Igκ in small pre-B cells. Second, it is unclear whether initial transient pre-BCR signaling is sufficient to execute the entire developmental program in small pre-B cells, or whether other signals are required.

Withdrawal of IL-7 in vitro was thought to be sufficient to induce pre-BCR activation, cell cycle exit, and subsequent Igκ recombination [22, 28, 37, 45]. However, these experiments were done using the stroma feeder cell line OP9, which expresses CXCL12, thus obscuring the role of CXCR4 signaling [77]. A recent study using high-resolution confocal microscopy on an Igκ reporter mouse model (Ck-YFP) found that most of the small pre-B cells were in tight contact with CXCL12+ stromal cells with high local accumulations of extracellular CXCL12 [49] (Figure 3A inset). Furthermore, the small pre-B cells clearly have CXCL12 in their cytoplasm, suggesting recent CXCL12 internalization. This observation prompted the idea that CXCR4 has an additional signaling role beyond the localization of pre-B cells.

CXCR4 is expressed at various levels in all stages of B cell development in BM from hematopoietic stem cells (HSCs) to mature B cells and plays a major role in the homing of B cell precursors (Figure 3B). It is associated with various cellular processes and malignancies

when it is deregulated [42, 78, 79, 80, 81]. Embryonic lethality in CXCR4 and CXCL12 deficient mice, as well as impaired hematopoiesis, made it difficult to understand the specific role of CXCR4 in the development of pre-B cells [80, 81]. Early studies of CXCR4-deficiency in B cells using either fetal liver $Cxcr4^{-/-}$ chimeras [80] or B lineage specific deletion of a loxP-flanked (floxed) Cxcr4 locus ($Ccxr4^{fl/fl}$) by CD19-Cre [82] revealed a developmental arrest at the pro-B and pre-B stages. However, the $Cxcr4^{-/-}$ fetal liver chimeras were limited by poor generation of B cells and deletion by CD19-Cre is incomplete in early B cell populations with up to 60% of $Cxcr4^{fl/fl}$ x CD19-Cre mice expressing CXCR4 in pre-B cells.

These issues were addressed in a recent study in which CXCR4 was conditionally deleted in developing large pre-B cells by mb1-Cre which provides near-complete deletion in all committed B cell progenitors. These studies revealed that CXCR4 was required to generate small pre-B cells and not earlier progenitor compartments [49]. Additionally, the epigenetic and transcriptional signature of small pre-B cells from CXCR4-deficient mice resembles the wild-type proliferative large pre-B cells, suggesting a developmental block at this stage [49]. To understand the true effect of CXCL12, a novel in vitro pre-B culture was also established that requires no stroma feeder cell line [49]. In vitro cultures of pre-B cells from CXCR4 sufficient and deficient mice with or without IL-7 and with or without CXCL12 confirmed the direct role of CXCR4 in both cell cycle exit and Igκ recombination [49].

A Positive Feedback Loop Drives CXCR4 Signaling

Successful μHC recombination and expression of the pre-BCR leads to the upregulation of the transcription factor IRF4, which in turn activates expression of CXCR4. Once active, CXCR4 acts on IRF4 and increases its expression. This pre-BCR induced IRF4/CXCR4 feed-forward loop makes the cells enter into a motile phase by turning the IL-7 responsive cells into CXCL12 sensitive ones. Thus, even in the presence of IL-7, CXCL12 will predominate

and cause developing B cells to be positioned in CXCL12-rich areas in the BM.

CXCR4 Signaling in Cell Proliferation and Survival

The CXCR4/CXCL12 signaling pathway serves to promote cell survival, as well as support the role of the pre-BCR in inhibiting proliferation. CXCR4 induces the expression of NF-κB, a promoter of cell survival. Additionally, it represses the pro-apoptotic proteins BIM and BID. CXCR4 also aids the pre-BCR in promoting cell cycle escape by amplifying a feed-forward loop of ERK activation. This induces Aiolos, which is necessary to cease cell cycle via repression of cyclin D3. Overall CXCR4/ CXCL12 signaling in pre-B cells complements almost all of the functions initiated by the pre-BCR. This includes promoting survival by maintaining the Mcl1 level and the continuation of the differentiation programs, even when the pre-BCR surface expression is attenuated by CXCR4 [49].

CXCR4 Promotes Igκ Recombination and Late B Lymphopoiesis

The CXCR4/CXCL12 signaling pathway (Figure 2A) coordinates the induction of both transcription factor networks and chromatin remodeling complexes that dictate which regulatory sites are open to transcription factor binding at late B lymphopoiesis [3, 83, 49]. It induces the expression of IRF4 and NF-κB, which are both critical for late B lymphopoiesis [9, 83, 49]. Additionally, CXCR4 signaling enhances accessibility at sites bound by multiple mediators of late lymphopoiesis, including IRF4, IRF8, E2A, SPIB, PAX5 and FOXO1. Conversely, sites bound by early mediators of B cell development, such as MYC and STAT5, are closed by CXCR4 signaling. As CXCR4 induces the expression of BRWD1, it is likely that at least some CXCR4-dependent changes in chromatin accessibility are mediated by BRWD1, which both opens enhancers of late lymphopoiesis and represses those targeted by early transcription factor developmental programs [83]. Additionally, CXCR4 facilitates the

cells reaching the IgM expressing immature state of development [49].

Along with promoting late B lymphopoiesis, CXCR4 also is essential for productive Igκ recombination. CXCR4 induces the transcription of the essential recombination proteins, RAG1 and RAG2. CXCR4 also induces the expression of BRWD1, which plays an essential role in increasing chromatin accessibility at Jκ, permitting RAG1 and RAG2 to bind and cleave DNA. Altogether, CXCR4 signaling supports signaling from the pre-BCR and is essential for developing B cells to enter the late stages of B lymphopoiesis.

1.1.6 V(D)J Recombination

For B cells to produce an antigen-specific BCR to any pathogen they might encounter, they need to produce a unique BCR for each antigen, requiring the capacity to produce, in effect, millions of unique proteins. However, in the central dogma of biology, one gene encodes one protein, and there are an estimated 30,000 genes in the human genome. This produces a problem for the creation of many unique antigen-specific receptors, as they cannot all be encoded by different genes. The solution to this problem lies in gene recombination. Rather than having a different gene for each antibody, the genome instead encodes only a few immunoglobulin (Ig) genes that are themselves composed of hundreds of gene segments that can be selected and recombined to make new, functional, protein-coding genes by a process known as V(D)J recombination. There is a so-called heavy chain immunoglobulin (IgH or Igμ), and a light chain immunoglobulin gene (IgL), which can be formed by recombining one of two possible genes Igκ or Igλ. Developing B cells will always try to recombine first at kappa, and if unsuccessful, will attempt to recombine the lambda chain.

Mechanics of V(D)J Recombination

Each functional gene segment in an immunoglobulin gene is flanked by a recombination signal sequence (RSS). The RSS is highly conserved, and consists of a heptamer and a nonamer DNA sequence, separated by a 12 or 23 base pair long linker sequence [73]. Generally, recombination is guided by the 12-23 rule, with recombination between a 12RSS and a 23RSS being strongly preferred [84]. The RSS is recognized and subsequently cleaved by the RAG recombinase [85]. The active RAG recombinase is complex made up of recombination activating gene 1(RAG1) and RAG2. RAG1 contains the primary DNA binding domain in the complex that is responsible for binding to the RSS nonamer sequence, as well as the active site for DNA cleavage [73]. RAG2 is an essential cofactor for recombination. It contains a plant homeodomain (PHD) finger that binds trimethylated histone H3 lysine 4 (H3K4me3) marks. Thus, it is thought to help guide the RAG complex to regions of active chromatin, and enhance the catalytic activity of the RAG complex [85, 86].

To initiate recombination, RAG must produce a double strand break at the RSS. In the case of Igκ, the RAG complex is first recruited to an RSS at the Jκ locus. It then captures a second RSS in the Vκ locus, forming a synaptic (or paired) complex [87, 88]. DNA cleavage then occurs through a two-step process by which RAG first produces a single strand nick between the heptamer and the gene segment. The 3' hydroxyl group on the free end then attacks the second strand to create a DNA double strand break, creating a DNA hairpin at the end of the gene segment [73]. In this fashion, the unwanted intervening sequence is spliced out. Following hairpin formation, factors involved in the nonhomogous end joining (NHEJ) DNA repair pathway are recruited to unite the DNA ends.

Sequence and regulation of V(D)J Recombination

V(D)J recombination occurs in a very strict order. Recombination first occurs at the heavy chain locus, with D to J recombination occurring first, followed by V to D(J) recombination. Once the heavy chain has been productively recombined, the light chain undergoes V(J) recombination at the Igκ locus. If this rearrangement is unsuccessful, the cell can then move to the lambda light chain to attempt rearrangement. The use of lambda is relatively uncommon in mouse B cells (10%), but for unknown reasons, it is much more commonly used in human B cells (50%).

One of the major purposes of gene recombination is to produce an incredible diversity of antigen receptors. Thus, it is important that there be a degree of stochasticity built into the choosing of different V, D, and J segments during recombination. However, the requirement for randomness in the system must be balanced against very tight regulation. This degree of regulation of V(D)J recombination is incredibly important and occurs at several different levels. Gene recombination involves the intentional production of double strand breaks, which in other contexts are very dangerous to a cell. Double strand breaks can lead to genomic instability and chromosomal translocations that can lead to the development of lymphomas and leukemias [89, 90, 91]. The first, and perhaps simplest, level of regulation is in the temporal and tissue-specific expression of the RAG proteins. The RAGs are expressed at high levels only in early lymphocyte progenitors, preventing gene recombination in other cell types [73]. The other regulation of gene recombination comes at the epigenetic level. Chromatin accessibility, transcription, and three-dimensional chromatin architecture all play important roles both in regulating gene recombination, and in contributing to the diversity of the rearrangements, and will be discussed at length.

1.2 Epigenetic Regulation of Chromatin Accessibility and Transcriptional Activity

1.2.1 The Primary Structure of Chromatin is Essential for Determining the Accessibility of DNA

The immunoglobulin genes do not exist within the cell merely as linear stretches of DNA. Rather, they are present in dynamic chromatin structures comprised of nucleosome-bound chromatin fibers organized into complex looping structures. Epigenetic regulation is capable not only of making certain regions within the immunoglobulin genes more accessible to cutting by RAG, but also making it possible for the entire locus to contract, which is essential for productive gene recombination.

DNA posed an early question to structural biologists, as in its uncompacted form, the amount of DNA necessary to comprise the human genome would never be able to fit in the nucleus. The solution to this problem is that DNA is compacted into chromatin, the primary repeating unit of which is the nucleosome. A nucleosome is comprised of a 146bp sequence of DNA wound around a histone octamer. Nucleosomes are then connected by short segments of linker DNA into nucleosomal arrays [92]. The nucleosome array defines the primary structure of chromatin, and is classically thought of in terms of the "beads on a string" model. Nucleosomes are arranged in a non-uniform manner throughout the genome. They are highly dense in inactive regions of heterochromatin, but much more variably dispersed in regions of transcribed euchromatin [93]. Histones are depleted both at transcribed gene bodies, as well as regulatory loci such as enhancers, promoters and insulators [94, 95]. Additionally, nucleosome position and occupancy are not set, but are rather subject to regulation, and in some chromatin regions are highly dynamic [96, 97, 98, 99].

1.2.2 Nucleosome Structure

Nucleosome structure is one of the most important factors in determining the accessibility of chromatin. Chromatin accessibility can be thought of as the extent to which DNA is physically able to be contacted by other macromolecules and proteins in the nucleus, such as transcription factors [93]. The degree of accessibility of a region of DNA has a large impact on the regulation and activity of that region. Transcription factors promote the transcription of regions of DNA and help to recruit polymerases, and most cannot bind at inaccessible, nucleosome-dense regions of chromatin. To help promote transcription, most promoters of active genes have a well-defined nucleosome-depleted region (NDR), consisting of two strongly positioned nucleosomes up and downstream of the transcription start site (TSS) [100]. The NDR is able to remain free of nucleosomes through the concerted work of a number of factors including BRG1/BRM-associated factor (BAF) and promoter-proximally paused RNA polymerase [93, 101]. NDRs can also be found at CTCF-bound insulators and transcription-factor bound enhancers. It is thought that a combination of ATP-dependent chromatin remodelers and occupancy by transcription factors can help regions of DNA remain free of nucleosomes [102, 103]. Additionally, DNA sequence is a strong intrinsic determinant for nucleosome positioning, due to the ability of various DNA sequence patterns to bend around histones [104].

1.2.3 Histone Modifications and the Epigenetic Landscape

In addition to the presence and density of nucleosomes, additional considerations can affect the practical accessibility of chromatinized DNA to transcription factors and other proteins. One critical component is the histone itself. Histones can be post-translationally modified, and these modifications have a number of effects upon chromatin and transcriptional regulation. Histone modifications can add or neutralize charge, making DNA bind

more tightly or loosely to the histone itself. Such charges can also encourage or discourage chromatin from associating into more complex higher order structures, such as 30nm fibers. Finally, epigenetic readers can recognize and be recruited to chromatin by posttranslational modifications on histones. The number and variety of extant histone modifications is extensive, but two of the more common types of modifications are histone acetylation and histone methylation.

Histone acetylation occurs on lysine residues in the histone amino terminal tail. A family of enzymes known as histone acetyltransferases (HATs) catalyse the transfer of an acetyl group onto the lysine side chain [105]. The effect of this is to neutralize the lysine's positive charge, which can weaken the interactions between the histone and negatively charged DNA. Thus, generally, histone acetylation is a mark of transcriptionally active chromatin. Histone deacetylases (HDACs) oppose the actions of the HATs by reversing lysine acetylation. This acts to stabilize chromatin and repress transcription [105].

Histone methylation primarily takes place on the side chains of lysine and arginine residues of N-terminal histone tails [105]. However, unlike acetylation, methylation does not impact the charge of a histone, making its impact on transcription somewhat more complex. Methylation of histones can result either in an increase or decrease of transcription of the underlying DNA, depending on the particular modification. Histones are methylated by two major classes of enzymes: histone lysine methyltransferases (HKMTs) and arginine methyltransferases (PRMTs). Also, unlike acetylation, methylation can involve the transfer of one, two, or three methyl groups [105]. These marks can then be differentiated by different histone readers.

1.2.4 Chromatin Remodeling Complexes

NURF

NURF is a large, highly conserved, multi-subunit complex that is embryonic lethal when knocked out [106]. It was first identified in Drosophila as an ATP-dependent factor that enhanced the ability of GAGA factor to increase the accessibility of chromatin in vitro [107, 108]. Drosophila NURF has four subunits, three of which have human homologs, NURF301, ISWI, NURF55 and NURF38 [106]. NURF301 contains an AT hook, which is thought to facilitate direct interaction with DNA, a poly-glutamate region, tow PHD fingers, and a bromodomain, which is essential for binding histone H3 di/trimethyl-K4 (H3K4me2/3) and histone H4 K16-acetyl (H4K16ac) marks [109, 110, 111, 112, 113]. ISWI is notable for its ATPase domain, which is essential in all chromatin remodeling proteins [114]. Together, NURF301 and ISWI are sufficient for the chromatin remodeling activity of NURF [109]. The final two subunits are NURF55, which is characterized by its WD repeat domain, and NURF38, which evinces a strong inorganic pyrophosphatase activity [115]. The WD-repeat domain is a highly conserved domain that is commonly found in chromatin-associated complexes [106]. NURF38 has no human homolog, and its significance to the complex remains unclear [106].

In Drosophila, NURF remodels chromatin by sliding nucleosome in 10 bp bidirectional steps, during which there is minimal unwrapping of the DNA from the nucleosome [116, 117]. In vitro, this remodeling is affected by the length of linker DNA and the inherent strength of the DNA positioning sequence for the nucleosome [118, 119]. NURF will slide nucleosomes either into a thermodynamically stable position, or to the end of a particular DNA fragment [116, 117, 118]. This nucleosome sliding can be impeded by various DNA binding factors or adjacent nucleosomes acting as barriers. Additionally, post-translational modifications to

the histones themselves play a role. NURF can be recruited via H3K4me2/3 and H4K16ac marks via its NURF301 domain; however, it has also been noted that H4K16 and K4K8 marks inhibit the ATPase activity of ISWI [120, 121, 122].

SWI/SNF

The SWI/SNF complex represents a second class of chromatin remodeler. It was originally identified in a genetic screen Saccharomyces cerevisiae for altered gene expression [123]. The complex is conserved in eukaryotes, and is known in mammals as the BRG1/BRM associated factor (BAF). In vitro studies showed that SWI/SNF caused ATP-dependent disruption of nucleosome structure, and that this disruption allowed for increased binding of transcription factors to the DNA template [124]. SWI/SNF complexes are very large, and consist of eight or more proteins [124]. In humans, the BRG1 protein in the complex comprises the minimum functional unit, with some, albeit low, level of ATP-dependent nucleosome remodeling functionality.

SWI/SNF does not require ATP to bind to nucleosomes or DNA, and it has been observed that ATP-independent binding of SWI/SNF leads can lead to the formation of loop domains in DNA and nucleosomal arrays [125]. SWI/SNF does require ATP for nucleosome remodeling, however. In contrast to the NURF complex, which slides nucleosomes without significant unwrapping of the DNA, the SWI/SNF complex is capable of remodeling nucleosomes such that DNA is unwound from the surface of the histones. Such remodeled nucleosomes seem to have twice as much DNA wound much more loosely around the histone core, and show increased sensitivity to DNaseI [124]. In addition to this form of remodeling, SWI/SNF is also capable of sliding histones to other locations on the DNA strand, much like the NURF complex [126]. However, unlike NURF, SWI/SNF is also able to evict histones from DNA [127, 128]. SWI/SNF plays an important role in transcriptional regulation by

making promoters accessible to transcription factors [129]. Indeed, there is evidence that SWI/SNF complexes are required to maintain active levels of transcription in vivo [130, 131].

GAGA Factor

The active regulation of chromatin structure and accessibility is critical for transcription factor binding and gene expression. The role of GAGA DNA motifs in chromatin regulation was first noted in Drosophila. The Trl gene was found to be required for normal expression of homeotic genes, as well as in modifying position-effect variegation [132]. Position-effect variegation is a phenomenon by which genes become abnormally juxtaposed with heterochromatin, resulting in the gene being silenced in some of the cells in which it is normally active. This is an effect of its position in the genome, rather than a change in the genetic sequence of the gene, and genes that mediate position-effect variegation include many epigenetic factors and chromatin remodelers [133]. GAF was initially shown to bind to GA-rich DNA sequences in the Ultrabithorax (Ubx) promoter [134]. It was then discovered that GAF plays a role in the establishment and maintenance of DNase hypersensitive sites [135]. GAF is able remodel nucleosomes in concert with NURF in an ATP-dependent manner, disrupting histone octamers located over GAGA/CTCT sites resulting in nucleosome-free regions of DNA [107, 108]. Thus, GAF plays a critical role in gene regulation by increasing accessibility of gene promoters and other gene elements to transcription factors.

GAF is comprised of three major domains, namely the zinc finger DNA-binding domain, and BTB/POZ domain, and a polyglutamine-rich domain [136, 137, 138]. The BTB/POZ domain plays a role in the self-oligomerization of the GAF complex [139]. The polyglutamine-rich domain is the least understood domain, but may play a role in transcriptional activation [136]. It is the zinc finger domain that is most crucial to the role of GAF. The domain is a classical C2-H2 zinc finger preceded by a basic helix and flanked by 3 short basic resides:

BR1, BR2 and BR3 [140, 141]. The consensus binding sequence of GAF was determined to be GAGAG, although there is also evidence that the trinucleotide GAG is sufficient for GAF recognition and binding [140, 142].

1.2.5 BRWD1 is an Important Histone Reader and Epigenetic Modulator

Brwd1 (Bromodomain and WD Repeat Domain Containing 1) is a member of the dual bromodomain and WD40 repeat protein families [143]. BRWD1 has several domains of note. The bromodomain is a domain common to epigenetic readers, and facilitates binding to acetylated lysine residues in histone tails. Brwd1 contains two central bromodomains which are predicted to recognize histone 3 lysine 8 acetylation (H3K8Ac), histone 3 serine 10 phosphorylation (H3S10p) and histone 3 lysine 14 acetylation (H3K14Ac) [144]. Additionally, Brwd1 has eight N terminal WD repeat domains. WD repeat domains are generally thought to serve as a scaffold for protein interactions, and help to coordinate multi-protein complex assemblies. Brwd1 also contains a polyglutamine region that in vitro shows activity as a transcriptional activator [143]. Though its activity was poorly understood prior to 2015, it had been shown to coimmunoprecipitate with BRG1, a component of the SWI/SNF complex, furthering the possibility of its role as a transcriptional or epigenetic regulator [143].

In 2015, ChIP-seq experiments found that Brwd1 binds at regions of the genome containing H3S10pK14Ac and H3K9Ac marks [72]. Additionally, de novo motif analysis found that BRWD1, when binding coincident with H3S10pK14Ac marks, was enriched at DNA with stretches of repetitive GA sequences ("GAGA motifs") [72]. Throughout the genome, BRWD1 binding is associated with increased DNA accessibility and nucleosome depletion, as measured by ATAC-seq. In particular, peaks coincident for BRWD1, H3K9Ac, and H3S10pK14Ac had a strong tendency to be free of nucleosomes. Interestingly, the effect

of nucleosome depletion seems to be increased at GAGA motifs. In WT cells, when BRWD1 bound in regions enriched in H3K9Ac and H3S10pK14Ac marks, local GAGA motifs were generally free of nucleosomes. However, in Brwd1KO cells, these GAGA motifs tended to be occupied by nucleosomes [72]. This suggests that Brwd1 may be acting as a sort of eukaryotic GAGA factor, playing a role in sliding or evicting nucleosomes from GAGA motifs.

Brwd1 has two notable associations with human disease. First, it lies on chromosome 21, within one of the Down syndrome critical regions [143]. Down syndrome is known to cause heart disease, craniofacial abnormalities, and intestinal and immune issues, in addition to the characteristic mental and developmental disabilities. Additionally, mutations in Brwd1 have been associated with hypogammaglobulinemia, suggesting that it can cause B cell deficiencies [83].

The Role of Brwd1 in B Cell Lymphopoiesis

Brwd1 is critical for targeting Igκ for recombination in small pre-B cells [72]. The role of Brwd1 in regulating Igκ recombination was first postulated during an examination of the genes that are repressed by STAT5 in pro-B cells. STAT5 generally stably-silences its target genes, even through multiple successive stages of development. However, two STAT5 target genes are rapidly induced at the small pre-B stage: Igκ and Brwd1. Brwd1KO mice display a developmental defect in B lymphopoiesis, with a significant block starting at the small pre-B stage [72]. Further studies showed the germline transcription of Igκ was decreased twofold, and Igκ recombination was reduced approximately fivefold. In the transition from the pro-B to the small pre-B stage, Igκ, which is devoid of active histone marks in the pro-B stage, becomes decorated with H3K9Ac and H3S10pK14Ac at the Jκ, iEκ and Cκ loci at the small pre-B stage. ChIP experiments showed that Brwd1 is recruited to the Jκ and Cκ chromatin regions containing H3K9Ac and H3S10pK14Ac. These marks are deposited

in both WT and Brwd1KO cells, suggesting that Brwd1 does not lead to the deposition of these marks, but rather is recruited to a specific pre-existing epigenetic landscape [72]. This landscape is likely established by ERK downstream of the pre-BCR.

Brwd1 plays an important role in establishing the accessibility at Igκ required for gene recombination [72]. In the absence of Brwd1, Jκ in small pre-B cells is roughly half as accessible as it is in WT cells, as assessed by ATAC-seq. Furthermore, when examining the actual positioning of nucleosomes at Igκ, it was found that in WT small pre-B cells, nucleosomes were positioned in between Jκ segments, with the RSSs and gene segment bodies largely free of nucleosomes. However, in Brwd1KO small pre-B cells, nucleosomes were positioned over the RSSs and gene bodies at Jκ, making them less accessible to recombination machinery. Indeed, RAG recruitment was fourfold lower in Brwd1KO cells as compared to WT [72].

In addition to facilitating recombination, Brwd1 plays an important role in establishing the epigenetic landscape required to enable late B lymphopoiesis [83]. RNA-seq analysis of Brwd1KO and WT small pre-B cells revealed that over 7000 genes were dysregulated in the KO cells. Upon further examination, it was found that BRWD1 induced B-cell activation and transcription genes, and repressed proliferation and metabolism genes, suggesting that Brwd1 is important in controlling the transition between the proliferative program that dominates early B cell development, and the differentiative program that characterizes later developmental stages [83]. ATAC-seq data revealed that BRWD1 closes the chromatin regions involved in driving early B-cell developmental stages, while opening regions active at the small pre-B cell stage. The ability to both close and open chromatin appears to be related to how far away a site is from the BRWD1 binding site. Near a binding site, it appears that BRWD1 increases accessibility and transcription of chromatin. However, at loci farther than approximately 5kb from a BRWD1 binding site, BRWD1 appears to be playing a role in repressing chromatin accessibility and transcriptional activity [83]. Brwd1

primarily plays a role in altering the accessibility of enhancer regions. This is important because though Brwd1 does not play a role in regulating the expression of transcription factors, it does alter the accessible regions in which that transcription factor can bind [83]. This is yet another example of the many subtle and powerful ways developmental programs can fine-tune transcriptional regulation.

1.2.6 Methods of assessing chromatin accessibility

DNaseI-Seq

It has been recognized since the 1970s that chromatin, when treated with an endonuclease such as DNaseI, will be digested down to fragments of 100-200bp long, representing DNA protected from digestion by nucleosomes [145, 146, 147]. This method evolved into DNA footprinting, in which it was recognized that DNaseI digestion could be impeded by the binding of other proteins, like transcription factors, which would protect the DNA from fragmentation, leaving characteristic "footprints" in the digestion pattern [146]. This method was later combined with PCR, and later next generation sequencing, to be able to quantitatively assess chromatin accessibility throughout the genome [148, 149, 150, 151, 152].

DNaseI hypersensitivity evolved into DNase-seq with the advent of next-gen sequencing methods [153, 154]. In DNase-seq, chromatin is first digested with DNaseI. The cut fragments are then assembled into a sequencing library and subjected to high-throughput sequencing. The cut fragments are then mapped back to the genome, and for the identification of regions of open chromatin. It can be further inferred that these sites are likely active regulatory sites [93].

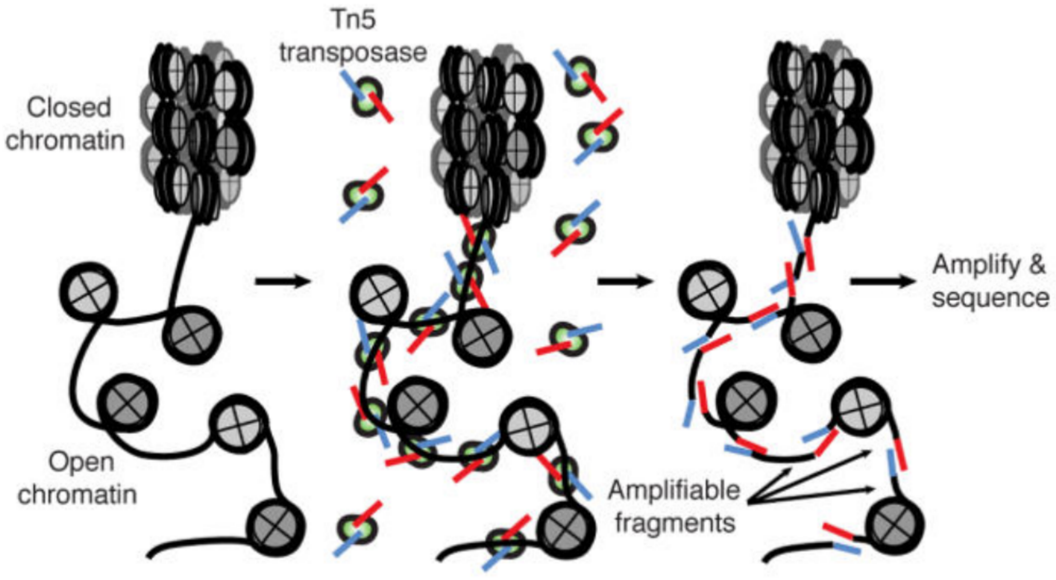

Figure 1.7: Assay for transposase-accessible chromatin using sequencing (ATAC-seq)
A schematic of the library preparation for ATAC-seq. Chromatin is isolated from cells, and incubated with Tn5 transposase, which will insert sequencing tags into regions of accessible chromatin throughout the genome. These tags are then amplified and sequenced, allowing for the identification of regions of open chromatin. [Adapted from Buenrostro *et al*, 2015]

ATAC-seq

Assay for transposase-accessible chromatin using sequencing (ATAC-seq) is another commonly used method to assess chromatin accessibility genome-wide [155] (Figure 1.7). This method uses a hyperactive Tn5 transposase to insert sequencing tags into regions of accessible chromatin. ATAC-seq has gained in popularity due to its ease and ability to perform on relatively small populations of cells. The reads generated from the inserted sequencing tags identify areas of the genome that are more accessible, and thus likely to be regulatory regions. Furthermore, ATAC-seq can be used to determine nucleosome occupancy and positioning.

ChIP-seq

ChIP-Seq, or Chromatin Immunoprecipitation with Sequencing, is the primary tool for researchers to determine not only where transcription factors and other proteins bind throughout the genome, but also to determine where in the genome various histone modifications occur [156]. ChIP was first developed as a method to study DNA-protein interactions [157]. In this method, the proteins are covalently cross-linked to the DNA with formaldehyde or another cross-linking agent to preserve interactions. The DNA is then sonicated to shear it into smaller fragments. Next, an antibody against the protein of interest is used to pull down complexes of DNA bound to the protein of interest. Finally, the cross-links are reversed, and DNA is purified for further analysis. More recently, the advent of cost-effective next generation sequencing has allowed for the high throughput sequencing of all of the pulled down DNA fragments [158]. These fragments are then aligned to the genome, allowing for researchers to understand where different proteins are binding throughout the genome. This technique has also been adapted to study histone modifications [159, 160]. Using antibodies raised against specific histone modifications, the ChIP-seq protocol can be modified to pull down DNA sequences associated with a histone with a particular modification. This technique has been incredibly useful in determining which histone marks are activating and which are repressive, and in turn, which areas of the genome are active in a given cell type.

1.2.7 Epigenetic Regulation of Igκ Recombination

Epigenetic regulation is essential for coordinating the timing of recombination at the Igκ locus. Prior to recombination, while developing B cells are actively proliferating, it is important to prevent cleavage of the Igκ locus by the RAG proteins to avoid genomic instability. One level of regulation comes in at the temporal and tissue-specific expression of RAG. However, to prevent unintentional cuts at unintended sites (like cryptic RSSs),

genome accessibility plays an important role in the regulation of RAG. RSSs bound within nucleosomes are resistant to cleavage by RAG [161, 162]. On the basis of sequence alone, the RSS, particularly the nonamer component, should be a preferred nucleosome binding region. Indeed, in vitro and on plasmids in vivo, the RSS does serve as a nucleosome-positioning sequence [163]. Thus, one would expect that the RSS should exhibit a high frequency of nucleosome occupancy in the immunoglobulin loci. However, there seems to be no clear correlation between RSSs and nucleosome occupancy [164]. This observation lead to the development of a model in which ATP-dependent nucleosome remodeling complexes are actively involved in increasing accessibility at RSSs prior to recombination [165].

Additionally, an intricate signaling network regulates the deposition of activating and repressive histone marks that regulate the accessibility of the Igκ locus. IL-7R signaling in pro-B cells not only drives proliferation, but also results in the deposition of repressive epigenetic marks at the Igκ locus. IL-7R signaling leads to the phosphorylation and activation of STAT5, which then directly binds within the Igκ intronic enhancer (iEκ) as a tetramer and recruits the Polycomb repressive complex 2 (PRC2), which includes the methyltransferase subunit enhancer of zesta homologue 2 (EZH2) [37] (Figure 1.8i). EZH2 then modifies nucleosomes at the iEκ, Jκ, and Cκ regions with the histone H3 lysine 27 trimethylation (H3K27me3) mark. This mechanism of repression does not extend to Vκ regions, which are devoid of either activating or repressive histone marks [3, 36]. Rather, another downstream target of IL-7R signaling, Cyclin D3, serves as a potent repressor of Vκ transcription [35]. The unique action of cyclin D3 on Vκ appears to be due to the differential compartmentalization of cyclin D3 with Vκ genes within the nuclear matrix. By an unknown mechanism, the capture of Vκ gene-containing topologically associating domains by transcription factories is prevented by cyclin D3, resulting in transcriptional repression [36].

The kappa locus must be made epigenetically accessible to the RAG recombination machinery for successful gene recombination of the light chain. IL-7R signaling is overcome by

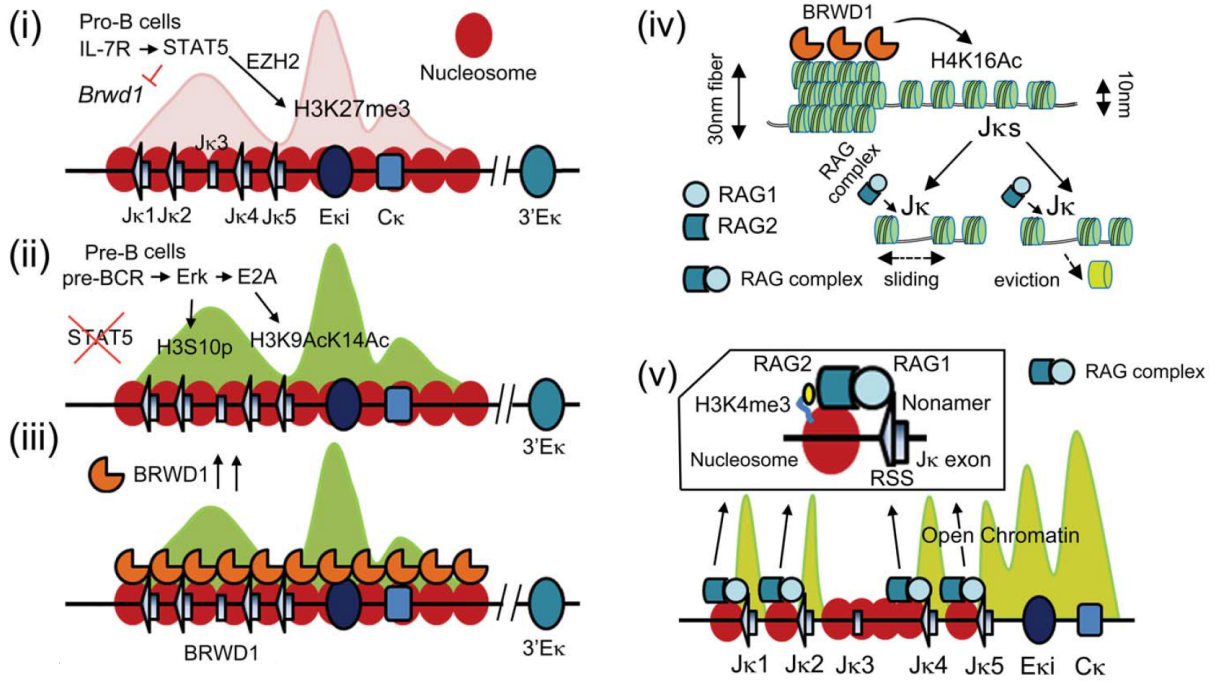

Figure 1.8: Epigenetic Regulation of Igκ recombination
During the transition from the pro-B to differentiating pre-B cell stage, significant epigenetic remodeling occurs in the Igκ locus. **(i)** At the pro-B and large pre-B cell stages, IL-7R-driven STAT5 signaling represses Brwd1 and upregulates EZH2, which lays down repressive H3K27me3 histone marks. **(ii)** In differentiating pre-B cells, when ERK signaling is activated, repressive histone marks are replaced by H3S10p (directly via ERK) and H3K9AcK14Ac epigenetic marks (via E2A). **(iii)** BRWD1 is then recruited by this specific epigenetic landscape to the Ig recombination center. **(iv)** BRWD1 regulates the positioning of nucleosomes. **(v)** this allows the recruitment of the RAG complex and subsequent recombination. Abbreviations:H3K9/K14Ac, histone H3 acetylation on lysine 9/14; H3K27me3, histone H3 lysine 27 trimethylation; H3S10p, histone H3phosphorylation on serine 10; RSS, recombination signal sequence. Adapted from McLean *et al*, 2020.

the concerted activity of the pre-BCR and CXCR4 signaling [49]. In the absence of IL-7R driving STAT5, there is rapid derepression and induction of two of STAT5's target genes, Igκ and Brwd1, at the small pre-B stage [72] (Figure 1.8ii). The pre-BCR and CXCR4 upregulate the RAS-ERK pathway, which induces the expression not only of the RAGs but also of E2A [28, 66, 67]. Binding of E2A at the intronic enhancer of Igκ locus (iEκ) activates this regulatory element and allows the recruitment of co-transcriptional activators protein (CBP) and p300 to decorate the H3 histones present in the flanking Cκ and Jκ regions with acetyl groups (H3Ac), thereby making the region accessible to the recombination machinery [28, 68, 70, 71]. Furthermore, the RAS-ERK pathway directly induces phosphorylation of serine 10 in histone H3 (H3S10p), which in combination with E2A mediated acetylation of H3K9 and H3K14 (H3K4AcK14Ac) sets up a specific epigenetic landscape (H3K9AcS10pK14Ac) for recruitment of the epigenetic reader BRWD1 (Bromodomain and WD repeat-containing protein 1) at the putative recombination center at Jκ [72, 73] (Figure 1.8iii). Binding of BRWD1 at Jκ leads to increased local chromatin accessibility, which exposes the RSS, enabling RAG recruitment and subsequent Igκ recombination [72] (Figure 1.8iv-v). Additionally, BRWD1 inhibits proliferation by coordinately repressing Myc and MYC's downstream targets [83]. In addition to E2A, pre-BCR signaling induces the expression of IRF4 and IRF8 required for Igκ recombination, silencing pre-B cell proliferation and suppressing the SLC [62, 45, 75, 76]. IRF4 and IRF8 also complement the function of E2A by binding and activating both of the Igκ enhancers, iEκ and the 3' enhancer located 9kb downstream of iEκ [45]. Altogether, this intricate epigenetic regulation of the Igκ locus ensures the proper temporal and spatial recruitment of the recombination machinery to allow for productive recombination while protecting the integrity of the cell's genome.

1.3 The Role of Genome Organization in Genetic Regulation and Gene Recombination

1.3.1 Three-Dimensional Chromatin Architecture

Once DNA has been packaged into chromatin, that chromatin is then organized into a highly structured three-dimensional architecture within the nucleus, and this architecture is crucial to the expression and regulation of genes [166]. This 3D organization of the genome can be studied and understood using both microscopy, through techniques such as fluorescence in situ hybridization (FISH), and sequencing, as with chromosome conformation capture (3C) technologies [167].

At the genomic scale, chromatin can be characterized as belonging to one of two compartments. Compartment A generally contains the euchromatin regions [168]. DNA in this compartment is enriched for genes and activating histone marks, and is generally more highly expressed and accessible [168]. Compartment B, by contrast, is more consistent with heterochromatin, and contains more densely packed regions of chromatin. Genes within compartment B are generally inactive and contain more repressive histone marks. Additionally, active and inactive regions of chromatin are spatially segregated, with more actively transcribed regions more often being associated with the nuclear periphery, while heterochromatic regions are more often found in the nuclear center [169].

1.3.2 Topologically Associating Domains

Compartments can be further broken down into topologically associating domains, or TADs. Within the nucleus, chromosomes fold into domains, forming regions of DNA that preferentially interact with themselves over other regions of the chromosome [170, 171].

TADs play an important role in gene regulation, as supported by the fact that genes located within the same TAD tend to be similarly regulated [171]. TADs themselves can contain small subTADs of regions of heavy DNA interaction within a TAD. This DNA interaction can represent stochastic interactions or meaningful enhancer-promoter contacts. One proposed role of TADs in lineage and stage specific gene regulation is by providing a boundary for enhancers and promoters, so that enhancers are able to more easily interact with the appropriate gene-specific promoter.

TADs are separated by boundaries that are rich in CTCF (CCCTC-binding factor) sites [172]. TADs are defined by the results of Hi-C experiments, but many TADs represent a physical loop domain anchored by CTCF and cohesin. Elegant experiments have demonstrated the ability to form stable chromatin loops via a process known as loop extrusion [173] (Davidson 2019). Cohesin, a ring shaped ATPase, encircles DNA and extrudes it as a loop. Cohesin will extrude the DNA loop until it encounters a boundary element, commonly CTCF. For CTCF to block loop extrusion, it must encounter a properly oriented CTCF site [174].

There is strong evidence that loop extrusion plays an important role in D_H-J_H recombination at the Ig_H locus [175]. The D_H-J_H region is contained in a CTCF anchored loop domain. In this loop extrusion model, RAG scans linearly upstream from J_H to D_H as the DNA is extruded. Impediments to loop extrusion, including CTCF-binding elements and transcription, may bias toward the usage of some D segments over others [175]. While there is strong evidence for this model in D_H to J_H recombination, it is unclear if the same principles also apply in V_H to $D_H(J_H)$ recombination in heavy chain, or in $V\kappa(J\kappa)$ recombination in the light chain. Indeed, this model is unlikely to be able to completely explain Ig recombination, due to the variation in orientation of CTCF binding sites and $V\kappa$ gene loci.

1.3.3 Locus contraction

One of the most dramatic examples of the role of 3D chromatin organization in gene recombination is that of locus contraction. In both the heavy and light chain loci, over 100 functional V segments span megabases. Thus, for a distal V segment located far from J segments to successfully recombine, it must be brought in proximity with these segments. This explains the major contraction event of the recombining allele. The uncontracted gene can stretch over 200nm in its uncontracted state, but compresses to half that size or smaller in preparation for gene recombination [36]. For Igκ, this contraction event occurs at the pre-B cell stage [176].

The mechanisms controlling contraction of the Igκ locus have not been completely elucidated; however, there is strong evidence for the role of CTCF, cyclin D3, and RNA polymerase (RNAP). In one study, CTCF was conditionally knocked out in mouse B cell progenitors using the mb1-cre. In these mice, repertoire skewed strongly toward usage of proximal Vκ gene segments, indicating that CTCF is likely important in structuring the Igκ locus so that distal Vκ segments can be recombined [177]. Additionally, FISH experiments showed that cyclin D3 regulates the monoallelic association of RNAP with Vκ, and that it is the allele where Vκ and RNAP have colocalized that undergoes contraction [36].

1.3.4 Transcription plays an important role in immunoglobulin gene accessibility and recombination

One largely open question within the study of Igκ recombination is that of the role of transcription, and particularly, antisense transcription in the regulation of Igκ accessibility, recombination, and regulation of 3D chromatin architecture. Clues exist from work at the heavy chain locus, but it is uncertain whether similar principles extend to the light chain

locus.

One of the first studies to speculate at the role of antisense transcription in gene recombination was a study of heavy chain recombination in which the authors observed that antisense transcription across the V_H regions correlated temporally with recombination [178]. They found that the V_H region was heavily transcribed across what appeared to be multiple V_H segments specifically in pro-B cells. This transcription occurred after D to J_H rearrangement, and ceased following V_H to $D(J_H)$ rearrangement. The authors proposed that this transcription could be playing a role in temporally opening up the V_H regions to recombination, ensuring that recombination at V_H follows recombination between D and J_H loci [178].

Further support for the role of antisense transcription in V(D)J recombination came from a study in which the authors deleted the 100kb section of DNA in between V_H and D= gene loci in mice [179]. They seem to have successfully removed important control elements, as they found high levels of antisense transcription from within the D locus extending into V_H. Furthermore, they observed V_H to $D(J_H)$ rearrangements in developing T lymphocytes. It is not uncommon to observe $D(J_H)$ rearrangements in T cells, but V_H to $D(J_H)$ rearrangements generally do not occur. This supports the role of antisense transcription in facilitating recombination, at least in the heavy chain.

As was previously discussed, the Jκ regions are heavily epigenetically regulated through the post-translational modifications of histones, and the remodeling of chromatin. However, Vκ regions seem to be largely devoid of activating or repressive histone marks [3, 36]. Thus, while the Jκ regions seem to regulate accessibility epigenetically, it is possible that the Vκ regions regulate accessibility through transcription. It remains to be seen if and how those mechanisms are related.

CHAPTER 2

MATERIALS AND METHODS

Mice

Wild type (C57BL/6), *Jκ1-GAGA-Δ40* (C57BL/6), *Jκ1-GAGA-Δ57* (C57BL/6), *Jκ2-GAGA-del* (C57BL/6), and *Brwd1KO* (C57BL/6) mice were housed in a clean animal facility at the University of Chicago. Experiments were conducted when the mice were 6-12 weeks of age, according to IACUC protocol. Both male and female mice were used.

Crispr-Cas9 gRNA Design

Site-specific crispr-Cas9 guides were designed flanking the region targeted for deletion (Table 2.1). Guides for the targeted regions were designed using the Crispr-Cas9 design resource hosted by MIT (http://crispr.mit.edu) as well as the site CHOPCHOP (https://chopchop.cbu.uib.no), which identify possible guides as well as provide predicted efficiency and off-target scores. Guides were optimized for high efficiency and low off-target scores.

Guide	Sequence
Jκ1-GAGA-del1	5'-ACCCCCTCTCCAAGCATGCG-3'
Jκ1-GAGA-del2	5'- TGCTCTGTTCCTCTTCAGTG-3'
Jκ2-GAGA-del1	5'- CTTACTCTGAAACCAGATTC -3'
Jκ2-GAGA-del2	5'- ACTTGTGTTAATTATTACAC -3'

Table 2.1: CRISPR-Cas9 guide RNAs

gRNA-Crispr-Cas9 Ribonucleoprotein Production

The designed guides were ordered from IDT as Alt-R CRISPR-Cas9 crRNA oligos, along with a tracrRNA (IDT 1072532), a universal 67mer that hybridizes to the site-specific crRNA, and includes the binding site for the Cas9 enzyme. To generate an active guide RNA, these two RNAs were annealed at a 1:1 molar concentration. This guide was then incubated with the Alt-R S.p. Cas9 Nuclease (IDT 1081058) to produce the active ribonucleoprotein (RNP) which was injected into mouse embryos.

To prepare the RNP, the crRNA and tracrRNA were first resuspended in injection buffer (1mM Tris-HCl, pH 7.5; 0.1mM EDTA) to a final concentration of 1μg/μL. The crRNA and tracrRNA were then mixed at a 1:2 ratio by mass (5μg crRNA and 10μg tracrRNA), and annealed in a thermocycler (95°C for 5 min, and ramping down to 25°C at 5°C/min). A 100μL injection mix was prepared containing the annealed guide RNA and the cas9 protein. For the Jκ1-GAGA-del mouse injection, the mix was prepared with 25 ng/μL of each guide, and 100 ng/μL Cas9. This was later adjusted to 300ng/μL of guide RNA and 300ng/μL Cas9 for the Jκ2-GAGA-del injection. The mixture was incubated at room temperature for 15 minutes to allow for the formation of RNP complexes. For the Jκ2-GAGA-del injection, cas9 mRNA was also added, at a concentration of 100ng/μL. To remove solid particles that could clog the microinjection needles, the injection mix was then centrifuged at 13,000 RPM at room temperature, and the top 80μL was taken off and kept on ice until injected.

CRISPR-Cas9 Injection and Screening

The CRISPR-Cas9 injections were carried out by the University of Chicago Transgenics Core. Briefly, 1 cell-stage fertilized embryos were injected with the RNP mix described above. The fertilized and injected embryos were then implanted into a pseudopregnant mouse.

Due to the heterogenous nature of Crispr-Cas9 gene editing, it was necessary to screen both alleles for edits separately in the F0 generation of injected mice. Thus, the region surrounding the targeted DNA was amplified by PCR, and products were cloned into a TA vector. Genotyping vectors were then transformed, and a minimum of 8 individual colonies were screened. This ensured that there would be less than a 0.5% chance that a heterzygous mutation went undetected. The genotyping primers are displayed in Table 2.2.

Primer	Sequence
Jκ1Seq489F	5'-AGAGGCTGTCAGATTCCTTGC-3'
Jκ1Seq489R	5'-AGCCACAGACATAGACAACGG-3'
Jκ2Seq393F	5'-TTGTACAGCCAGACAGTGGAG-3'
Jκ2Seq393R	5'-GTACACACACTGGTGTCCCTT-3'

Table 2.2: Genotyping Primers for CRISPR-Cas9 edited mice

Assay of transposase-accessible chromatin using sequencing

The assay of transposase-accessible chromatin using sequencing (ATAC-Seq) protocol was adapted from Beuenrostro et al 2013 and optimized [155]. Approximately 120,000 primary mouse small pre-B and immature B cells were isolated by FACS sorting and harvested by centrifuging at 500g for 5 minutes at 4 °C . To prepare the nuclei for ATAC-seq, the cell pellet was washed once with ice-cold PBS and centrifuged again at 500g for 5 min. Cells were lysed with cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2 and 0.1% IGEPAL CA-630). Immediately after lysis, nuclei were spun at 500g for 10 min at 4 °C. Supernatant was carefully pipetted away from the pellet after centrifugation. Immediately after the nuclei prep, the pellet was resuspended in the transposase reaction mix (25 μl 2 Tagment buffer, 2.5 μl Tagment DNA enzyme (Illumina, FC-121-1030) and 22.5 μl nuclease-free water). The transposition reaction was carried out at 37 °C for 30 min.

After transposition, the sample was purified with a Qiagen MinElute kit. After purification, library fragments were amplified using Nextera PCR Primers (IlluminaNextera Index

kit) and NEBnext PCR master mix (New England BioLabs, 0541) for a total of 5 cycles followed by purification with a Qiagen PCR cleanup kit. The amplified, adaptor-ligated libraries were size-selected with Life Technologies' E-Gel SizeSelect gel system in the range of 150–650 bp. We quantified the size-selected libraries with an Agilent Bioanalyzer and via qPCR in triplicate using the KAPA Library Quantification Kit on the Life Technologies Step One System. Libraries were sequenced on the Illumina HiSeq2000 system to generate 7.5 107 to 10 107 50-bp paired-end reads.

Quality control and DNA alignment

All raw sequence data were quality-trimmed to a minimum Phred score of 20 with Trimmomatic51. Alignment to reference genome mm9 was done with BWA52. For ATAC-Seq data, read pairs where one pair passed quality trimming but the other did not were aligned separately and merged with the paired-end alignments.

ATAC-Seq analysis

Read-alignment positions were adjusted according to their strand: +4 bp for + strand alignments, and 5 bp for – strand alignments. We called open chromatin regions using Zinba53 with a window size of 300 bp, an offset of 75 bp, and a posterior probability threshold of 0.8.

For nucleosome positioning, we filtered properly paired alignments by their fragment size. Fragments less than 100 bp in size were considered nucleosome free and were replaced with a single BED region and used as a background. Those with sizes between 180 and 247 bp were considered mononucleosomes and were replaced with a single BED region; those with sizes between 315 and 473 bp were considered dinucleosomes and were replaced with two

BED regions, each spanning half the overall fragment length; and those with sizes between 558 and 615 bp were considered trinucleosomes and were replaced with three BED regions, each spanning one-third of the overall fragment length. The mono-, di- and trinucleosome regions were concatenated and used as the nucleosome signal. The resulting BED regions were analyzed with DANPOS54 with the parameters –p 1 –a 1 –d 20 –clonalcut 0 to identify regions enriched or depleted for nucleosomes.

We obtained DNA footprinting data by combining bigWig enrichment tracks for ATAC-Seq data over specified BED regions (combinations of peak calls or motif hits). Open chromatin enrichment data from ATAC-Seq were generated from the read-adjusted alignments with custom scripts and normalized to reads per million alignments, and nucleosome-positioning enrichment data were obtained from DANPOS54. DNA footprinting scores were averaged over 10-bp bins from enrichment tracks with custom scripts.

For comparative analysis of chromatin accessibility between samples, we used the bedtools bigWigAverageOverBed tool, with the open chromatin bigwig file and a bed file containing an annotation of regions of interest as inputs. The samples were normalized to accessibility GAPDH. For comparative analysis of nucleosome occupancy between samples, we used the bedtools bigWigAverageOverBed tool, with the nucleosome signal bigwig file and a bed file containing an annotation of regions of interest as inputs.

Isolation of Total RNA

Cells were pelleted by centrifugation at 14,000 RPM for 30 seconds. Supernatant was removed, and cells were resuspended in 500μL Trizol (Invitogen 15596026). Cells were incubate at room temperature for 5 minutes to permit complete dissociation of the nucleoproteins, then left at -80°C overnight. The following day, the samples were then allowed to thaw, and 0.1mL of chloroform was added, and incubated at room temperature for 2-3 minutes.

The sample was then centrifuged for 15 minutes at 12,000 g at 4°C. The colorless upper aqueous phase was then removed to a new tube. To the aqueous phase was added .25mL of isopropanol, and the sample was stored at -80°C overnight. The next day, the sample was centrifuged for 10 min at 12,000 g at 4°C. The RNA pellet was then resuspended in 1mL 75% ethanol, and centrifuged for 5 minutes at 7500 x g at 4°C. The supernantant was discarded, and the pellet was allowed to air dry. The dry pellet was then resuspended in 20μL DNase/Rnase-free water.

RNA-seq: quality control and quantification

Total RNA was isolated using Trizol as described. Libraries were prepared using the standard Illumina library protocol (Kit, RS-122-2101 TruSeq Stranded mRNA LT-SetA) before sequencing on the Illumina Hiseq2500. Raw reads were aligned to reference genome mm9 in a splice-aware manner using STAR51. Gene expression was quantified using FeatureCounts52 against UCSC genes, with Ensembl IG genes from mm10 converted to mm9 coordinates with UCSC liftOver.

RNA-seq: Differential expression analysis

Differential expression statistics (fold-change and P value) were computed using edgeR58, on raw expression counts obtained from quantification (either genes or peaks). Pairwise comparisons were computed using exactTest, and multigroup comparisons using the generalized linear modeling capability in edgeR. In all cases, P values were adjusted for multiple testing using the FDR correction of Benjamini and Hochberg. Significant genes were determined based on an FDR threshold of 5% (0.05) in the multigroup comparison.

Clustering, heatmaps and pathway analysis

We performed complete linkage hierarchical clustering of the gene and peak expression levels and plotted the data in a heatmap using the 'hclust' and 'heatmap.2' functions in R. Metascape data portal59 was used for pathway analyses.

Quantitative PCR analysis

Total cellular RNA was isolated with Trizol as described and RNA was reverse-transcribed with SuperScript III reverse transcriptase (Thermofisher 18080051). For qPCR, a total volume of 25 μL containing 1 μL cDNA template, 0.5 μM of each primer (Supplementary Table 3.3) and SYBR Green PCRMaster Mix (Applied Biosystems) was analyzed in triplicate. Gene expression was analyzed with an ABI PRISM 7300 Sequence Detector and ABI Prism Sequence Detection Software version 1.9.1 (Applied Biosystems). Normalized results were calculated using the $\Delta\Delta$Ct method, by which the differences between the test gene and the housekeeping gene in the wild type and knock out samples were calculated, and then the difference between these values was calculated (the $\Delta\Delta$Ct). Then, the expression fold change was given by 2 raised to the negative $\Delta\Delta$Ct. B2M was used as the housekeeping gene.

$$\text{Expression Fold Change} = 2^{-((\text{KO}_{\text{TEST}} - \text{KO}_{\text{HOUSEKEEPING}}) - (\text{WT}_{\text{TEST}} - \text{WT}_{\text{HOUSEKEEPING}})}$$

Primer	Sequence
Degenerate Vκ-F	5'-AGCTTCAGTGGCAGTGGRTCWGGRA-3'
Jκ1-R	5'- AGCATGGTCTGAGCACCGAGTAAAGG- 3'
Jκ1-2-R	5'-CCAACCTCTTGTGGGACAGTT-3'
Jκ2-R	5'-GTACACACACTGGTGTCCCTT-3'

Table 2.3: Quantitative PCR Primers

PCR analysis of Igκ rearrangements

Quantification of Jκ usage by PCR was performed by sequencing the products of Igκ rearrangements. Total cellular RNA from primary WT, Jκ1-GAGA-del, or Jκ2-GAGA-del small pre-B and immature B cells was isolated with Trizol as described and RNA was reverse-transcribed with SuperScript III reverse transcriptase (Invitrogen). Degenerate Vκ and Cκ primers were used along with 2μL of cDNA in a 25μL reaction using the Platinum Taq DNA polymerase (ThermoFisher). The cycling conditions. Then, 2μL of the PCR product was cloned into the pCRII-TOPO TA vector, and transformed into DH5α cells. The resulting colonies were then miniprepped and sequenced. Unique sequences were analyzed for Jκ usage by alignment to the mouse Jκ (Table 3.4) and Vκ sequences using the NCBI BLAST alignment tool (https://blast.ncbi.nlm.nih.gov/).

Jκ	**Sequence**
Jκ1	5'-GTGGACGTTCGGTGGAGGCACCAAGCTGGAAATCAAACG-3'
Jκ2	5'-TGTACACGTTCGGAGGGGGGACCAAGCTGGAAATAAAACG-3'
Jκ3	5'-CACTGTAAATCACATTCAGTGATGGGACCAGACTGGAAATAAAACC-3'
Jκ4	5'-ATTCACGTTCGGCTCGGGGACAAAGTTGGAAATAAAACG-3'
Jκ5	5'-GCTCACGTTCGGTGCTGGGACCAAGCTGGAGCTGAAACG-3'

Table 2.4: Jκ Sequences

Flow Cytometry and Flow Activated Cell Sorting (FACS)

Total bone marrow was extracted from the hind leg bones of mice, suspended in media (RPMI + 10% FBS), and passed through a 70μm filter. Cells were pelleted and resuspended in ACK lysis buffer for 5 minutes to lyse red blood cells. Cells were then washed and resuspended in media. They were blocked with 2.5μL FC block for 30 minutes prior to staining. To obtain total cell counts, a 10μL aliquot of cells were taken for counting. A master mix of fluorescently conjugated antibodies was prepared, and an appropriate quantity was added so that each sample was stained with 2μL of each labeled antibody. Each sample

was incubated with the antibody mix for 30 minutes, then washed. The panel of antibodies for sorting small pre-B cells and immature B cells is listed in Table XX. Small pre-B cells (LinB220+CD43IgMFSClo) and immature B cells (LinB220+CD43IgM+) were isolated by cell sorting with a FACSAria II (BD Biosciences).

Marker	Color	Company	Cat #	Lot #
B220	PerCp Cy5.5	BD Pharmigen	552771	7011541
CD19	APC-Cy7	BD Pharmigen	557655	0007778
CD43	PE	BD Pharmigen	553271	7297616
IgM	APC	BD Pharmigen	550676	7159889
Igκ	BV510	BD Biosciences	742834	9301993
Igl	FITC	BD Pharmigen	553434	5096711

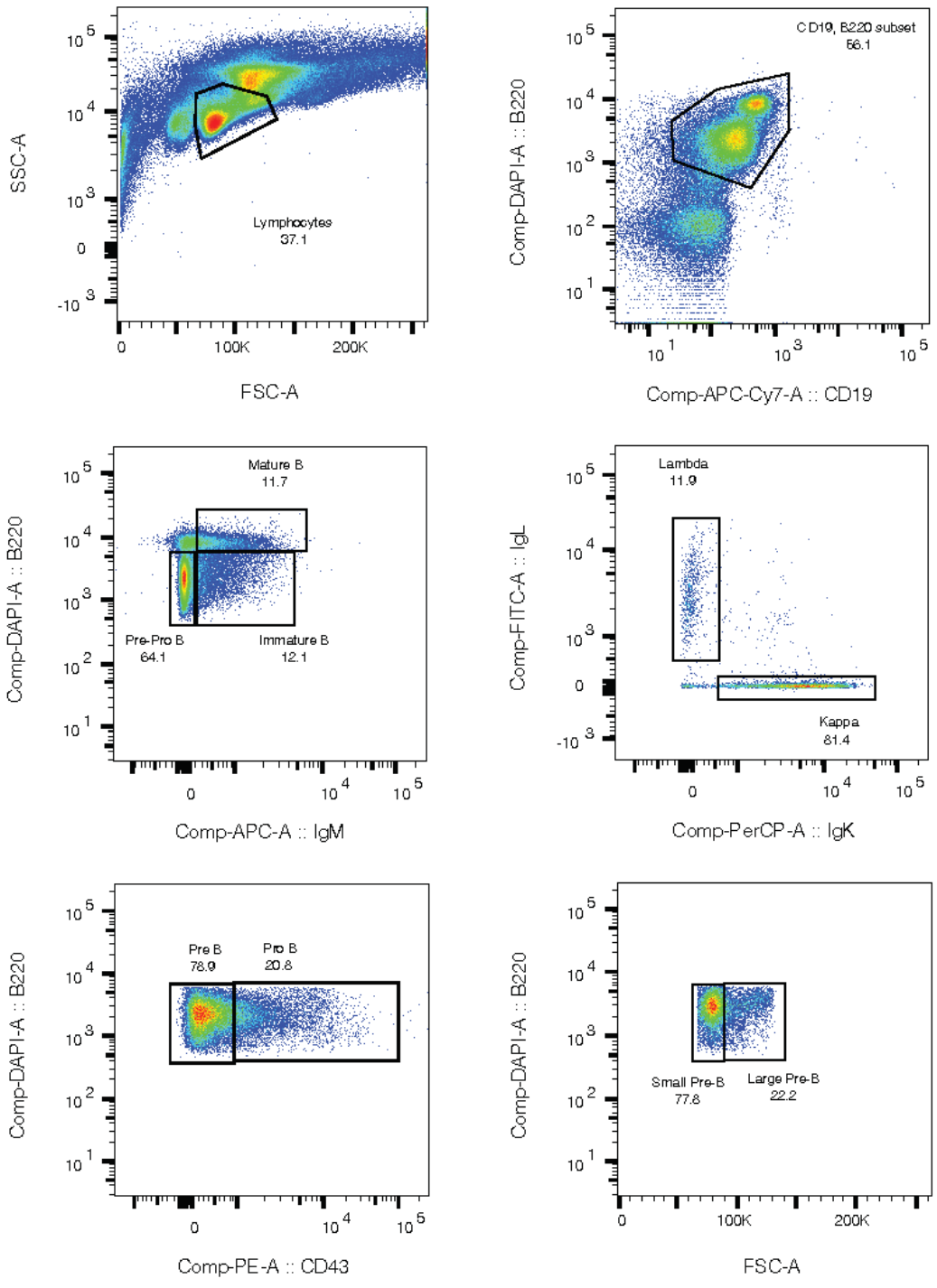

Figure 2.1: Flow gating strategy for bone marrow B cells

HiC Sample Preparation and analysis

Flow sorted WT small pre-B ($5x10^6$ cells per sample) were crosslinked with 1% formaldehyde following wash with ice cold PBS. Then the membranes were lysed keeping the nuclei as intact as possible following restriction digestion with 100U of MboI (NEB R0147). The DNA ends were then marked with biotin, ligated proximally and reversed the crosslinks. DNA shearing and size selection were then performed for fragments 300-500bp. DNA was quantified by Qubit dsDNA High Sensitivity Assay (Life technologies, Q32854) and biotin pull down was performed to prepare final in-situ Hi-C library to be quantified and sequenced using an Illumina sequencing platform.

Sequenced paired-end DNA reads were aligned to the mouse mm9 genome with Bowtie2 and reads were filtered further, with the HiC-Pro and HiC-bench workflow, to verify which read pairs should be used for downstream analysis [180, 181]. HiC-contact files were then generated with Juicer tools. Further analyses were performed using JuiceBox, Homer HiC software, and Hi-C domain caller [180, 182, 183, 184]. The genome was divided into active (A) versus inactive (B) chromatin compartments based on the correlation matrix ("runHiCpca.pl"). Changes in chromatin compartments between B cell subpopulation were determined using "findHiCCompartments.pl". A/B compartments are based on the sign of eigenvector sign, with compartment A defined as a positive eigenvector value, and B defined as a negative eigenvector value [170, 185, 186, 187, 188]. Bed files of compartments were visualized with the IGV browser.

cRSS identification and analysis

To identify cRSSs, we used the online Recombination Signal Seuqences Site (https://www.itb.cnr.it/rss/) [189]. The RIC algorithm was used to identify cRSSs passing

RIC score thresholds of -38.81 for 12- RSSs and -58.45 for 23-RSSs [190]. RIC scores of functional RSSs have been shown to be highly correlated with recombination effiencies [190]. Using this algorithm, for each chromosome, we identified all sites with passing RIC scores. This provided us with the sequences, start and end sites, and RIC scores of all cryptic RSSs.

To analyze nucleosome occupancy, we combined the data of the cRSS start, end, chromosome and unique identifier into an annotation bed format file. For our analysis, we were interested in looking exclusively at cRSSs in small pre-B cells, so we excluded Igκ and Igl RSSs. To accomplish this, using a bed file of Igκ and Igl RSSs, we removed all overlapping sites using the bedtools intersect tool with option -v. This left a bed file of 3,559,481 total cryptic RSSs.

Additionally, we wanted a non-cRSS outgroup to which we could compare the cRSS. We used the bedtools random tool to produce 1,000,000 random tracks of length 27 (equivalent to cRSS-12) and 1,000,000 random tracks of length 38 (equivalent to cRSS-23), and set the seed to 1. These random tracks were then run through bedtools overlap -v with the cryptic RSS bed files to remove any sequences from the random tracks that contained a cryptic RSS. This left us with 923,381 random 27bp sequences, and 911,267 random 38bp sequences.

Next, using the bedtools bigWigAverageOverBed tool, we calculated nucleosome average and sum density over all cRSSs and random sequences, using bigwig files of nucleosome occupancy produced by the DANPOS tool as described for WT small pre-B cells. The average mean nucleosome occupancy score with 95

To correlate the strength of the RIC score and the mean nucleosome occupancy, it was first necessary to separate 12 and 23 cRSSs, as their passing cRSS score ranges differ. Then, separately for the 12 and 23 cRSS datasets, the data was broken into 15 bucketed RIC score ranges, from poor to very good. For each range, the mean nucleosome occupancy was calculated, then the data was reaggregated. This data was plotted, and a linear regression

was run to produce a trendline.

Finally, we performed a cRSS compartment analysis. We first separated the compartment annotations into two bed files: one for compartment A and one for compartment B. Using bedtools intersect with the cRSS bed file, and the option -wa, we first made a new bed file with only cRSSs located in compartment A. Then we did the same for compartment B. From this, we were able to ascertain how many cRSSs were located in compartment A (1793006) vs compartment B (1766138). We then summed up the total length of genome comprising each compartment (59471.819 kb for compartment A and 58121.675kb for compartment B). From this, we calculated the density of cRSSs in compartment A (0.03002462), compartment B (0.03036774), and in the whole genome (0.03026937). Using the density of cRSSs in the whole genome and the size of compartments A and B, we calculated the predicted number of cRSSs per compartment, and found that there were significantly fewer cRSSs than expected in compartment A, and more than expected in compartment B ($p = 1.75 \times 10\text{-}31$).

CHAPTER 3

RESULTS

GAGA motifs are required for Brwd1-mediated nucleosome remodeling and Igκ recombination

3.1 Introduction

The adaptive immune system is tasked with the production of millions of different antibodies that will help the organism to fight off pathogens and develop an immunological memory, and this is achieved in large part through the recombination of the immunoglobulin genes. However, gene recombination presents a very real danger to the integrity of the genome due to the necessity of producing double-strand breaks (DSBs), as off-target cleavage can result in genomic translocations and malignant transformation [191, 192, 193]. Indeed, the process of immunoglobulin cleavage by RAG is tightly regulated. The DSBs involved in recombination occur only at particular regions known as recombination signal sequences (RSSs), which are defined by the presence of a highly conserved nonamer and heptamer sequence, separated by 12 or 23 base pairs. The "12/23 Rule" dictates that recombination occur between one gene segment with an RSS containing a 12bp spacer, and one RSS containing a 23bp spacer. However, the restriction of cleavage to RSSs would be insufficient to maintain genomic integrity, due to the inevitable presence throughout the genome of so-called cryptic RSSs, which are sequences that closely resemble a functional RSS, and thus, could perhaps be cleaved by the RAG proteins.

As a second level of regulation, the immunoglobulin genes are heavily epigenetically regulated. Prior to Igκ recombination, a recombination center is established at Jκ, which is

characterized by high levels of germline transcription, activating histone marks, and increased accessibility [3]. The presence of modified nucleosomes is crucial to the recruitment of the RAG recombinase, because RAG2 binds specifically to trimethylated histone H3 lysine 4 (H3K4me3) via a plant homeodomain (PHD) finger domain [194, 195]. Thus, RAG-mediated recombination is dependent on the local presence of nucleosomes containing the H3K4me3 modification.

Conversely, nucleosomes are also thought to be able to impede RAG mediated recombination. When bound within nucleosomes, RSSs are resistant to cleavage by RAG [162]. *In vitro*, however, the RSS serves as a nucleosome-positioning sequence [161, 162, 163] Thus, for efficient recombination, one would expect a mechanism to have evolved to ensure not only that histones carrying the H3K4me3 modification are present at Jκ, but also that RSSs are free of nucleosomes prior to recombination. However, this mechanism has long remained elusive.

Some clues of the underpinnings of this process came to light in a recent paper on the role of an epigenetic reader and modifier, Brwd1, in regulating Igκ recombination [72]. It was shown that Brwd1 is recruited to a specific epigenetic landscape at Igκ, where it increased local gene accessibility and positioned nucleosomes 5 to each Jκ RSS. The loss of Brwd1 resulted in a dramatic reduction of Igκ recombination, suggesting that the chromatin remodeling it effects at Igκ is crucial to making the region accessible to cleavage by the RAG complex. Genome-wide, there was a notable association between Brwd1-dependent chromatin remodeling and the presence of GA-repeats in DNA ("GAGA motifs"), but it was unclear whether the presence of these GAGA motifs was required for the chromatin remodeling activity exhibited by Brwd1.

There is a biological precedent for the importance of GAGA motifs in local nucleosome remodeling. In Drosophila, an epigenetic modifier known as GAGA Factor, or GAF, plays a critical role in the expression of homeotic genes [132]. GAF binds to GA-rich DNA sequences,

with a GAGAG 5mer being the minimum required consensus sequence [134]. GAF works with NURF in an ATP-dependent manner, remodeling nucleosomes by disrupting histone octamers located over GAGA/CTCT sites [107, 108]. However, it is unknown whether Brwd1 exhibits GAGA Factor-like activity.

Here, we sought to address whether GAGA motifs are required for Brwd1-dependent nucleosome remodeling, and subsequently for recombination at Igκ. To this end, we developed several novel mouse models of the Jκ locus with various GAGA-motif deletions. We demonstrate that GAGA motifs are required for nucleosome remodeling by Brwd1 and efficient usage of Jκ segments in Igκ recombination. Furthermore, our studies shed light on the process of the selection of a Jκ segment for recombination, as well as provide an additional mechanism explaining the extremely low rate of off-target cleavage of cryptic RSSs.

3.2 Results

3.2.1 Establishing a mouse model to examine the role of GAGA motifs in Igκ recombination

A fine examination of the sequence at the Jκ locus revealed that each functional Jκ gene segment is preceded by a GAGA motif, located an average of 80 nucleotides 5' to each RSS, less than the length of one nucleosome (Figure 3.1a). A comparison of the 5' sequence of the Igκ Jκ1 segment across various vertebrate species shows that this 5' GAGAG sequence is generally conserved within a few hundred nucleotides of the Jκ1 RSS (Supplementary Figure 3.1). If this motif is important for Brwd1-mediated nucleosome positioning and Igκ recombination, its removal should result in reduced usage of the corresponding Jκ in the expressed Igκ repertoire. As Jκ1 is the most heavily used Jκ segment, we chose to remove the GAGA motif 5' to Jκ using CRISPR-Cas9 [196].

CRISPR-Cas9 (Clustered regularly interspaced palindromic repeats) is a remarkably powerful and flexible technology that can be used to introduce specific germline gene edits into

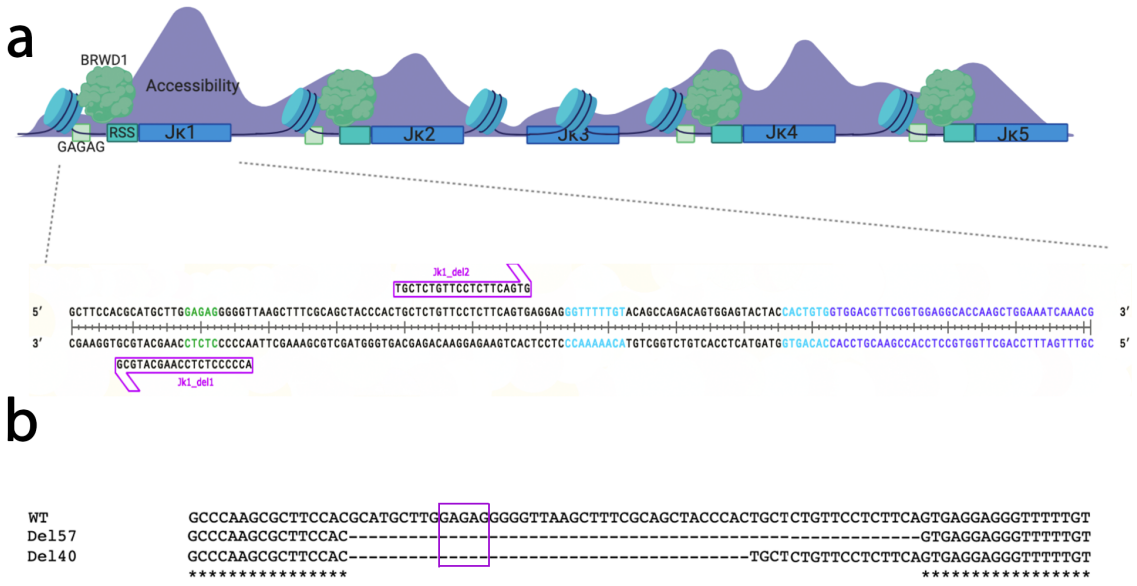

Figure 3.1: Removal of the GAGA Motif at Jκ1
(a) A graphical representation of the Jκ locus. The Jκ locus consists of 5 gene segments (blue): 4 functional gene segments (Jκ1, Jκ2, Jκ4, and Jκ5) and one nonfunctional segment (Jκ3). Each functional segment is preceded by a recombination signal sequence (RSS; teal) consisting of a highly conserved nonamer and heptamer. Additionally, further examination reveals the presence of a GAGA motif (green) located within one nucleosome 5' of the start of each Jκ segment. In wild type cells, just preceding recombination, Brwd1 binds throughout the Jκ locus, and chromatin accessibility (purple) over the RSS and gene segment increases. We designed crispr guides to target the regions up and downstream of the GAGA motif preceding Jκ1 to excise that small and specific region. **(b)** The results of sequencing the CRISPR-Cas9 edited mice. We sequenced each allele of the CRISPR-Cas9 edited mouse pups for gene edits. Displayed are the mutations of the founder lines of Jκ1-GAGA-del mice. Only founders with edits that removed the GAGA motif without disturbing the RSS or Jκ gene body were selected and bred to homozygosity. [Created with BioRender.com and SnapGene]

a variety of model systems. Using a 20 nucleotide guide RNA complementary to a DNA region upstream of a PAM sequence (NGG), Cas9 can be recruited to produce site-specific double strand breaks, which are then repaired using the error prone non-homologous end joining (NHEJ) pathway. This allows one to reliably produce indels (insertions and deletions) at the cut site if one guide is used, or to excise segments of DNA if two guides are used.

Because we wanted to specifically remove a particular segment of DNA, we designed guide RNAs to cut both 5' and 3' of the Jκ1 GAGA motif, and thereby excise the sequence. Guides were designed to minimize off-target effects and maximize cutting efficiency. The target-specific guide RNA was annealed to a tracrRNA (trans-activating crispr RNA), which is bound by the Cas9 protein. This complete guide RNA was incubated with CAS9 protein and injected in single cell mouse embryos as a ribonucleoprotein (RNP) complex.

```
Goat            TGTAAGGAGAGTAG …78bp…GGTTTTTGTTAAAGGGGATATGTCAGGGACGTCACTGTG
Sheep           AAGGGGGAGAGATG …51bp…GGTTTTCGTTAAGGGGGTATGTCAGGAAC-GTCACTGTG
Mouse           TGCTTGGAGAGGGG …48bp…GGTTTTTGTACAGCCAGACAGTGGAGTACTACCACTGTG
Human           ATGTCAGAGAAGAA …78bp…GGTTTTTGTTCAGCAAGACAATGGAGAGCTCTCACTGTG
Horse           TTCACTGAGAGAAG…357bp…GGTGTCTGTGCAGTGACACAGTGGAGAGCTCTCACTGTG
Dog             TTAACTGAGAGAAA…359bp…GGTTTTTGCCCAGTGACACAATGAGGAGCTATCACTGTG
Cat             TTAACGGAGAGAAA…338bp…GGTTTTTGTATATTGACACAATGGGGAGCTATCACTGTG
                      ****           *****  *                          *******

Goat            ATATGCATTCGGCGGTGGGACCAAGGTGGAAATCAAAC-
Sheep           ATATGCATTCGGCGGTGGGACCAAGGTGGAAATCAAAC-
Mouse           GTGGACGTTCGGTGGAGGCACCAAGCTGGAAATCAAACG
Human           GTGGACGTTCGGCCAAGGGACCAAGGTGGAAATCAAAC-
Horse           GTGGACGTTCGGTGCCGGGACCAAGCTGGAAATCAAAC-
Dog             GTGGACGTTCGGAGCAGGAACCAAGGTGGAGCTCAAAC-
Cat             GTGGACGTTTGGCCAAGGAACCAAGCTGGAAGTCAAAC-
```

Figure 3.2: Conservation of the Jκ1 GAGA motif
The alignment of the Igκ region in various vertebrate species. The Jκ1 is shown in green, and the nonamer and heptamer of the Jκ1 RSS are shown in blue. All species show a conserved GAGAG motif (shown in red) 5' to the RSS.

The resulting pups displayed a variety of heterozygous mutations surrounding the Jκ1 GAGA motif (Supplementary Figure 3.2). We selected two pups with mutations that removed the GAGA sequence without disturbing the RSS or Jκ1 gene body to breed to homozygosity. The final result was two different mouse lines, one with a 40 base pair (bp) deletion ("Jκ1Δ40") and one with a 57 bp deletion ("Jκ1Δ57") (Fig 1b). These two lines

carry very similar deletions, but their existence as two separate lines derived from different founders guards against the possibility of any resulting phenomena existing in both lines being the result of an off-target effect.

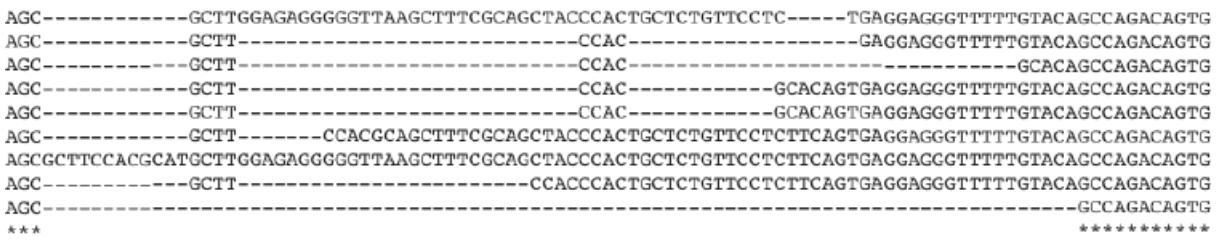

Figure 3.3: Results of Jκ1-GAGA CRISPR-Cas9 Injection
An alignment of the genotyping of all resultant genetically edited pups with the WT sequence.

3.2.2 The 5' GAGA motif domain is required for recombination to Jκ1

To determine how the loss of the 5' Jκ1 GAGA motif would affect repertoire in developing B cells, we sorted small pre-B cells from Jκ1-GAGA deletion mice, and performed quantitative PCR for the recombination product of Vκ-Jκ1 in the GAGA deletion mice relative to WT using a degenerate Vκ primer in conjunction with a Jκ1 primer. We found that in both Jκ1-GAGA-deletion lines, expression of the Vκ-Jκ1 recombination product was down nearly 40 fold (Figure 3.2a).

To assess whether recombination at other Jκ segments was also impacted, we performed quantitative PCR of the Vκ-Jκ4 recombination product in the GAGA deletion mice relative to WT. Surprisingly, Jκ4 usage was also diminished in the GAGA deletion mice, suggesting general recombination defects (Figure 3.2b). However, though the defect seemed to extend throughout the Jκ locus, the effect was most profound at Jκ1. We found that the ratio of Vκ-Jκ1: Vκ-Jκ4 recombination was nearly 3 times smaller in the Jκ1-GAGA deletion mice as in the wild type mice (Figure 3.2c). This indicates that although recombination in general is lower in the Jκ1-GAGA deletion mice, the effect is most pronounced at Jκ1. Both

knockout lines, the 40bp and the 57bp deletion, produced similar results, indicating that the phenotype is more likely attributable to the deletion including the GAGA motif than to any potential off target effect.

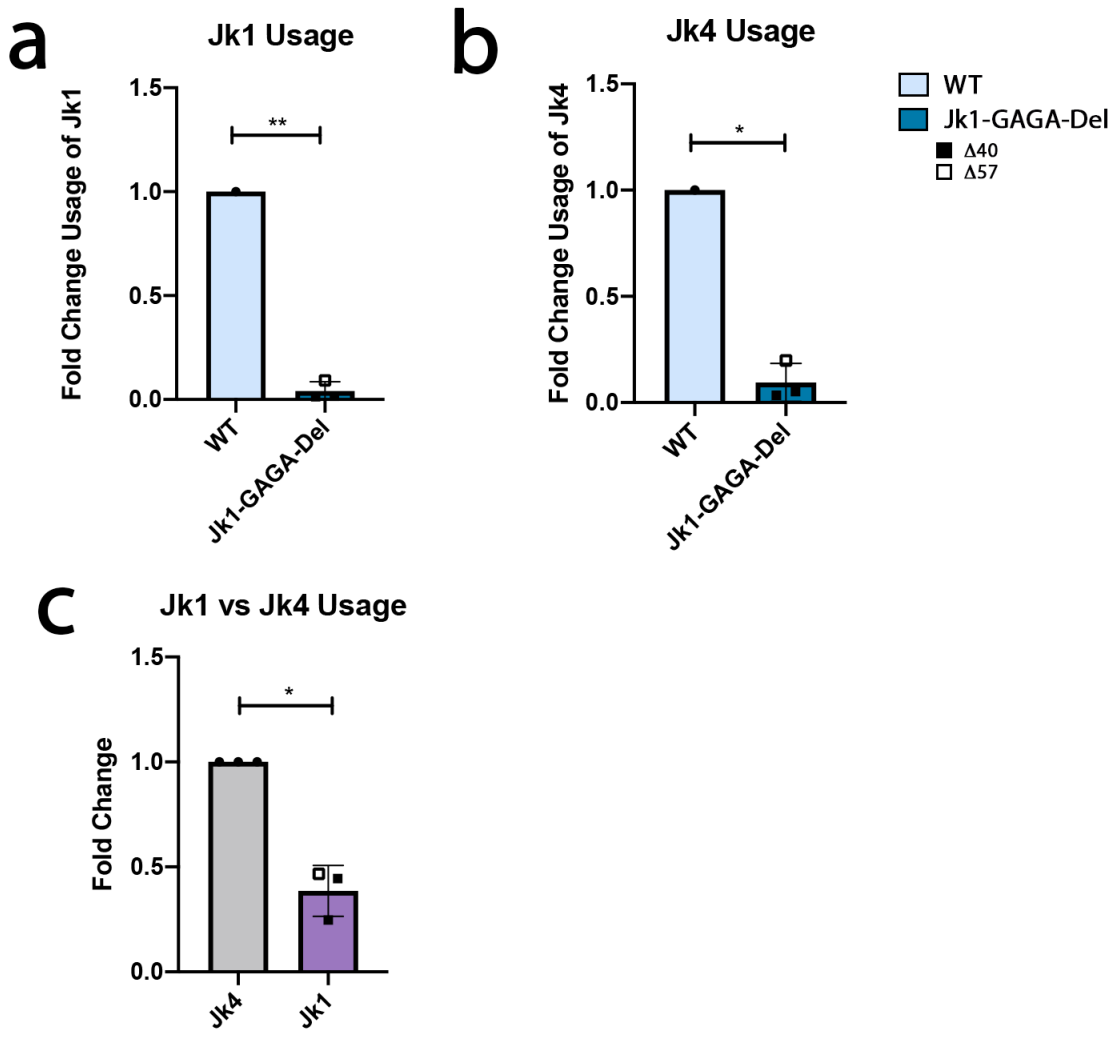

Figure 3.4: Deletion of the Jκ1-GAGA motif Reduces Usage of Jκ1
(a) Quantitative RT-PCR for the Vκ-Jκ1 recombination product in flow-sorted small pre-B cells isolated from wild-type and Jκ1-GAGA deletion mice. **(b)** Quantitative RT-PCR for the Vκ-Jκ1 and Vκ-Jκ4 recombination products in flow-sorted small pre-B cells isolated from wild-type and Jκ1-GAGA deletion mice. **(c)** Quantification of relative usage of Jκ1 to Jκ4 in WT and Jκ1-GAGA deletion mice, as measured by quantitative qPCR of recombination products. (n = 3). *p <0.05 **p <0.01.

3.2.3 Alterations in chromatin architecture at the Jκ locus in Jκ1-GAGA-deletion small pre-B cells

In non-recombining cells, the RSSs in Igκ are obscured by nucleosomes, which likely serves to help prevent premature recombination. In small pre-B cells, to facilitate Igκ recombination, the nucleosomes are shifted off of the RSS as a result of Brwd1 activity. We wanted to ascertain whether the loss of the 5' Jκ1 GAGA motif domain would result in a failure to shift nucleosome density away from the Jκ1 RSS.

We performed ATAC-seq (assay for transposase-accessible chromatin with sequencing) on flow sorted small pre-B cells from wild-type, Brwd1-KO, Jκ1Δ40, and Jκ1Δ57 mice, and analyzed nucleosome structure in the Igκ locus. Strikingly, we could see the specific change in position of one crucial nucleosome in the Jκ1 region (Figure 3.3a). In the WT cells, at the pro-B cell stage, the Jκ1 gene segment and RSS is occupied by a nucleosome. Then, in WT small pre-B cells that are preparing for and undergoing Igκ recombination, this nucleosome seemed to move away from the Jκ1 RSS, which is predicted to make it accessible to cleavage. Instead, we saw signal from a nucleosome positioned just upstream of the Jκ1 RSS.

However, in both the Jκ140 and Jκ157 small pre-B cells, we saw that the Jκ1 RSS remained obscured by a nucleosome (Figure 3.3c), and the region that is normally occupied by a nucleosome in the WT remains largely clear in the Jκ1-GAGA deletion mice (Figure 3.3b), following a pattern similar to the Brwd1-KO small pre-B cells. From this, we can infer that without the Jκ1 GAGA motif domain, Brwd1 is unable to properly shift nucleosomes, which likely contributes to recombination defects at Jκ1.

We also analyzed local accessibility around Jκ1 in the wild-type, Jκ1Δ40, and Jκ1Δ57 small pre-B cells (Figure 3.4a). While we saw no significant changes in accessibility at the gene body of Jκ1 (Figure 3.4b), we did see a marked change in accessibility at the RSS

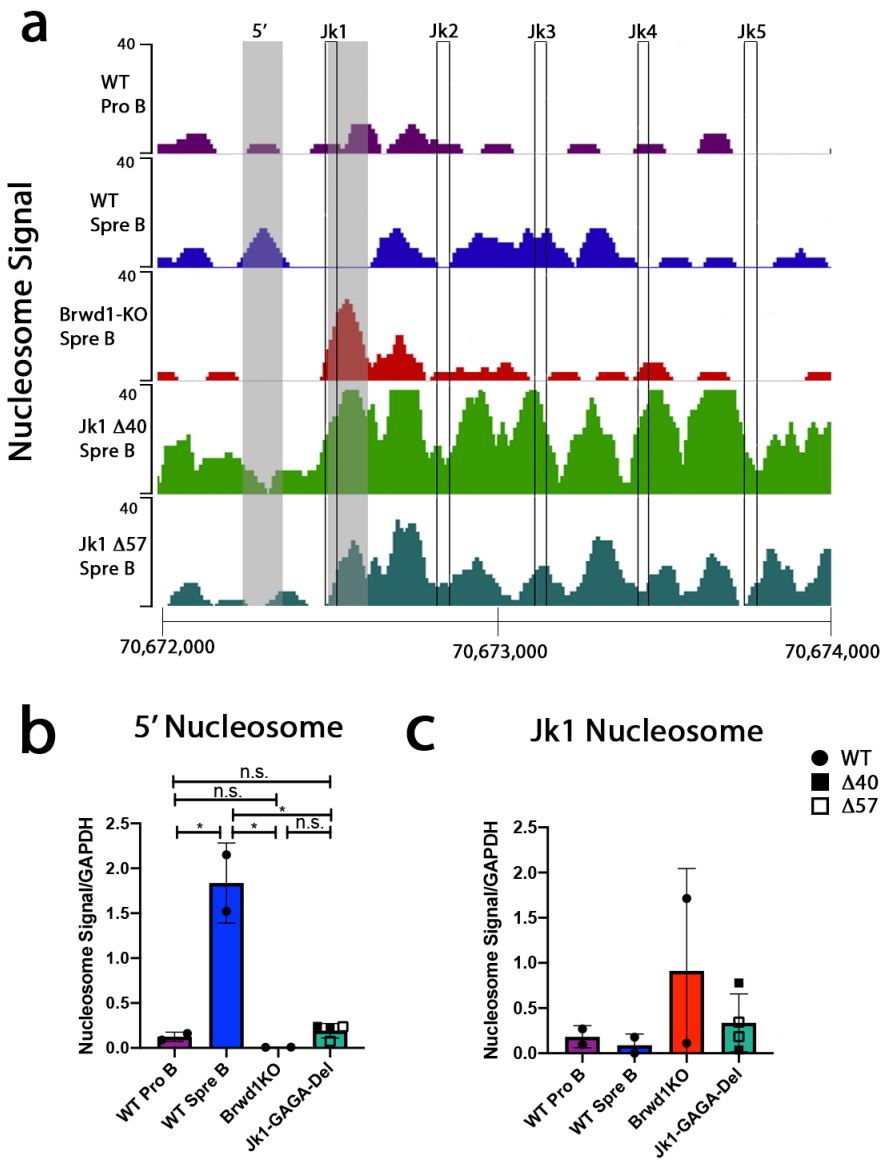

Figure 3.5: Altered nucleosome positioning at the Jκ1 RSS in Jκ1-GAGA deletion mice **(a)** Nucleosome positioning at the Jκ locus in wild-type pro B, wild-type small pre-B, Brwd1-KO small pre-B, Jκ1Δ40 small pre-B, and Jκ1Δ57 small pre-B cells. Nucleosome signal represents the difference in normalized density between the simulated signal and background data, with signal defined as from read pairs with large insert sizes, and background defined as from read pairs with short insert sizes. Data are representative of two independent experiments. **(b)** Quantification of nucleosome signal at the indicated site 5' of the Jκ1 RSS, noted by a shaded box. Nucleosome signal is significantly higher in wild-type small pre-B cells as compared to wild-type pro B cells, and Brwd1-KO and Jκ1-GAGA-del small pre-B cells. **(c)** Quantification of nucleosome signal at the Jκ1 gene segment plus RSS. *P <0.05

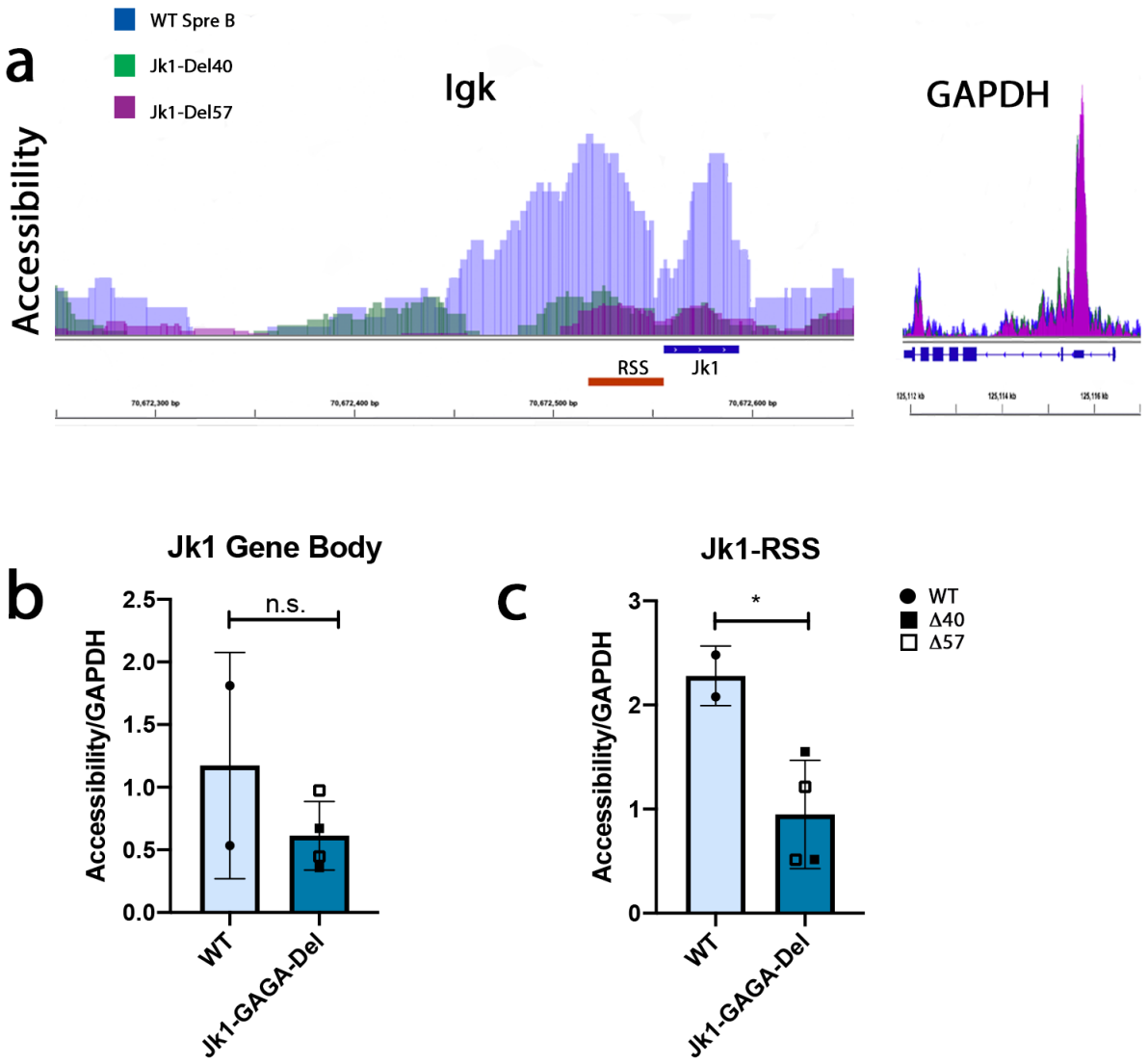

Figure 3.6: Chromatin accessibility is reduced at the Jκ1 RSS in Jκ1-GAGA-deleted small pre-B cells

a. Accessibility at the Jκ locus as measured by ATAC-seq. The y axis represents tags per million reads. Data from two independent experiments (10^5 cells per sample). **b.** Quantification of chromatin accessibility at the Jκ1 gene body. **c.** Quantification of chromatin accessibility at the Jκ1 RSS. *p <0.05

(Figure 3.4c). The Jκ1 RSS is significantly less accessible in the Jκ1Δ40 and Jκ1Δ57 mice compared to WT, reflecting the change in nucleosome structure that we observed.

To determine whether the effects on nucleosome positioning we saw were the result of a change in recruitment of Brwd1 to the Jκ1 region, we performed ChIP-qPCR (chromatin immunoprecipitation and quantitative PCR) on wild-type and Jκ1-GAGA deletion mice (Figure 3.5). While Jκ1 trended toward less binding of Brwd1, the effect was not significant. We also did not see a significant change in Brwd1 blinding in the Jκ4 region. Thus, the change in nucleosome positioning points to the necessity of GAGA motifs in directing the nucleosome positioning activity of Brwd1, rather than in recruitment of Brwd1.

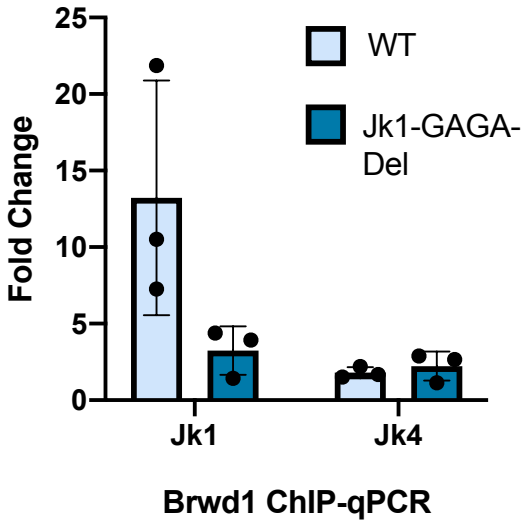

Figure 3.7: Brwd1 recruitment to Igκ is not dependent on the Jκ1-GAGA motif
Quantitative PCR performed on DNA products of Brwd1-ChIP (chromatin immunoprecipitation) performed in wild-type and Jκ1-GAGA-deletion small pre-B cells. Target regions of ChIP are the Jκ1 and Jκ4 regions (n = 3).

3.2.4 Germline Transcription of Jκ is reduced in Jκ1-GAGA deletion small pre-B cells

It has long been known that germline transcription of Igκ increases prior to recombination. It is unclear whether that transcription is merely the result of increased accessibility of the locus, or whether the transcription itself plays a role in biasing certain areas of the locus for recombination. Thus, given that we saw defects in recombination following removal of the Jκ1 GAGA motif domain, we were interested to see if this deletion would also result in altered germline transcription of the Jκ region.

To interrogate this, we performed RNA-sequencing (RNA-seq) on our Jκ1Δ40 and Jκ1Δ57 small pre-B cells, and compared them to RNA-seq data from Brwd1-KO small pre-B cells, as well as wild-type small pre-B, large-pre-B and immature B cells. We first performed PCA analysis to determine with which cell populations our Jκ1-GAGA-del small pre-B cells clustered (Figure 3.6a). We found that the Jκ1-GAGA-del cells cluster in between wild-type small and large pre-B cells, and near the Brwd1-KO small pre B cells. Additionally, the heatmap of these cell populations supports this analysis, with the Jκ1-GAGA-del small pre-B cells clustering mostly closely with the Brwd1-KO small pre-B cells in the dendogram (Figure 3.6b). Finally, we performed a gene ontological analysis of differentially expressed genes between the Jκ1-GAGA-del small pre-B and the wild-type small pre-B (Figure 3.6c). Here, we found several pathways that we would expect to see differentially expressed if there were to be a defect in recombination and developmental progression; namely, cellular response to DNA damage, mitosis, transcription, chromatin organization, and cellular stress response.

Next, we analyzed differences in germline transcription between wild-type and Jκ1-GAGA-del small pre-B cells. To specifically analyze germline transcription, and not transcription of recombination products, we removed multimapping reads from our analysis.

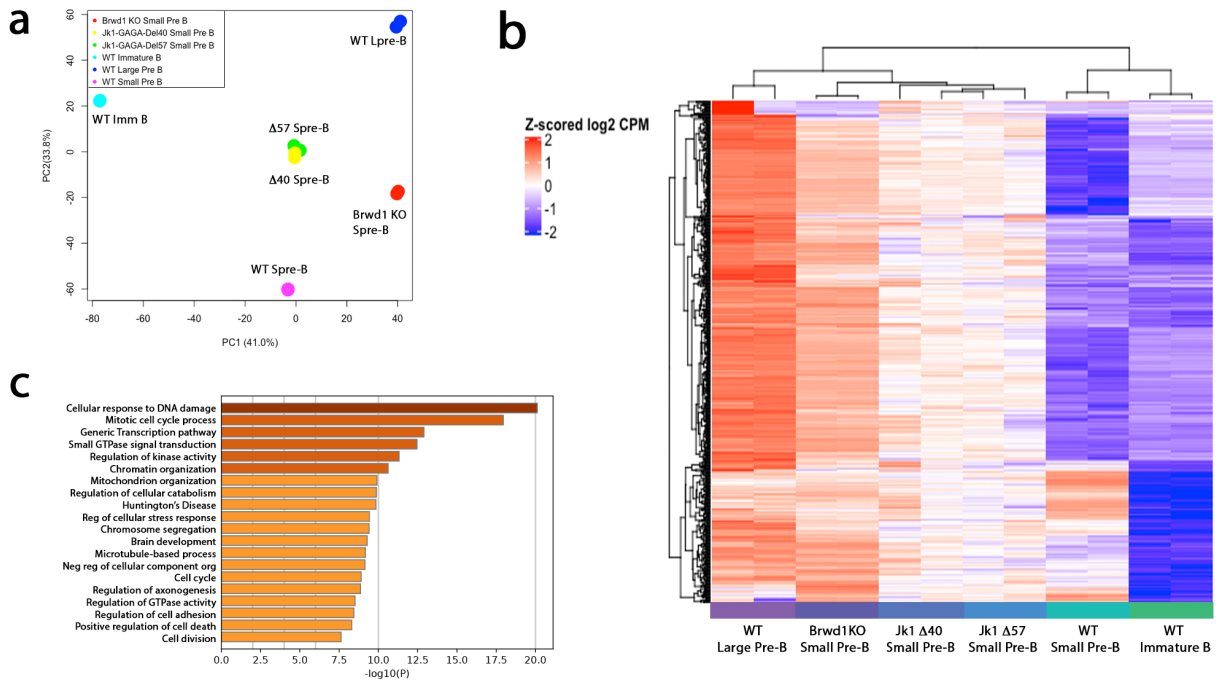

Figure 3.8: RNA-seq reveals genetic programming characteristic of recombination and developmental defects in the Jκ1-GAGA-del small pre-B cells

(a) Principle component analysis of the RNA-sequencing data from sorted Jκ1-GAGA-del and Brwd1-KO small pre-B cells, as well as wild-type small pre-B, large pre-B, and immature B cells. **(b)** Heatmap of differentially expressed genes between Jκ1-GAGA-del and Brwd1-KO small pre-B cells, as well as wild-type small pre-B, large pre-B, and immature B cells. **(c)** Gene ontological analysis of differentially expressed pathways between Jκ1-GAGA-del and WT small pre-B cells.

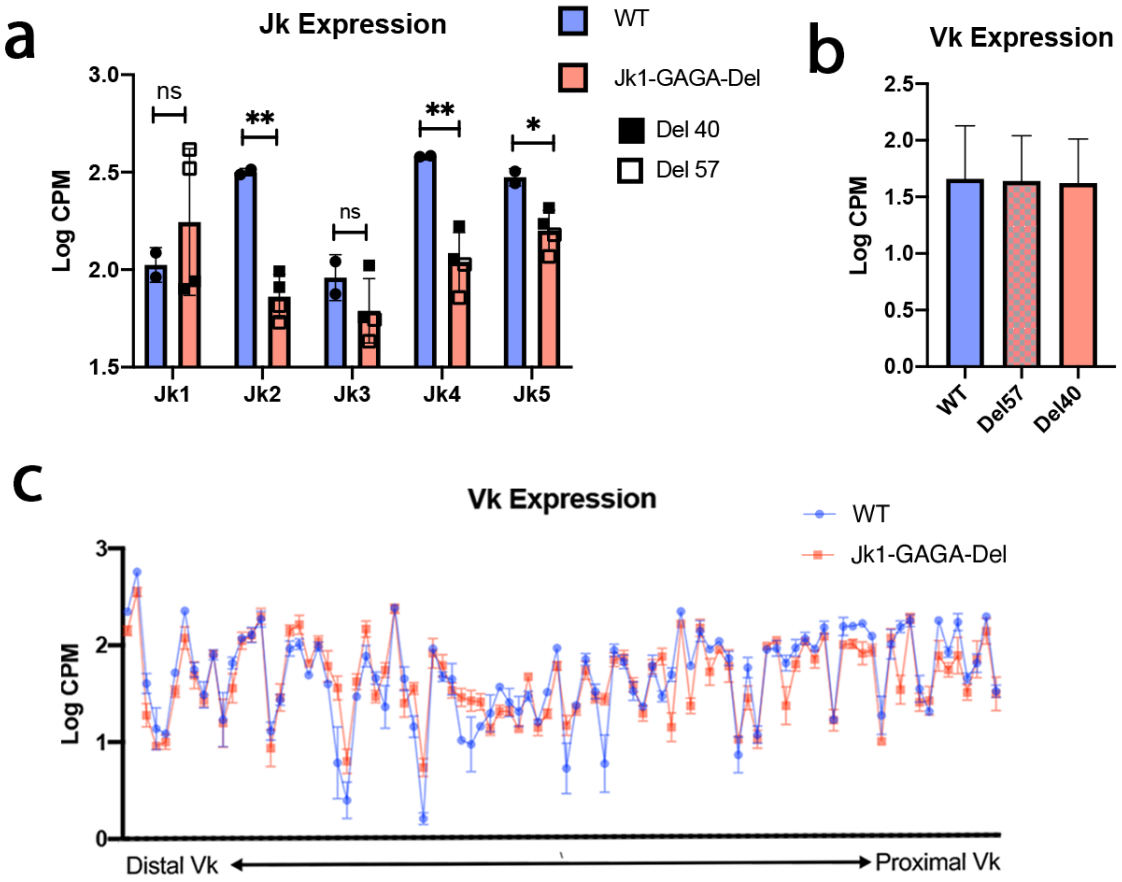

Figure 3.9: Germline Transcription of the Jκ locus is diminished in the Jκ1-GAGA-deletion mice

(a) Quantification of germline Jκ transcription from RNA-seq of wild type vs Jκ1-GAGA-deletion small pre-B cells in counts per million. **(b)** Quantification of germline Vκ transcription from RNA-seq of wild type vs Jκ1Δ57 and Jκ1Δ40 small pre-B cells in counts per million. **(c)** Germline transcription of individual Vκs in wild-type vs Jκ1-GAGA-del small pre-B cells. (n= 2). *P <0.05 and **P <0.005. Data presented as average +/- S.D.

This analysis revealed that throughout much of the Jκ region germline transcription of the unrecombined locus is down greater than two-fold (Figure 3.7a). The exceptions to this trend are at the non-functional Jκ3 locus and at the Jκ1 locus. The lack of change in expression at Jκ3 is perhaps not surprising, as it is a non-functional segment. However, germline transcription at Jκ1 is normal, and actually trends somewhat upward. However, we see a distinct recombination defect at Jκ1. Previously it has been hypothesized that one of the functions of germline transcription is to make the Iκk locus more accessible to recombination. However, this data suggests that transcription alone is insufficient to make Jκ accessible.

We also looked to see if there was an effect on germline transcription throughout the Vκ locus in the Jκ1-GAGA deletion mice. As expected, transcription at Vκ was unchanged in the Jκ1-GAGA deletion mice (Figure 3.7b and c). Thus, the impact of the Jκ1-GAGA deletion on recombination appears to be localized to Jκ, and not directly affect transcription at Vκ.

3.2.5 Deletion of the Jκ1 GAGA motif results in a B cell Developmental Defect

To determine whether the GAGA motif preceding Jκ1 is important for B lymphopoiesis, we harvested bone marrow (BM) and spleens from $J\kappa1^{\Delta40}$ and $J\kappa1^{\Delta57}$ mice, as well as wild type littermate control mice, and analyzed B lymphopoiesis by flow cytometry (Figure 3.8a). We found statistically significantly decreased cell numbers at the developmental stages surrounding Igκ recombination (Figure 3.8b-f). The stage at which recombination occurs, small pre-B (B220+CD19+CD43IgMFSC^{lo}), was the most significantly reduced in the Jκ1-GAGA-deletion mice as compared to wild type ($p = 0.0027$). Cell counts were also down in the large pre-B (B220+CD19+CD43IgMFSC^{hi}) ($p = 0.0200$) and immature

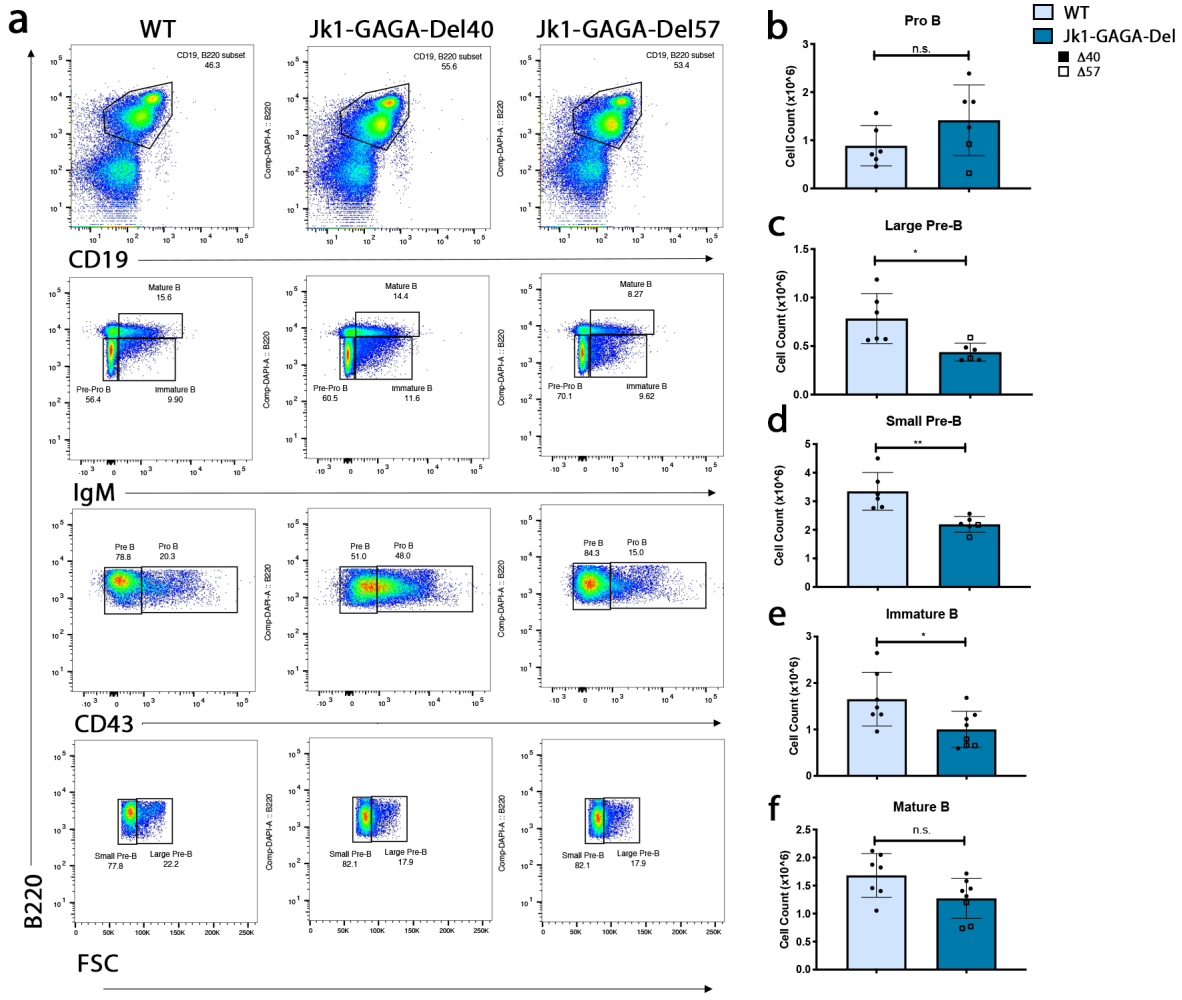

Figure 3.10: Developmental B cell defects in GAGA deletion mice
(a) Flow cytometric analysis of different developmental stages of B cell lymphopoesis in the bone marrow of wild type, Jκ1$^{\Delta 40}$, and Jκ1$^{\Delta 57}$ mice. **(b-f)** Absolute numbers of cells per mouse at different stages of B cell development in the bone marrow of wild-type and Jκ1 GAGA-deletion mice (Jκ1$^{\Delta 40}$ and Jκ1$^{\Delta 57}$ were combined for this analysis) (n= 6). *P <0.05 and **P <0.005 compared to the respective wild-type control (unpaired t-test). Data presented as average ± S.D.

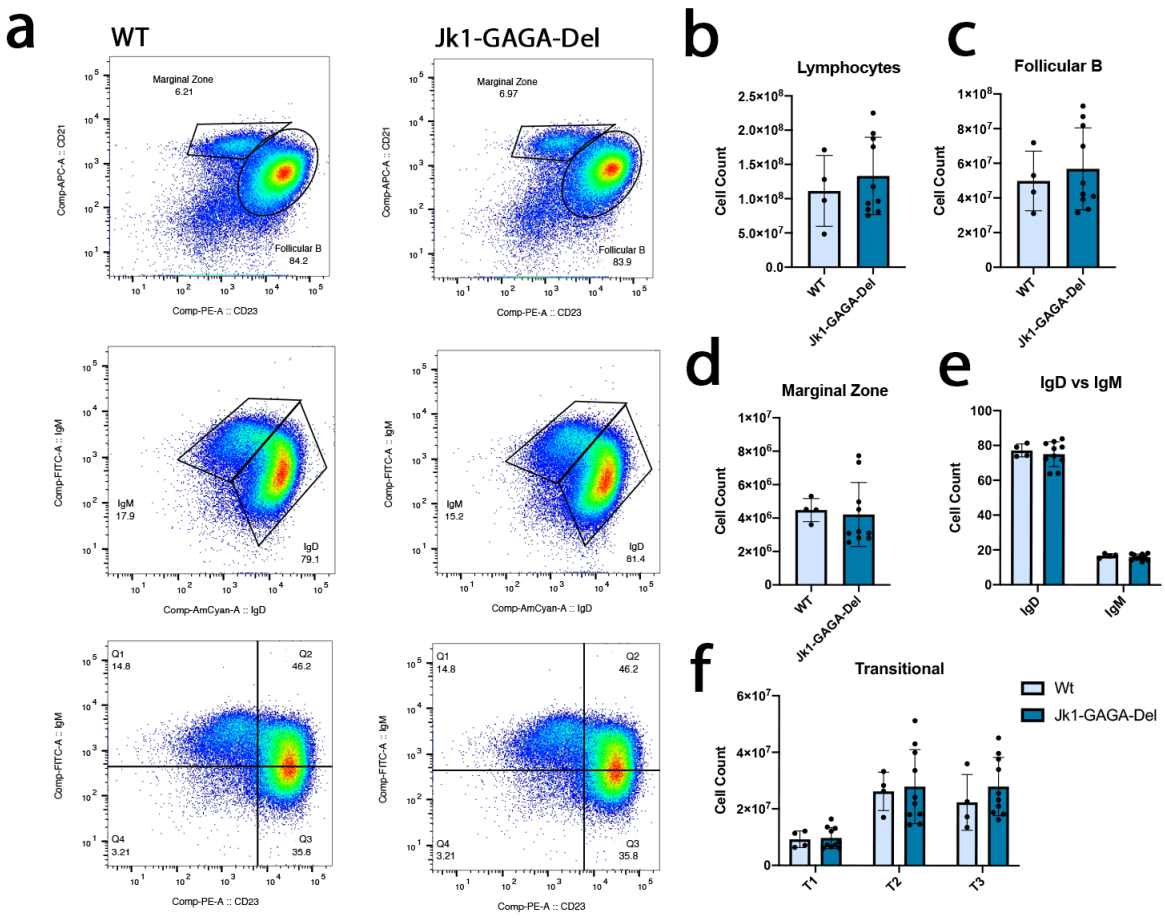

Figure 3.11: Peripheral B populations are unaltered in GAGA deletion mice
Flow cytometric analysis of various peripheral splenic B cell populations in WT vs Jκ1-GAGA-del mice. **a** Representative flow plots from WT and Jκ1-GAGA-del mice. **b** Total lymphocytes. **c** Absolute numbers of follicular B cells. **d** Absolute numbers of marginal zone B cells. **e** Absolute numbers of IgD vs IgM positive cells. **f** Absolute numbers of T1-T3 transitional zone B cells. There are no significant numbers in any of these population of cells between WT and Jκ1-GAGA-del mice.

B stages (B220loCD19+CD43-IgM+) (p = 0.0227), which flank Igκ recombination. While pro-B (B220+CD19+CD43+IgM) and mature B (B220hiCD19+CD43-IgM+) were not significantly impacted, they followed the expected trend for an Igκ defect; namely, numbers of pro-B cells were increased and mature B cells were somewhat decreased, though cell numbers had largely recovered by this stage.

These developmental defects were localized to the early stages of B cell development in the bone marrow, and did not persist to the periphery. We examined splenic transitional, immature, mature, follicular, and mariginal zone B cells, but found no significant differences between Jκ1-GAGA-deletion mice and wild type mice (Supplementary Figure 3.3).

3.2.6 Increased Igλ usage in Jκ1-GAGA-Deletion mice

During light chain recombination, developing B cells will attempt first to recombine at the Igκ locus. However, if this recombination is unsuccessful, developing B cells can then attempt recombination at the Igλ locus. While in humans, the usage of kappa to lambda is roughly 50:50, in mice, lambda is only used very rarely (approximately 10% of the time). However, we reasoned that if the GAGA motif 5' of Jκ1 is important for the activity of Brwd1 in opening up the kappa locus for recombination, we might see a skewing of repertoire toward Igλ.

We analyzed by flow cytometry how many immature B cells in the bone marrow expressed Igκ vs Igλ (Figure 3.9a). We found that by flow, a higher percentage of immature B cells were lambda positive in the Jκ1-GAGA deletion mice than in wild type. A quantification of the absolute cell numbers confirmed this observation (Figure 3.9b). The ratio of absolute number of lambda: kappa positive cells was statistically significantly higher in the Jκ1-GAGA deletion mice than in the wild type littermate controls. Additionally, the absolute number of kappa positive cells was higher in the wild type mice than in the Jκ1-GAGA

deletion mice. The number of lambda positive cells was not significantly higher in the Jκ1-GAGA deletion mice than in wild type. However, this is likely due to the fact that the overall number of immature cells is lower in the Jκ1-GAGA deletion mice than in wild type. Overall, these results indicate that rather than specifically impacting recombination at only Jκ1, as we had originally hypothesized, removing the GAGA at Jκ1 might have broader effects upon the entire Jκ locus.

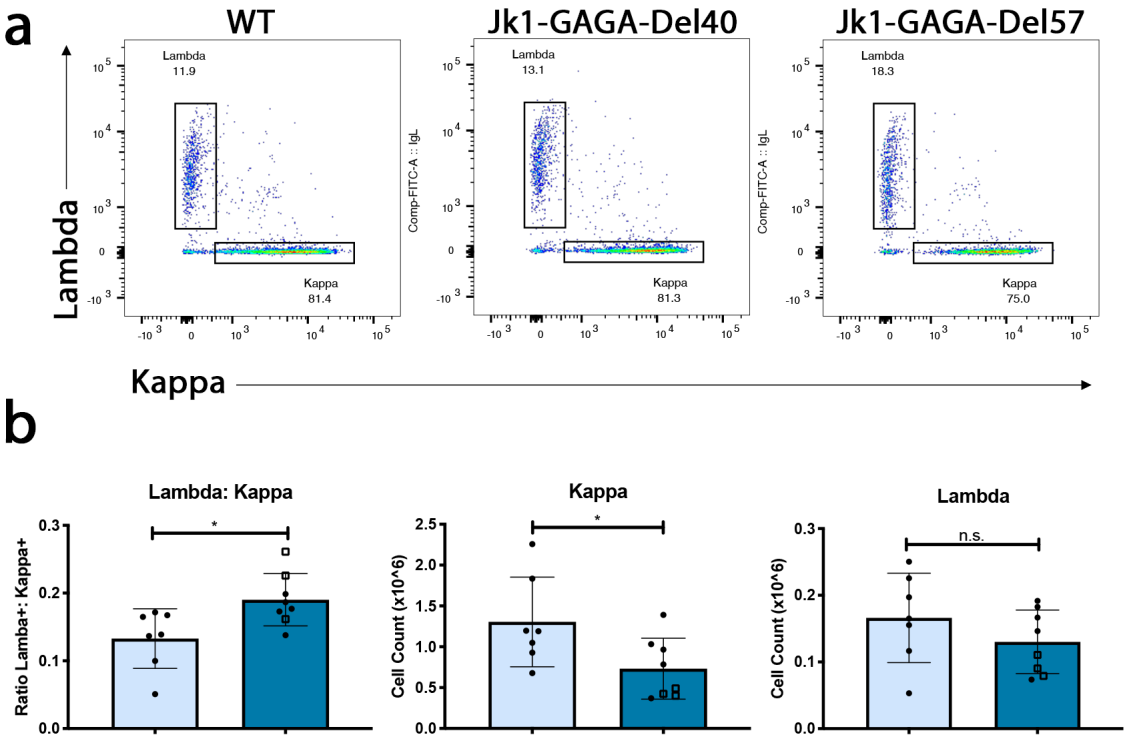

Figure 3.12: Lambda usage in Jκ1-GAGA deletion mice
A. Flow cytometric analysis of kappa and lambda expression in immature B cells in the bone marrow of wild type, $J\kappa1^{\Delta40}$, and $J\kappa1^{\Delta57}$ mice. **B.** Absolute numbers of cells per mouse of kappa and lambda expressing cells, as well as a quantification of the lambda to kappa positive ratio of immature B cells of wild-type and Jκ1 GAGA-deletion mice ($J\kappa1^{\Delta40}$ and $J\kappa1^{\Delta57}$ were combined for this analysis) (n= 6). *P <0.05 compared to the respective wild-type control (unpaired t-test). Data presented as average ± S.D.

3.2.7 Deletion of 5' GAGA motif at Jκ2 reduces usage of Jκ2 without significant developmental defects

Our finding that removing the GAGA motif at Jκ1 not only reduced usage of Jκ1, but reduced usage of the kappa light chain in general, seemed to support a model of sequential usage of Jκ gene segments. By this model, RAG likely attempts to recombine first at Jκ1, rather than stochastically choosing any of the four functional J gene segments. Thus, disrupting chromatin remodeling at Jκ1 would have larger implications for Jκ rearrangement as a whole. To test this model, we removed the GAGA at Jκ2. If usage of Jκ segments is sequential, then this should produce a less dramatic phenotype, and not affect usage of Jκ1.

To remove the GAGA motif from Jκ2, we again utilized Crispr-Cas9. We designed primers 5' and 3' to the GAGA motif preceding Jκ2, optimizing for high efficiency and low off target rates. The target specific guides were then was incubated with CAS9 protein and injected in single cell mouse embryos as a ribonucleoprotein (RNP) complex. This injection produced two different founder lines with homozygous deletions that removed the GAGA motif without interfering with any RSS or Jκ gene sequences.

First, we examined usage of Jκ2 in small pre-B cells from these mice using quantitative PCR for the recombination products of Vκ-Jκ1, Vκ-Jκ2, and Vκ-Jκ4 in the GAGA deletion mice relative to WT. The recombination product of Vκ-Jκ2 is significantly diminished in these mice; however, the products for Jκ1 and Jκ4 usage are unchanged (Figure 3.10a). Next, we performed a semi-quantitative PCR using a degenerate Vκ primer and a Cκ primer, such that only recombination products would be amplified, and we would capture recombination products from all Jκs (Figure 3.10b-c). The product of this PCR was cloned and sequenced, then from this we were able to identify individual recombination products. There was a striking difference in the number of Jκ2 recombination products identified in WT vs Jκ1-GAGA-del small pre-B cells (Figure 3.10b). In the WT small pre-B cells, Jκ1 was used in

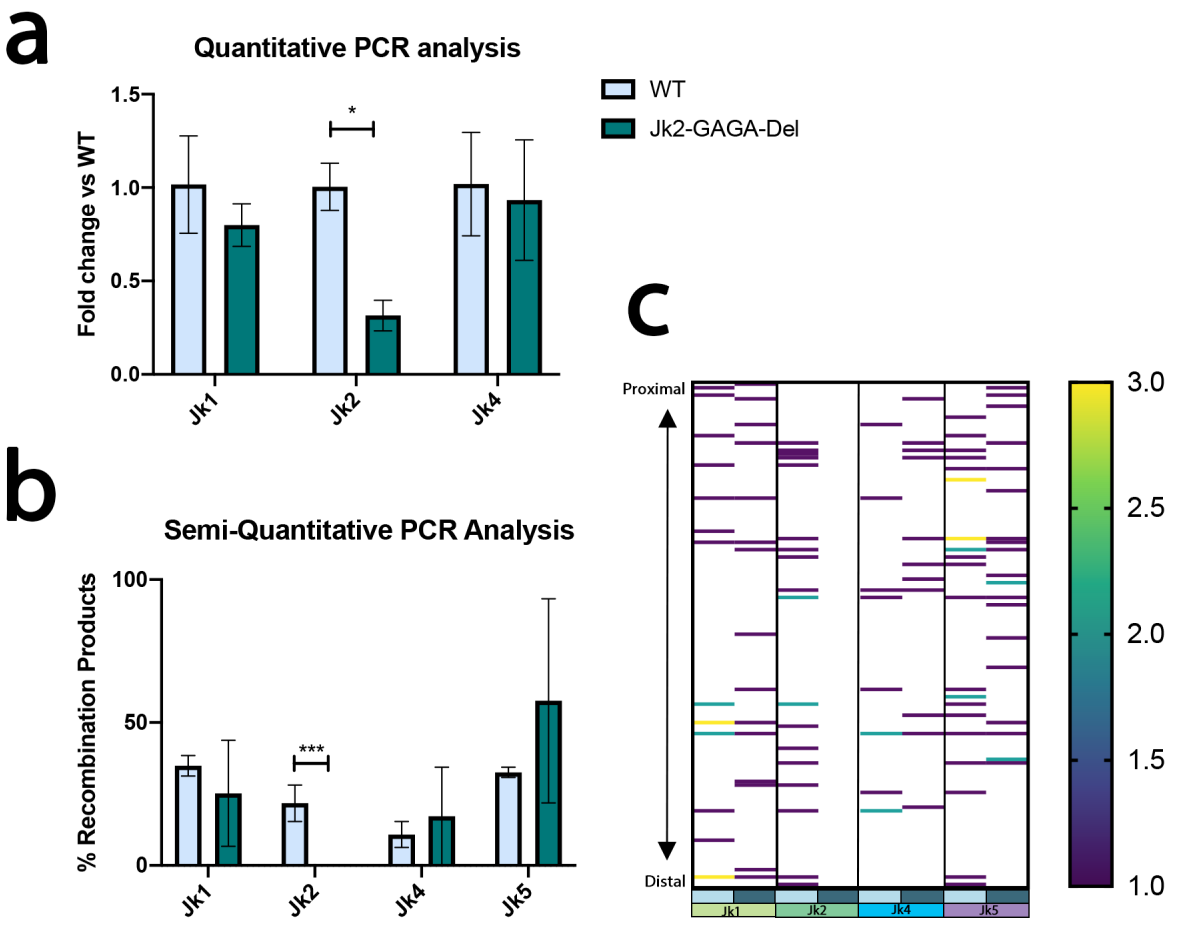

Figure 3.13: Removal of 5' GAGA-sequence results in decreased usage of Jκ2
(a) Quantitative RT-PCR for the Vκ-Jκ1, Vκ-Jκ2 and Vκ-Jκ4 recombination products in flow-sorted small pre-B cells isolated from wild-type and Jκ2-GAGA deletion mice (n = 3). **(b-c)** Semi-quantitative PCR analysis of recombination products using a degenerate Vκ primer and Cκ primer. PCR products were cloned and individually sequenced to determine Jκ and Vκ usage. (n = 48 sequences from 2 mice) **(b)** Percentage usage of Jκ segment, as determined by sequencing recombination products. p = 0.0016 by Chi-Squared test. **(c)** A representation of Vκ segment usage in WT vs Jκ2-GAGA-del mice. The location on the heat map indicates the location of the Vκ segment relative to Jκ (with Vκ-1 being most proximal and Vκ-137 being most distal), and the color indicating in how many sequenced recombination products that particular Vκ was used. *p <0.05, **p <0.01, ***p <0.005

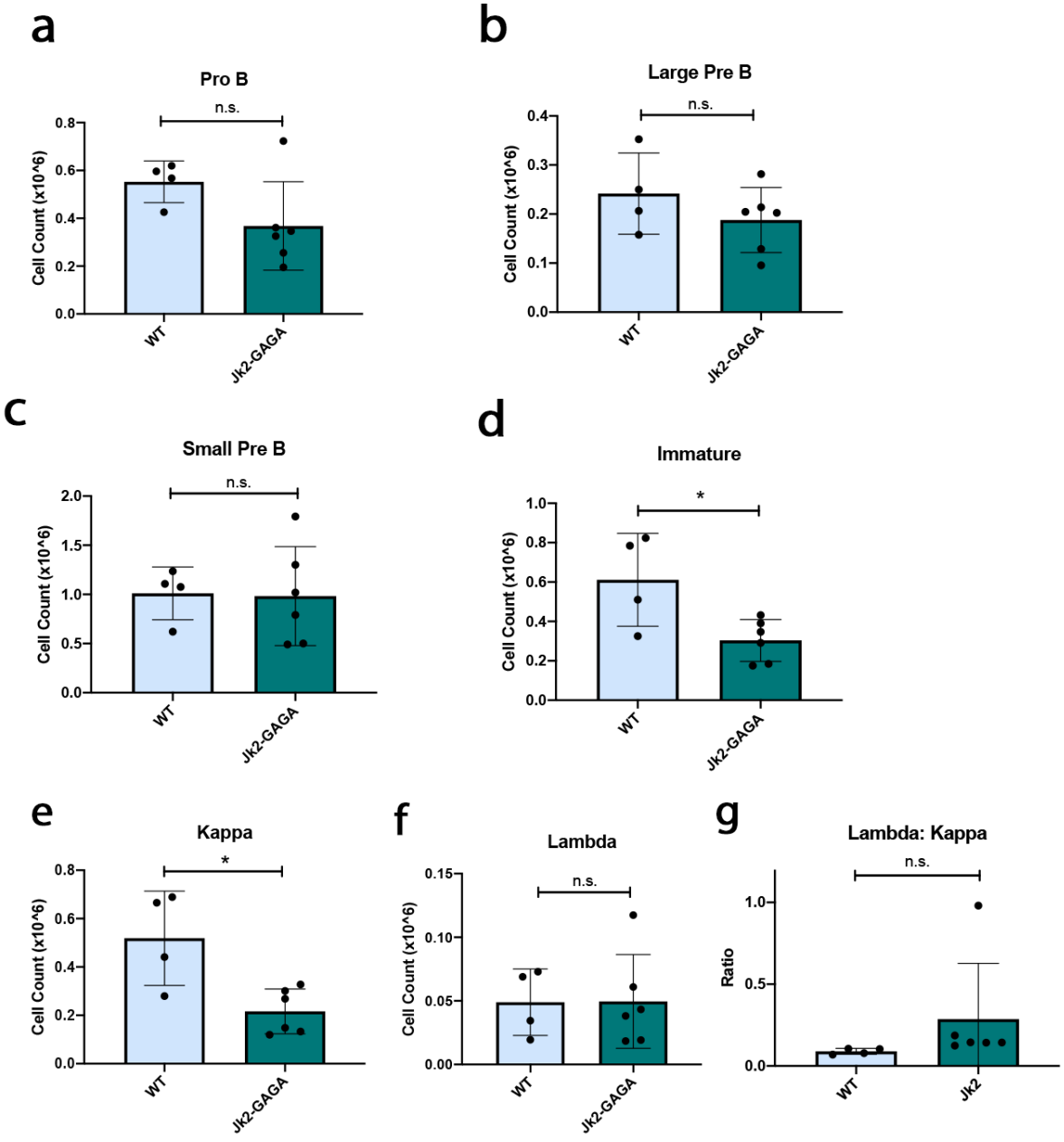

Figure 3.14: Deletion of the Jκ2-GAGA produces a less dramatic developmental phenotype **(a-d)** Absolute numbers of cells per mouse at different stages of B cell development in the bone marrow of wild-type and Jκ2-GAGA-deletion mice (n= 6). **(e-g)** Absolute numbers of cells per mouse of kappa and lambda expressing cells, as well as a quantification of the lambda to kappa positive ratio of immature B cells of wild-type and Jκ2 GAGA-deletion mice (n= 6). *p <0.05

35% of products, Jκ2 was used in 21% of products, Jκ4 in 11% of products, and Jκ5 in 33% of products (n = 45). However, in the Jκ2-GAGA-Del small pre-B cells, we were unable to detect any Jκ2 recombination products. In these cells, Jκ1 was used in 32% of products, Jκ2 was used in 0% of products, Jκ4 in 23% of products, and Jκ5 in 45% of products (n = 47). Because this analysis also allowed us to identify the V to which J had recombined, we examined the distribution throughout the Vκ locus to see if we could identify any marked differences in Vκ usage (Figure 3.10c). We did not notice any appreciable differences in proximal vs distal V preference in the Jκ2-GAGA-del small pre-B cells compared to WT.

We next harvested BM from the Jκ2-GAGA-del and wild-type mice, and analyzed B lymphopoiesis by flow cytometry as we did with the Jκ1-GAGA-del (Fig 3.11a-d). We found statistically significantly decreased cell numbers at the immature B cell stage, indicating a temporary developmental block, likely reflecting less-efficient Igκ recombination. Indeed, we saw a significant decrease in kappa-positive cells in these mice; however, this did not translate into a significant increase in the lambda: kappa ratio, indicating a less severe decrease in Igκ recombination (Figure 3.10e-g).

3.2.8 Increased nucleosome occupancy at cryptic RSSs

Our data supports a model in which nucleosomes are actively moved from the RSSs over which they were positioned in preparation for recombination. This model also suggests that nucleosome occupancy is an effective means of protecting RSSs from cleavage at inappropriate times, such as when a cell is cycling. These findings led us to consider whether similar mechanisms might discourage the cleavage of the many cryptic RSSs scattered throughout the genome.

HiC techniques have revealed that regions genome can fall broadly into two distinct categories: compartment A and compartment B. Compartment A consists of more actively

transcribed regions of DNA, and is characterized by higher chromatin accessibility, higher gene density, and more active histone marks. Compartment B, in contrast, consists of heterochromatin that is less actively transcribed, and marked by low chromatin accessibility and more repressive histone marks. Additionally, it's been shown that chromatin compartments are cell-type specific [168].

We found that decreased accessibility at the RSS correlated with less recombination. Thus, it seemed possible that evolutionary selective pressure may have resulted in a relative depletion of cryptic RSSs from the active Compartment A regions of cell types actively expressing RAG, and thus at higher risk of errant recombination at cryptic RSSs, such as small pre-B cells. Using Hi-C on sorted primary cells, we were able to determine the Compartment A and Compartment B boundaries in small pre-B cells.

Next, we compiled an annotation of all cryptic 12 and 23-RSSs passing the established threshold (RIC = -38.81 for 12-RSSs, and RIC = -58.45 for 23-RSSs) throughout the entire mouse genome, using the RSS Information Content (RIC) algorithm [197]. In all, we were able to identify approximately 3.55x 106 cryptic RSSs (cRSSs). These cRSSs were divided into Compartment A and Compartment B cRSSs. We then calculated the density of cRSSs in Compartment A and Compartment B compared to the density of cRSS throughout the entire genome. As predicted, cRSS density was highest in Compartment B and lowest in Compartment A (Figure 3.12a). Using the density of cRSSs throughout the entire genome, we were able to calculate predicted numbers of cRSSs in Compartment A and Compartment B were cRSSs to be distributed entirely randomly throughout the genome. This revealed that cRSSs are significantly depleted from Compartment A and enriched in Compartment B ($p = 1.75$ x 10-31) (Figure 3.12b).

Next, we wanted to determine whether cryptic RSSs are more likely to be occupied by nucleosomes, and thus more shielded from errant recombination. To do this analysis, we compared our annotated cRSSs with one million randomly generated genomic regions

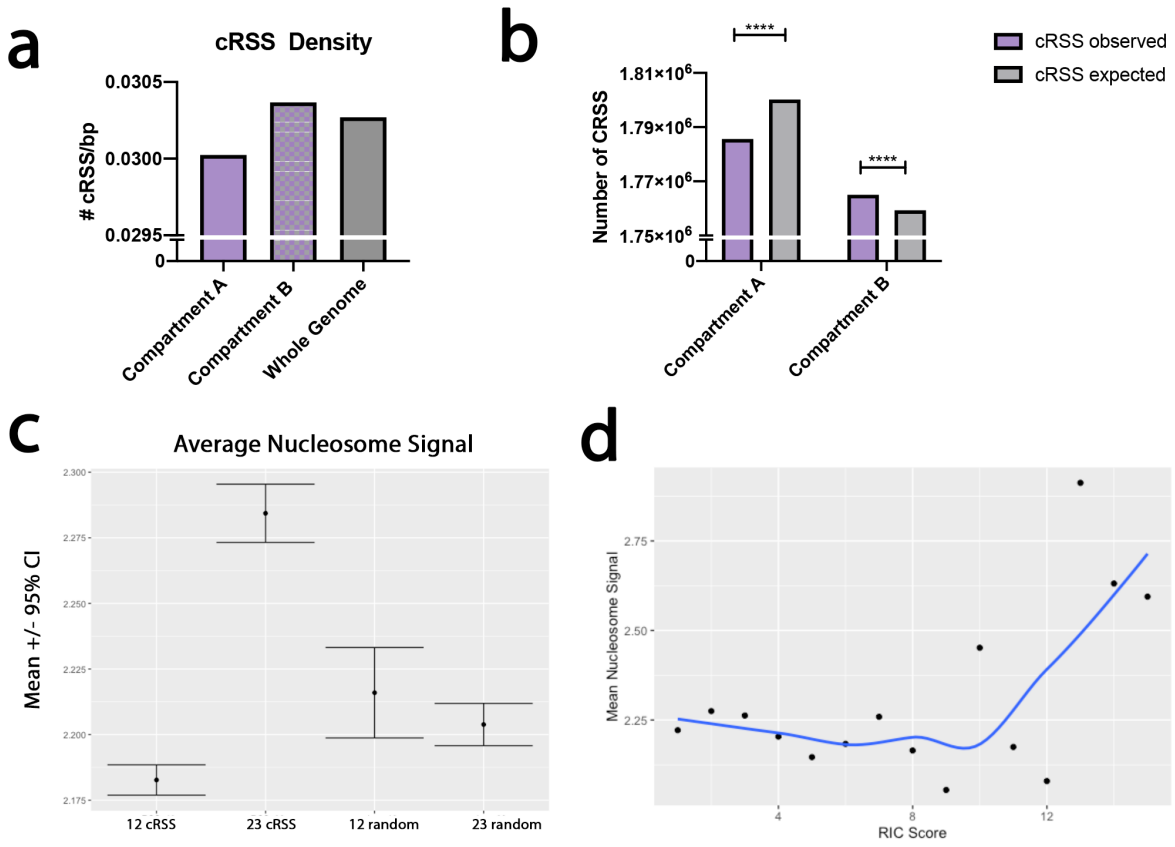

Figure 3.15: Cryptic RSSs are depleted from regions of active chromatin and are more likely to be bound within nucleosomes

(a) cRSS Density in different genomic compartments; calculated by dividing the number of cRSSs in the compartment by the total length of the compartment in base pairs. **(b)** Total observed cRSSs per compartment compared to the expected number of cRSSs based on genomic density of cRSSs. p = 1.75E-31. **(c)** Mean nucleosome signal for 12 and 23 cRSSs compared to randomly generated sequences of the same length. Plotted as mean +/- a 95% confidence interval. Mean nucleosome signal calculated using the bedtools bigWigAverageOverBed. **(d)** Mean Nucleosome Signal vs RIC Score. Because the range of poor to good RIC scores are different for 12 and 23 cRSSs, we bucketed and ranked the RIC scores from worst (RIC score = 1) to best (RIC score = 15) for both 12 and 23 cRSS separately. The data were then recombined and then plotted against the mean nucleosome score.

corresponding to the length of either a 12 or a 23 RSS. We then used the bedtools bigWigAverageOverBed tool to determine the mean nucleosome occupancy at each cRSS or randomly generated region (Figure 3.12c). This analysis revealed several notable findings. First, the 23-cRSSs, as expected, had a significantly higher mean nucleosome signal than either the randomly generated 12 or 23 regions. However, interestingly, the 12-cRSSs had a significantly lower mean nucleosome signal. This indicates the possibility that 12-RSSs (like those at Vκ) and 23 RSSs (like those at Jκ) are epigenetically regulated in somewhat different manners.

Finally, we wanted to determine if nucleosome occupancy correlated with the strength of the RIC score. The algorithm which determines the RIC score ranks a sequence on how closely it resembles a canonical RSS, with higher (less negative) scores being closer to canonical. The majority of RSSs are low-scoring, and only poorly resemble an RSS. Thus, to determine if nucleosome occupancy increased along with the RIC score, we bucketed and ranked the RIC scores from worst (RIC score = 1) to best (RIC score = 15), and then plotted the mean nucleosome score. This analysis revealed that as an RSS becomes stronger, the mean nucleosome signal increases (Figure 3.12d). This suggests that cryptic RSSs, particularly strong cryptic RSSs, are under evolutionary pressure to be protected from cleavage by RAG by being bound within nucleosomes.

3.3 Discussion

Overall, our results support a model in which short GA repeats, or "GAGA motifs", play an essential role in the chromatin remodeling activity of Brwd1, thereby permitting efficient Igκ recombination. We found that the deletion of a small bit of 5' sequence to which no function had previously been attributed was sufficient to not only interfere with the usage of its most proximal Jκ segment, but with Igκ recombination in general.

As predicted based on previous studies of Brwd1, the mechanism to which we could attribute this recombination defect was a failure to properly reposition nucleosomes at the Jκ1 locus. This remodeling failure resembled the phenotype seen in Brwd1KO mice at Jκ1. Rather than being moved upstream past the GAGA motif, the nucleosome at Jκ1 continued to bind the recombination signal sequence (RSS), thus obscuring the nucleosome from cleavage by RAG. Correspondingly, we also found an overall decrease in accessibility at the Jκ1 RSS. This supports the idea that nucleosome architecture is a critical component of the tight and fined tuned process involving the dangerous process of intentionally producing DNA double-strand breaks that is necessary for recombination.

While these experiments add to our mechanistic knowledge of the process of regulating gene recombination, the mechanism that remains somewhat mysterious is how Brwd1 interacts with GAGA motifs. While our data suggest that GAGA motifs are not required for the recruitment of Brwd1, limitations in our method make it difficult for us to conclusively rule out this possibility. However, other data suggests that GAGA motifs are not required for recruitment of Brwd1. Mandal *et al* found that genome-wide, only 64% of Brwd1 binding sites contain an extended GAGA motif. Though nucleosome remodeling was most pronounced at those sites containing GAGA motifs, they did not in their analysis or in ours seem to be required for binding [72].

Surprisingly, not only did the removal of the 5' Jκ1-GAGA motif reduce recombination at Jκ1, but it also interfered with Igκ recombination overall. Usage of another Jκ segment, Jκ5, was significantly diminished. Analysis of the bone marrow of Jκ1-GAGA-deletion mice showed that there were significantly fewer kappa positive immature B cells, and that the ratio of kappa:lambda positive cells was decreased in these mice. To determine if this was a principle general to the GAGA motifs of other Jκ1 loci, or something particular to Jκ1, we then removed the GAGA motif 5' to Jκ2. This mutant showed diminished usage of Jκ2 without diminishing the usage of other Jκ segments. While there were fewer kappa positive

and immature B cells, we did not see the more general effects on B cell development observed in the Jκ1-GAGA-del cells.

These experiments suggest the possibility that the recombination signal sequence is not the only primary DNA sequence crucial to recombination. The RSS, consisting of highly conserved nonamer and heptamer sequence separated by either 12 or 23 nucleotides is sufficient to enable recombination [198]. However, by deleting the GAGA motif without interring with the RSS, we see a large drop in recombination efficiency at the related Jκ locus. The GAGA motif seems essential for regulating accessibility through chromatin architecture and accessibility. It is unclear if GAGA motifs are important for recombination at the Vκ segments. Of 158 annotated Vκ segments, 140 have a GAGA motif 5' of the RSS. However, Brwd1 does not bind strongly throughout the Vκ locus. Thus, it will be left to future studies to determine whether GAGA motifs are critical at 12-RSSs, or only at 23-RSSs in Igκ.

There are three main models that describe the selection of a Jκ gene segment for recombination. In the first, the choice is purely stochastic. By this model, any one segment should be as likely to be chosen as any other, and the selection of one segment is independent of the other segments. A second model holds that Jκ selection is sequential, with recombination at Jκ1 being the default, and the order of selection necessarily proceeding from the 5' to the 3' segments. A third model holds that selection rates are mainly dependent on the quality of the RSS. When initially conceiving of our model, we envisioned a stochastic model of Jκ recombination. However, the more dramatic general Igκ recombination defect in the Jκ1-GAGA-deletion mice does not hold with this model, nor with the RSS quality model. In opposition to the sequential model, however, was the fact that in the Jκ2-GAGA-deletion mice, we only saw a significant decrease in recombination at the Jκ2 locus.

None of these simple models is able to satisfactorily predict the recombination results in our selective knockout lines. Thus, we suggest a more nuanced gatekeeper model. By this model, establishment of a permissive chromatin landscape is dependent on the actions of

BRWD1 at the Jκ1 5' GAGA motif. Following this chromatin organization, recombination will occur most frequently first at Jκ1. If this is unproductive, further rearrangements will occur largely stochastically, biased by the quality of the RSS. While the role of BRWD1 at other gene segments is important for their efficient usage, it is not critical for the locus as a whole. The reasons for Jκ1 acting as a seeming gatekeeper are unclear and will require future study. Additionally, we have not ruled out the possibility that our Jκ1-GAGA-deletion interfered with contraction of the Igκ locus, which is a possible alternate mechanism that could explain the severity of the phenotype.

Finally, our observations at Jκ led us to explore the importance of nucleosome positioning genome-wide in the context of cryptic recombination signal sequences (cRSSs). The innovation of a highly conserved sequence at which the double strand break very specifically occurs allows for a high degree of regulation over the process of gene recombination. However, due to the sheer size of the genome, sequences similar enough to a canonical RSS to be a possible substrate for RAG are predicted to occur once per 600bp (Lewis 1997). Additionally, the RAG complex is known to bind at thousands of sites throughout the lymphocyte genome, prompting the very real possibility of catastrophic levels of off-target DNA cleavage (Teng 2015). However, off-target recombination by RAG occurs only very rarely. Based on our results, we wanted to explore the possibility that higher levels of nucleosome occupancy at these sequences. Using ATAC-seq data and a comprehensive annotation of all cryptic RSSs in the mouse genome, we found that cryptic RSSs are significantly more likely to be occupied by nucleosomes than random sequences of the same length. Additionally, cRSSs are depleted from the active compartment of chromatin, and instead are more likely to be located in less active, less accessible regions of the genome. This adds another important layer onto our understanding of how lymphocytes are protected from the genomic instability that would result from high levels of off-target RAG-mediated recombination.

CHAPTER 4

DISCUSSION

The foremost goal of this study was to elucidate the importance of a DNA motif, the GAGA motif, on the function of an important epigenetic reader and modulator, Brwd1. Initially identified for being genetically regulated in a pattern similar to Igκ in developing B cells, knock out studies showed that Brwd1 was required for B lymphopoesis, and particularly for Igκ recombination [72].

Brwd1 is named for two hallmark features: tandem bromodomains and a WD repeat domain (Huang 2003). As predicted by its bromodomains, Brwd1 is recruited to the active histone marks H3K9Ac, H3S10p, and H3K14Ac, and a motif analysis revealed that approximately 64% of these Brwd1 binding sites co-localized with an extended GA repeat motif [72]. Prompting our study was the finding by Mandal et al that Brwd1 localizes to the Jκ region of Igκ, where there were marked changes in accessibility and nucleosome positioning [72].

Though it was clear that Brwd1 played a crucial role in chromatin remodeling at Igκ prior to recombination, the mechanism for this remodeling was less clear, particularly the role of GAGA motifs. The authors had striking data that showed that though Brwd1 bound in my locations throughout the genome, it was only at binding sites containing GAGA motifs where there was a large change in nucleosome occupancy between the WT and KO [72]. However, it was unclear whether GAGA motifs were required for chromatin remodeling, or whether this finding was correlative.

To mechanistically investigate the importance of GAGA motifs on Brwd1 function, we employed the use of CRISPR-Cas9 gene editing at the Igκ Jκ locus. Because we know that Brwd1 actively remodels chromatin at this region in small pre-B cells, and because each

functional Jκ locus is preceded by a 5mer GAGAG motif, we had a relatively simple system in which to test the importance of the GAGA motif on Brwd1 functioning.

Removal of the Jκ1 GAGA motif, which is located approximately 50bp 5' to the Jκ1 recombination signal sequence (RSS), was accomplished using two targeting guide RNAs on either side of the GAGA sequence, and resulted in the creation of two independent lines, both with small deletions encompassing the GAGA motif. As anticipated, this deletion resulted in a failure to remodel nucleosomes at Jκ1 and a corresponding decrease in accessibility at the Jκ1-RSS. Correspondingly, we observed a decrease in recombination at Jκ1. These findings seemed the result of a functional failure of Brwd1, rather than a failure to recruit Brwd1.

However, there were also several results that we found surprising. Not only did the removal of the Jκ1 5' GAGA sequence affect recombination of Jκ1, but it resulted in more broad effects throughout the locus, including lower expression of recombination products of other Jκ segments, as well as a slight developmental block at the small pre-B cell stage, and an increased ratio of lambda:kappa usage. Also unexpectedly, though germline transcription was reduced at Jκ2-5, it was unaltered at Jκ1. These results necessitated the creation of a second knockout model to determine whether this was a general feature, or something particular to the Jκ1 GAGA. For our next model, we chose to remove the 5' GAGA motif from Jκ2. This model exhibited the expected results in decreased usage of Jκ2, but without larger developmental defects.

Finally, using the insight we gained on the function of nucleosome occupancy and positioning in Igκ recombination, we decided to investigate whether this method of regulation might also be employed throughout the genome at so-called cryptic RSSs (cRSSs). Our data suggests that cRSSs are indeed more likely to be occupied by nucleosomes. Additionally, we found that cRSSs are depleted from active regions of the genome. Both of these findings help to explain how the genome is able to protect itself from off-target recombination events and thereby preserve genomic integrity.

4.0.1 The Accessibility Hypothesis

One longstanding question in the field of lymphocyte development is that of ordered recombination. In B cells, the recombination of different immunoglobulin subunits is strictly ordered and segregated, with heavy chain (Igμ) recombination always preceding light chain (Igκ) recombination [199]. Additionally, at the heavy chain, DH_H to J_H recombination always precedes V_H to $D_H(J_H)$ recombination. However, recombination at all of these loci occurs through the actions of the same RAG recombinase. Thus, it was an early mystery as to how RAG was instructed to first target the heavy chain, then later the light chain.

An explanation for this phenomenon came in the form of the accessibility hypothesis, devised by Yancopoulos and Alt [200]. They observed that unrearranged V_H segments, rather than being unexpressed as previously thought, were expressed at a high level, but only in a strict developmental and tissue-specific manner. At very early stages of B cell differentiation in cells undergoing V_H to $D_H(J_H)$ recombination, unrecombined germline V_H transcripts were expressed at a high level. This led them to suggest that ordered V(D)J is regulated by the differential chromatin accessibility of the different gene segment loci [201].

More support for the accessibility hypothesis came from in vitro studies showing that RAG proteins would cleave RSSs in purified nuclei in a manner specific to the cell type of the source of the nuclei [38]. RAG would cleave at Igμ in cell nuclei extracted from B progenitor cells, but not from T progenitor cells, and vice versa, suggesting that different chromatin patterns in different cells played an important role in whether or not they would be cleaved by RAG. Later studies showed the importance of histone acetylation in promoting the accessibility of chromatin for recombination [202].

The strong correlation between germline transcription and V(D)J recombination first observed by Yancopoulos and Alt has led to much research into the role of transcriptional

control elements like as enhancers and promoters, transcription factors, and of transcription itself in making RSSs accessible for recombination [203, 204, 205, 206, 207, 69]. Perhaps the most conclusive evidence for the role of the necessity of transcription in gene recombination came from studies in T cells. Abarrategui et al showed that the blocking transcription elongation at the mouse T cell receptor-α locus suppressed V_α-to-J_α recombination and chromatin remodeling of the J_α segments [208].

In our data, one somewhat surprising finding was that, despite the clear drop in recombination at Jκ1 in the Jκ1-GAGA-deletion small pre-B cells, there was not a corresponding decrease in germline transcription at Jκ1. Using paired-end RNA seq, we were able to specifically examine unrecombined, germline transcription by excluding multimapping reads (for example, reads that mapped to both a Vκ and a Jκ segment). This data showed that transcription was significantly diminished at Jκ2, Jκ4 and Jκ5, but not at Jκ3 and Jκ1. Indeed, at Jκ1, transcription actually seemed to trend upwards. Furthermore, when examining the Jκ locus as a whole in Jκ1-GAGA-KO small pre-B cells compared to WT, we found that germline transcription was not significantly changed. Our data suggests that while transcription may be necessary for V(D)J recombination, it is not sufficient. This conclusion is supported by another study that showed that in PAX5-deficient pro-B cells, VH gene segments failed to recombine despite having normal levels of germline transcription [209].

We also examined germline transcription throughout the Vκ locus and found no significant differences in transcription. Given that no edits were made that were expected to alter chromatin accessibility at the Vκ locus, this was not surprising, and supports existing data that the upregulation in germline transcription precedes recombination, rather than occurring because of recombination.

In our analysis, we observed that the different Jκ gene segments differed in levels of germline transcription, which was somewhat surprising given how small each segment is, and given that the entire Jκ region is less than 1.5 kb in total. Because Igκ recombination

in general was so diminished in the Jκ1-GAGA-deletion small pre-B cells, it's difficult to conjecture how or indeed whether germline transcription would be impacted in the case of diminished recombination of just one segment. To explore this, germline transcription at each Jκ segment in the Jκ2-GAGA-deletion small pre-B cells could be analyzed, as these cells exhibited overall more normal recombination, with only Jκ2 usage being significantly down.

4.0.2 Nucleosome Positioning

Our studies reveal that the GAGA DNA motif 5' to Jκ1 is crucial for nucleosome positioning around the Jκ1 RSS. Prior to recombination in wild type (WT) pro-B cells, the Jκ1 RSS is occupied by a nucleosome, which likely helps prevent its cleavage by RAG. In WT small pre-B cells, the nucleosome density is no longer present at the Jκ1-RSS, but a new nucleosome is present within a few hundred base pairs upstream of the RSS. The Jκ1-RSS in Brwd1-KO small pre-B cells remain obscured by a nucleosome, suggesting that Brwd1 is crucial to pre-recombination nucleosome remodeling. To that model, we can now add the necessity of an upstream GAGA sequence for effective Brwd1 nucleosome remodeling. Our very small targeted deletions encompassing the 5' upstream GAGAG sequence proved sufficient to disrupt chromatin remodeling at Jκ1.

Moreover, this disruption in chromatin remodeling produced a tangible effect upon gene recombination itself. In small pre-B cells containing a deletion of the GAGA motif (Jκ1-GAGA-del and Jκ2-GAGA-del), we saw marked differences in the quantity of recombination and the selection of Jκ segments. This effect was most dramatic in the Jκ1-GAGA-del mice. There, the deletion of the GAGA motif resulted not only in a reduction in usage of Jκ1, but overall recombination defects at the Igκ locus overall, as reflected in fewer overall kappa positive immature B cells and more lambda positive immature B cells. The effects of GAGA

deletion at Jκ2 were more specific, and rather than disrupting recombination generally, the effect seemed highly localized to the Jκ2 locus. Indeed, we were unable to clone any Jκ2 recombination products from the Jκ2-GAGA-deletion mice; a finding that was corroborated by qPCR data showing significantly lower expression of Jκ2 recombination products.

The functional consequences of nucleosome positioning at Jκ present a powerful argument for the importance of an underappreciated element of chromatin organization. Much of the focus on primary chromatin structure has been at the level of accessibility. However, the absolute position of nucleosomes, as demonstrated with the Jκ-GAGA-deletion models, can have important functional consequences.

The system of the immunoglobulin kappa gene locus presents a rare opportunity to study both the mechanistic and functional consequences of a particular DNA motif on chromatin organization. From studies of transcription, we know that the specific positions of nucleosomes in the genome can be greatly consequential. This is because the DNA wrapped around nucleosomes is occluded, and hence the accessibility of functional DNA binding sites can be affected [210]. Nucleosomes can block the binding of crucial transcription factors, necessitating an entire class of so-called pioneer transcription factors that are essential during development to helping the transcription machinery gain access to regions of high nucleosome density [211]. Additionally, complexes such as SWI/SNF have evolved to remodel chromatin, sliding nucleosomes away from loci fated for active transcription in an ATP-dependent manner [212]. The presence of nucleosomes can even cause transcriptional pausing, and thus slow the transcription of genes occupied by nucleosomes [213].

Nucleosome positioning likewise plays a critical role in RAG recruitment and Igκ recombination. V(D)J recombination is dependent on the cleavage of recombination signal sequences (RSSs) by the RAG recombinase, which is composed of two subunits: RAG1 and RAG2 [214, 215]. RAG1 binds the nonamer of the RSS, as well as contains the active site for DNA cleavage [73]. The RAG2 subunit is also crucial for DNA cleavage activity. RAG2

is recruited to regions of active chromatin via a plant homeodomain (PHD) finger that binds specifically to trimethylated histone H3 lysine 4 (H3K4me3) [194, 195]. Because the RAG recombinase requires both subunits, RAG2 recruitment, and consequently the function of the recombinase, is dependent on the local presence of nucleosomes containing the H3K4me3 modification. However, when bound within nucleosomes, RSSs are resistant to cleavage by RAG [162]. Thus, for efficient recombination, one would expect a mechanism to have evolved to ensure not only that histones carrying the H3K4me3 modification present at Jκ, but also that RSSs are free of nucleosomes prior to recombination. However, this mechanism has long remained elusive.

One possible way for precise nucleosome positioning to be encoded could be at the level of the sequence itself. Indeed, whether nucleosome positioning is a property intrinsic to DNA sequences remains an area of active debate. Some scholars contend that there is a genomic code that determines nucleosome positioning, and this allows for the prediction of nucleosome occupancy based on genetic sequence [216]. This model is based on an experimental model by which nucleosomes were assembled and mapped in vitro, and then compared with nucleosome maps in vivo. Based on similarities between these maps, researchers concluded that nucleosome positioning was strongly dependent on intrinsic interactions between histones and DNA nucleotides.

On the basis of sequence alone, the RSS should be a preferred nucleosome binding region, and in in vitro experiments, the RSS does serve as a nucleosome-positioning sequence [163]. Thus, one would expect that the RSS should exhibit a high frequency of nucleosome occupancy in the immunoglobulin loci. However, this runs contrary to necessity of the immunoglobulin RSSs to be accessible to RAG, and conflicts with ATAC-seq data showing high accessibility at the RSSs in cells poised for recombination. Taken together, this suggests that ATP-dependent nucleosome remodeling complexes are likely actively involved in altering nucleosome positioning and increasing accessibility at RSSs prior to recombination

[165].

Significant evidence in support of this notion was presented in Mandal 2015. The authors found that while Jκ RSSs are highly accessible in WT small pre-Bs, there was a marked decrease in accessibility in the Brwd1KO cells [72]. Furthermore, there was a seeming failure to shift nucleosomes off of the RSS of Jκ segments in the Brwd1KO cells. When looking genome-wide, regions where Brwd1 was found coincident with the activating histone marks H3S10pK14Ac and H3K9Ac were generally free of nucleosomes. However, significantly more active remodeling was found where these peaks were also associated with stretches of GA-repeats. These regions accumulated nucleosomes when Brwd1 was knocked out; however, the presence of Brwd1 lead to a decrease in nucleosome density at these regions [72]. Igκ emerged as an ideal test locus to determine the importance of GAGA motifs at Igκ.

Mechanistically, it is mostly likely that GAGA motifs influence the sliding or eviction of nucleosomes from chromatin. There is evidence that Brwd1 binds a component of the SWI/SNF complex, which is able both to slide nucleosomes and evict them from chromatin [143]. In Drosophila, GAF is known to interact with NURF, a chromatin remodeling complex that acts by sliding nucleosomes [107, 108]. From our data, we are unable to distinguish between these two mechanisms. However, we favor a sliding model, as we observe the new placement of a 5' nucleosome, not just the loss of a nucleosome over the RSS.

One important question to consider was what the minimum consensus sequence for enhanced Brwd1 activity. Previously, Brwd1 had been associated with chromatin remodeling activity at extended regions of GAGA repeats (GA_{11}) [72]. However, this did not necessarily represent a biological minimum consensus sequence, but merely a minimum motif that could be bioinformatically determined with confidence genome-wide. Indeed, another example of GAGA motifs in nucleosome remodeling suggested that perhaps a smaller stretch of repeats could suffice. Extensive studies of the Drosophila GAGA Factor (encoded by Trl) had shown that the minimum required consensus sequence for GAGA factor to be able to bind and ef-

fect nucleosome sliding was GAGAG [140, 142]. The Igκ-J locus has no extended stretches of GA repeats. However, the functional Jκ segments are preceded by a shorter GA repeat domain, GAGAG.

Our studies suggest that like GAGA Factor, Brwd1 requires a minimum consensus sequence of GAGAG for robust remodeling activity. The deletion of this motif resulted in a pronounced effect both on nucleosome positioning and on usage of the associated Jκ segment. If enhancement of Brwd1 activity is dependent on extensive GA repeats, we would expect the removal of this 5mer to have little effect. Additionally, from our studies we can infer that the 5mer is the likely minimal sequence. The deletion in Jκ1-GAGA edited mice brings in closer proximity to the RSS a GAGA 4mer located upstream of the deleted GAGA 5mer. Were a 4mer sufficient for activity, we likely would not see a strong phenotype in our mice.

Another question raised by our studies is the importance of the location of the GAGA motif relative to the RSS. While our experiments that remove the most proximal GAGA motif suggest that proximity to the RSS is important, it is unclear what range of proximity is acceptable for proper function, and how narrow or broad is that range. We can gain some insights from the proximity of GAGA motifs to the 5' Jκ1-RSS across and within species. Within the mouse Igκ locus, the GAGA ranges from 51 to 172 bp 5' of the RSS, averaging 90 bp. This all falls within a relatively narrow acceptable range, and furthermore is within the 146bp contained within one nucleosome. However, an examination of GAGA motifs at Jκ across several mammalian species suggests the possibility either of more flexibility in placement of the GAGA motif, or different spacing requirements in different species. In the seven species we examined, four had the Jκ1 GAGA motif located 50-80bp upstream of the RSS [217]. The other three had the Jκ1 GAGA motif located approximately 350bp upstream of the RSS. Thus, while we can speculate that the GAGA motif must be located within the space of approximately two nucleosomes length of DNA, the stringency of this spacing is

unclear.

One significant point of consideration in our findings is that while our targeted deletions were small, the genetic edits were not strictly limited to the five base pair GAGA motif, and removed elements of sequence both 5' and 3' of the targeted GAGA. Thus, it is possible that the effects we are seeing are not strictly the result of deletion of the GAGA motif. They could reflect an accidental deletion or creation of a consensus motif for a transcription factor or other chromatin remodeling complex. In the case of the Jκ2 phenotype, it is possible that altering the distance between Jκ1 and Jκ2 impacts recombination frequency.

To address this concern, a more specific genetic edit would be required. This could be accomplished using a CRISPR-Cas9 method with homology directed repair, rather than non-homologous end joining. In this method, a repair DNA template is provided homologous to the region surrounding the cut site. To properly address all concerns, an ideal repair template would replace the GAGAG sequence with a sequence of the same length and GC content, such as CCCTT, for example. If the phenotype of such an edit recapitulated the Jκ-GAGA-deletion phenotype, the model we have presented would be supported.

These findings extend our current mechanistic understanding of the epigenetic regulation of recombination at the Igκ locus. As was known before, prior to recombination, the Jκ locus is modified with the activating histone marks H3K9Ac and H3S10pK14Ac, making the region more accessible and transcriptionally active [3]. Then, Brwd1 is recruited to these activating histone marks, and repositions nucleosomes off the RSSs in the region to make them accessible to RAG, a process in which GAGA motifs play a critical role [72]. What remains unclear, however, is how Brwd1 recognizes and interacts with GAGA motifs in the process of remodeling chromatin.

Conjectures can be made as to possible mechanisms. Based on Brwd1's tandem bromodomains, and our finding that loss of GAGA does not abolish recruitment of Brwd1,

it is likely that Brwd1 is primarily recruited by binding active histones [218]. Furthermore, there are many Brwd1 recruitment sites across the genome that are decorated with H3K9AcH3S10pK14Ac but do not have GAGA motifs [83]. Rather, GAGA motifs are associated with function and not recruitment. However, there must be a means by which Brwd1 recognizes or interacts with GAGA motifs. One plausible mechanism is by direct binding.

Brwd1 is a very large protein (approximately 260 kDa) of which few domains, namely the tandem bromodomains and the WD repeat domain, are at all understood [143]. Only the tandem bromodomains and the WD repeat have clear domain homology. Thus it is not impossible that it could have a DNA binding domain (DBD) as well as bromodomains. However, a simple domain analysis of Brwd1 does not find any likely DBDs. An alignment of the Drosophila Trl and mouse Brwd1 does reveal some homology between the zinc finger DBD and a region near the C terminus of BRWD1. However, it is unclear if this homology is sufficiently strong to form a functional DBD.

We have begun to examine the role of this domain of Brwd1. First, in cell lines, we targeted the Trl homologous region in Brwd1, which lies in exon 41, with a CRISPR guide targeted to exon 39, and produced a frameshift mutation. We then performed ChIP-qPCR from these cells, but found that binding of Brwd1 to a control region was entirely ablated in the edited cells. This pointed to the likelihood that our edit had caused nonsense-mediated decay of the mRNA, and was thus not useful in studying the role of the C terminal domain.

Our next attempt, which is ongoing, attempted a smaller deletion using paired guide RNAs up and downstream of the putative zinc finger domain in Brwd1. The goal in this case was not to truncate the protein, but simply to disrupt the suspected DBD. This was injected into mice, and produced several different heterozygous mice, which are in the process of being bred to homozygosity. It will be interested to see whether this mutation disrupts the activity of Brwd1. If the Brwd1 in these mice is still able to localize to active histone marks throughout the genome, but does not exhibit chromatin remodeling capacity, then

we will have likely identified the domain by which Brwd1 interacts with GAGA motifs, and thus, how GAGA motifs are able to influence the chromatin remodeling activity associated with Brwd1. If this mutation has no impact on chromatin remodeling, then it is more likely that Brwd1 only binds histones, not DNA, and it is the recruitment of other factors like SWI/SWF that is responsible for sliding histones, and possibly another factor entirely that is interacting with GAGA motifs.

The actual process of actively sliding or evicting nucleosomes is generally ATP-dependent, and Brwd1 does not contain an ATPase. Thus, it is likely recruiting other factors to perform the actual remodeling of chromatin. A strong candidate for this is Brg1, which was shown to interact with Brwd1 [143]. Brg1 is a component of the SWI/SNF complex, which serves as a chromatin remodeler responsible for sliding and evicting nucleosomes. One possible means of testing this hypothesis would be through a conditional knockdown of Brg1 in pre-B cells to see if the Brwd1KO phenotype was recapitulated.

Additionally, we would like to perform more extensive immunoprecipitation experiments to add to those performed by Huang et al. To this end, we chose a 43 amino acid C-terminal section of the Brwd1 protein that did not share homology with other mouse proteins, but would be affected by our truncation mutation, and cloned it into the pGex-3x GST-tagged plasmid. The antigen was then overexpressed in bacteria, and the GST-tagged protein was collected on a glutathione sepharose column, washed, and eluted using reduced glutathione. The purified GST-tagged antigen was then injected in rabbits, and polyclonal sera was collected.

To purify the polyclonal anti-Brwd1, we first depleted GST-specific antibodies. We crosslinked GST to a glutathione sepharose column using DSP. The serum was loaded onto the column and allowed to incubate overnight. The flow through was then collected, and anti-GST antibodies were eluted by low pH. On a second column, we crosslinked the GST-Brwd1-fusion antigen to a glutathione sepharose column using DSP. The anti-GST depleted

serum was loaded onto the column and allowed to incubate overnight. The serum was then allowed to flow through the column. Anti-Brwd1 antibodies remained bound in the column, and were antibodies were eluted by low pH [219]. The anti-Brwd1 antibodies were pH neutralized, and quantified.

We then attempted an immunoprecipitation using our purified polyclonal antibodies in WT splenic B and bone marrow B cells. Though we identified a promising band at 260kD, the mass spectrometry did not identify Brwd1 in the pull down. It is likely that either the antibodies need to be further purified and concentrated, or that we did not adequately lyse the nuclei in our immunoprecipitation, and thus did not allow sufficient access to Brwd1.

Overall, these findings concur with and add new layers to the understanding of the role of nucleosome positioning in Igκ recombination. We now have evidence in vivo to supplement the in vitro evidence that RSSs bound within nucleosomes are protected from cleavage by RAG. However, though it is important that nucleosomes not obscure the RSS, it nevertheless is important that there be nucleosomes in the vicinity of the RSS, as RAG2 is recruited to the activating histone mark H3K4me3. The GAGA factor domain appears to be important for the regulation of both of these critical aspects of nucleosome positioning. Our data suggest that Brwd1 is recruited via its tandem bromodomains to the activating histone marks H3K4me3 and H3S10pK14Ac present throughout the Jκ locus. The upstream GAGA sequence then assists Brwd1 in shifting these active histone-mark-containing nucleosomes upstream. Based on the binding mechanism of RAG, these nucleosomes are crucial to the recruitment of RAG to the Jκ locus, where it initiates recombination. Simultaneously, the movement of these nucleosmes upstream clears the RSS of nucleosome occupancy, which likely makes it much more accessible to cleavage by RAG1.

4.0.3 Brwd1 recruitment to Igκ

The Jκ2-GAGA-deletion caused a significant decrease in usage of Jκ2, and the Jκ1-GAGA-deletion caused not only a decrease in Jκ1 usage, but a serious recombination defect at the Igκ locus more broadly. These defects were reflected in the nucleosome structure of the locus. In the Jκ1-GAGA-deletion mouse, nucleosome structure across the entire Jκ locus was disrupted. Within the limits of our data, this suggests that the positioning of the first nucleosome at Jκ1 is likely important for the ordering of nucleosome structure across all of Jκ. By contrast, the effect on Jκ2 recombination was localized to Jκ2, which suggests that the Jκ2 GAGA is only important for the establishment of nucleosome structure at Jκ2. We will use ATAC-seq in the Jκ2-GAGA-deletion small pre-B cells to demonstrate whether this is indeed the case.

There are two obvious potential mechanisms by which the deletion of the GAGA motif could effect this change. The first possibility is that Brwd1 is unable to effectively remodel chromatin in the absence of these GAGA motifs. The second possibility is that without a GAGA motif, Brwd1 is unable to be properly recruited to the locus. Our data suggest that it is the first of these mechanisms at play. To test these models, we performed a chromatin immunoprecipitation and quantitative PCR (ChIP-qPCR) of Brwd1 in Jκ1-GAGA-deletion and WT cells for the Jκ1 and Jκ4 loci. We did not see a significant difference in Brwd1 binding at Jκ1 in the Jκ1-GAGA-deletion cells compared to WT. To rule out the possibility that recombination defects are due to a failure to recruit RAG, a RAG ChIP-qPCR in Jκ1-GAGA-deletion cells would be required.

The chromatin immunoprecipitation showed that Brwd1 is still recruited to Jκ1 in cells lacking a 5' Jκ1 GAGA motif. However, a combination of limitations of the method and spread in the data leave some room for interpretation. To first address the data, while it is true that there was no significant difference in binding between WT and KO cells at Jκ1,

there was a trend toward less binding at Jκ1. This stands in marked contrast to Brwd1 binding at Jκ4, where Brwd1 binding is somewhat higher in the KO compared to the WT. It is possible that with more replicates, a statistically significant difference in binding at Jκ1 between the WT and KO would have emerged.

There are also limitations in resolution inherent to the method. The degree to which binding can be resolved to a particular locus is limited by the degree of shearing of the chromatin prior to immunoprecipitation. Our protocol is optimized to shear to approximately 500bp fragments. For most applications, this resolution is more than sufficient. However, given that the entire Jκ locus is only 1400bp long, and each Jκ segment is separated from the next by approximately 300bp, it is hard to be absolutely confident that this method provides an accurate picture of Brwd1 recruitment to the Jκ1 locus specifically. To address this limitation, the resolution of the ChIP-qPCR could potentially be increased by more extensive sonication, or by using an MNase digestion [220].

4.0.4 Sequential vs Stochastic Recombination

The question of whether recombination is sequential or stochastic has long been of interest to immunologists. The general process of construction of an immunoglobulin was early recognized to be highly ordered, with V(D)J recombination of the heavy chain strictly preceding V(J) recombination at the light chain [201]. What was less clear for a time was whether recombination at the kappa and lambda light chains was stochastic or sequential. Elegant work demonstrated that in regard to the light chains broadly, the sequential model was correct, with kappa recombination preceding lambda [221, 222]. This also revealed an interesting feature of Igκ recombination, which is that multiple different recombination events can occur on the same allele [223]. If the first rearrangement proves nonproductive, the cell can then attempt a second recombination using a Jκ segment 3' to the first attempt, thus

removing the failed junction in the process.

Now that this first question of stochastic vs sequential usage of the light chain has been conclusively solved, another, more intricate question remains: that of the sequential or stochastic selection of Jκ segments themselves. Jκ segments are ordered 5' to 3', with Jκ1 being first and also closest to the Vκ segments. In a strictly sequential model, recombination would occur first at Jκ1, with subsequent recombination attempts proceeding to Jκ2, Jκ4, then Jκ5. In this model, we would predict the majority of recombination products to use Jκ1, and the fewest to use Jκ5. Additionally, interfering with the Jκ1 segment would be expected to have a large effect on recombination as a whole. In a purely stochastic model, Jκ segments would be chosen at random, and thus we would expect to see each segment being used in approximately 25% of recombination products. In such a case, interference with any one segment would be unlikely to have a large effect on recombination as a whole.

The ability of a small pre-B cells to undergo multiple rearrangements on one allele presents a potential evolutionary benefit of a sequential model of recombination on the Igκ allele, because it presents the cell with more opportunities to productively recombine kappa. Additionally, the sequential model potentially helps to reconcile the difference in kappa vs lambda usage. In mice, kappa is used approximately 90% of the time, while lambda is used only 10% of the time. Multiple chances per cell to recombine at kappa could help explain why cells so infrequently are forced to recombine at lambda.

However, WT rearrangement frequencies of Jκ stand in contrast to the logic of the sequential model. Published frequencies of usage of Jκ record Jκ1 being used in 33% of products, Jκ2 in 25% of products, Jκ4 in 13% of products, and Jκ5 in 28% of products (Prak 1994). Our data in WT small pre-B cells supports this reported recombination frequency (Jκ1- 34%, Jκ2- 21%, Jκ4- 11%, and Jκ5- 32%). However, these frequencies are not what we would expect to see with a sequential model. Jκ5 is used nearly as often as Jκ1, and Jκ2, which one would expect to be used the second most frequently, is actually used the least

frequently.

The stochastic model of Jκ selection also has potential evolutionary benefits. One of the most basic purposes of V(D)J recombination is to create an enormous diversity of antigens. Here, a stochastic model would have the benefit of promoting diversity in recombination products. However, there are also issues with reconciling the observed recombination frequencies with a stochastic model, the most obvious of which is that the Jκ segments are not used at equal rates.

Some of the variation in usage seen can be explained by the quality of RSS. Both Jκ2 and Jκ4 have slightly non-canonical RSSs. At first glance, Jκ2 appears to have a perfectly canonical RSS; however, the canonical nonamer and heptamer are separated by 24, not 23 base pairs. Thus, if following the 12-23 rule, to use Jκ2 would require the recognition of a one-base-pair off nonamer. One could also speculate that recombination at Jκ2 is probably more frequently non-productive due the RSS structure. Additionally, Jκ4 is used less frequently, but this can be at least partially explained by an RSS that varies from canonical by one nucleotide.

Our knock out Jκ1-GAGA lines provide a nice model for investigating whether recombination at Jκ is sequential or stochastic. If choice of Jκ segment is entirely stochastic, and removing the GAGA sequence interferes with recombination at Jκ1, then we would expect to see a skewing in the repertoire away from Jκ1, and towards Jκ2, Jκ4 and Jκ5. We would not expect an impact on the overall efficiency of kappa recombination. By contrast, if Jκ recombination is sequential, we would expect to see a larger impact on kappa recombination overall, and likely a decrease in the kappa: lambda ratio. The Jκ2-GAGA-deletion mouse model is also helpful in teasing out this question. In a stochastic model, this mouse should have decreased usage of Jκ2, and increased usage of all other segments. In a sequential model however, Jκ1 would be expected to be the dominantly used segment, with little contribution from any other segment.

When initially planning this experiment, we envisioned a stochastic model of Jκ recombination, which is why the more dramatic general Igκ recombination defect we observed in the Jκ1-GAGA-deletion mice was surprising. This phenotype was not consistent with stochastic model, and furthermore could not be explained away by considerations of RSS quality. However, the results from our Jκ2-GAGA-deletion mice also forced us to reject a strictly sequential model, as we only saw a significant decrease in recombination at the Jκ2 locus.

Neither of these simple models was able to predict or explain the recombination results in our selective knockout lines. In opposition to the stochastic model, we found a large recombination defect across all of Jκ in the Jκ1-GAGA-deletion small pre-B cells, which also exhibited a decreased ratio of lambda to kappa. In opposition to the sequential model was our finding that in the Jκ2 mouse, we saw virtually no change in percent usage of Jκ1 (34% vs 32%), but an increase in percent usage of Jκ4 (11% vs 21%) and Jκ5 (32% vs 45%).

Thus, we propose a more nuanced gatekeeper model. By this model, establishment of a permissive chromatin landscape is dependent on the actions of BRWD1 at the Jκ1 5' GAGA motif. Chromatin remodeling at Jκ1 is essential, and has a ripple effect in chromatin organization across the entire locus. Following this chromatin organization, recombination will occur most frequently first at Jκ1. If this is unproductive, further rearrangements will occur largely stochastically, biased by the quality of the RSS. While the role of BRWD1 at other gene segments is important for their efficient usage, it is not critical for the locus as a whole.

Finally, we have not ruled out the possibility that our Jκ1-GAGA-deletion interfered with contraction of the Igκ locus, which is a possible alternate mechanism that could explain the severity of the phenotype. One effective means of exploring contraction is through florescence in vitro hybridization, using florescent probes against Vκ and Ck. We are currently in the process of undertaking these experiments, which will help shape the final interpretation of

the Jκ selection and recombination defects.

4.0.5 The Power and Limitations of the CRISPR-Cas9 technology

CRISPR-Cas9 gene editing is a relatively new and very powerful method that in theory offers researchers the flexibility to target nearly any region of the genome [224]. Derived from a sort of adaptive immune system in yeast, CRISPR-Cas9 has been adapted for many cell culture systems and animal models, and promises the ability to easily knock out genes via frameshift mutations, remove segments of DNA via non-homologous end joining (NHEJ), or even replace segments of DNA with a template DNA of the researchers choosing via homology directed repair (HDR). Our study demonstrates the power of this technique to investigate the effects of small motif deletions that would have been impractical or impossible prior to the wide-spread application of CRISPR-Cas9.

One concern we had when designing our study was that of off-target effects. That is, could we be certain that our phenotypes were the result of our targeted deletions, and not of some unknown edit elsewhere in the genome? It was impractical to perform full genome sequencing to look for such off-target edits, so we approached this issue through a two-pronged solution. First, we used bioinformatic tools to predict off-target rates, which in recent years have become very powerful. Second, we developed two lines that were produced and maintained separately. Because it was highly unlikely that both lines would have the same off-target edits, we could be confident that the phenotypes we saw were the result of our targeted edit.

However, we did experience difficulties with producing successful CRISPR-Cas9 edits. We spent a considerable amount of time attempting these experiments in a vAbl-transformed pro-B cell line (A70.2) [225]. However, we had significant difficulty successfully transfecting these cells, and when we did, we were rarely able to successfully isolate cells containing the

desired genetic edit. Our only real success with this technique was in the truncation Brwd1 mutant. However, that mutant was not very mechanistically informative, and we suspected that the mutation caused nonsense-mediated decay of Brwd1.

This prompted us to attempt CRISPR in mice, where we have had generally mixed results. About half of the pups of the Jκ1-GAGA-deletion injection displayed any edits, and all were heterozygous. Of these, only a few contained the desired mutation. Then, after greatly increasing the concentration of guide RNA and Cas9 protein, as well as adding Cas9 mRNA, we had a very successful round of editing at the Jκ2 5' GAGA sequence, and were able to produce several correct homozygously edited mice in the F0 generation.

We had several unsuccessful injection attempts as well, most notably when attempting add-backs. We first tried to see if we would be able to successfully target the Jκ3 segment, and add back a GAGA plus RSS sequence, but we were unsuccessful. We were also unsuccessful in the attempt to very specifically edit the Jκ1 5' sequence, and use a repair template so as to only remove the GAGA 5mer, even at the higher concentration of guide and Cas9. To rule out the possibility that the guide RNA was defective or inefficient, we tested guides in vitro, and saw high levels of cleavage. It is possible that we had so much difficulty in targeting the Igκ locus due to the high degree of nucleosome structure throughout this locus, which may make it less accessible to Cas9.

4.0.6 A potential method for preventing recombination at cryptic RSSs

The recombination signal sequence is composed of a highly conserved nonamer and heptamer sequence separated by a random sequence of 12 or 23 base pairs. While this genetic structure is critical to the function of an RSS, it is unfortunately not unique to the RSS. Sequences known as cryptic RSSs (cRSSs) are similar enough to the canonical RSS to be a viable substrate for RAG. These sequences are predicted to occur once every 600 bp,

and given the size of the mouse (or indeed the human) genome, this equates to millions of sites spread throughout the genome at which off-target RAG recombination could occur [84]. This along with the fact that the RAG complex binds at thousands of sites throughout the lymphocyte genome has alarming implications for the genomic stability of lymphocytes. However, off-target RAG-mediated recombination is only very rarely seen [191].

Teng et al hypothesized that there would have been selective evolutionary pressure to deplete cRSSs from RAG binding sites to lower the risk of off target sites. Indeed, an analysis of RAG binding sites showed a significant depletion of cRSSs. However, this depletion was by no means complete, and there were still many RAG binding sites overlapping with cRSSs, suggesting that there are likely other means of discouraging off-target recombination at these sites.

Thus, in a similar line of reasoning to Teng et al, we hypothesized that cRSSs should be depleted from the actively transcribed regions of the genome (Compartment A), and should conversely be enriched in Compartment B, which contains the largest inactive heterochromatin regions of the genome. Compartment B heterochromatin is generally more nucleosome-dense and contains less accessible chromatin. Additionally, Compartment B contains significantly fewer histones with activating modifications, such as the H3K9Ac to which RAG2 binds. Because cRSSs in Compartment B seemed less likely to be errantly recombined, we hypothesized that there would be evolutionary selective pressure for these sequences to be preferentially located in Compartment B.

An analysis of HiC data from WT mice enabled us to demarcate the boundaries of Compartment A and Compartment B. Compartments were determined by the correlation matrix, with the A/B compartments based on the sign of the eigenvector. We then aligned the cRSSs within the compartments, and calculated the density of cRSSs within each compartment, and in the genome as a whole. From this analysis, we were able to conclude that cRSSs are indeed depleted from the active compartment of chromatin, and instead are more likely

to be located in less active, less accessible regions of the genome. This expands upon our understanding of how lymphocytes are protected from high rates of off-target RAG-mediated recombination.

It would be interesting to determine if the same logic holds in non-recombining cell types. Compartments are cell type specific, and non-recombining cells would not have the same evolutionary pressure to deplete cryptic RSSs from active compartments. Future studies could examine cell types such as embryonic stem cells and follicular B cells to determine the density of cryptic RSSs in compartment A vs compartment B.

Additionally, our results suggest that increased nucleosome occupancy of the RSS in vivo lowers recombination frequency by limiting accessibility to RAG. There is some evidence for a similar mechanism occurring for transcription factor binding sites (Segal 2006). Segal et al, who subscribe to the still controversial model of a genomic code for nucleosome positioning, noted that for an given transcription factor, there exist throughout the genome many more binding sites than the transcription factor actually occupies. It seems likely that many of these sites that are unbounded by transcription factor occur by mere chance. They hypothesized that the intrinsic nucleosome organization encoded by genomes allows for the positioning of stable nucleosomes over non-functional transcription factor binding sites, thereby decreasing their accessibility to transcription factors. To test this hypothesis, they examined predicted nucleosome occupancy at functional vs non-functional binding sites, and found that the occupancy was predicted to be lower at the functional sites. In the case of transcription factors, the likely consequence would be biasing transcription factor binding to functional sites by excluding them from the random, non-functional sites.

To determine whether cRSSs are analogous to random transcription factor binding sites in this way, we wanted to explore the possibility that crptic RSSs have higher than expected average nucleosome occupancy. Using nucleosome occupancy data derived from ATAC-seq data, we indeed found that cryptic RSSs are significantly more likely to be occupied by

nucleosomes than random sequences of the same length. Thus, it is possible that though these chance cRSSs are largely prevented from accidental recombination by being obscured by nucleosomes.

Finally, we found that as the quality of the cryptic RSS improved, which is to say the closer it resembled a canonical RSS, the more striking the increase in mean nucleosome occupancy. There was a steady progression in nucleosome occupany values as the cRSS score increased. This suggests that evolutionarily, there is more incentive to protect those sites that are at the highest risk of being subject to off-target cleavage.

4.0.7 Future directions

Our studies raise several outstanding questions and directions for future experiments. In regards to the role of germline transcription in the accessibility hypothesis, our results suggest that transcription is necessary but not sufficient. However, we did not investigate whether altering the usage of one segment corresponds to a decline in speicifc drop in germline transcription at that locus. To investigate this, one could use qPCR to analyze germline transcription at Jκ2 vs the other Jκ2 loci in the Jκ2-GAGA-deletion mouse.

Our studies also provided new insight on the role of GAGA motifs influencing nucleosome positioning at Jκ. However, a more specific edit that strictly removed the GAGAG motif would be required to conclusively prove that it is specifically the GAGA motif domain, and not any of the surrounding DNA, that is critical. Similarly, it would be interesting to look at the effect of the positioning of the GAGA relative to the RSS, which could be studied by adding or deleting intervening sequence. Additionally, to rule out other possible explanations for the Igκ recombination defect we observed, a RAG ChIP-qPCR would be required to prove that it is not a result of a failure to recruit RAG. Similarly, an MNase ChIP-qPCR for Brwd1 at Jκ1 could be helpful in providing stronger evidence that the phenotype is not the result

of a Brwd1 binding defect. Finally, FISH for Vκ and Ck in the Jκ1-GAGA-Deletion small pre-B cells should be performed to rule out the possibility of the GAGA deletion resulting in a contraction defect.

Our experiments do not delve into the mechanism of Brwd1-mediated chromatin remodeling. We cannot conclude from our studies whether Brwd1 itself interacts with and slides or evicts nucleosomes, or whether it recruits a chromatin-remodeling complex to effect these changes. The Brwd1 potential DNA binding domain mutation and the immunoprecipitation experiments in progress in the lab should hopefully shed light on this mechanism in the future. Finally, our cRSS analysis was restricted to our data in small pre-B cells. It would be interesting to carry out the compartment analysis in non-recombining B cells.

4.0.8 Conclusion

Our work on the role of GAGA motif domains in chromatin remodeling of Igκ during B cell development add new dimensions to the accessibility hypothesis of ordered recombination, and illustrates how crucial specific nucleosome positioning is at Jκ in preparation for recombination. Furthermore, it provides evidence that GAGA motif domains are required for the role of Brwd1 in chromatin remodeling. This shows that the means of recruitment and activity of Brwd1 are segregated, allowing for more refined control over Brwd1-mediated chromatin remodeling. Finally, we shed more light on how genomic integrity in cells undergoing gene recombination is maintained by providing a mechanism for how cRSSs are shielded from off-target cleavage by RAG.

REFERENCES

[1] K. Murphy and C. Weaver, *Janeway's immunobiology*. Garland science, 2016.

[2] M. A. Rieger and T. Schroeder, "Hematopoiesis," *Cold Spring Harbor Perspectives in Biology*, vol. 4, dec 2012.

[3] M. R. Clark, M. Mandal, K. Ochiai, and H. Singh, "Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling," feb 2014.

[4] M. S. Schlissel, "Regulating antigen-receptor gene assembly," 2003.

[5] L. Erlandsson, S. Licence, F. Gaspal, P. Lane, A. E. Corcoran, and I. L. Mårtensson, "Both the pre-BCR and the IL-7Rα are essential for expansion at the pre-BII cell stage in vivo," *European Journal of Immunology*, vol. 35, pp. 1969–1976, jun 2005.

[6] J. K. Geier and M. S. Schlissel, "Pre-BCR signals and the control of Ig gene rearrangements," feb 2006.

[7] P. F. van Loo, G. M. Dingjan, A. Maas, and R. W. Hendriks, "Surrogate-Light-Chain Silencing Is Not Critical for the Limitation of Pre-B Cell Expansion but Is for the Termination of Constitutive Signaling," *Immunity*, vol. 27, pp. 468–480, sep 2007.

[8] D. Nemazee, "Receptor editing in lymphocyte development and central tolerance," oct 2006.

[9] E. Derudder, E. J. Cadera, J. C. Vahl, J. Wang, C. J. Fox, S. Zha, G. van Loo, M. Pasparakis, M. S. Schlissel, M. Schmidt-Supprian, and K. Rajewsky, "Development of immunoglobulin λ-chain-positive B cells, but not editing of immunoglobulin κ-chain, depends on NF-κB signals," *Nature Immunology*, vol. 10, no. 6, pp. 647–654, 2009.

[10] F. Melchers, "Checkpoints that control B cell development," *Journal of Clinical Investigation*, vol. 125, pp. 2203–2210, jun 2015.

[11] S. Herzog, M. Reth, and H. Jumaa, "Regulation of B-cell proliferation and differentiation by pre-B-cell receptor signalling," mar 2009.

[12] C. H. Bassing and F. W. Alt, "The cellular response to general and programmed DNA double strand breaks," aug 2004.

[13] G. Petkau and M. Turner, "Signalling circuits that direct early B-cell development," *Biochemical Journal*, vol. 476, no. 5, pp. 769–778, 2019.

[14] M. Noguchi, Y. Nakamura, S. M. Russell, S. F. Ziegler, M. Tsang, X. Cao, and W. J. Leonard, "Interleukin-2 receptor γ chain: A functional component of the interleukin-7 receptor," *Science*, vol. 262, pp. 1877–1880, dec 1993.

[15] A. E. Corcoran, A. Riddell, D. Krooshoop, and A. R. Venkitaraman, "Impaired immunoglobulin gene rearrangement in mice lacking the IL-7 receptor," *Nature*, vol. 391, pp. 904–907, feb 1998.

[16] S. A. Corfe and C. J. Paige, "The many roles of IL-7 in B cell development; Mediator of survival, proliferation and differentiation," jun 2012.

[17] A. Puel, S. F. Ziegler, R. H. Buckley, and W. J. Leonard, "Defective IL7R expression in T-B+NK+ severe combined immunodeficiency," *Nature Genetics*, vol. 20, pp. 394–397, dec 1998.

[18] S. Giliani, L. Mori, G. De Saint Basile, F. Le Deist, C. Rodriguez-Perez, C. Forino, E. Mazzolari, S. Dupuis, R. Elhasid, A. Kessel, C. Galambrun, J. Gil, A. Fischer, A. Etzioni, and L. D. Notarangelo, "Interleukin-7 receptor α (IL-7Rα) deficiency: Cellular and molecular bases. Analysis of clinical, immunological, and molecular features in 16 novel patients," feb 2005.

[19] J. J. O'Shea and R. Plenge, "JAK and STAT Signaling Molecules in Immunoregulation and Immune-Mediated Disease," apr 2012.

[20] Z. Yao, Y. Cui, W. T. Watford, J. H. Bream, K. Yamaoka, B. D. Hissong, D. Li, S. K. Durum, Q. Jiang, A. Bhandoola, L. Hennighausen, and J. J. O'Shea, "Stat5a/b are essential for normal lymphoid development and differentiation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 1000–1005, jan 2006.

[21] C. A. Goetz, I. R. Harmon, J. J. O'Neil, M. A. Burchill, and M. A. Farrar, "STAT5 Activation Underlies IL7 Receptor-Dependent B Cell Development," *The Journal of Immunology*, vol. 172, pp. 4770–4778, apr 2004.

[22] K. Ochiai, M. Maienschein-Cline, M. Mandal, J. R. Triggs, E. Bertolino, R. Sciammas, A. R. Dinner, M. R. Clark, and H. Singh, "A self-reinforcing regulatory network triggered by limiting IL-7 activates pre-BCR signaling and differentiation," *Nature Immunology*, vol. 13, pp. 300–307, mar 2012.

[23] F. Ramadani, D. J. Bolland, F. Garcon, J. L. Emery, B. Vanhaesebroeck, A. E. Corcoran, and K. Okkenhaug, "The PI3K isoforms p110α and p110δ are essential for pre-B cell receptor signaling and B cell development," *Science Signaling*, vol. 3, p. ra60, aug 2010.

[24] D. A. Fruman, S. B. Snapper, C. M. Yballe, L. Davidson, J. Y. Yu, F. W. Alt, and L. C. Cantley, "Impaired B cell development and proliferation in absence of phosphoinositide 3-kinase p85α," *Science*, vol. 283, pp. 393–397, jan 1999.

[25] Q. Jiang, W. Q. Li, R. R. Hofmeister, H. A. Young, D. R. Hodge, J. R. Keller, A. R. Khaled, and S. K. Durum, "Distinct Regions of the Interleukin-7 Receptor Regulate Different Bcl2 Family Members," *Molecular and Cellular Biology*, vol. 24, pp. 6501–6513, jul 2004.

[26] S. Malin, S. McManus, C. Cobaleda, M. Novatchkova, A. Delogu, P. Bouillet, A. Strasser, and M. Busslinger, "Role of STAT5 in controlling cell survival and immunoglobulin gene recombination during pro-B cell development," *Nature Immunology*, vol. 11, pp. 171–179, feb 2010.

[27] A. B. Cooper, C. M. Sawai, E. Sicinska, S. E. Powers, P. Sicinski, M. R. Clark, and I. Aifantis, "A unique function for cyclin D3 in early B cell development," *Nature Immunology*, vol. 7, pp. 489–497, may 2006.

[28] M. Mandal, S. E. Powers, K. Ochiai, K. Georgopoulos, B. L. Kee, H. Singh, and M. R. Clark, "Ras orchestrates exit from the cell cycle and light-chain recombination during early B cell development," *Nature Immunology*, vol. 10, no. 10, pp. 1110–1117, 2009.

[29] N. N. Danial and S. J. Korsmeyer, "Cell Death: Critical Control Points," jan 2004.

[30] A. Essafi, S. Fernández De Mattos, Y. A. Hassen, I. Soeiro, G. J. Mufti, N. S. B. Thomas, R. H. Medema, and E. W. Lam, "Direct transcriptional regulation of Bim by FoxO3a mediates STI571-induced apoptosis in Bcr-Abl-expressing cells," *Oncogene*, vol. 24, pp. 2317–2329, mar 2005.

[31] E. Castellano and J. Downward, "Ras interaction with PI3K: More than just another effector pathway," mar 2011.

[32] S. Herzog, E. Hug, S. Meixlsperger, J. H. Paik, R. A. DePinho, M. Reth, and H. Jumaa, "SLP-65 regulates immunoglobulin light chain gene recombination through the PI(3)K-PKB-Foxo pathway," *Nature Immunology*, vol. 9, pp. 623–631, jun 2008.

[33] R. H. Amin and M. S. Schlissel, "Foxo1 directly regulates the transcription of recombination-activating genes during B cell development," *Nature Immunology*, vol. 9, pp. 613–622, jun 2008.

[34] M. A. Oettinger, D. G. Schatz, C. Gorka, and D. Baltimore, "RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination," *Science*, vol. 248, no. 4962, pp. 1517–1523, 1990.

[35] S. E. Powers, M. Mandal, S. Matsuda, A. V. Miletic, M. H. Cato, A. Tanaka, R. C. Rickert, S. Koyasu, and M. R. Clark, "Subnuclear cyclin D3 compartments and the coordinated regulation of proliferation and immunoglobulin variable gene repression," *Journal of Experimental Medicine*, vol. 209, pp. 2199–2213, nov 2012.

[36] S. Karki, D. E. Kennedy, K. Mclean, A. T. Grzybowski, M. Maienschein-Cline, S. Banerjee, H. Xu, E. Davis, M. Mandal, C. Labno, S. E. Powers, M. M. Le Beau, A. R. Dinner, H. Singh, A. J. Ruthenburg, and M. R. Clark, "Regulated Capture of Vκ Gene Topologically Associating Domains by Transcription Factories," *Cell Reports*, vol. 24, pp. 2443–2456, aug 2018.

[37] M. Mandal, S. Powers, M. Maienschein-Cline, E. Bartom, K. Hamel, B. Kee, A. Dinner, and M. Clark, "Epigenetic repression of the Igk locus by STAT5-mediated Ezh2 recruitment," *Nat Immunol*, vol. 12, no. 12, pp. 1212–1220, 2011.

[38] P. Stanhope-Baker and K. M. Hudson, "Cell Type-Specific Chromatin Structure Determines the Targeting of V(D)J Recombinase Activity In Vitro," tech. rep., 1996.

[39] S. Ma, S. Pathak, M. Mandal, L. Trinh, M. R. Clark, and R. Lu, "Ikaros and Aiolos Inhibit Pre-B-Cell Proliferation by Directly Suppressing c-Myc Expression," *Molecular and Cellular Biology*, vol. 30, pp. 4149–4158, sep 2010.

[40] M. J. Parker, S. Licence, L. Erlandsson, G. R. Galler, L. Chakalova, C. S. Osborne, G. Morgan, P. Fraser, H. Jumaa, T. H. Winkler, J. Skok, and I. L. Mårtensson, "The pre-B-cell receptor induces silencing of VpreB and $\lambda 5$ transcription," *EMBO Journal*, vol. 24, pp. 3895–3905, nov 2005.

[41] F. Melchers, E. ten Boekel, T. Seidl, X. C. Kong, T. Yamagami, K. Onishi, T. Shimizu, A. G. Rolink, and J. Andersson, "Repertoire selection by pre-B-cell receptors and B-cell receptors, and genetic control of B-cell development from immature to mature B cells.," *Immunological reviews*, vol. 175, pp. 33–46, jun 2000.

[42] K. Tokoyoda, T. Egawa, T. Sugiyama, B. I. Choi, and T. Nagasawa, "Cellular niches controlling B lymphocyte behavior within bone marrow during development," *Immunity*, vol. 20, pp. 707–718, jun 2004.

[43] S. Zehentmeier and J. P. Pereira, "Cell circuits and niches controlling B cell development," *Immunological Reviews*, vol. 289, pp. 142–157, may 2019.

[44] A. Cordeiro Gomes, T. Hara, V. Y. Lim, D. Herndler-Brandstetter, E. Nevius, T. Sugiyama, S. Tani-ichi, S. Schlenner, E. Richie, H. R. Rodewald, R. A. Flavell, T. Nagasawa, K. Ikuta, and J. P. Pereira, "Hematopoietic Stem Cell Niches Produce Lineage-Instructive Signals to Control Multipotent Progenitor Differentiation," *Immunity*, vol. 45, pp. 1219–1231, dec 2016.

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

[46] C. Fistonich, S. Zehentmeier, J. J. Bednarski, R. Miao, H. Schjerven, B. P. Sleckman, and J. P. Pereira, "Cell circuits between B cell progenitors and IL-7+ mesenchymal progenitor cells control B cell development," *The Journal of experimental medicine*, vol. 215, pp. 2586–2599, oct 2018.

[47] M. Decker, L. Martinez-Morentin, G. Wang, Y. Lee, Q. Liu, J. Leslie, and L. Ding, "Leptin-receptor-expressing bone marrow stromal cells are myofibroblasts in primary myelofibrosis," *Nature Cell Biology*, vol. 19, pp. 677–688, may 2017.

[48] A. N. Tikhonova, I. Dolgalev, H. Hu, K. K. Sivaraj, E. Hoxha, Á. Cuesta-Domínguez, S. Pinho, I. Akhmetzyanova, J. Gao, M. Witkowski, M. Guillamot, M. C. Gutkin, Y. Zhang, C. Marier, C. Diefenbach, S. Kousteni, A. Heguy, H. Zhong, D. R. Fooksman, J. M. Butler, A. Economides, P. S. Frenette, R. H. Adams, R. Satija, A. Tsirigos, and

I. Aifantis, "The bone marrow microenvironment at single-cell resolution," *Nature*, vol. 569, pp. 222–228, may 2019.

[49] M. Mandal, M. K. Okoreeh, D. E. Kennedy, M. Maienschein-Cline, J. Ai, K. C. McLean, N. Kaverina, M. Veselits, I. Aifantis, F. Gounari, and M. R. Clark, "CXCR4 signaling directs Igk recombination and the molecular mechanisms of late B lymphopoiesis," *Nature Immunology*, vol. 20, pp. 1393–1403, oct 2019.

[50] K. Ohnishi and F. Melchers, "The nonimmunoglobulin portion of λ5 mediates cell-autonomous pre-B cell receptor signalling," *Nature Immunology*, vol. 4, pp. 849–856, sep 2003.

[51] R. Übelhart, M. P. Bach, C. Eschbach, T. Wossning, M. Reth, and H. Jumaa, "N-linked glycosylation selectively regulates autonomous precursor BCR function," *Nature Immunology*, vol. 11, pp. 759–765, aug 2010.

[52] Y. Kawano, S. Yoshikawa, Y. Minegishi, and H. Karasuyama, "Pre-B Cell Receptor Assesses the Quality of IgH Chains and Tunes the Pre-B Cell Repertoire by Delivering Differential Signals," *The Journal of Immunology*, vol. 177, pp. 2242–2249, aug 2006.

[53] H. Karasuyama, A. Kudo, and F. Melchers, "The proteins encoded by the VpreB and λ5 Pre-B cell-specific genes can associate with each other and with heavy chain," *Journal of Experimental Medicine*, vol. 172, pp. 969–972, sep 1990.

[54] T. H. Winkler and I. L. Martensson, "The role of the pre-b cell receptor in b cell development, repertoire selection, and tolerance," nov 2018.

[55] R. A. Keenan, A. De Riva, B. Corleis, L. Hepburn, S. Licence, T. H. Winkler, and I. L. Mårtensson, "Censoring of autoreactive B cell development by the pre-B cell receptor," *Science*, vol. 321, pp. 696–699, aug 2008.

[56] T. Kurosaki, "Functional dissection of BCR signaling pathways.," *Current opinion in immunology*, vol. 12, pp. 276–81, jun 2000.

[57] H. Flaswinkel and M. Reth, "Dual role of the tyrosine activation motif of the Ig-alpha protein during signal transduction via the B cell antigen receptor.," *The EMBO Journal*, vol. 13, pp. 83–89, jan 1994.

[58] S. Kabak, B. J. Skaggs, M. R. Gold, M. Affolter, K. L. West, M. S. Foster, K. Siemasko, A. C. Chan, R. Aebersold, and M. R. Clark, "The Direct Recruitment of BLNK to Immunoglobulin Couples the B-Cell Antigen Receptor to Distal Signaling Pathways," *Molecular and Cellular Biology*, vol. 22, pp. 2524–2535, apr 2002.

[59] M. Turner, P. J. Mee, P. S. Costello, O. Williams, A. A. Price, L. P. Duddy, M. T. Furlong, R. L. Geahlen, and V. L. Tybulewicz, "Perinatal lethality and blocked B-cell development in mice lacking the tyrosine kinase Syk," *Nature*, vol. 378, pp. 298–302, nov 1995.

[60] H. Jumaa, B. Wollscheid, M. Mitterer, J. Wienands, M. Reth, and P. J. Nielsen, "Abnormal development and function of B lymphocytes in mice deficient for the signaling adaptor protein SLP-65.," *Immunity*, vol. 11, pp. 547–54, nov 1999.

[61] S. Xu, K. G. Lee, J. Huo, T. Kurosaki, and K. P. Lam, "Combined deficiencies in Bruton tyrosine kinase and phospholipase Cγ2 arrest B-cell development at a pre-BCR+ stage," *Blood*, vol. 109, pp. 3377–3384, apr 2007.

[62] S. Ma, S. Pathak, L. Trinh, and R. Lu, "Interferon regulatory factors 4 and 8 induce the expression of Ikaros and Aiolos to down-regulate pre-B-cell receptor and promote cell-cycle withdrawal in pre-B-cell development," *Blood*, vol. 111, pp. 1396–1403, feb 2008.

[63] J. H. Um, A. L. Brown, S. K. Singh, Y. Chen, M. Gucek, B. S. Lee, M. A. Luckey, M. K. Kim, J. H. Park, B. P. Sleckman, M. Gellert, and J. H. Chung, "Metabolic sensor AMPK directly phosphorylates RAG1 protein and regulates V(D)J recombination," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 9873–9878, jun 2013.

[64] F. Young, B. Ardman, Y. Shinkai, R. Lansford, T. K. Blackwell, M. Mendelsohn, A. Rolink, F. Melchers, and F. W. Alt, "Influence of immunoglobulin heavy- and light-chain expression on B-cell differentiation," *Genes and Development*, vol. 8, pp. 1043–1057, may 1994.

[65] A. Flemming, T. Brummer, M. Reth, and H. Jumaa, "The adaptor protein SLP-65 acts as a tumor suppressor that limits pre-B cell expansion," jan 2003.

[66] A. C. Shaw, W. Swat, L. Davidson, and F. W. Alt, "Induction of Ig light chain gene rearrangement in heavy chain-deficient B cells by activated ras," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, pp. 2239–2243, mar 1999.

[67] A. C. Shaw, W. Swat, R. Ferrini, L. Davidson, and F. W. Alt, "Activated Ras signals developmental progression of recombinase- activating gene (RAG)-deficient pro-B lymphocytes," *Journal of Experimental Medicine*, vol. 189, pp. 123–129, jan 1999.

[68] A. S. Lazorchak, M. S. Schlissel, and Y. Zhuang, "E2A and IRF-4/Pip Promote Chromatin Modification and Transcription of the Immunoglobulin Locus in Pre-B Cells," *Molecular and Cellular Biology*, vol. 26, pp. 810–821, feb 2006.

[69] M. S. Schlissel, "Regulation of activation and recombination of the murine Igkappa locus," *Immunological Reviews*, vol. 200, pp. 215–223, aug 2004.

[70] K. Beck, M. M. Peak, T. Ota, D. Nemazee, and C. Murre, "Distinct roles for E12 and E47 in B cell specification and the sequential rearrangement of immunoglobulin light chain loci," *Journal of Experimental Medicine*, vol. 206, pp. 2271–2284, sep 2009.

[71] S. Sakamoto, K. Wakae, Y. Anzai, K. Murai, N. Tamaki, M. Miyazaki, K. Miyazaki, W. J. Romanow, T. Ikawa, D. Kitamura, I. Yanagihara, N. Minato, C. Murre, and Y. Agata, "E2A and CBP/p300 Act in Synergy To Promote Chromatin Accessibility of the Immunoglobulin κ Locus," *The Journal of Immunology*, vol. 188, pp. 5547–5560, jun 2012.

[72] M. Mandal, K. M. Hamel, M. Maienschein-Cline, A. Tanaka, G. Teng, J. H. Tuteja, J. J. Bunker, N. Bahroos, J. J. Eppig, D. G. Schatz, and M. R. Clark, "Histone reader BRWD1 targets and restricts recombination to the Igk locus," *Nature Immunology*, vol. 16, pp. 1094–1103, oct 2015.

[73] D. G. Schatz and Y. Ji, "Recombination centres and the orchestration of V(D)J recombination," apr 2011.

[74] W. W. Tee, S. S. Shen, O. Oksuz, V. Narendra, and D. Reinberg, "Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs," *Cell*, vol. 156, pp. 678–690, feb 2014.

[75] R. Lu, K. L. Medina, D. W. Lancki, and H. Singh, "IRF-4,8 orchestrate the pre-B-to-B transition in lymphocyte development," *Genes and Development*, vol. 17, pp. 1703–1708, jul 2003.

[76] E. C. Thompson, B. S. Cobb, P. Sabbattini, S. Meixlsperger, V. Parelho, D. Liberg, B. Taylor, N. Dillon, K. Georgopoulos, H. Jumaa, S. T. Smale, A. G. Fisher, and M. Merkenschlager, "Ikaros DNA-Binding Proteins as Integral Components of B Cell Developmental-Stage-Specific Regulatory Circuits (DOI:10.1016/j.immuni.2007.02.010)," apr 2007.

[77] A. Lagergren, R. Månsson, J. Zetterblad, E. Smith, B. Basta, D. Bryder, P. Åkerblad, and M. Sigvardsson, "The Cxcl12, Periostin, and Ccl9 genes are direct targets for early B-cell factor in OP-9 stroma cells," *Journal of Biological Chemistry*, vol. 282, pp. 14454–14462, may 2007.

[78] T. Egawa, K. Kawabata, H. Kawamoto, K. Amada, R. Okamoto, N. Fujii, T. Kishimoto, Y. Katsura, and T. Nagasawa, "The earliest stages of B cell development require a chemokine stromal cell-derived factor/pre-B cell growth-stimulating factor," *Immunity*, vol. 15, no. 2, pp. 323–334, 2001.

[79] T. Nagasawa, S. Hirota, K. Tachibana, N. Takakura, S. I. Nishikawa, Y. Kitamura, N. Yoshida, H. Kikutani, and T. Kishimoto, "Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1," *Nature*, vol. 382, pp. 635–638, aug 1996.

[80] Q. Ma, D. Jones, P. R. Borghesani, R. A. Segal, T. Nagasawa, T. Kismimoto, R. T. Bronson, and T. A. Springer, "Impaired B-lymphopoiesis, myelopoiesis, and derailed cerebellar neuron migration in CXCR4- and SDF-1-deficient mice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, pp. 9448–9453, aug 1998.

[81] M. Z. Ratajczak, E. Zuba-Surma, M. Kucia, R. Reca, W. Wojakowski, and J. Ratajczak, "The pleiotropic effects of the SDF-1-CXCR4 axis in organogenesis, regeneration and tumorigenesis," aug 2006.

[82] Y. Nie, J. Waite, F. Brewer, M. J. Sunshine, D. R. Littman, and Y. R. Zou, "The role of CXCR4 in maintaining peripheral B cell compartments and humoral immunity," *Journal of Experimental Medicine*, vol. 200, pp. 1145–1156, nov 2004.

[83] M. Mandal, M. Maienschein-Cline, P. Maffucci, M. Veselits, D. E. Kennedy, K. C. McLean, M. K. Okoreeh, S. Karki, C. Cunningham-Rundles, and M. R. Clark, "BRWD1 orchestrates epigenetic landscape of late B lymphopoiesis," *Nature Communications*, vol. 9, no. 1, 2018.

[84] S. M. Lewis, "The mechanism of V(D)J joining: Lessons from molecular, immunological, and comparative analyses," 1994.

[85] P. C. Swanson, "The bounty of RAGs: recombination signal complexes and reaction outcomes," *Immunological Reviews*, vol. 200, pp. 90–114, aug 2004.

[86] S. D. Fugmann and D. G. Schatz, "Identification of basic residues in RAG2 critical for DNA binding by the RAG1-RAG2 complex.," *Molecular cell*, vol. 8, pp. 899–910, oct 2001.

[87] C. L. Mundy, N. Patenge, A. G. W. Matthews, and M. A. Oettinger, "Assembly of the RAG1/RAG2 Synaptic Complex," *Molecular and Cellular Biology*, vol. 22, pp. 69–77, jan 2002.

[88] J. M. Jones and M. Gellert, "Ordered assembly of the V(D)J synaptic complex ensures accurate recombination," *EMBO Journal*, vol. 21, no. 15, pp. 4162–4171, 2002.

[89] K. D. Mills, D. O. Ferguson, and F. W. Alt, "The role of DNA breaks in genomic instability and tumorigenesis," aug 2003.

[90] M. R. Lieber, K. Yu, and S. C. Raghavan, "Roles of nonhomologous DNA end joining, V(D)J recombination, and class switch recombination in chromosomal translocations," *DNA Repair*, vol. 5, pp. 1234–1245, sep 2006.

[91] A. G. Tsai, H. Lu, S. C. Raghavan, M. Muschen, C. L. Hsieh, and M. R. Lieber, "Human Chromosomal Translocations at CpG Sites and a Theoretical Basis for Their Lineage and Stage Specificity," *Cell*, vol. 135, pp. 1130–1142, dec 2008.

[92] K. Luger, M. L. Dechassa, and D. J. Tremethick, "New insights into nucleosome and chromatin structure: An ordered state or a disordered affair?," *Nature Reviews Molecular Cell Biology*, vol. 13, no. 7, pp. 436–447, 2012.

[93] S. L. Klemm, Z. Shipony, and W. J. Greenleaf, "Chromatin accessibility and the regulatory epigenome," apr 2019.

[94] C. K. Lee, Y. Shibata, B. Rao, B. D. Strahl, and J. D. Lieb, "Evidence for nucleosome depletion at active regulatory regions genome-wide," *Nature Genetics*, vol. 36, pp. 900–905, aug 2004.

[95] R. E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E. Giste, A. K. Johnson, E. M. Johnson, T. Kutyavin, B. Lajoie, B. K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E. Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J. Dekker, G. E. Crawford, and J. A. Stamatoyannopoulos, "The accessible chromatin landscape of the human genome," *Nature*, vol. 489, pp. 75–82, sep 2012.

[96] R. B. Deal, J. G. Henikoff, and S. Henikoff, "Genome-wide kinetics of nucleosome turnover determined by metabolic labelinq of histones," *Science*, vol. 328, pp. 1161–1164, may 2010.

[97] D. E. Schones, K. Cui, S. Cuddapah, T. Y. Roh, A. Barski, Z. Wang, G. Wei, and K. Zhao, "Dynamic Regulation of Nucleosome Positioning in the Human Genome," *Cell*, vol. 132, pp. 887–898, mar 2008.

[98] A. M. Deaton, M. Gómez-Rodríguez, J. Mieczkowski, M. Y. Tolstorukov, S. Kundu, R. I. Sadreyev, L. E. Jansen, and R. E. Kingston, "Enhancer regions show high histone H3.3 turnover that changes during differentiation," *eLife*, vol. 5, jun 2016.

[99] M. A. Lever, J. P. Th'ng, X. Sun, and M. J. Hendzel, "Rapid exchange of histone H1.1 on chromatin in living human cells," *Nature*, vol. 408, pp. 873–876, dec 2000.

[100] A. Almer and W. Hörz, "Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the PHO5/PHO3 locus in yeast.," *The EMBO Journal*, vol. 5, pp. 2681–2687, oct 1986.

[101] D. A. Gilchrist, G. Dos Santos, D. C. Fargo, B. Xie, Y. Gao, L. Li, and K. Adelman, "Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation," *Cell*, vol. 143, pp. 540–551, nov 2010.

[102] S. John, P. J. Sabo, R. E. Thurman, M. H. Sung, S. C. Biddie, T. A. Johnson, G. L. Hager, and J. A. Stamatoyannopoulos, "Chromatin accessibility pre-determines glucocorticoid receptor binding patterns," mar 2011.

[103] A. N. Schep, J. D. Buenrostro, S. K. Denny, K. Schwartz, G. Sherlock, and W. J. Greenleaf, "Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions," *Genome Research*, vol. 25, pp. 1757–1770, nov 2015.

[104] K. Struhl and E. Segal, "Determinants of nucleosome positioning," mar 2013.

[105] A. J. Bannister and T. Kouzarides, "Regulation of chromatin by histone modifications," *Nature Publishing Group*, vol. 21, pp. 381–395, 2011.

[106] S. G. Alkhatib and J. W. Landry, "The nucleosome remodeling factor," 2011.

[107] T. Tsukiyama, P. B. Becker, and C. Wu, "ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor," *Nature*, vol. 367, no. 6463, pp. 525–532, 1994.

[108] T. Tsukiyama and C. Wu, "Purification and properties of an ATP-dependent nucleosome remodeling factor," *Cell*, vol. 83, pp. 1011–1020, dec 1995.

[109] H. Xiao, R. Sandaltzopoulos, H. M. Wang, A. Hamiche, R. Ranallo, K. M. Lee, D. Fu, and C. Wu, "Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions," *Molecular Cell*, vol. 8, no. 3, pp. 531–543, 2001.

[110] R. Reeves and M. S. Nissen, "The A·T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure," *Journal of Biological Chemistry*, vol. 265, no. 15, pp. 8573–8582, 1990.

[111] S. Y. Kwon, H. Xiao, C. Wu, and P. Badenhorst, "Alternative splicing of NURF301 generates distinct NURF chromatin remodeling complexes with altered modified histone binding specificities," *PLoS Genetics*, vol. 5, p. e1000574, jul 2009.

[112] J. Wysocka, T. Swigut, H. Xiao, T. A. Milne, S. Y. Kwon, J. Landry, M. Kauer, A. J. Tackett, B. T. Chait, P. Badenhorst, C. Wu, and C. D. Allis, "A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling," *Nature*, vol. 442, pp. 86–90, jul 2006.

[113] A. J. Ruthenburg, H. Li, T. A. Milne, S. Dewell, R. K. McGinty, M. Yuen, B. Ueberheide, Y. Dou, T. W. Muir, D. J. Patel, and C. D. Allis, "Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions," *Cell*, vol. 145, pp. 692–706, may 2011.

[114] W. Dang and B. Bartholomew, "Domain Architecture of the Catalytic Subunit in the ISW2-Nucleosome Complex," *Molecular and Cellular Biology*, vol. 27, pp. 8306–8317, dec 2007.

[115] D. A. Gdula, R. Sandaltzopoulos, T. Tsukiyama, V. Ossipow, and C. Wu, "Inorganic pyrophosphatase is a component of the Drosophila nucleosome remodeling factor complex," *Genes and Development*, vol. 12, pp. 3206–3216, oct 1998.

[116] R. Schwanbeck, H. Xiao, and C. Wu, "Spatial contacts and nucleosome step movements induced by the NURF chromatin remodeling complex," *Journal of Biological Chemistry*, vol. 279, pp. 39933–39941, sep 2004.

[117] A. Hamiche, R. Sandaltzopoulos, D. A. Gdula, and C. Wu, "ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF," *Cell*, vol. 97, pp. 833–842, jun 1999.

[118] A. Flaus and T. Owen-hughes, "Dynamic Properties of Nucleosomes during Thermal and ATP-driven mobilization," *Molecular and Cellular Biology*, vol. 23, no. 21, pp. 7767–7779, 2003.

[119] K. Rippe, A. Schrader, P. Riede, R. Strohner, E. Lehmann, and G. Längst, "DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 15635–15640, oct 2007.

[120] M. Shogren-Knaak, H. Ishii, J. M. Sun, M. J. Pazin, J. R. Davie, and C. L. Peterson, "Histone H4-K16 acetylation controls chromatin structure and protein interactions," *Science*, vol. 311, pp. 844–847, feb 2006.

[121] C. R. Clapier, G. Längst, D. F. Corona, P. B. Becker, and K. P. Nightingale, "Critical role for the histone H4 N terminus in nucleosome remodeling by ISWI.," *Molecular and cellular biology*, vol. 21, pp. 875–83, feb 2001.

[122] D. F. Corona, C. R. Clapier, P. B. Becker, and J. W. Tamkun, "Modulation of ISWI function by site-specific histone acetylation," *EMBO Reports*, vol. 3, no. 3, p. 242, 2002.

[123] F. Winston and M. Carlson, "Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection," nov 1992.

[124] P. Sudarsanam and F. Winston, "The Swi/Snf family: Nucleosome-remodeling complexes and transcriptional control," aug 2000.

[125] D. P. Bazett-Jones, J. Côté, C. C. Landel, C. L. Peterson, and J. L. Workman, "The SWI/SNF Complex Creates Loop Domains in DNA and Polynucleosome Arrays and Can Disrupt DNA-Histone Contacts within These Domains," Tech. Rep. 2, 1999.

[126] I. Whitehouse, A. Flaus, B. R. Cairns, M. F. White, J. L. Workman, and T. Owen-Hughes, "Nucleosome mobilization catalysed by the yeast SWI/SNF complex," *Nature*, vol. 400, no. 6746, pp. 784–787, 1999.

[127] Y. Lorch, M. Zhang, and R. D. Kornberg, "Histone octamer transfer by a chromatin-remodeling complex," *Cell*, vol. 96, pp. 389–392, feb 1999.

[128] T. Owen-Hughes, R. T. Utley, J. Côté, C. L. Peterson, and J. L. Workman, "Persistent site-specific remodeling of a nucleosome array by transient action of the SWI/SNF complex," *Science*, vol. 273, no. 5274, pp. 513–516, 1996.

[129] R. E. Kingston, C. A. Bunker, and A. N. Imbalzano, "Repression and activation by multiprotein complexes that alter chromatin structure," apr 1996.

[130] S. R. Biggar, "Continuous and widespread roles for the Swi-Snf complex in transcription," *The EMBO Journal*, vol. 18, no. 8, pp. 2254–2264, 1999.

[131] P. Sudarsanam, Y. Cao, L. Wu, B. C. Laurent, and F. Winston, "The nucleosome remodeling complex, Snf/Swi, is required for the maintenance of transcription in vivo and is partially redundant with the histone acetyltransferase, Gcn5," *EMBO Journal*, vol. 18, no. 11, pp. 3101–3106, 1999.

[132] G. Farkas, J. Gauszt, M. Galloni, G. Reuter, H. Gyurkovicst, I. Be Fran, and I. Karch, "The Trithorax-like gene encodes the Drosophila GAGA factor," Tech. Rep. 6, 1994.

[133] S. C. Elgin and G. Reuter, "Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila," *Cold Spring Harbor Perspectives in Biology*, vol. 5, aug 2013.

[134] D. S. Gilmour, G. H. Thomas, and S. C. Elgin, "Drosophila nuclear proteins bind to regions of alternating C and T residues in gene promoters," *Science*, vol. 245, pp. 1487–1490, sep 1989.

[135] Q. Lu, L. L. Wallrath, H. Granok, and S. C. Elgin, "(CT)n (GA)n repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the Drosophila hsp26 gene.," *Molecular and Cellular Biology*, vol. 13, pp. 2802–2814, may 1993.

[136] H. Granok, B. A. Leibovitch, C. D. Shaffer, and S. C. Elgin, "Chromatin: Ga-ga over GAGA factor," *Current Biology*, 1995.

[137] S. Zollman, D. Godt, G. G. Privé, J. L. Couderc, and F. A. Laski, "The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in Drosophila," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, pp. 10717–10721, oct 1994.

[138] V. J. Bardwell and R. Treisman, "The POZ domain: A conserved protein-protein interaction motif," *Genes and Development*, vol. 8, pp. 1664–1677, jul 1994.

[139] M. L. Espinás, E. Jiménez-García, A. Vaquero, S. Canudas, J. Bernués, and F. Azorín, "The N-terminal POZ domain of GAGA mediates the formation of oligomers that bind DNA with high affinity and specificity," *Journal of Biological Chemistry*, vol. 274, pp. 16461–16469, jun 1999.

[140] P. V. Pedone, R. Ghirlando, G. M. Clore, A. M. Gronenborn, G. Felsenfeld, and J. G. Omichinski, "The single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, pp. 2822–2826, apr 1996.

[141] J. G. Omichinski, P. V. Pedone, G. Felsenfeld, A. M. Gronenborn, and G. M. Clore, "The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode," *Nature Structural Biology*, vol. 4, pp. 122–132, feb 1997.

[142] R. C. Wilkins and J. T. Lis, "GAGA factor binding to DNA via a single trinucleotide sequence element.," *Nucleic acids research*, vol. 26, pp. 2672–8, jun 1998.

[143] H. Huang, I. Rambaldi, E. Daniels, and M. Featherstone, "Expression of theWdr9 gene and protein products during mouse development," *Developmental Dynamics*, vol. 227, pp. 608–614, aug 2003.

[144] P. Filippakopoulos, S. Picaud, M. Mangos, T. Keates, J. P. Lambert, D. Barsyte-Lovejoy, I. Felletar, R. Volkmer, S. Müller, T. Pawson, A. C. Gingras, C. H. Arrowsmith, and S. Knapp, "Histone recognition and large-scale structural analysis of the human bromodomain family," *Cell*, vol. 149, pp. 214–231, mar 2012.

[145] D. R. Hewish and L. A. Burgoyne, "Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease.," *Biochemical and biophysical research communications*, vol. 52, pp. 504–10, may 1973.

[146] C. Wu, Y. C. Wong, and S. C. Elgin, "The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity," *Cell*, vol. 16, pp. 807–814, apr 1979.

[147] R. D. Kornberg, "Chromatin structure: A repeating unit of histones and DNA," *Science*, vol. 184, no. 4139, pp. 868–871, 1974.

[148] R. K. Saiki, S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim, "Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia," *Science*, vol. 230, pp. 1350–1354, dec 1985.

[149] P. R. Mueller and B. Wold, "In vivo footprinting of a muscle specific enhancer by ligation mediated PCR," *Science*, vol. 246, pp. 780–786, nov 1989.

[150] S. Rao, E. Procko, and M. F. Shannon, " Chromatin Remodeling, Measured by a Novel Real-Time Polymerase Chain Reaction Assay, Across the Proximal Promoter Region of the IL-2 Gene ," *The Journal of Immunology*, vol. 167, pp. 4494–4503, oct 2001.

[151] G. E. Crawford, S. Davis, P. C. Scacheri, G. Renaud, M. J. Halawi, M. R. Erdos, R. Green, P. S. Meltzer, T. G. Wolfsberg, and F. S. Collins, "DNase-chip: A high-resolution method to identify DNase I hypersensitive sites using tiled microarrays," *Nature Methods*, vol. 3, pp. 503–509, jul 2006.

[152] P. J. Sabo, M. S. Kuehn, R. Thurman, B. E. Johnson, E. M. Johnson, H. Cao, M. Yu, E. Rosenzweig, J. Goldy, A. Haydock, M. Weaver, A. Shafer, K. Lee, F. Neri, R. Humbert, M. A. Singer, T. A. Richmond, M. O. Dorschner, M. McArthur, M. Hawrylycz,

R. D. Green, P. A. Navas, W. S. Noble, and J. A. Stamatoyannopoulos, "Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays," *Nature Methods*, vol. 3, pp. 511–518, jul 2006.

[153] A. P. Boyle, S. Davis, H. P. Shulha, P. Meltzer, E. H. Margulies, Z. Weng, T. S. Furey, and G. E. Crawford, "High-Resolution Mapping and Characterization of Open Chromatin across the Genome," *Cell*, vol. 132, pp. 311–322, jan 2008.

[154] J. R. Hesselberth, X. Chen, Z. Zhang, P. J. Sabo, R. Sandstrom, A. P. Reynolds, R. E. Thurman, S. Neph, M. S. Kuehn, W. S. Noble, S. Fields, and J. A. Stamatoyannopoulos, "Global mapping of protein-DNA interactions in vivo by digital genomic footprinting," *Nature Methods*, vol. 6, pp. 283–289, mar 2009.

[155] J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf, "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position," *Nature Methods*, vol. 10, pp. 1213–1218, dec 2013.

[156] E. R. Mardis, "ChIP-seq: Welcome to the new frontier," *Nature Methods*, vol. 4, pp. 613–614, aug 2007.

[157] M. J. Solomon, P. L. Larsen, and A. Varshavsky, "Mapping proteinDNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene," *Cell*, vol. 53, pp. 937–947, jun 1988.

[158] G. Robertson, M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. Euskirchen, B. Bernier, R. Varhol, A. Delaney, N. Thiessen, O. L. Griffith, A. He, M. Marra, M. Snyder, and S. Jones, "Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing," *Nature Methods*, vol. 4, pp. 651–657, aug 2007.

[159] A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao, "High-Resolution Profiling of Histone Methylations in the Human Genome," *Cell*, vol. 129, pp. 823–837, may 2007.

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

[161] J. Kwon, A. N. Imbalzano, A. Matthews, and M. A. Oettinger, "Accessibility of nucleosomal DNA to V(D)J cleavage is modulated by RSS positioning and HMG1," *Molecular Cell*, vol. 2, no. 6, pp. 829–839, 1998.

[162] A. Golding, S. Chandler, E. Ballestar, A. P. Wolffe, and M. S. Schlissel, "Nucleosome structure completely inhibits in vitro cleavage by the V(D)J recombinase," 1999.

[163] M. Baumann, A. Mamais, F. McBlane, H. Xiao, and J. Boyes, "Regulation of V(D)J recombination by nucleosome positioning at recombination signal sequences," 2003.

[164] H. D. Kondilis-Mangum, R. M. Cobb, O. Osipovich, S. Srivatsan, E. M. Oltz, and M. S. Krangel, "Transcription-Dependent Mobilization of Nucleosomes at Accessible TCR Gene Segments In Vivo," 2010.

[165] J. Kwon, K. B. Morshead, J. R. Guyon, R. E. Kingston, and M. A. Oettinger, "Histone acetylation and hSWI/SNF remodeling act in concert to stimulate V(D)J cleavage of nucleosomal DNA," *Molecular Cell*, vol. 6, no. 5, pp. 1037–1048, 2000.

[166] M. J. Rowley and V. G. Corces, "Organizational principles of 3D genome architecture," *Nature Reviews Genetics*, vol. 19, pp. 789–800, dec 2018.

[167] A. Miele, N. Gheldof, T. M. Tabuchi, J. Dostie, and J. Dekker, "Mapping chromatin interactions by chromosome conformation capture," *Current protocols in molecular biology*, vol. 74, no. 1, pp. 21–11, 2006.

[168] E. Lieberman-Aiden, N. L. Van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, *et al.*, "Comprehensive mapping of long-range interactions reveals folding principles of the human genome," *science*, vol. 326, no. 5950, pp. 289–293, 2009.

[169] W. A. Bickmore, "The spatial organization of the human genome," *Annual review of genomics and human genetics*, vol. 14, pp. 67–84, 2013.

[170] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, "Topological domains in mammalian genomes identified by analysis of chromatin interactions," *Nature*, vol. 485, no. 7398, pp. 376–380, 2012.

[171] E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, *et al.*, "Spatial partitioning of the regulatory landscape of the x-inactivation centre," *Nature*, vol. 485, no. 7398, pp. 381–385, 2012.

[172] C.-T. Ong and V. G. Corces, "Ctcf: an architectural protein bridging genome topology and function," *Nature Reviews Genetics*, vol. 15, no. 4, pp. 234–246, 2014.

[173] I. F. Davidson, B. Bauer, D. Goetz, W. Tang, G. Wutz, and J.-M. Peters, "Dna loop extrusion by human cohesin," *Science*, vol. 366, no. 6471, pp. 1338–1345, 2019.

[174] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, "Formation of chromosomal domains by loop extrusion," *Cell reports*, vol. 15, no. 9, pp. 2038–2049, 2016.

[175] Y. Zhang, X. Zhang, Z. Ba, Z. Liang, E. W. Dring, H. Hu, J. Lou, N. Kyritsis, J. Zurita, M. S. Shamim, *et al.*, "The fundamental role of chromatin loop extrusion in physiological v (d) j recombination," *Nature*, vol. 573, no. 7775, pp. 600–604, 2019.

[176] E. Roldán, M. Fuxa, W. Chong, D. Martinez, M. Novatchkova, M. Busslinger, and J. A. Skok, "Locus' decontraction'and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene," *Nature immunology*, vol. 6, no. 1, pp. 31–41, 2005.

[177] C. R. de Almeida, R. Stadhouders, M. J. de Bruijn, I. M. Bergen, S. Thongjuea, B. Lenhard, W. van IJcken, F. Grosveld, N. Galjart, E. Soler, *et al.*, "The dna-binding protein ctcf limits proximal vκ recombination and restricts κ enhancer interactions to the immunoglobulin κ light chain locus," *Immunity*, vol. 35, no. 4, pp. 501–513, 2011.

[178] D. J. Bolland, A. L. Wood, C. M. Johnston, S. F. Bunting, G. Morgan, L. Chakalova, P. J. Fraser, and A. E. Corcoran, "Antisense intergenic transcription in v (d) j recombination," *Nature immunology*, vol. 5, no. 6, pp. 630–637, 2004.

[179] C. C. Giallourakis, A. Franklin, C. Guo, H.-L. Cheng, H. S. Yoon, M. Gallagher, T. Perlot, M. Andzelm, A. J. Murphy, L. E. Macdonald, *et al.*, "Elements between the igh variable (v) and diversity (d) clusters influence antisense transcription and lineage-specific v (d) j recombination," *Proceedings of the National Academy of Sciences*, vol. 107, no. 51, pp. 22207–22212, 2010.

[180] N. Servant, N. Varoquaux, B. R. Lajoie, E. Viara, C.-J. Chen, J.-P. Vert, E. Heard, J. Dekker, and E. Barillot, "Hic-pro: an optimized and flexible pipeline for hi-c data processing," *Genome biology*, vol. 16, no. 1, p. 259, 2015.

[181] C. Lazaris, S. Kelly, P. Ntziachristos, I. Aifantis, and A. Tsirigos, "Hic-bench: comprehensive and reproducible hi-c data analysis designed for parameter exploration and benchmarking," *BMC genomics*, vol. 18, no. 1, pp. 1–16, 2017.

[182] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass, "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities," *Molecular cell*, vol. 38, no. 4, pp. 576–589, 2010.

[183] Y. C. Lin, C. Benner, R. Mansson, S. Heinz, K. Miyazaki, M. Miyazaki, V. Chandra, C. Bossen, C. K. Glass, and C. Murre, "Global changes in the nuclear positioning of genes and intra-and interdomain genomic interactions that orchestrate b cell fate," *Nature immunology*, vol. 13, no. 12, p. 1196, 2012.

[184] N. C. Durand, J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov, E. S. Lander, and E. L. Aiden, "Juicebox provides a visualization system for hi-c contact maps with unlimited zoom," *Cell systems*, vol. 3, no. 1, pp. 99–101, 2016.

[185] F. Serra, D. Baù, M. Goodstadt, D. Castillo, G. J. Filion, and M. A. Marti-Renom, "Automatic analysis and 3d-modelling of hi-c data using tadbit reveals structural features of the fly chromatin colors," *PLoS computational biology*, vol. 13, no. 7, p. e1005665, 2017.

[186] F. Ramírez, V. Bhardwaj, L. Arrigoni, K. C. Lam, B. A. Grüning, J. Villaveces, B. Habermann, A. Akhtar, and T. Manke, "High-resolution tads reveal dna sequences underlying genome organization in flies," *Nature communications*, vol. 9, no. 1, pp. 1–15, 2018.

[187] J. Wolff, L. Rabbani, R. Gilsbach, G. Richard, T. Manke, R. Backofen, and B. A. Grüning, "Galaxy hicexplorer 3: a web server for reproducible hi-c, capture hi-c and single-cell hi-c data analysis, quality control and visualization," *Nucleic Acids Research*, 2020.

[188] J. Wolff, V. Bhardwaj, S. Nothjunge, G. Richard, G. Renschler, R. Gilsbach, T. Manke, R. Backofen, F. Ramírez, and B. A. Grüning, "Galaxy hicexplorer: a web server for reproducible hi-c data analysis, quality control and visualization," *Nucleic acids research*, vol. 46, no. W1, pp. W11–W16, 2018.

[189] I. Merelli, A. Guffanti, M. Fabbri, A. Cocito, L. Furia, U. Grazini, R. J. Bonnal, L. Milanesi, and F. McBlane, "Rsssite: a reference database and prediction tool for the identification of cryptic recombination signal sequences in human and murine genomes," *Nucleic acids research*, vol. 38, no. suppl_2, pp. W262–W267, 2010.

[190] M. Davila, F. Liu, L. G. Cowell, A. E. Lieberman, E. Heikamp, A. Patel, and G. Kelsoe, "Multiple, conserved cryptic recombination signals in vh gene segments: detection of cleavage products only in pro–b cells," *The Journal of experimental medicine*, vol. 204, no. 13, pp. 3195–3208, 2007.

[191] G. Teng, Y. Maman, W. Resch, M. Kim, A. Yamane, J. Qian, K.-R. Kieffer-Kwon, M. Mandal, Y. Ji, E. Meffre, *et al.*, "Rag represents a widespread threat to the lymphocyte genome," *Cell*, vol. 162, no. 4, pp. 751–765, 2015.

[192] R. A. Gladdy, M. D. Taylor, C. J. Williams, I. Grandal, J. Karaskova, J. A. Squire, J. T. Rutka, C. J. Guidos, and J. S. Danska, "The rag-1/2 endonuclease causes genomic instability and controls cns complications of lymphoblastic leukemia in p53/prkdc-deficient mice," *Cancer cell*, vol. 3, no. 1, pp. 37–50, 2003.

[193] E. Papaemmanuil, I. Rapado, Y. Li, N. E. Potter, D. C. Wedge, J. Tubio, L. B. Alexandrov, P. Van Loo, S. L. Cooke, J. Marshall, *et al.*, "Rag-mediated recombination is the predominant driver of oncogenic rearrangement in etv6-runx1 acute lymphoblastic leukemia," *Nature genetics*, vol. 46, no. 2, pp. 116–125, 2014.

[194] S. D. Fugmann, A. I. Lee, P. E. Shockett, I. J. Villey, and D. G. Schatz, "The rag proteins and v (d) j recombination: complexes, ends, and transposition," *Annual review of immunology*, vol. 18, no. 1, pp. 495–527, 2000.

[195] P. De and K. K. Rodgers, "Putting the pieces together: identification and characterization of structural domains in the v (d) j recombination protein rag1," *Immunological reviews*, vol. 200, no. 1, pp. 70–82, 2004.

[196] M. Adli, "The crispr tool kit for genome editing and beyond," *Nature communications*, vol. 9, no. 1, pp. 1–13, 2018.

[197] L. Cowell, M. Davila, T. Kepler, and G. Kelsoe, "A statistical model of rss structure that identifies cryptic rss and predicts recombination efficiency in vivo," in *FASEB JOURNAL*, vol. 16, pp. A1073–A1073, FEDERATION AMER SOC EXP BIOL 9650 ROCKVILLE PIKE, BETHESDA, MD 20814-3998 USA, 2002.

[198] R. J. Aguilera, S. Akira, K. Okazaki, and H. Sakano, "A pre-b cell nuclear protein that specifically interacts with the immunoglobulin vj recombination sequences," *Cell*, vol. 51, no. 6, pp. 909–917, 1987.

[199] S. Tonegawa, "Somatic generation of antibody diversity," *Nature*, vol. 302, no. 5909, pp. 575–581, 1983.

[200] G. D. Yancopoulos and F. W. Alt, "Developmentally controlled and tissue-specific expression of unrearranged vh gene segments," *Cell*, vol. 40, no. 2, pp. 271–281, 1985.

[201] D. Jung and F. W. Alt, "Unraveling v (d) j recombination: insights into gene regulation," *Cell*, vol. 116, no. 2, pp. 299–311, 2004.

[202] M. T. McMurry and M. S. Krangel, "A role for histone acetylation in the developmental regulation of v (d) j recombination," *Science*, vol. 287, no. 5452, pp. 495–498, 2000.

[203] R. M. Cobb, K. J. Oestreich, O. A. Osipovich, and E. M. Oltz, "Accessibility control of v (d) j recombination," *Advances in immunology*, vol. 91, pp. 45–109, 2006.

[204] D. G. Hesslein and D. G. Schatz, "Factors and forces controlling v (d) j recombination," *Advances in immunology*, vol. 78, pp. 169–232, 2001.

[205] D. Jung, C. Giallourakis, R. Mostoslavsky, and F. W. Alt, "Mechanism and control of v (d) j recombination at the immunoglobulin heavy chain locus," *Annu. Rev. Immunol.*, vol. 24, pp. 541–570, 2006.

[206] M. S. Krangel, "Mechanics of t cell receptor gene rearrangement," *Current opinion in immunology*, vol. 21, no. 2, pp. 133–139, 2009.

[207] S. Jhunjhunwala, M. C. van Zelm, M. M. Peak, and C. Murre, "Chromatin architecture and the generation of antigen receptor diversity," *Cell*, vol. 138, no. 3, pp. 435–448, 2009.

[208] I. Abarrategui and M. S. Krangel, "Regulation of t cell receptor-α gene recombination by transcription," *Nature immunology*, vol. 7, no. 10, pp. 1109–1115, 2006.

[209] D. G. Hesslein, D. L. Pflugh, D. Chowdhury, A. L. Bothwell, R. Sen, and D. G. Schatz, "Pax5 is required for recombination of transcribed, acetylated, 5 igh v gene segments," *Genes & development*, vol. 17, no. 1, pp. 37–42, 2003.

[210] T. J. Richmond and C. A. Davey, "The structure of dna in the nucleosome core," *Nature*, vol. 423, no. 6936, pp. 145–150, 2003.

[211] M. Iwafuchi-Doi and K. S. Zaret, "Cell fate control by pioneer transcription factors," *Development*, vol. 143, no. 11, pp. 1833–1837, 2016.

[212] C. W. Roberts and S. H. Orkin, "The swi/snf complex—chromatin and cancer," *Nature Reviews Cancer*, vol. 4, no. 2, pp. 133–142, 2004.

[213] L. S. Churchman and J. S. Weissman, "Nascent transcript sequencing visualizes transcription at nucleotide resolution," *Nature*, vol. 469, no. 7330, pp. 368–373, 2011.

[214] D. G. Schatz, M. A. Oettinger, and D. Baltimore, "The v (d) j recombination activating gene, rag-1," *Cell*, vol. 59, no. 6, pp. 1035–1048, 1989.

[215] T. Leu and D. G. Schatz, "rag-1 and rag-2 are components of a high-molecular-weight complex, and association of rag-2 with this complex is rag-1 dependent.," *Molecular and cellular biology*, vol. 15, no. 10, pp. 5657–5670, 1995.

[216] N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo, Y. Field, E. M. LeProust, T. R. Hughes, J. D. Lieb, J. Widom, *et al.*, "The dna-encoded nucleosome organization of a eukaryotic genome," *Nature*, vol. 458, no. 7236, pp. 362–366, 2009.

[217] M.-P. Lefranc, V. Giudicelli, C. Ginestoux, J. Bodmer, W. Müller, R. Bontrop, M. Lemaitre, A. Malik, V. Barbié, and D. Chaume, "Imgt, the international immunogenetics database," *Nucleic acids research*, vol. 27, no. 1, pp. 209–212, 1999.

[218] P. Filippakopoulos and S. Knapp, "Targeting bromodomains: epigenetic readers of lysine acetylation," *Nature reviews Drug discovery*, vol. 13, no. 5, pp. 337–356, 2014.

[219] L. Zalazar, C. A. I. Alonso, R. E. De Castro, and A. Cesari, "An alternative easy method for antibody purification and analysis of protein–protein interaction using gst fusion proteins immobilized onto glutathione–agarose," *Analytical and bioanalytical chemistry*, vol. 406, no. 3, pp. 911–914, 2014.

[220] P. J. Skene and S. Henikoff, "A simple method for generating high-resolution maps of genome-wide protein binding," *Elife*, vol. 4, p. e09225, 2015.

[221] A. W. Langerak, B. Nadel, A. de Torbal, I. L. Wolvers-Tettero, E. J. van Gastel-Mol, B. Verhaaf, U. Jäger, and J. J. van Dongen, "Unraveling the consecutive recombination events in the human igk locus," *The Journal of Immunology*, vol. 173, no. 6, pp. 3878–3888, 2004.

[222] M. van der Burg, T. Tumkaya, M. Boerma, S. de Bruin-Versteeg, A. W. Langerak, and J. J. van Dongen, "Ordered recombination of immunoglobulin light chain genes occurs at the igk locus but seems less strict at the igl locus," *Blood, The Journal of the American Society of Hematology*, vol. 97, no. 4, pp. 1001–1008, 2001.

[223] R. M. Feddersen, D. J. Martin, and B. G. Van Ness, "The frequency of multiple recombination events occurring at the human ig kappa l chain locus.," *The Journal of Immunology*, vol. 144, no. 3, pp. 1088–1093, 1990.

[224] E. Charpentier and J. A. Doudna, "Biotechnology: Rewriting a genome," *Nature*, vol. 495, no. 7439, pp. 50–51, 2013.

[225] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang, "Genome engineering using the crispr-cas9 system," *Nature protocols*, vol. 8, no. 11, pp. 2281–2308, 2013.